

Summer 2019

Ordinary Causation: A Study in Experimental Statutory Interpretation

James Macleod

Brooklyn Law School, james.macleod@brooklaw.edu

Follow this and additional works at: <https://brooklynworks.brooklaw.edu/faculty>



Part of the [Other Law Commons](#), and the [Torts Commons](#)

Recommended Citation

94 Ind. L.J. 957 (2019)

This Article is brought to you for free and open access by BrooklynWorks. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of BrooklynWorks.

ORDINARY CAUSATION: A STUDY IN EXPERIMENTAL STATUTORY INTERPRETATION

JAMES A. MACLEOD*

In a series of recent split decisions interpreting criminal and tort-like legislation, the Supreme Court has purported to give statutory causation requirements their ordinary, plain meaning. Armed with dictionaries, examples from everyday speech, and commonsense intuitions, the Court's majority has explained that statutory phrases like "because of" and "results from" entail but-for causation as a matter of ordinary usage. There's just one problem: The Court's majority (and the many state and federal courts following its lead) is wrong on the facts—specifically, the facts about how people ordinarily interpret, understand, and use causal language.

This Article considers a novel approach to ordinary meaning statutory interpretation, using these recent causation cases as a proof of concept: To find how people would ordinarily construe statutory language in context, ask a lot of people to apply the disputed language, and observe what they do. In short, to find public meaning, ask the public. As a demonstration, the Article reports the results of a nationally representative survey of nearly 1500 jury-eligible laypeople. It tests the Supreme Court's recent pronouncements about the ordinary meaning of causal language in Title VII, the Hate Crimes Prevention Act, the Controlled Substances Act, and jury instructions in similar criminal and statutory tort settings. The results reveal clear and consistent patterns of causal attribution and ordinary usage—patterns that squarely contradict the Court's ordinary meaning determinations. The results also demonstrate that certain alternative causation standards, though rejected by the Court as inconsistent with ordinary linguistic, conceptual, and moral intuitions, come closer to tracking all three.

These discoveries raise serious concerns about the outcomes in recent criminal and tort causation cases, and possibly about ordinary and plain meaning interpretation more broadly. After discussing the implications for causation doctrine and statutory interpretation, the Article considers whether similar experimental methodologies might shed light on additional interpretation controversies in criminal and tort settings, on theories of common law doctrinal development, and on philosophical analyses of causation in criminal and tort theory.

* Associate in Law, Columbia Law School; Assistant Professor of Law, Brooklyn Law School (beginning fall 2019). For helpful discussions and comments, I thank Omri Ben-Shahar, Lisa Bernstein, Anu Bradford, Harold Edgar, Lee Fennell, Victor Goldberg, Neal Goldfarb, David Gray, Philip Hamburger, Bernard Harcourt, Bert Huang, Aziz Huq, Avery Katz, Shlomo Klapper, Joshua Knobe, Jody Kraus, Genevieve Lakier, Brian Leiter, Melissa Longgrass, Richard McAdams, Janice Nadler, Christina D. Ponza-Kraus, David Pozen, Dan Richman, Brian Slocum, Larry Solan, Lior Strahilevitz, Eric Talley, Kristen Underhill, and participants in the Associates & Fellows Workshop at Columbia Law School. I especially thank Bert Huang, Joshua Knobe, Janice Nadler, and Kristen Underhill for discussions regarding experimental design and data analysis. All mistakes are my own.

INTRODUCTION.....	958
I. COURTS ON THE FOLK CONCEPT OF CAUSATION AND ORDINARY USAGE.....	966
A. THE TRIUMPH OF THE BUT-FOR TEST IN RECENT ORDINARY MEANING CASE LAW.....	966
1. STATUTORY TORTS	966
2. CRIMINAL LAW	969
B. POTENTIAL PROBLEMS AND ROADS NOT TAKEN.....	971
1. UNNECESSARY “CAUSES”?	972
2. ALTERNATIVE CAUSATION TESTS	974
II. WHEN AND WHY COURTS CARE ABOUT ORDINARY USAGE AND FOLK CONCEPTS.....	977
A. STATUTORY INTERPRETATION GENERALLY	978
B. CRIMINAL AND TORT LAW GENERALLY	980
C. CAUSATION IN CRIMINAL AND TORT LAW	982
III. HOW COURTS FIND ORDINARY MEANING	985
A. INTROSPECTION AND INTUITION PUMPS	985
B. DICTIONARIES AND CORPUS LINGUISTICS	986
IV. EXPERIMENTAL EVIDENCE OF ORDINARY MEANING AND CAUSAL ATTRIBUTION.....	991
A. ADVANTAGES AND LIMITATIONS OF EXPERIMENTAL STATUTORY INTERPRETATION.....	991
B. THE EXPERIMENTS	994
1. PARTICIPANTS	994
2. DESIGN AND MATERIALS	995
3. RESULTS	998
A. STATUTORY CAUSATION AND THE FAILURE OF THE BUT-FOR TEST	998
B. ALTERNATIVE CAUSATION TESTS	1001
C. TRACKING BLAME, BUT NOT MERELY BLAMING	1002
D. INTERPRETER CONFIDENCE AS NOISE.....	1005
C. SUMMARY AND IMPLICATIONS	1006
V. FURTHER IMPLICATIONS AND EXTENSIONS.....	1012
A. LEGAL INTERPRETATION BEYOND STATUTORY CAUSATION AND PLAIN MEANING	1012
B. PREDICTING THE COMMON LAW: LEGAL REALISM, THE NEW DOCTRINALISM, AND EXPERIMENTAL JURISPRUDENCE	1015
C. EXPERIMENTAL PHILOSOPHY OF CRIMINAL AND TORT LAW: CAUSATION AS A CASE STUDY	1018
CONCLUSION.....	1021
APPENDIX	1023

INTRODUCTION

Statutory interpretation cases sometimes turn on the plain, ordinary meaning of everyday words. A series of recent split Supreme Court decisions concerning causation is representative. Armed with dictionaries, thought experiments, intuition-pumping examples from everyday speech, and common sense, the Court's majority

has explained that statutory phrases like “because of” and “results from” entail but-for causation as a matter of ordinary usage. So, for example, employment discrimination plaintiffs alleging that they were fired “because of” a discriminatory motive under various provisions of Title VII must show that the employer would not have fired them absent the unlawful consideration.¹ Otherwise, as a matter of “common talk” and “ordinary meaning,” we would not say that they were fired “because of” the unlawful consideration.² Similarly, in applying the words of the Controlled Substances Act (CSA) according to their plain, ordinary meaning, we would not say that a drug user’s death “result[ed] from” a drug sold by the defendant if, absent that drug, he would have died at the same time anyway from the other drugs he had taken.³ Again, this is simply a matter of “the common understanding of cause,” and what “is natural to say” about causal relations.⁴

The stakes of these ordinary meaning debates are high. Most obviously, the statutory causation standard in the cases described above matters for discrimination plaintiffs and for criminal defendants accused of selling deadly products. But the Court’s ordinary meaning pronouncements reverberate far beyond the particular statutes they explicitly address because similar causal language appears throughout federal and state codes, in a wide range of substantive settings, and courts often follow the Supreme Court’s lead in importing ordinary meaning determinations from one statute to another without regard to context. As the Fifth Circuit explained in a recent Employee Retirement Income Security Act (ERISA) case, “[t]he Court in *Burrage* [the CSA case described above] was interpreting a drug crime statute,” but the “word ‘results’ retains its ordinary meaning, regardless of whether it appears in Title 21 or Title 29 of the United States Code.”⁵

1. *Gross v. FBL Fin. Servs., Inc.*, 557 U.S. 167, 176 (2009); *see also* *Univ. of Tex. Sw. Med. Ctr. v. Nassar*, 570 U.S. 338 (2013).

2. *Gross*, 557 U.S. at 176 (quoting *Safeco Ins. Co. of Am. v. Burr*, 551 U.S. 47, 63–64, n.14 (2007)); *Nassar*, 570 U.S. at 362 (emphasizing “the plain textual meaning[] of the word ‘because’”). *But see* *Gross*, 557 U.S. at 180 (Stevens, J., dissenting); *id.* at 190 (Breyer, J., dissenting); *Nassar*, 570 U.S. at 363 (Ginsburg, J., dissenting).

3. *Burrage v. United States*, 571 U.S. 204, 216 (2014).

4. *Id.* at 211–12; *see also id.* at 213 (discussing additional examples of statutory phrases that entail but-for causation as a matter of ordinary usage). *But see id.* at 219 (Ginsburg, J., concurring) (noting “room for debate” regarding the meaning of “results from” and concurring on grounds of lenity).

5. *Tolbert v. RBC Capital Markets Corp.*, 758 F.3d 619, 624–25 n.4 (5th Cir. 2014); *see also, e.g.,* *Changzhou Trina Solar Energy Co. v. U.S. Int’l Trade Comm’n*, 879 F.3d 1377, 1380 (Fed. Cir. 2018) (relying on *Burrage* in interpreting federal provision levying duties on imports where “domestic industry was, in the statutory phrase, ‘materially injured . . . by reason of imports’” (emphasis added)); *United States v. Yihao Pu*, 814 F.3d 818, 824 (7th Cir. 2016) (federal fraud sentencing enhancement guidelines provision); *United States v. Alphas*, 785 F.3d 775, 783 (1st Cir. 2015) (same); *Annex Med., Inc. v. Burwell*, 769 F.3d 578, 584–85 (8th Cir. 2014) (Colloton, J., concurring) (for purposes of federal regulation, if “unavailability of a group health plan without the objected-to coverage is ‘a result of the Mandate,’ then it follows in ordinary usage that the Health and Human Services (HHS) mandate is a but-for cause of the desired plan’s unavailability.” (emphasis added)); *United States v. Ramos-Delgado*, 763 F.3d 398, 401–02 (5th Cir. 2014) (federal sentencing guidelines provision); *Rogers v. Bromac Title Servs., LLC*, 755 F.3d 347, 352 (5th Cir. 2014) (federal Jury System

And in addition to these *transsubstantive* effects, the *transprocedural* effects of the Court's ordinary meaning pronouncements raise the stakes further still. For example, after *Burrage*, courts increasingly approve jury instructions that parrot the statute's "results from" language but include no explicit but-for causation instruction. After all, "[a]s the Supreme Court made clear in *Burrage*, the 'ordinary meaning' of 'results from' is that the outcome or harm would not have occurred were it not for the defendant's conduct. . . . The jury could have relied on its general understanding of the phrase 'results from.'"⁶ Indeed, as another court explained, a but-for causation instruction would be redundant: "Generally understood words of ordinary usage need not be defined We are convinced [that] the term[] . . . 'resulted' . . . [is] readily understood by the layperson and [is] not [a] technical or legal term[] requiring further definition."⁷ In short, as these and many other decisions assure us,⁸ juries convicting

Improvement Act); *McNely v. Ocala Star-Banner Corp.*, 99 F.3d 1068, 1077 (11th Cir. 1996) (construing the Americans with Disabilities Act's causal language according to "the ordinary, everyday meaning" of causation); *Micks v. Gurstel Law Firm, P.C.*, No. 17-cv-4659, 2019 WL 418850, at *11–12 (D. Minn. Feb. 1, 2019) (relying on *Burrage*'s analysis of the ordinary meaning of "results from" for purposes of interpreting the Fair Debt Collection Practices Act's use of the same phrase); *Kemply v. Cashcall, Inc.*, No. 08-cv-03174-MEJ, 2016 WL 1055251, at *10 (N.D. Cal. Mar. 16, 2016) (Federal Electronic Funds Transfer Act); *People v. Nere*, 82 N.E.3d 730, 730 (Ill. Ct. App. 2017) (state homicide statute); *Commonwealth v. Kakhankham*, 132 A.3d 986, 992–93 & n.6 (Pa. Super. Ct. 2015) (state sentencing statute); *Goree v. United Parcel Serv., Inc.*, 490 S.W.3d 413, 439 (Tenn. Ct. App. 2015) (relying on plain meaning of "because of" for purposes of retaliation claims under Tennessee Human Rights Act); *Wagoner v. Commonwealth*, 756 S.E.2d 165, 175–76 (Va. Ct. App. 2014), *aff'd*, 770 S.E.2d 479 (Va. 2015) ("We are persuaded by *Burrage* that, where the [Virginia] legislature has not clarified otherwise, this Court should give the phrase 'results in' its ordinary meaning, which imports 'but for' causation."); *cf.* *City of W. Liberty v. Employers Mut. Cas. Co.*, No. 16-1972, 2018 WL 1182764, at *5 (Iowa Ct. App. Mar. 7, 2018) (relying on *Burrage*'s pronouncements concerning the "ordinary meaning of the phrase 'results from'" to construe insurance contract's use of the same phrase to entail but-for causation). *But see* *United States ex rel. Greenfield v. Medco Health Sols., Inc.*, 880 F.3d 89, 95 (3d Cir. 2018) (holding that the phrase "resulting from" in the federal Anti-Kick Back Statute is "arguably not plain," and a but-for causation requirement would produce "incongruous results" under the statute).

6. *United States v. Thomas*, No. 7:10CR00016, 2016 WL 1070868, at *4 (W.D. Va. Mar. 16, 2016) (citations omitted). A similar transprocedural effect occurs where courts interpret defendants' factual admissions in written plea agreements and verbal plea colloquies. *See, e.g.,* *Donavan v. United States*, No. 14-1310, 2016 WL 424946, at *8 (C.D. Ill. Feb. 3, 2016) (explaining that defendant's admission, prior to *Burrage*, that the drug he sold "caused" death, constituted an admission that the drug was a but-for cause of that death and thus satisfied the causation requirement articulated in *Burrage*), *vacated on other grounds*, 865 F.3d 472 (7th Cir. 2017).

7. *State v. Knight*, No. 150568, 2016 WL 3269596, at *8 (Iowa Ct. App. June 15, 2016).

8. *See, e.g.,* *United States v. Alvarado*, 816 F.3d 242, 248, 252 (4th Cir. 2016), *cert. denied*, 137 S. Ct. 492 (2016); *Linder v. Kreuger*, No. 1:15-cv-01055-SLD, 2017 WL 5011879, at *4 (C.D. Ill. Nov. 2, 2017); *Hancock v. United States*, Nos. 2:08-CR-75-JRG-DHI, 2:13-CV-252-JRG, 2016 WL 4290468, at *7 (E.D. Tenn. Aug. 15, 2016); *Doyle v. United States*, No. 14-0364-DRH, 2015 WL 2106179, at *3 (S.D. Ill. May 4, 2015); *Smith v. Wilson*, No. 14-CV-0510 (PJS/HB), 2014 WL 5489244, at *2, *8 (D. Minn. Oct. 30, 2014); *Knight*, 2016 WL 3269596, at *7; *see also* *United States v. Piekarsky*, 687 F.3d 134, 146 (3d

defendants under the CSA's "death results from" provision, which carries a mandatory minimum sentence of twenty years, are using a but-for causation standard, because the phrase "results from" plainly entails but-for causation as a matter of ordinary language and common sense. The Supreme Court, after all, has said so.⁹

There's just one problem: in these and other similar cases, the Court's majority (like the many other courts following its lead) was wrong on the facts—specifically, the facts about how people ordinarily think and talk about causation. The majority was wrong, in other words, about the ordinary and allegedly plain meaning of these statutes' language, and about the "commonsense" concept of causation that language invoked.

But how could we know? The answer is surprisingly simple: ask a lot of people to apply these same phrases and concepts in context and observe what they do. In other words, to find public meaning, ask the public.¹⁰

Cir. 2012) ("'[B]ecause' is a word with an ordinary meaning. As a result, the addition of a clarifying definition—as proposed by the Defendants—was not necessary to the jury's understanding of the instructions it received."); *United States v. Hatfield*, 591 F.3d 945, 949 (7th Cir. 2010) (upholding undefined statutory "results from" instruction because the ordinary meaning was "clear enough" and explaining that "[e]laborating on a term often makes it less rather than more clear"); *cf. Knight*, 2016 WL 3269596, at *7; *Thomas*, 2016 WL 1070868, at *4; *Czimmer v. Janssen Pharm., Inc.*, 122 A.3d 1043, 1054 (Pa. Super. Ct. 2015) (holding that the jury charge's "substantial factor" causation instruction was not erroneous because it clarified that "substantial factor" was synonymous with "factual cause," and the plain meaning of the phrase "factual cause" entails but-for causation as a matter of ordinary usage). *But see United States v. Schneider*, 112 F. Supp. 3d 1197, 1215 (D. Kan. 2015) (holding error not harmless where jury instruction did not define statutory "results from" language); *United States v. MacKay*, 20 F. Supp. 3d 1287, 1294 (D. Utah 2014), *aff'd*, 610 F. App'x 797 (10th Cir. 2015) (same).

9. Of the cases cited above, *Alvarado* may be the most remarkable. There, the Fourth Circuit upheld the jury's conviction after the jury had asked the trial court judge *twice* during deliberation "whether 'death resulted from the use of the heroin' means *solely* from the use of heroin, or that heroin '*contributed to [the deceased's] death.*'" *Alvarado*, 816 F.3d at 248, 252 (emphasis in original). The trial judge had responded to both written notes by simply telling the jury to continue deliberating. As the Fourth Circuit explained in upholding the conviction, "[t]he *Burrage* Court held that 'results from' in [the CSA] invokes the 'ordinary, accepted meaning' of the phrase. And the ordinary meaning of 'results from' is but-for causation—i.e., that death would not have occurred in the absence of heroin." *Id.* at 248.

10. A note regarding terminology: the Article uses "ordinary meaning" and "public meaning" interchangeably. *See* Victoria F. Nourse, *Two Kinds of Plain Meaning*, 76 BROOK. L. REV. 997, 1004 (2011); Lawrence B. Solum, *Communicative Content and Legal Content*, 89 NOTRE DAME L. REV. 479, 480 (2013). Unless otherwise noted, in this Article both refer to what scholars sometimes call "communicative content" or "audience meaning" and, for the most part, also track what scholars sometimes call "linguistic meaning," "semantic meaning," and "contextual meaning." *See* BRIAN SLOCUM, *ORDINARY MEANING: A THEORY OF THE MOST FUNDAMENTAL PRINCIPLE OF LEGAL INTERPRETATION* 4–8 (2015); Richard H. Fallon Jr., *The Meaning of Legal "Meaning" and Its Implications for Theories of Legal Interpretation*, 82 U. CHI. L. REV. 1235 (2015).

As will hopefully become apparent in Parts I–III, this Article's construal of "ordinary meaning" is not an arbitrary stipulation; rather, it reflects the concept of ordinary meaning that

That is the approach this Article proposes and demonstrates in action. The Article reports the results of a nationally representative survey of nearly 1500 jury-eligible laypeople. As it turns out, most people do *not* interpret common causal phrases to imply but-for causation, even in the linguistic and factual contexts at issue in these and similar cases. In fact, for example, most people *would* say that the drug user's death "resulted from" his use of the drug at issue, even if he still would have died from other drugs without it. And most people *would* say that an employer who fires an employee for multiple reasons, only one of which is unlawful, fired the employee "because of" the unlawful reason, even where the alternative benign motivations were sufficient in themselves to cause the firing. Furthermore, most people would express a great deal of *confidence* in claiming that these statutory causation requirements are satisfied, even while simultaneously (and confidently) acknowledging that but-for causation is absent.¹¹ And in doing so, they would not be merely expressing antipathy toward discriminatory employers, drug dealers, and the like; instead, they would be using their commonsense understanding of causation. (For example, most would interpret and apply these causation requirements in the same way in contexts nearly identical to the cases described above, but subtly altered to eliminate the impulse to morally blame or punish.)¹²

In addition to the above findings, the survey reveals that some alternative causation standards rejected by the Court's majority come closer to tracking ordinary usage (and, independently, assessments of moral blameworthiness) than does the majority's "but-for" test, while other alternative standards favored by some courts and commentators fare even worse. Specifically, the "substantial factor" standard for causation comes much closer to tracking common sense and statutory causality attribution than does the "but-for" test, the "contributing factor" test, or the "sole factor" test, and the *sufficiency* of the relevant "cause" is more predictive of causality attribution (and blameworthiness assessments) than the Court's "but-for" standard.¹³

One final finding may help put the rest into perspective. While the Court majority's "commonsense" concept of causation appears far less common than the Court majority seemed to think, there's an important respect in which both the

courts themselves typically appear to invoke in the relevant case law, as evidenced by (1) what courts explicitly say about their conception of "ordinary meaning" (e.g., that it is reflective of how ordinary people ordinarily talk), (2) what evidence courts turn to in order to find it (e.g., nonlegal dictionaries and examples from everyday speech), (3) what courts say about their reasons for caring about ordinary meaning (e.g., it shows what notice the public received), and (4) what inferences courts draw based on ordinary meaning determinations (e.g., if a meaning is ordinary then it is the meaning reasonable juries are lost likely to presume absent further instruction). See *infra* Parts I–III; cf. William Baude & Stephen E. Sachs, *The Law of Interpretation*, 130 HARV. L. REV. 1079, 1134 (2017). On similar construal of "ordinary meaning" in scholarly statutory interpretation literature, see SLOCUM, *supra*, at ch. 2 (surveying scholarly literature concerning ordinary meaning in interpretation). On the meaning of "plain meaning," i.e., "obvious" or "clear" meaning, and the "plain meaning rule," see William Baude & Ryan D. Doerfler, *The (Not So) Plain Meaning Rule*, 84 U. CHI. L. REV. 539, 541 (2017) ("The plain meaning rule says that otherwise-relevant information about statutory meaning is forbidden when the statutory text is plain or unambiguous.").

11. *Infra* Section IV.B.3.a.

12. *Infra* Section IV.B.3.c.

13. *Infra* Section IV.B.3.b.

majority and dissenting Justices—along with the many other judges and legal theorists similarly divided over these issues—appear to share the survey participants’ outlook: nearly everyone thinks that his or her own interpretation is shared by most people, even when it is not.¹⁴

In light of these findings, the Article considers the role survey experiments might usefully play in reconsidering the causation standards addressed in recent statutory tort and criminal case law¹⁵ (including, most immediately, in cases approving jury instructions premised on the Court’s misleading pronouncements about ordinary meaning and causation).¹⁶ The Article goes on to consider the role such experimental

14. *Infra* Section IV.B.3.d.

15. In arguing for the relevance of experimental methodologies in legal interpretation, the Article builds on important recent work advocating the use of surveys and experiments in contract interpretation. See Omri Ben-Shahar & Lior Jacob Strahilevitz, *Interpreting Contracts Via Surveys and Experiments*, 92 N.Y.U. L. REV. 1753 (2017). Ben-Shahar and Strahilevitz convincingly demonstrate that lay survey respondents are capable of providing reliable and consistent interpretations of vague language in consumer contracts. *Id.* at 1766–1801.

But to be clear, while Ben-Shahar and Strahilevitz advocate fully “outsourcing” or “delegating” to survey participants the determination of *contractual* meaning—thereby effectively “oust[ing] interpretations based on non-textual approaches,” *id.* at 1802, 1808–09—the present Article does not argue for an analogous wholesale outsourcing approach to *statutory* meaning determinations, nor does it argue in favor of ridding *statutory* interpretation of “non-textual” considerations. (Indeed, this Article does not contend that “textualist” considerations like ordinary and plain meaning ought to receive any greater weight than they already receive vis-à-vis “non-textual” sources of statutory meaning. See, e.g., *infra* Section IV.C. See generally William N. Eskridge, Jr., *The New Textualism and Normative Canons*, 113 COLUM. L. REV. 531, 532 (2013); Fallon, *supra* note 10.)

In any event, although this Article is to my knowledge the first published piece to argue for the relevance of experimental methodologies in statutory interpretation, an excellent unpublished student paper argues for the use of surveys to find ordinary meaning. See J.P. Sevilla, *Measuring Ordinary Meaning Using Surveys* (Sept. 10, 2014) (unpublished working paper), <https://ssrn.com/abstract=2466667> [<https://perma.cc/CPX3-27QX>]. That paper, unlike Ben-Shahar and Strahilevitz’s article and unlike this Article, does not use an experimental design to test survey respondents’ sensitivity to alterations in the language being interpreted. See *infra* Sections IV.B.1–3; Ben-Shahar & Strahilevitz at 1766–1801. And an early review of Larry Solan’s *The Language of Judges* includes discussion of a survey questionnaire administered to 116 university students and thirty-seven judges for purposes of interpreting the word “enterprise” as used in the federal RICO statute. Clark D. Cunningham, Judith N. Levi, Georgia M. Green & Jeffrey P. Kaplan, *Plain Meaning and Hard Cases*, 103 YALE L.J. 1561, 1598–1601 (1994) (concluding that the term is ambiguous in certain respects relevant to interpretation of the relevant statutory provision).

More generally, in arguing for the relevance of empirical methods in legal interpretation, this Article builds on the recent and rapidly expanding literature concerning law and corpus linguistics. See *infra* Section III.B. Although this Article is somewhat critical of certain aspects of that literature, it shares that literature’s goal of rendering judicial claims about ordinary and plain meaning less suppositious and more readily amenable to constructive and transparent debate.

16. See, e.g., cases discussed *supra* notes 6–9 and accompanying text (relying on *Burrage* to uphold CSA convictions premised on jury instructions lacking explicit but-for instructions).

evidence might play in resolving other interpretive disputes that arise in criminal and tort law more generally (e.g., in applying substantive canons and interpretive doctrines like lenity and void-for-vagueness, as well as standards for reasonable mistake of law excuses and qualified immunity).¹⁷

While the idea of treating survey experiments like the ones reported here as relevant in tort and criminal litigation may initially seem like a radical departure from current legal practice, it is worth noting that the use of such experiments in these areas is already more common than many realize, though it goes on entirely behind closed doors: in mass torts, products liability, and white-collar and corporate criminal matters, wealthy and repeat-player litigants and their firms sometimes undertake and collect survey-based research (the results of which are nonpublic and confidential) for their own use in making strategic decisions from forum selection and settlement through trial and jury instruction proposals.¹⁸ These litigants are thus able to observe, for example, how people interpret statutory causation language compared to “but-for,” “contributing factor,” and “substantial factor” causation instructions.¹⁹ In short, then, survey-based empirical research already plays a role in shaping litigation even in the context of statutory causation requirements, just not necessarily a visible role or one that benefits most litigants.²⁰ To slightly revise the slogan form of this Article’s approach: to find public meaning, ask the public—then make the results public.

Moving from legal practice to theory, the Article argues that legal theorists and philosophers, like judges and litigants, can usefully draw on empirical data concerning ordinary language and common intuitions. First, experimental research concerning concepts like “causation,” “reasonableness,” etc. can help flesh out prominent theoretical accounts of common law adjudication associated with American Legal Realism, on the one hand, and self-styled New Doctrinalists on the other.²¹ Second, similar empirical data can usefully aid in the philosophical analysis of legal concepts, at least as such analysis is conducted in much of modern analytic

17. See *infra* Section V.A.

18. Interviews with Anonymous Jury and Trial Consultants and Anonymous Lawyers (2016–2017) (notes on file with author) [hereinafter Interviews]. To be sure, one frequent use of survey data in litigation is publicly visible and well known: in several substantive areas of law, most prominently trademarks and unfair competition cases concerning consumer confusion, parties routinely present survey evidence to the court via expert witnesses. See Ben-Shahar & Strahilevitz, *supra* note 15.

19. See Interviews, *supra* note 18.

20. There is surprisingly little publicly available information about the methods and practices of jury and trial consulting firms. The few scholarly treatments of the topic focus exclusively on jury consultants’ role in selecting individual jurors during voir dire. See Andrew Guthrie Ferguson, *The Big Data Jury*, 91 NOTRE DAME L. REV. 935 (2016); Franklin Strier & Donna Shestowsky, *Profiling the Profilers: A Study of the Trial Consulting Profession, Its Impact on Trial Justice and What, If Anything, To Do About It*, 1999 WISC. L. REV. 441. As is apparent from publicly available business profiles, these firms are already numerous and oft-used, and they continue to grow as big data technologies become cheaper and more informative. In any event, it is no secret that these firms provide services—often including sophisticated empirical research—that litigants and firms use to make strategic litigation decisions far beyond voir dire.

21. See *infra* Section V.B.

legal philosophy.²² For example, despite many important differences, the two preeminent modern philosophical works on causation in torts and criminal law—H.L.A. Hart and Tony Honore’s *Causation in the Law*, and Michael Moore’s *Causation and Responsibility*²³—both frequently appeal to allegedly common linguistic, conceptual, and moral intuitions to support their arguments (often using the same sorts of intuition-pumping examples from ordinary language that permeate modern statutory interpretation case law).²⁴ Philosophers’ contentions regarding common intuitions, like judges’, ought to be tested using a larger sample size than the author him- or herself. And here again, although conducting experiments might initially sound like a radical departure from current practice, it is more common in contemporary philosophy than many realize,²⁵ even if it is not (yet) common in philosophy of law.²⁶

The Article proceeds as follows. Part I examines recent case law concerning the ordinary meaning of statutory causation language and the “commonsense,” or “folk,” concept²⁷ of causation. Parts II and III, while continuing to use these cases as illustrative examples, address two more fundamental issues underlying courts’ ordinary meaning interpretation. Part II explains *why* courts do, should, and inevitably must care about ordinary meaning in legal interpretation, especially in criminal and tort law, and especially in construing core common law concepts like causation. Part III then examines *how* judges ascertain the ordinary meaning of statutory language and the ordinary conceptual understandings it reflects. Judges’ usual tools—introspection, intuition-pumping examples, and dictionaries, as well as

22. See *infra* Section V.C.

23. H.L.A. HART & TONY HONORÉ, *CAUSATION IN THE LAW* (1959); MICHAEL MOORE, *CAUSATION AND RESPONSIBILITY* (2009); cf. *Burrage v. United States*, 571 U.S. 204, 210 (2014) (citing HART & HONORÉ, *supra*, at 104).

24. See, e.g., *infra* notes 308–10.

25. For a general introduction, see Joshua Knobe & Shaun Nichols, *Experimental Philosophy*, in *THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY* (2017).

26. See Alex Langlais & Brian Leiter, *The Methodology of Legal Philosophy*, in *THE OXFORD HANDBOOK OF PHILOSOPHICAL METHODOLOGY* (H. Cappelen, T. Gendler & J. Hawthorne eds., 2013); BRIAN LEITER, *NATURALIZING JURISPRUDENCE: ESSAYS ON AMERICAN LEGAL REALISM AND NATURALISM IN LEGAL PHILOSOPHY* 4 (2007) [hereinafter LEITER, *NATURALIZING JURISPRUDENCE*]; cf. Brian Leiter, *Naturalizing Jurisprudence: Three Approaches*, in *THE FUTURE OF NATURALISM* 9 (J. Shook & P. Kurtz eds., 2009) [hereinafter Leiter, *Three Approaches*].

27. By “concept” and “folk concept,” I mean what Richard Fallon, following Frank Jackson, meant by it, as explained in Fallon, *supra* note 10: “[T]he term ‘concept’ refers to ‘the possible situations covered by the words we use to ask our questions.’ . . . A folk concept, roughly speaking, is one rooted in the understandings and usages of ordinary people. This assumption makes linguistic intuitions relevant because ‘[i]n as much as [one’s] intuitions are shared by the folk, they reveal the folk theory’ that presumptively defines a folk concept’s extension.” *Id.* at 1254 n.66 (emphasis in original) (internal citation omitted) (quoting FRANK JACKSON, *FROM METAPHYSICS TO ETHICS: A DEFENCE OF CONCEPTUAL ANALYSIS* 33, 37 (1998)).

Importantly, I adopt that account because it tracks what the courts, in the decisions discussed below and elsewhere, appear to adopt when they discuss “ordinary,” “commonsense” concepts, including factual causation.

recently proposed corpus linguistics methodologies—all run into serious problems of bias, inaccuracy, and indeterminacy. Part IV therefore proposes the survey experiment method, explains why it holds promise as an improvement over other traditional methods, and demonstrates it in action. It shows that courts have often reached the wrong outcomes (according to their own stated ordinary language criteria) in recent statutory causation case law. The Article then briefly considers possible explanations for why they arrived at the particular wrong answers they reached, and what normative lessons we might draw with respect to causation doctrine and with respect to textualist statutory interpretation more broadly. Part V considers further implications and potential extensions of the experimental methodology.

I. COURTS ON THE FOLK CONCEPT OF CAUSATION AND ORDINARY USAGE

Section I.A will provide a brief survey of a series of recent cases construing statutory causation requirements. In these cases, the Supreme Court (and other state and federal courts following its lead) has consistently found that the ordinary, plain meaning of causal language, and the common, everyday concept of causation it invokes, entails but-for causation.²⁸ But while the Justices in the majority derived this conclusion from seemingly commonsense intuitions about the folk concept of causation, intuition-pumping examples of uncontroversial everyday linguistic usage, and dictionary definitions objectively attesting to the ordinary meaning of the relevant statutory language, some other courts and commentators (as well as some dissenting Justices) have derived a variety of contrary conclusions from many of these same resources. Section I.B therefore examines a few types of cases that arguably pose problems for the Court's but-for entailment thesis. It then briefly describes three alternative causation tests that various courts and commentators—despite disagreeing amongst themselves about the proper scope and interpretation of each test—nonetheless claim more closely track the folk concept of causation and the ordinary meaning of causal language than does the Court's but-for test.

A. *The Triumph of the But-For Test in Recent Ordinary Meaning Case Law*

1. Statutory Torts

Mixed-motivation discrimination cases, concerning what courts and scholars often label “statutory torts,”²⁹ comprise some of the Supreme Court's most important

28. More specifically, in these cases the Court has held that this language entails *at least* but-for causation. These cases concern “factual” causation (as opposed to “proximate” or “legal” causation) as does this Article except where otherwise noted. For an examination of recent statutory case law concerning ordinary meaning and *proximate* causation requirements, see Sandra F. Sperino, *Statutory Proximate Cause*, 88 NOTRE DAME L. REV. 1199 (2013).

29. See Charles A. Sullivan, *Tortifying Employment Discrimination*, 92 B.U. L. REV. 1431, 1431 (2012) (“Title VII is often described as a ‘statutory tort’”); W. Jonathan Cardi, *The Role of Negligence Duty Analysis in Employment Discrimination Cases*, 75 OHIO ST. L.J. 1129, 1129 (2014) (“The tortification of employment discrimination law has been thoroughly documented and theorized.”) (citing examples); Sandra F. Sperino, *The Tort Label*,

ordinary meaning causation cases. “Motive,” at least as the courts have consistently treated it, “is a causal concept”³⁰; it concerns somebody’s reason for acting—i.e., something that made them decide to (caused them to) do something. Causal language comprises some of the most consequential clauses of Title VII,³¹ the Age Discrimination in Employment Act (ADEA),³² the Americans with Disabilities Act (ADA),³³ the Fair Housing Act (FHA),³⁴ and other modern anti-discrimination and civil rights laws.³⁵

Consider the ADEA, which makes it unlawful to, *inter alia*, discharge an employee “because of” the employee’s age. In *Gross v. FBL Financial Services*, the Supreme Court held that the term “because of” in the ADEA entails “but-for” causation.³⁶ In other words, to prevail on an age discrimination claim, plaintiffs must show that, but for the employer’s consideration of the employee’s age, the employer would not have discharged the employee. In so holding, the Court’s opinion focused entirely on “the ordinary meaning” of the phrase “because of.”³⁷

To support its ordinary meaning determination, the *Gross* Court first turned to dictionaries, quoting two definitions of the phrase “because of” (each of which included among its definitions “by reason of” and “on account of,”)³⁸ as well as one definition of the word “because” (defined as “by reason; on account”).³⁹ In light of these definitions, the Court explained, “[T]he plain language of the ADEA” (i.e., the phrase “because of”) requires that age be “the ‘but-for’ cause of the employer’s adverse decision.”⁴⁰ The opinion next turned to the Court’s own precedent interpreting the ordinary meaning of similar causal language found in other statutes across various substantive contexts, observing, for example, that, “[i]n common talk, the phrase ‘based on’ indicates a but-for causal relationship,” and that “based on”

66 FLA. L. REV. 1051, 1070 (2014). For more on civil rights and discrimination law as statutory torts, see, e.g., the *Ohio State Law Journal*’s recent symposium: Martha Chamallas & Sandra F. Sperino, *Torts and Civil Rights Law: Migration and Conflict*, 75 OHIO ST. L. REV. 1021 (2014).

30. D. Don Welch, *Removing Discriminatory Barriers: Basing Disparate Treatment Analysis on Motive Rather than Intent*, 60 S. CAL. L. REV. 733, 739 (1987); see also Andrew Verstein, *The Jurisprudence of Mixed Motives*, 127 YALE L.J. 1106, 1117 (2018) (stating that in mixed motive case law, the “causation-focused approach,” i.e., “tort causation analysis, the notion that plaintiffs must show that their injury was caused by the defendant’s illicit motives[,] has been terrifically influential” (emphasis in original)).

31. 42 U.S.C. § 2000 (2012).

32. 29 U.S.C. §§ 621–34 (2012).

33. 42 U.S.C. §§ 12101–12213 (2012).

34. 42 U.S.C. § 3604 (2012).

35. See, e.g., cases cited *supra* notes 5–9, *infra* note 47; sources cited *supra* notes 29–30, *infra* note 69.

36. 557 U.S. 167, 176 (2009).

37. *Id.* at 175–76.

38. *Id.* at 176 (quoting 1 WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 194 (1966); 1 OXFORD ENGLISH DICTIONARY 746 (1933)).

39. *Id.* (quoting THE RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE 132 (1966)).

40. *Id.*

has the same ordinary meaning as “because of.”⁴¹ In short, as a matter of ordinary understanding and linguistic usage, the phrase “because of” entails but-for causation, and the ADEA therefore does too.⁴²

In *University of Texas Southwestern Medical Center v. Nassar*, the Court “returned again to the meaning of ‘because,’” here addressing Title VII’s anti-retaliation provision.⁴³ As in *Gross*, where the Court had “[c]oncentrat[ed] first and foremost on the meaning of the phrase ‘because of’” in ordinary usage, the Court again considered, among other things, the “‘plain textual meaning’ of the word ‘because’”⁴⁴ and claimed that it entails but-for causation,⁴⁵ citing the earlier plain meaning causation cases on which the *Gross* Court relied.⁴⁶

And *Gross* and *Nassar* are just the tip of the iceberg: In cases concerning the ADA, the FHA, and numerous other “statutory torts,” federal circuit courts—relying on their own intuitions, along with the occasional dictionary definition—have likewise determined that terms like “because” and “because of” entail but-for

41. *Id.* (quoting *Safeco Ins. Co. of Am. v. Burr*, 551 U.S. 47, 63–64 & n.14 (2007)); *see also id.* (quoting *Bridge v. Phx. Bond & Indem. Co.*, 553 U.S. 639, 652–55 (2008), which “recogniz[ed] that the phrase, ‘by reason of,’ requires at least a showing of ‘but for’ causation”); *id.* at 176–77 (“‘An act or omission is not regarded as a cause of an event if the particular event would have occurred without it.’” (quoting W. PAGE KEETON, DAN B. DOBBS, ROBERT E. KEETON & DAVID G. OWEN, PROSSER AND KEETON ON LAW OF TORTS 265 (5th ed. 1984)); *see also Price Waterhouse v. Hopkins*, 490 U.S. 228, 281–82 (1989) (Kennedy, J., dissenting) (“By any normal understanding, the phrase ‘because of’ conveys the idea that the motive in question made a difference to the outcome. We use the words this way in everyday speech. . . . Any standard less than but-for . . . simply represents a decision to impose liability without causation.”); *id.* at 282 (Kennedy, J., dissenting) (stating the same with respect to “result from,” “based on,” and “made on the basis of”).

42. *Gross v. FBL Fin. Servs., Inc.*, 557 U.S. 167, 176–77 (2009). Justices Stevens and Breyer wrote separate dissents. Justice Stevens deemed Thomas’s dictionary definitions a *non sequitur*: the fact that “dictionaries define ‘because of’ as ‘by reason of’ or ‘on account of[.]’ [c]ontrary to the majority’s bald assertion . . . does not establish that the term denotes but-for causation.” *Id.* at 183 n.4 (Stevens, J., dissenting); *see also id.* at 183 n.4 (“The dictionaries the Court cites do not, for instance, define ‘because of’ as ‘solely by reason of’ or ‘exclusively on account of.’” (emphasis in original)); *id.* at 191 (“[T]he fact that a jury has found that age did play a role in the decision justifies the use of the word ‘because.’”). And Justice Breyer likewise contended that, as a matter of ordinary language, “[t]he words ‘because of’ do not inherently require a showing of ‘but-for’ causation.” *Id.* at 190 (Breyer, J., dissenting).

43. 570 U.S. 338, 349 (2013).

44. *Id.* at 350, 362 (emphasis in original) (citing *Gross*, 557 U.S. at 176–77).

45. *Id.* at 350 (“Concentrating first and foremost on the meaning of the phrase “‘because of . . . age,’” the Court in *Gross* explained that the ordinary meaning of “‘because of’” is “‘by reason of’” or “‘on account of.’”); *id.* (noting that in *Gross*, “the ‘requirement that an employer took adverse action ‘because of’ age [meant] that age was the “reason” that the employer decided to act,’ or, in other words, that ‘age was the “but-for” cause of the employer’s adverse decision.’” (emphasis added) (quoting *Gross*, 557 U.S. at 176)).

46. *Id.* (citing *Safeco Ins. Co. of Am. v. Burr*, 551 U.S. 47, 63–64 & n. 14 (2007) (noting that “because of” means “based on” and that “‘based on’ indicates a but-for causal relationship”); *Holmes v. Sec. Inv’r Prot. Corp.*, 503 U.S. 258, 265–66, (1992) (equating “by reason of” with “‘but for’ cause”)).

causation as a matter of ordinary usage, and that plaintiffs must therefore allege and prove but-for causation.⁴⁷

2. Criminal Law

Moving now from statutory torts to criminal law,⁴⁸ these same ordinary causal language issues recently resurfaced in *Burrage v. United States*, involving interpretation of the CSA.⁴⁹ The defendant, Burrage, had sold heroin to Josh Banka, who died of “mixed drug intoxication” shortly after injecting it.⁵⁰ The statute subjects heroin sellers to enhanced sentences “if death or serious bodily injury *results from* the use of such substance.”⁵¹ Because Banka had taken multiple drugs, experts were unable to testify that Banka would not have died “but for” his use of the heroin Burrage supplied.⁵² Burrage was convicted under the enhanced sentencing provision nonetheless.⁵³

The Supreme Court reversed, holding that the enhanced sentence applies only where the substance the defendant provided was a but-for cause of death.⁵⁴ Since the statute “does not define the phrase ‘results from,’” the Court explained, “we give it its ordinary meaning.”⁵⁵ and in “common talk,” phrases like “results from,” “because of,” and “by reason of . . . indicate[] a but-for causal relationship.”⁵⁶ This but-for requirement, the Court explained, “is part of the common understanding of cause,”⁵⁷

47. See, e.g., *Gentry v. E. W. Partners Club Mgmt. Co.*, 816 F.3d 228, 235–36 (4th Cir. 2016) (analyzing the language of the ADA); *United States v. Pickarsky*, 687 F.3d 134, 143–44 (3d Cir. 2012) (analyzing the language of the FHA); *Bolmer v. Oliveira*, 594 F.3d 134 (2d Cir. 2010) (analyzing the language of the ADA); *Serwatka v. Rockwell Automation, Inc.*, 591 F.3d 957, 962 (7th Cir. 2010) (same); *United States v. Craft*, 484 F.3d 922, 926 (7th Cir. 2007) (analyzing the language of the FHA) (citing *United States v. Magleby*, 241 F.3d 1306, 1310 (10th Cir. 2001) (same)); *United States v. Johns*, 615 F.2d 672 (5th Cir. 1980) (same); *United States v. Ellis*, 595 F.2d 154 (3d Cir. 1979) (construing 18 U.S.C. § 241, which criminalizes conspiracies targeting individuals “because of [their] having exercised” certain civil rights); see also William N. Eskridge, Jr., *Reneging on History? Playing the Court/Congress/President Civil Rights Game*, 79 CALIF. L. REV. 613, 677–79 (1991) (discussing textualism in civil rights cases).

48. Courts and commentators traditionally treat factual causation identically in the criminal and torts (including statutory torts) context, frequently drawing from both areas in the same decision. See, e.g., *Burrage v. United States*, 571 U.S. 204, 214 (2014); *People v. Jennings*, 237 P.3d 474, 496 n.13 (Cal. 2010) (“It is well established that the principles of causation, as they apply to tort law, are equally applicable to criminal law.”) (citing cases); *State v. Burton*, 370 S.W.3d 926, 931 n.3 (Mo. Ct. App. 2012); *State v. Christman*, 249 P.3d 680, 687 (Wash. Ct. App. 2011); MOORE, *supra* note 23.

49. 571 U.S. at 204.

50. *Id.* at 207.

51. *Id.* at 206 (emphasis added); see 21 U.S.C. § 841(b)(1)(C) (2012).

52. 571 U.S. at 207.

53. *Id.* at 208.

54. *Id.* at 219.

55. *Id.* at 210.

56. *Id.* at 213.

57. *Id.* at 211.

as revealed by examples from ordinary usage—e.g., what “is natural to say” and what “makes little sense to say” about the “cause” of a baseball team having won a game.⁵⁸ The Court additionally cited a dictionary definition of “results,”⁵⁹ discussed its own prior cases from various substantive contexts concerning the ordinary meaning of other causal language,⁶⁰ and explained that the government’s “policy discussions are beside the point. The role of this Court is to apply the statute as it is written.”⁶¹

An important post-*Burrage* case brings us back, as in *Gross* and *Nassar*, to motives as causes, this time in the context of the federal Hate Crimes Prevention Act (HCPA).⁶² In *United States v. Miller*, the Sixth Circuit considered whether the HCPA, which, *inter alia*, provides enhanced sentences for attacking someone “because of” that person’s religion, requires a showing of but-for causation.⁶³ The Sixth Circuit held that it does, on grounds of plain, ordinary meaning, and therefore reversed the conviction.⁶⁴ Judge Sutton, writing for the majority, provided his own ordinary meaning analysis: “In everyday usage, the phrase ‘because of’ indicates a but-for causal link between the action that comes before it and the circumstance that comes afterwards. John carried an umbrella because of the rain. Jane stayed home from school because of her fever.”⁶⁵ He quoted dictionary definitions for support: “Dictionary definitions of the phrase reflect this common-sense understanding: ‘Because of’ means ‘by reason of’ or ‘on account of.’”⁶⁶ And he cited the Supreme Court’s and other courts’ recent ordinary meaning cases construing similar phrases,⁶⁷ before concluding once again that “‘because of’ in brief means what it says: The

58. *Id.* at 212.

59. *Id.* at 210 (“A thing ‘results’ when it ‘[a]rise[s] as an effect, issue, or outcome *from* some action, process or design.” (emphasis in original) (quoting 2 THE NEW SHORTER OXFORD ENGLISH DICTIONARY 2570 (1993))). The Court therefore used the phrases “caused by” and “resulting from” interchangeably in the opinion. *E.g., id.* (“death caused by (‘resulting from’) the use of that drug”).

60. *Id.* at 212–13.

61. *Id.* at 218. In a one-paragraph concurrence, Justice Ginsburg (joined by Justice Sotomayor) reiterated that she “do[es] not read ‘because of’ in the context of antidiscrimination laws to mean ‘solely because of.’” *Id.* at 219. Nonetheless, because “there is room for debate” in interpreting the words “results from,” Ginsburg applied the rule of lenity and joined the majority’s judgment. *Id.*

62. See 18 U.S.C. § 249 (2012 & Supp. 2018).

63. 767 F.3d 585 (6th Cir. 2014).

64. *Id.* at 589. Defendants had argued that, at most, religion-based animosity was one among several motivations for their assaults. *Id.* at 590–91. But they had been convicted under a jury instruction that stated that the statute’s causation requirement could be met if the victims’ faith was a “significant motivating factor” in the assaults. *Id.* at 591. At least prior to *Burrage*, a similar standard and jury instruction—“substantial factor” or “substantial motivating factor”—was the most common instruction for federal statutes with a motive element. See LEONARD B. SAND, JOHN S. SIFFERT, WALTER P. LOUGHLIN, STEVEN A. REISS, STEVE ALLEN, JED S. RAKOFF & DAVID M. EPSTEIN, MODERN FEDERAL JURY INSTRUCTIONS chs. 17.02, 17.06, 17.15, 17.29 (2018) (collecting cases).

65. *Miller*, 767 F.3d at 591.

66. *Id.* (quoting WEBSTER’S SECOND NEW INTERNATIONAL DICTIONARY 242 (1950); OXFORD ENGLISH DICTIONARY (2012)).

67. See *id.* 592–93.

prohibited act or motive must *be* an actual *cause* of the specified outcome” (by which he meant a “but-for” cause).⁶⁸

Again, *Burrage* and *Miller*, like *Gross* and *Nassar*, are only the tip of the iceberg. Other courts, interpreting federal and state statutes, jury instructions, and even judge-made common law standards, continue to expand the reach of the Court’s reasoning concerning the commonsense meaning of these causal phrases.⁶⁹

B. Potential Problems and Roads Not Taken

A quick skim of courts’ intuition-pumping examples (“John carried an umbrella because of the rain,” etc.⁷⁰) might convince one that these courts have it right; perhaps, one might think, terms like “because of” and “results from” really do entail but-for causation as a matter of ordinary usage. But there may be problems with these courts’ but-for entailment thesis. There may be cases where we know that *x* is not a but-for cause of *y*, yet at least some of us feel the urge to say that *y* happened “because of” *x*, that *y* “resulted from” *x*, and so forth. Even if you don’t share the urge, it is worth understanding where at least some courts and commentators feel it. Here, we’ll look at two categories of such cases (specifically, for those familiar with the jargon, two common kinds of “overdetermination” cases).⁷¹ Then we’ll examine a few alternative causation tests (the “contributing factor,” “substantial factor,” and “sole

68. *Id.* at 591–92 (emphasis in original); *accord, e.g.*, *United States v. Doggart*, No. 1:15-cr-39, 2016 WL 6205804, at *6 (E.D. Tenn. Oct. 24, 2016); *United States v. Metcalf*, No. 15-CR-1032-LRR, 2016 WL 1599485, at *3 (N.D. Iowa Apr. 20, 2016).

69. *See, e.g.*, cases cited *supra* notes 5–9, 47. Time will tell what additional statutes and doctrines courts will import these ordinary language holdings into next. Here, for example, are five relatively prominent possibilities: (1) criminal defense of provocation, *see, e.g.*, *United States v. Begay*, 673 F.3d 1038, 1045 (9th Cir. 2011) (“The defendant must produce evidence ‘that the defendant was acting out of passion *rather than* malice’ before the burden shifts to the government to ‘prove beyond a reasonable doubt the *absence* of sudden quarrel or heat of passion.’” (emphasis added)); (2) criminal defense of self-defense, *see* WAYNE R. LAFAYE, *CRIMINAL LAW* 394 (5th ed. 2012); (3) whistleblower retaliation claims, *see* Nancy M. Modesitt, *Causation in Whistleblowing Claims*, 50 U. RICH. L. REV. 1193 (2016); (4) securities fraud and other fraud claims, *see* Jill E. Fisch, *Cause for Concern: Causation and Federal Securities Fraud*, 94 IOWA L. REV. 811, 813 (2009); David A. Fischer, *Insufficient Causes*, 94 KY. L.J. 277, 297–98 (2005); (5) civil rights claims against individuals and municipalities under 42 U.S.C. § 1983 (2012 & Supp. 2018), *see* Barbara Kritchevsky, “Or Causes to Be Subjected”: *The Role of Causation in Section 1983 Municipal Liability Analysis*, 35 UCLA L. REV. 1187 (1988); Sheldon Nahmod, *Mt. Healthy and Causation-in-Fact: The Court Still Doesn’t Get It!*, 51 MERCER L. REV. 603 (2000); Hillel Levin & Michael Wells, *Qualified Immunity and Statutory Interpretation: A Response to William Baude*, CAL. L. REV. ONLINE (forthcoming) (unpublished manuscript on file with the *Indiana Law Journal*) (noting that, while the Supreme Court rarely turns to the text of 42 U.S.C. § 1983 to resolve cases brought pursuant to it, one exception is where the statute’s “subject, or cause to be subjected” phrase is at issue, as in *Monell v. Department of Social Services of New York*, 436 U.S. 658, 691–92 (1978)).

70. *E.g.*, *Miller*, 767 F.3d at 591; *see also* *Burrage v. United States*, 571 U.S. 204, 215 (2014); *Gross v. FBL Fin. Servs., Inc.*, 557 U.S. 167, 174 (2009).

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

cause” tests) that some courts and commentators claim more closely track commonsense causal intuitions, in these and other cases, than does the but-for test.

1. Unnecessary “Causes”?

Here is one situation where but-for causation may not capture our ordinary notions of causation: the focal event or motive is unnecessary to the outcome in question (so it is not a but-for cause) but it would also have been fully sufficient, all by itself, to bring about the outcome in question, i.e., absent whatever other (non-focal) events or motives happened to play a role. For example, imagine two shooters, Al and Ben: each simultaneously shoots the same victim in the head; the victim would have died from either shot by itself; the prosecutor charges Al with murder. But did Al cause the victim’s death? That hypothetical may seem far-fetched, but there are numerous real-life examples of multiple perpetrators generating a legally cognizable harm where their individual actions were unnecessary and independently sufficient to bring about the harm (and, in motive-as-cause cases, there are likewise numerous real-life examples where an actor has multiple motivations for his or her action and each motivation is unnecessary and independently sufficient to act).⁷² Indeed, all the cases above are easily turned into examples.⁷³ And while there are examples from drug overdose,⁷⁴ discrimination,⁷⁵ and hate crime⁷⁶ cases like the ones described

72. See *id.* (“Contrary to what proponents of the [but-for] criterion as the exclusive criterion [for causation] often assume, instances of overdetermined causation are not rare.”) (discussing examples); see also Eric A. Johnson, *Wrongful-Aspect Overdetermination: The Scope-of-the-Risk Requirement in Drunk-Driving Homicide*, 46 CONN. L. REV. 601, 644 (2013). Some common examples that have recently received attention involve acquittals of multiple police officers for their simultaneous use of force on a single victim. See, e.g., Ben Gifford, *State v. Brelo and the Problem of Actual Causation*, 44 AM. J. CRIM. L. 157, 157 (2017) (discussing case in which “the judge relied explicitly on the theory that the lethality of the gunshot wounds inflicted by each officer rendered every other officer’s gunshot wounds unnecessary to the victims’ deaths”); Barbara A. Spellman & Alexandra Kincannon, *The Relation Between Counterfactual (“But For”) and Causal Reasoning: Experimental Findings and Implications for Jurors’ Decisions*, 64 LAW & CONTEMP. PROBS. 241, 256–57 (2001) (discussing acquittal of four police officers who shot single victim forty-one times). Prosecutors sometimes avoid these causation problems by charging one of the relevant actors as a direct cause and charging the others as accomplices or co-conspirators. Although that charging decision is often available where multiple actors concurrently cause harm, it is not available in motive-as-cause cases (i.e., cases concerning the causal role of a single actor’s multiple motivations).

73. Cf. *infra* Part IV.

74. E.g., *United States v. Smith*, No. 6:13–43–KKC, 2015 WL 4458891 (E.D. Ky. July 21, 2015).

75. E.g., *Hendon v. Kamtek, Inc.*, 117 F. Supp. 3d 1325, 1332 (N.D. Ala. 2015); cf. Rebecca Hanmer White & Linda Hamilton Krieger, *Whose Motive Matters?: Discrimination in Multi-Actor Employment Decision Making*, 61 LA. L. REV. 495 (2001).

76. See, e.g., *State v. Hennings*, 791 N.W.2d 828, 833–35 (Iowa 2010) (construing state hate crime statute in case where driver was prosecuted for allegedly running over African American person both “because” she was in the street and “because” she was African American).

above, there are also others throughout criminal and tort law.⁷⁷ Courts treat these cases inconsistently.⁷⁸ Some hold fast, maintaining the requirement of but-for causation and finding no liability.⁷⁹ Others find causation requirements satisfied, often invoking “a kind of legal fiction or construct” or a “policy based departure” from what they consider to be the plain, commonsense meaning of the relevant causal language.⁸⁰

A second possible problem area for the but-for test involves factors that were neither necessary *nor independently sufficient* for bringing about the result in question, yet still played some role—contributed in some way—to its coming about. Consider Banka’s death in the *Burrage* case: if the heroin was unnecessary (the other drugs would have killed Banka by themselves), *and independently insufficient* (absent the other drugs, Banka would have lived), would we say that the use of the heroin caused Banka’s death? The Article will return to this and other examples in Part IV, but for now, the important point is that these sorts of scenarios are also pervasive in criminal and tort law, and courts once again lack a unified approach. Sometimes courts find insufficient and unnecessary contributors like these to be

77. See Gerald W. Boston, *Apportionment of Harm in Tort Law: A Proposed Restatement*, 21 U. DAYTON L. REV. 267 (1996) (providing examples).

78. See MOORE, *supra* note 23, at 115, 427–38 (discussing cases in criminal law and torts and noting different approaches both within and between U.K. and U.S. law).

79. See, e.g., *Hendon*, 117 F. Supp. 3d at 1332 (reasoning that employment discrimination plaintiff, by alleging multiple independently sufficient motivations on the part of defendant, defeated her own causation claims); *Donald v. UAB Hosp. Mgmt., LLC*, No. 2:14-cv-727-WMA, 2015 WL 3952307, at *2–3 (N.D. Ala. June 29, 2015) (same); Gifford, *supra* note 72; Spellman & Kincannon, *supra* note 72; see also *Burrage v. United States*, 571 U.S. 204, 215 (2014) (noting that the Model Penal Code (MPC) requires necessary (not merely sufficient) cause, while some state courts instead use “substantial” or “contributing” factor rules, and noting the debate that occurred during the MPC’s drafting over whether to instead adopt the “substantial factor” test); MODEL PENAL CODE § 2.03(1)(a) (AM. LAW INST., Proposed Official Draft 1985); MODEL PENAL CODE § 2.03 cmt., 2 at 259 (AM. LAW INST. 1985) (explaining the MPC’s attempted preservation of liability in overdetermination scenarios via a but-for test that asks whether the forbidden result would have arisen “when and as it did” absent the factor in question); AM. LAW INST., 39TH ANNUAL MEETING PROCEEDINGS 135–41 (1962); JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 187–88 (2018) (arguing in favor of this approach); PAUL ROBINSON, INTUITIONS OF JUSTICE AND THE UTILITY OF DESERT 392 (2013) (MPC drafters “apparently thought they had to preserve the necessary cause test because the necessary cause test was so well entrenched in the popular psyche”); Eric A. Johnson, *Cause-in-Fact After Burrage v. United States*, 68 FLA. L. REV. 1727, 1736 (2016) (noting that the MPC’s “when and as it did” approach “never really caught on with judges”).

80. *Paroline v. United States*, 572 U.S. 434, 452 (2014); see also *id.* at 458 (citing *Burrage*, 571 U.S. at 212); *Wright & Puppe*, *supra* note 71, at 474 (citing cases and commentaries). In *Burrage*, the majority explicitly declined to address whether some such departure from plain meaning would be warranted in the event that the defendant sold a drug that was unnecessary but independently sufficient to bring about the user’s death. *Burrage*, 571 U.S. at 215 (“We need not accept or reject the special rule developed for these cases, since there was no evidence here that Banka’s heroin use was an independently sufficient cause of his death.”).

causal, other times not.⁸¹ Both courts and commentators are likewise divided over whether and when treating insufficient and unnecessary contributions as causes tracks folk notions of causation and ordinary usage of causal terms.⁸²

2. Alternative Causation Tests

Some courts and commentators favor alternative factual causation tests in place of the but-for test. Some do so only in circumstances like those described directly above; others do so more generally, regardless of whether those circumstances are present or considered problematic. Courts' alternative tests can be divided into the following three: (1) "contributing factor" tests; (2) "substantial factor" tests; and (3) "sole" cause tests.⁸³ (The Article will come back to these tests in Part IV below; for now, the main point is simply to see what they are, and to get a sense of courts' and commentators' widespread disagreement about them.)

The first alternative to a "but-for" standard (and the one that is most inclusive in what it deems a "cause") was explicitly rejected in *Burrage* and many of the other decisions described above: the "contributing factor" test.⁸⁴ It would label as causes both types of contributors considered above (i.e., all contributing factors regardless of whether they are independently sufficient to bring about the harm). It would, in

81. See *Burrage*, 571 U.S. at 215; Fischer, *supra* note 69, at 277–79, 284–87; RESTATEMENT (THIRD) OF TORTS § 27(i) & cmt. f (Am. Law. Inst. 2005) (“[D]ecided cases [concerning unnecessary and insufficient contributors] do not support clear rules.”); MOORE, *supra* note 23, at 115; Joshua Dressler, *Reassessing the Theoretical Underpinnings of Accomplice Liability: New Solutions to an Old Problem*, 37 HASTINGS L.J. 91, 131–32 (1985); David J. Karp, *Causation in the Model Penal Code*, 78 COLUM. L. REV. 1249, 1761 (1978); Richard W. Wright, *Causation in Tort Law*, 73 CALIF. L. REV. 1735, 1974, 1980 (1985); Richard W. Wright, *Legal Responsibility*, 40 SAN DIEGO L. REV. 1425, 1443–45 (2003); cf. RESTATEMENT (THIRD) OF TORTS § 27 cmt. f, illus. 3 (AM. LAW INST. 2005); Eric A. Johnson, *Criminal Liability for Loss of a Chance*, 91 IOWA L. REV. 59, 67–68 (2005).

82. See, e.g., RESTATEMENT (THIRD) OF TORTS § 27(i) cmt. f (AM. LAW INST. 2005) (“Intuition might suggest that the [unnecessary and insufficient contributor] is not a cause,” but “significantly, [the intuition] may not be widely shared. . . . Several scholars have tried to provide algorithms to explain these cases. Unfortunately, they do not always agree on the appropriate outcomes.”); Brief of Amici Curiae States of Alaska, Colorado, Hawai’i, Kansas, New Mexico, South Dakota, Tennessee, Wisconsin, and Wyoming in Support of Respondent, *Burrage v. United States*, 571 U.S. 204 (2014) (No. 12-7515), 2013 WL 5616723, at *18 n.5 (arguing that “[t]here is no reason to suppose . . . that the sufficient-by-itself stipulation bears on” whether conduct “seems intuitively to qualify as a cause” (citing LAFAVE, *supra* note 69, at 468)); MOORE, *supra* note 23, at 86 (claiming that non-liability in cases of unnecessary and insufficient contributors is “contrary to common intuition as well as some legal authority” (citing *People v. Lewis*, 57 P. 470 (Cal. 1899))); Fischer, *supra* note 69, at 282–83, 289 (surveying Hart and Honore’s, Richard Wright’s, and the author’s own differing intuitions concerning whether unnecessary and insufficient contributors constitute causes); David Hamer, “Factual Causation” and “Scope of Liability”: *What’s the Difference?*, 77 MODERN L. REV. 155, 172 (2014) (describing differing intuitions in cases of unnecessary and/or insufficient contribution).

83. Cf. Wright & Puppe, *supra* note 71, at 480 (citing cases and commentary).

84. E.g., *Burrage*, 571 U.S. at 216; *Gross v. FBL Fin. Servs., Inc.*, 557 U.S. 167, 176 (2009); *United States v. Miller*, 767 F.3d 585, 592 (6th Cir. 2014).

short, label every contributor a cause, no matter how small its contribution. This strikes some courts and commentators as simply too permissive to cohere with ordinary usage and intuitions regarding causation.⁸⁵ Others, however, claim that it does reflect common intuition.⁸⁶ Courts attracted to this approach but leery of its alleged overpermissiveness sometimes add qualifiers like “*significant* contributing factor.”⁸⁷ That impulse brings us to the less permissive, but also less clear, “substantial factor” test.

The second test some courts apply—the “substantial factor” test—was also explicitly rejected in *Burrage* and many of the other decisions discussed above.⁸⁸ Whereas the “contributing factor” test tells jurors, without further elaboration, that they are to consider whether *x* was a “contributing factor” in bringing about *y*, the substantial factor test, similarly without elaboration, simply instructs jurors to consider whether *x* was a “substantial factor” in bringing about *y*. As one can imagine, much turns on how people interpret the term “substantial.”⁸⁹ Courts and commentators have increasingly criticized the test for its seeming lack of clarity or predictability.⁹⁰ As the *Burrage* Court indicated, courts differ widely in their understanding of (a) what the test actually requires jurors to find, and (b) when, if ever, it more closely tracks ordinary causal intuitions than does the but-for test.⁹¹

As for *what* the test requires jurors to find, all agree that it entails more than mere contribution.⁹² But what else it requires is not clear.⁹³ Most agree that it requires a

85. *E.g.*, *Burrage*, 571 U.S. at 217; sources cited *supra* note 82.

86. *See, e.g.*, *Holsemback v. State*, 443 So. 2d 1371, 1382 (Ala. Crim. App. 1983); *Cox v. State*, 808 S.W.2d 306, 309 (Ark. 1991); *People v. Jennings*, 237 P.3d 474, 496 (Cal. 2010); *People v. Lewis*, 57 P. 470, 473 (Cal. 1899); *Commonwealth v. Osachuk*, 681 N.E.2d 292, 294 (Mass. App. Ct. 1997); *People v. Bailey*, 549 N.W.2d 325, 334 (Mich. 1996); *State v. Christman*, 249 P.3d 680, 687 (Wash. Ct. App. 2011); ROLLIN M. PERKINS & RONALD N. BOYCE, *CRIMINAL LAW* 771 (3d ed. 1982).

87. *See Miller*, 767 F.3d at 592 (citing *Burrage*, 571 U.S. at 218) (faulting such language for its lack of precision); *see also Burrage*, 571 U.S. at 217 (citing *Wilson v. State*, 24 S.W. 409, 410 (Tex. Crim. App. 1893) (“contributed *materially*”) (emphasis added)).

88. *See, e.g.*, *Burrage*, 571 U.S. at 216.

89. *See David A. Fischer, Causation in Fact in Omission Cases*, 1992 UTAH L. REV. 1335, 1347.

90. *See Burrage*, 571 U.S. at 217; *see also Miller*, 767 F.3d at 592. After the First and Second Restatements of Torts’ use of the test, the Restatement (Third) of Torts recently jettisoned it largely on the ground that it was too vague and promoted confusion. *See AARON D. TWERSKI & JAMES A. HENDERSON, JR., TORTS: CASES AND MATERIALS* 232 (2d ed. 2008) (“The [Third] Restatement has wisely rid itself of the substantial factor test. . . . [T]he substantial factor test caused confusion. . . . Few will mourn the passing of the ‘substantial factor’ test.”); David W. Robertson, *The Common Sense of Cause in Fact*, 75 TEX. L. REV. 1765, 1779–80 (1997); Joseph Sanders, Michael D. Green & William C. Powers, Jr., *The Insubstantiality of the “Substantial Factor” Test for Causation*, 73 MO. L. REV. 399, 418 (2008).

91. The oral argument in *Burrage* provides a particularly vivid demonstration of the confusion the test can generate among judges and litigants alike. *See Transcript of Oral Argument* at 13–14, 38–41, *Burrage v. United States*, 571 U.S. 204 (2013) (No. 12–7515).

92. *See, e.g.*, *Abney v. State*, 766 N.E.2d 1175 (Ind. 2002).

93. *E.g.*, *Bostic v. Georgia-Pacific Corp.*, 439 S.W.3d 332, 346, 351, 373–74 (Tex. 2014).

given contribution to be more than merely *de minimus* in amount.⁹⁴ Most also agree that it does *not* require but-for causation, though Joshua Dressler, along with several courts, disagree, based on intuitions about the term “substantial” and “the way we talk about causation.”⁹⁵

As for *when*, if ever, the “substantial factor” test should apply, some courts apply the test across the board, as an allegedly more accurate definition of causation for all cases,⁹⁶ including where but-for causation appears to be satisfied.⁹⁷ Other courts apply it only where the evidence is sufficient to prove either (a) that but-for causation is met, *or* (b) that the alleged “cause” in question was independently sufficient, even if unnecessary.⁹⁸ Yet other courts use the instruction only for non-necessary, independently *insufficient* contributing factors.⁹⁹ And many courts and commentators are not clear about when or why they apply the test.¹⁰⁰

94. Sanders et al., *supra* note 90, at 418; MOORE, *supra* note 23, at 120 (“Clearly a quantitative measure is intended here.”). In this way, it is sometimes more restrictive than the but-for test, which allows for tiny contributions to count as causes so long as the event in question would not have occurred without them. *See* People v. Caldwell, 681 P.2d 274, 280 (Cal. 1984); People v. Wells, 355 N.W.2d 105, 107 (Mich. 1984) (Levin, J., dissenting).

95. JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 187 (6th ed. 2012) (“[I]t is hard to comprehend how a person’s conduct can ever be a ‘substantial factor’ in causing a result that was going to occur when it did *even without his contribution*. The only way it may be sensibly be said that a concurrent sufficient cause . . . is a substantial factor in an outcome is to point out that the force *would have been* the cause of the harm if circumstances had been different (i.e., if the *other* sufficient cause . . . had not materialized). However this is not the way we ordinarily talk about causation.” (emphasis in original)); *accord* State v. Montoya, 61 F.3d 793, 799 (N.M. 2002) (quoting Dressler’s text); *see also* United States v. Jenkins, 120 F. Supp. 3d 650, 658 (E.D. Ky. 2013) (construing the HCPA and concluding that “because of” entails necessity—i.e., but-for causation—as a matter of ordinary language, and that the phrase “the substantial factor” captures this necessity requirement, since necessity is “a status reserved only for the most substantial of the motivating factors”).

96. *E.g.*, People v. Jennings, 237 P.3d 474, 496 (2010). *But see* Lawrence M. Solan & John M. Darley, *Causation, Contribution and Legal Liability: An Empirical Study*, 64 LAW & CONTEMP. PROBS. 265, 271 (2001) (“[I]t is not the least bit obvious that substantial factor defines causation other than stipulatively.”).

97. *E.g.*, Knodle v. Waikiki Gateway Hotel, Inc., 742 P.2d 377, 386–90 (Haw. 1987); Sumnicht v. Toyota Motor Sales, U.S.A., Inc., 360 N.W.2d 2, 11 (Wis. 1984). This is to the chagrin of numerous commentators. *See, e.g.*, Robertson, *supra* note 90, at 1779–80 (“When courts begin turning to the substantial factor vocabulary in a broader range of cases, valuable precision of analysis is lost and nothing is gained.”); Sanders et al., *supra* note 90, at 417–18.

98. *E.g.*, Jennings, 237 P.3d at 496.

99. *E.g.*, Mayhue v. Sparkman, 653 N.E.2d 1384 (Ind. 1995); Roberson v. Counselman, 686 P.2d 149 (Kan. 1984); Hake v. Manchester Twp., 486 A.2d 836 (N.J. 1985); State v. Christman, 249 P.3d 680, 687 (Wash. Ct. App. 2011). And some courts apply it only where it is clear that but-for causation is absent, yet there are multiple independently sufficient non-necessary causes. *E.g.*, *In re* Hanford Nuclear Reservation Litig., 497 F.3d 1005, 1028–29 (9th Cir. 2007) (“[S]ubstantial factor causation . . . applies when there have been ‘multiple, independent causes,’ each of which alone is sufficient to cause the injury.” (quoting Gausvik v. Abbey, 107 P.3d 98, 108 (Wash. Ct. App. 2005))).

100. *See, e.g.*, State v. Wassil, 658 A.2d 548, 551 (Conn. 1995); State v. William, 435 N.W.2d 174, 177 (Neb. 1989); State v. Lillie, 193 P.3d 1050, 1053 (Or. Ct. App. 2008);

A third and final test, while less prevalent, is worth mention in large part because of how frequently it gets mistaken for the others: the “sole” cause test. Justice Ginsburg, for example, in multiple opinions, registers disagreement with the notion that “because of” means “solely because of”¹⁰¹ and the notion that an event can only “result from” one cause,¹⁰² seemingly considering these points responsive to the Court’s adoption of the “but-for” test.¹⁰³ Yet but-for causes need not be, and rarely are, sole causes.¹⁰⁴ Justice Stevens likewise attacked the “sole” cause test in his *Gross* dissent, despite nobody advocating for it.¹⁰⁵ And other judges have at times echoed these same points.¹⁰⁶ In short, while criminal and tort law do occasionally use “sole” cause rules,¹⁰⁷ they are uncommon and not generally thought to reflect common usage or causal intuitions.

II. WHEN AND WHY COURTS CARE ABOUT ORDINARY USAGE AND FOLK CONCEPTS

As seen in the cases discussed above, courts often fixate on the ordinary meaning and folk concept of causation, evaluating competing interpretations and doctrinal tests according to how closely they map onto the ordinary meaning of the statute’s causal language. This Part takes up a more fundamental question: *Should* courts care about the ordinary understanding and folk concept of causation? It answers in the affirmative,¹⁰⁸ for reasons that have to do with (a) statutory interpretation generally,

KEETON ET AL., *supra* note 41, § 41, at 266, 268; LAFAYE, *supra* note 69, § 6.4, at 468, 479.

101. *Burrage v. United States*, 571 U.S. 204, 219 (Ginsburg, J., concurring) (citing *Univ. of Tex. Sw. Med. Ctr. v. Nassar*, 570 U.S. 338, 369 (2013) (Ginsburg, J., dissenting)).

102. *See id.*

103. *Id.* Occasionally courts do endorse a but-for test while also mistakenly drawing inferences that assume it is the same as a sole cause test. *E.g.*, *United States v. Hill*, 182 F. Supp. 3d 546, 555 n.8 (E.D. Va. 2016) (“[I]f the conduct underlying an HCPA prosecution has any sort of commercial motivation—if, for example, Hill had punched C.T. because of both C.T.’s sexual orientation and C.T. having taken Hill’s job as a packer—then the government could never obtain a conviction, as the HCPA requires ‘but-for’ causation.” (citing *United States v. Miller*, 767 F.3d 585, 594 (6th Cir. 2014))). And some judges have even relied on Ginsburg’s seeming equation of but-for and sole cause tests in order to deny relief to discrimination plaintiffs on the ground that, in pleading *multiple* unlawful motivations, the plaintiff had necessarily failed to allege but-for causation (since, on this mistaken reading, there can only be one but-for cause). *See, e.g.*, *Hendon v. Kamtek, Inc.*, 117 F. Supp. 3d 1325, 1332 (N.D. Ala. 2015); *Donald v. UAB Hosp. Mgmt.*, No. 2:14-cv-727-WMA, 2015 WL 3952307, at *2–3 (N.D. Ala. June 29, 2015).

104. *See, e.g.*, *Husted v. A. Philip Randolph Inst.*, No. 16–980, slip op. at 10–11 (U.S. June 11, 2018).

105. *Gross v. FBL Fin. Servs., Inc.*, 557 U.S. 167, 183 n.4 (2009) (Stevens, J., dissenting).

106. *E.g.*, *State v. Malone*, 819 P.2d 34, 36 (Alaska Ct. App. 1991); *State v. Christman*, 249 P.3d 680, 687 (Wash. Ct. App. 2011); cases cited *supra* note 103.

107. *See MOORE, supra* note 23, at 114; Verstein, *supra* note 30.

108. To reiterate: this Article does not claim that judges ought to resolve statutory disputes, even in the criminal and tort causation context, *exclusively* via determinations of ordinary meaning (as they presently purport to do). *See supra* note 16; *infra* Section IV.C.

(b) criminal and tort law specifically, and (c) the role of causation requirements in criminal and tort law.

A. Statutory Interpretation Generally

As the cases above and countless others make clear, judges “throughout the political spectrum”¹⁰⁹ purport to care a great deal about ordinary usage in statutory interpretation.¹¹⁰ They treat it as the default starting point for interpretation,¹¹¹ and where it clearly favors a given interpretation, many treat it as the sole determinant of statutory meaning.¹¹² Virtually all judges and commentators—including those who disfavor treating ordinary meaning as dispositive—agree that it is a necessary consideration in statutory interpretation.¹¹³ And in the near term at least, judicial interest in ordinary meaning appears likely to increase.¹¹⁴ But is ordinary meaning merely a necessary evil? Or are there reasons to *want* judges to consider it?

109. Lawrence M. Solan & Tammy Gales, *Finding Ordinary Meaning in Law: The Judge, the Dictionary or the Corpus?*, 1 INT’L J.L. DISCOURSE 253, 254 (2016) (collecting examples).

110. *Id.*; VALERIE C. BRANNON, CONG. RESEARCH SERV. R45153, STATUTORY INTERPRETATION: THEORIES, TOOLS, AND TRENDS 21 (2018) (quoting examples demonstrating that “all current members of the Supreme Court have regularly invoked this rule of ordinary meaning”); Brett M. Kavanaugh, Review, *Fixing Statutory Interpretation*, 129 HARV. L. REV. 2118 (2016); accord WILLIAM N. ESKRIDGE, INTERPRETING LAW: A PRIMER ON HOW TO READ STATUTES AND THE CONSTITUTION (2016); JOHN F. MANNING & MATTHEW C. STEPHENSON, LEGISLATION AND REGULATION 60 (2d ed. 2013); ANTONIN SCALIA & BRIAN GARNER, READING LAW: THE INTERPRETATION OF LEGAL TEXTS (2012); SLOCUM, *supra* note 10; LAWRENCE SOLAN, THE LANGUAGE OF STATUTES: LAWS AND THEIR INTERPRETATION (2010); Oliver Wendell Holmes, *The Theory of Legal Interpretation*, 12 HARV. L. REV. 417 (1899) (legal texts are to be interpreted by the standard of the “normal speaker of English”); Anita S. Krishnakumar, *Statutory Interpretation in the Roberts Court’s First Era: An Empirical and Doctrinal Analysis*, 62 HASTINGS L.J. 221, 251 (2010).

111. Abbe R. Gluck, *Congress, Statutory Interpretation, and the Failure of Formalism: The CBO Canon and Other Ways That Courts Can Improve on What They Are Already Trying to Do*, 84 U. CHI. L. REV. 177, 191 (2017); Kavanaugh, *supra* note 110, at 2118; Solan & Gales, *supra* note 109, at 254.

112. See, e.g., *Burrage v. United States*, 571 U.S. 204, 216 (2014); Baude & Doerfler, *supra* note 10 (collecting cases); Abbe R. Gluck, *The States as Laboratories of Statutory Interpretation: Methodological Consensus and the New Modified Textualism*, 119 YALE L.J. 1750, 1756–58 (2010) (concluding that state courts, like federal courts, give primacy to text and decline to look to external sources of meaning if they deem the text “plain”).

113. See Fallon, *supra* note 10, at 1252–53 (“In seeking to understand legal meaning, attention to the norms that govern ascriptions of meaning in ordinary conversation is almost self-evidently appropriate.”); SLOCUM, *supra* note 10, at 8–9 (“[T]he ordinary meaning doctrine is influential because it is difficult to conceive of a realistic methodology of interpretation in which it would not be influential,” hence “[e]ven critics who question the decisiveness of ordinary meaning concede the doctrine’s influence.”) (citing examples). But see Frederick Schauer, *Is Law a Technical Language?*, 52 SAN DIEGO L. REV. 501 (2015).

114. See Kavanaugh, *supra* note 110, at 2144–45, 2150 (advocating that judges “determine the best reading of the statute, not whether it is clear or ambiguous,” and explaining that under this “best reading” approach, “[c]ourts should try to read statutes as ordinary users of the English language might read and understand them”).

One such reason suggested by the recent “positive turn” in legal interpretation scholarship—and familiar to any practicing lawyer—is simply that the case law itself says to consider ordinary meaning.¹¹⁵ On this view, courts ought to care about ordinary meaning because binding authoritative texts direct courts to consider it (and indeed, to treat it as dispositive in the event it is clear or obvious).¹¹⁶

A second and more substantive reason for giving weight to ordinary meaning, often highlighted by courts and commentators, is that it frequently constitutes the most reliable indication of Congress’s understanding, intent, and purpose in passing the statutory provision whose meaning is in question. As an empirical matter, this rationale is plausible as applied to the understanding or intent of individual drafters of legislation: congressional drafters are more aware of the ordinary and plain meaning doctrines than any other interpretive canon, and while drafting legislation they rarely consult sources other than their own intuitive understanding of the words used.¹¹⁷ But perhaps more importantly, even where individual congresspersons or legislative drafters possess their own idiosyncratic and non-ordinary understandings, ordinary meaning might nonetheless remain the only plausible candidate meaning understood by Congress *as a whole*, i.e., as a collective body capable of possessing an intent, a purpose, or an understanding of what a word or phrase means.¹¹⁸ Of course, there are exceptions—instances where Congress clearly did not intend what the ordinary meaning of the statute’s words conveys (scrivener’s errors are the clearest example).¹¹⁹ But by and large, as the Court emphasized in *Gross*, courts purport to interpret statutes in accordance with “the language employed by Congress and the assumption that the ordinary meaning of that language accurately expresses the legislative purpose.”¹²⁰

115. See Baude & Sachs, *supra* note 10, at 1134.

116. *Id.*; see also *supra* note 10 (discussing plain meaning rule).

117. Abbe R. Gluck & Lisa Schultz Bressman, *Statutory Interpretation from the Inside—An Empirical Study of Congressional Drafting, Delegation, and the Canons: Part I*, 65 STAN. L. REV. 901, 995 (2013).

118. See, e.g., *Green v. Bock Laundry Mach. Co.*, 490 U.S. 504, 528 (1989) (Scalia, J., concurring); Frank H. Easterbrook, *The Role of Original Intent in Statutory Construction*, 11 HARV. J.L. & PUB. POL’Y 59, 65 (1988). This aspect of ordinary meaning—its being the meaning on which the largest number of people’s interpretations are likely to converge—is central to other prominent arguments in favor of ordinary meaning interpretation, beyond those concerning congressional purpose or intent. See, e.g., Frederick Schauer, *Statutory Construction and the Coordinating Function of Plain Meaning*, 1990 SUP. CT. REV. 231, 232 (advocating the plain meaning rule as a “second-best coordinating device”); David A. Strauss, *Why Plain Meaning?*, 72 NOTRE DAME L. REV. 1565, 1566 (1997) (arguing that an ordinary meaning default reduces otherwise costly disagreement).

119. See Ryan D. Doerfler, *The Scrivener’s Error*, 110 NW. U. L. REV. 811 (2016); John F. Manning, *The Absurdity Doctrine*, 116 HARV. L. REV. 2387, 2459 n.265 (2005).

120. *Gross v. FBL Fin. Servs., Inc.*, 557 U.S. 167, 175 (2009) (quoting *Engine Mfrs. Ass’n v. S. Coast Air Quality Mgmt. Dist.*, 541 U.S. 246, 252 (2004)); see also *Burrage v. United States*, 571 U.S. 204, 216 (2014); *Marx v. Gen. Revenue Corp.*, 568 U.S. 371, 375–76 (2013); *Watson v. United States*, 552 U.S. 74, 79 (2007) (stating that statutory meaning often “has to turn on the language as we normally speak it”—on “everyday meaning”—because “[w]ith no statutory definition or definitive clue, . . . there is no other source of a reasonable inference about what Congress understood when writing” the statute (emphasis added)); *Green*, 490

Finally, proponents of ordinary meaning interpretation, and of textualism more broadly, claim that it constrains judges,¹²¹ thereby promoting rule of law values¹²² such as adjudicative consistency, predictability, and neutrality.¹²³ These and other related rule of law concerns receive their greatest emphasis in the context of criminal statutes, to which we'll now turn.

B. Criminal and Tort Law Generally

Unlike many areas of law, criminal liability (along with some kinds of tort liability) expresses moral condemnation, often via extraordinary means like punitive fines or prison.¹²⁴ Here the Article considers three closely related outgrowths of these blaming and punitive treatment aspects of criminal and tort law, each of which makes ordinary meaning analysis especially appealing to courts and commentators: (1) the need for restraints on the government, (2) the need to provide fair notice to those subject to the law, and (3) the use of juries.

First, ordinary meaning analysis restrains the government's ability to use its punitive power improperly, including in lawmaking,¹²⁵ enforcement,¹²⁶ and adjudication.¹²⁷ As for lawmaking, ordinary meaning facilitates accountability for lawmakers by allowing citizens to understand the content of laws at the time they are passed.¹²⁸ With respect to enforcement, ordinary meaning diminishes enforcement authorities' (e.g., police officers') ability to claim mistake of law based on nonordinary interpretations, and it makes genuine mistakes of law less likely in the

U.S. at 528 (Scalia, J., concurring).

121. All agree that text and ordinary meaning provide some constraint; the disagreement among judges and scholars concerns how much. *See, e.g.*, sources cited *supra* notes 10, 113, *infra* note 301.

122. *See, e.g.*, William N. Eskridge, Jr. & Philip P. Frickey, *Statutory Interpretation as Practical Reasoning*, 42 STAN. L. REV. 321, 364 (1990).

123. *See* ESKRIDGE, *supra* note 110, at 39–40; Michael S. Moore, *A Natural Law Theory of Interpretation*, 58 S. CAL. L. REV. 277, 291 (1985).

124. *See* Joel Feinberg, *The Expressive Function of Punishment*, 49 MONIST 397 (1965); Scott Hershovitz, *Treating Wrongs as Wrongs: An Expressive Argument for Tort Law*, 10 J. TORT L. 1 (2017).

125. *See* Paul H. Robinson, *Fair Notice and Fair Adjudication: Two Kinds of Legality*, 154 U. PA. L. REV. 335, 348–53 (2005) (discussing the Constitution's prohibitions of certain *ex post facto* laws).

126. *See, e.g.*, *Alabama v. Smith*, 490 U.S. 794, 798 (1989) (discussing prohibition on vindictive prosecution).

127. *See* Robinson, *supra* note 125, at 353–56.

128. *See* Samuel W. Buell, *Culpability and Modern Crime*, 103 GEO. L.J. 547, 588 (2015); Nourse, *supra* note 10, at 1004. The point is not only about the importance of *timing* for purposes of holding lawmakers accountable but also about the importance of *concentrating, rather than diffusing, perceived responsibility for the law's content*. Insofar as ordinary meaning interpretation is perceived to constrain judges, it may help concentrate perceived responsibility for statutory law's content on the legislature itself, rather than allowing the legislature to diffuse that responsibility by blaming the content of an unpopular law on judges' subsequent interpretive decisions.

first place.¹²⁹ Finally, as to adjudication, ordinary meaning prevents judges from giving statutes less lenient interpretations than those most readily apparent to ordinary speakers.¹³⁰

A second rule of law concern paramount in criminal law and at least some areas of tort law is fair notice—a value that more directly implicates ordinary meaning analysis (as do the various doctrines associated with notice concerns, including void-for-vagueness,¹³¹ the prohibition on punitive ex post facto laws,¹³² and the doctrine of lenity¹³³). As courts often emphasize, ordinary citizens must have a fair opportunity to ascertain what the law permits, what it prohibits, and what legal consequences can follow from violating it.¹³⁴ This means articulating law in language understandable to normal English speakers whose conduct is subject to it.¹³⁵

129. See Richard H. McAdams, *Close Enough for Government Work? Heien's Less-Than-Reasonable Mistake of the Rule of Law*, 2015 SUP. CT. REV. 147. Relatedly, heavy reliance on pro se representation and overburdened public legal services in criminal and civil rights litigation likewise puts a thumb on the scale in favor of ordinary meaning as a check on the government's ability to overpower defendants bringing meritorious claims. *Cf. id.* at 188–90 (emphasizing the dangers implicit in allowing government to exploit ambiguities in laws it has passed).

130. See, e.g., *Burrage v. United States*, 571 U.S. 204, 218 (2014).

131. *Kolender v. Lawson*, 461 U.S. 352, 357 (1983) (“[T]he void-for-vagueness doctrine requires that a penal statute define the criminal offense with sufficient definiteness that ordinary people can understand what conduct is prohibited”); *Grayned v. City of Rockford*, 408 U.S. 104, 108 (1972) (“[W]e insist that laws give the person of ordinary intelligence a reasonable opportunity to know what is prohibited, so that he may act accordingly.”).

132. *Robinson*, *supra* note 125, at 348–56.

133. E.g., *Burrage*, 571 U.S. at 219 (Ginsburg, J., concurring).

134. See, e.g., cases cited *supra* note 131; *Sessions v. Dimaya*, 138 S. Ct. 1204 (2018); *Burrage*, 571 U.S. at 218; *United States v. Nat'l Dairy Prods. Corp.*, 372 U.S. 29, 36 (1963) (“[W]e are concerned with the vagueness of the statute ‘on its face’ because such vagueness may in itself deter constitutionally protected and socially desirable conduct.” (citing *Thornhill v. Alabama*, 310 U.S. 88, 98 (1940))); *Robinson*, *supra* note 125.

135. See *Burrage*, 571 U.S. at 217–18; *Grayned*, 408 U.S. at 108; *McBoyle v. United States*, 283 U.S. 25, 27 (1931) (Holmes, J.) (“Although it is not likely that a criminal will carefully consider the text of the law before he murders or steals, it is reasonable that a fair warning should be given to the world in language that the common world will understand, of what the law intends to do if a certain line is passed.”); Kent Greenawalt, *Punishment*, 74 J. CRIM. L. & CRIMINOLOGY 343, 345 (1983); *Robinson*, *supra* note 125, at 364.

The standard for fair notice as applied to police officers in the context of qualified immunity requires far more by way of legal clarity than does the requisite fair notice as applied to ordinary citizens. See *Kisela v. Hughes*, 138 S. Ct. 1148, 1152 (2018) (“[T]he focus is on whether the officer had fair notice that her conduct was unlawful,” which in turn requires that the “statutory or constitutional question [be] beyond debate” at the time the officer acted, so that “all but the plainly incompetent or those who knowingly violate the law” are protected from liability due to their having received insufficient notice of what the law required.); see also *id.* at 1158 (Sotomayor, J., dissenting) (“At its core, then, the ‘clearly established’ inquiry boils down to whether [defendant officer] had ‘fair notice’ that he acted unconstitutionally.”); David Louk, *The Audiences of Statutes: Statutory Interpretation from the Outside*, 104 CORNELL L. REV. (forthcoming 2019) (unpublished manuscript on file with the *Indiana*

Third and finally, criminal and tort law's extensive reliance on juries provides further reason for courts to consider ordinary usage. Juries construe instructions according to their ordinary meaning, and judges presume that they do so when reviewing and approving jury instructions.¹³⁶ Moreover, judges at the trial and appellate level, through various procedural mechanisms, must determine what a reasonable jury could (or in post-trial circumstances, could have) conclude(d) based on the instructions it received, often compared with some variation of them that was proposed and rejected.¹³⁷ In short, jury-related procedures necessitate frequent and consequential ordinary meaning determinations throughout tort and criminal law, whether judges like it or not.

C. Causation in Criminal and Tort Law

Throughout statutory and common law criminal and torts case law, judges claim that the law's concept of causation is the man on the street's concept of causation.¹³⁸ Many commentators, most famously philosophers of criminal and tort law, make the same claim, treating the concept of factual causation in law as a matter of "common sense."¹³⁹ "[I]t is the plain man's notions of causation," Hart and Honoré asserted in their seminal work, *Causation In the Law*, "with which the law is concerned."¹⁴⁰ But why *should* it be? Many of the reasons judges and philosophers provide echo the reasons considered above. Hart and Honoré, for example, appealed to rule of law values like those discussed above,¹⁴¹ and many of the other arguments discussed

Law Journal).

136. See, e.g., *Estelle v. McGuire*, 502 U.S. 62, 72 (1991).

137. This ordinary language interpretation and application task is especially fraught where—as happens frequently in criminal and tort law—the instructions concern mixed questions of law and fact, using open-ended and normativity-infused terms like "reasonable" or "reckless" (or perhaps "results from" or "because of"). See Ronald J. Allen & Michael S. Pardo, *The Myth of the Law-Fact Distinction*, 97 NW. L. REV. 1769 (2003); Suja A. Thomas, *The Fallacy of Dispositive Procedure*, 50 B.C. L. REV. 759 (2009).

138. *Burrage*, 571 U.S. at 215; *People v. Jennings*, 237 P.3d 474 (Cal. 2010); *State v. Burton*, 370 S.W.3d 926, 931 n.3 (Mont. Ct. App. 2012); *State v. Christman*, 249 P.3d 680, 687 (Wash. Ct. App. 2011); Moore, *supra* note 23; Richard W. Wright, *The Nightmare and the Noble Dream: Hart and Honoré on Causation and Responsibility*, in *THE LEGACY OF H.L.A. HART: LEGAL, POLITICAL AND MORAL PHILOSOPHY* 11 (Matthew H. Kramer et al. eds., 2008); Solan & Darley, *supra* note 96, at 272 ("[I]n making decisions about causation, courts at least purport to be applying 'common sense' intuitions, to use Hart & Honoré's standard for how causative notions enter the law.").

139. DRESSLER, *supra* note 95, at 160 ("The role of [factual] causality in the criminal law is the same as it is in the evaluation of any everyday event: to determine why something occurred."); KEETON ET AL., *supra* note 41, at 264 ("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."); Robertson, *supra* note 90; Solan & Darley, *supra* note 96, at 271 n.30 ("Contemporary torts theorists also look at causation in fact as a common sense notion.").

140. HART & HONORÉ, *supra* note 23, at 1.

141. See WILLIAM LUCY, *PHILOSOPHY OF PRIVATE LAW* 203 (2007); Jane Stapleton, *Unpacking Causation*, in *RELATING TO RESPONSIBILITY: ESSAYS FOR TONY HONORÉ ON HIS*

above likewise apply to criminal and statutory tort causation for the same reason they apply to other statutory terms in law generally and/or criminal and tort law specifically.¹⁴² But is there anything about the issue of *causation* in particular that gives additional weight to those earlier arguments for ordinary meaning, or that gives rise to new and distinct reasons to care about ordinary meaning?

One prominent reason in the philosophy of criminal and tort law begins along the following lines, as articulated by Michael Moore: (1) “criminal and tort liability must track moral responsibility,” (2) “moral responsibility depends in part on causal responsibility,” *therefore* (3) “‘cause’ in [at least criminal and tort] law must mean what it means in morality.”¹⁴³ To be sure, there are those who would argue against each of Moore’s three claims. But for present purposes, the key question is how, if one agrees with (3), one might get to caring about ordinary meaning and folk concepts in the context of causation. And here’s that route: (4) ordinary moral intuitions, and the ordinary linguistic usage through which we communicate those intuitions in terms of causal responsibility, are important evidence of what “cause” means, including in morality.¹⁴⁴ In short, we care about how people ordinarily think and talk about causation because it is evidence of what causation is, which in turn bears on moral responsibility.

Now, as it happens, Moore himself might not put much stock in (4) (for reasons we’ll return to in Section V.C).¹⁴⁵ But we needn’t get bogged down by that right now, because even assuming that lay usage and intuitions aren’t credible evidence of fundamental conceptual, moral, or metaphysical truths, there remain reasons why criminal and tort law ought nonetheless to track lay usage and intuitions about causation. Like Moore’s argument above, these arguments posit that causation is, at least in lay understanding, a morally important concept. But they are arguments for why, independent of the “true nature” of causation or morality as a matter of metaphysics,¹⁴⁶ or the “true content” of causal terms’ referents as a matter of philosophical semantics,¹⁴⁷ folk intuitions and usage have independent importance to law, at least when it comes to normatively infused folk concepts like causation.

One such argument comes from consequentialist theorists like Paul Robinson, who argue that criminal law¹⁴⁸ ought to track lay moral intuitions of justice, at least

EIGHTIETH BIRTHDAY 160–61 (2001) [hereinafter Stapleton, *Unpacking Causation*]; Tony Honoré, *Herbert Lionel Adolphus Hart*, 84 PROC. BRITISH ACAD. 295, 304 (1994); Jane Stapleton, *Choosing What We Mean by “Causation” in the Law*, 73 MO. L. REV. 433, 458 (2008) [hereinafter Stapleton, *Choosing What We Mean*].

142. See *supra* Sections II.A–B.

143. MOORE, *supra* note 23, at 4; see also, e.g., *id.* at 95 (“This corrective-justice view of tort law demands a robustly metaphysical interpretation of cause. For legal liability tracks moral responsibility on this view, and moral responsibility is for those harms we *cause*.” (emphasis in original)).

144. See HART & HONORÉ, *supra* note 23, at 1; Stapleton, *Choosing What We Mean*, *supra* note 141, at 458–59.

145. See *infra* notes 308–10 and accompanying text.

146. Michael S. Moore, *The Metaphysics of Causal Intervention*, 88 CALIF. L. REV. 827, 828 (2000).

147. *Id.* at 854–56.

148. The point arguably holds for tort law as well, though the torts literature is far less developed on this issue. See Theodore Eisenberg & Christoph Engel, *Unpacking Negligence*

in the vast majority of cases, because doing so fosters greater compliance with and deference to criminal law on the part of the citizenry¹⁴⁹ (not to mention juries with nullification power).¹⁵⁰ In other words, regardless of what “true” morality is, law ought to track lay intuitions of morality because doing so brings about good consequences.¹⁵¹

A second argument concerns democratic representativeness. On this view, law (or at least criminal and tort law) in a democracy ought to reflect the moral intuitions or preferences of its citizens.¹⁵² We can glean people’s moral intuitions or preferences from their intuitions about concepts that play important roles in moral evaluation;

Liability: Experimentally Testing the Governance Effect, 13 J. EMPIRICAL LEGAL STUD. 116 (2016).

149. See ROBINSON, *supra* note 79; PAUL H. ROBINSON & JOHN M. DARLEY, JUSTICE, LIABILITY, AND BLAME: COMMUNITY VIEWS AND THE CRIMINAL LAW (1995).

150. A subtler but more prevalent form of “nullification” may simply be jurors’ interpretation of ambiguous instructions in ways that conform to their preferred outcome. This need not be, and probably often is not, done consciously. See Ward Farnsworth, Dustin F. Guzor & Anup Malani, *Ambiguity About Ambiguity: An Empirical Inquiry into Legal Interpretation*, 2 J. LEGAL ANALYSIS 257 (2010); Ward Farnsworth, Dustin Guzor & Anup Malani, *Policy Preferences and Legal Interpretation*, 1 J.L. & CTS. 115 (2013).

151. To be sure, Robinson considers lay intuitions a generally good piece of evidence as to true morality. See Paul H. Robinson, *The Role of Moral Philosophers in the Competition Between Deontological and Empirical Desert*, 48 WM. & MARY L. REV. 1831 (2007). But his work does not focus on that as a reason for their validity as a guide to forming criminal law doctrine.

Of course, if intuitions and ordinary usage concerning *causation* are mere *proxies* for intuitions about justice and morality, then one might be tempted to skip the proxy and simply track lay moral intuitions without regard to causal language. Cf. WILLIAM D. POPKIN, STATUTES IN COURT: THE HISTORY AND THEORY OF STATUTORY INTERPRETATION 181 (1999). That, however, could run afoul of fair notice and other rule of law requirements. See *supra* Section II.B. And by running afoul of those intuitively just notice requirements, it could backfire, for there are consequentialist reasons to care about perceptions of procedural justice (including perceptions of fair notice). See, e.g., TOM TYLER, WHY PEOPLE OBEY THE LAW: PROCEDURAL JUSTICE, LEGITIMACY, AND COMPLIANCE (2d ed. 2006); Janice Nadler, *Flouting the Law*, 83 TEX. L. REV. 1399 (2005). Still, in a more realistic sense, notice concerns might favor direct consideration of lay intuitions of justice (in situations where they differ from statutory causation requirements) insofar as general moral intuitions and related *assumptions* about the law’s content are likely to be the only *actual* “notice” that criminal and tort defendants receive, and people tend to assume that law tracks their intuitions of justice. See ARDEN ROWELL, LEGAL RULES, BELIEFS AND ASPIRATIONS (2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903049 [<https://perma.cc/2W8E-WUAA>]. In any event, as a positive matter, direct consideration of lay intuitions of justice (or, more precisely, legal actors’ guesses as to the content of those lay intuitions) likely does drive many enforcement decisions in criminal and tort contexts. See *supra* notes 18–19 and accompanying text; Anna Offit, *Prosecuting in the Shadow of the Jury*, 113 NW. U. L. REV. (forthcoming 2019) (unpublished manuscript on file with the *Indiana Law Journal*).

152. See, e.g., ROBINSON & DARLEY, *supra* note 149; POPULAR PUNISHMENT: ON THE NORMATIVE SIGNIFICANCE OF PUBLIC OPINION (Jesper Ryberg & Julian Roberts eds., 2014); see also Joshua Kleinfeld, *Manifesto of Democratic Criminal Justice*, 111 NW. U. L. REV. 1367 (2017).

causation is one such concept.¹⁵³ So in order to effectively track lay intuitions of justice, and thereby give effect to laypeople's preferences and conceptions of justice to an extent befitting a representative democracy, we should, on this view, have causation in law track the folk concept of causation.

III. HOW COURTS FIND ORDINARY MEANING

So, to recap, we've seen *that* courts care about ordinary usage of causal language that appears in statutes. And we've seen *why* they care—for reasons stemming from statutory interpretation generally, criminal and tort law more specifically, and causation's role in criminal and tort law. Now we'll examine *how* courts figure out the content of these statutory terms and the folk concepts they reflect, using tools common throughout modern statutory interpretation. Each of these tools—introspection and intuition pumps (beloved by judges and philosophers alike, as we'll see in Part V), as well as dictionaries and corpus linguistics—runs into serious problems of bias, indeterminacy, and inaccuracy. The deficiencies in courts' (and philosophers') methods help motivate the experimental study reported in Part IV.

A. Introspection and Intuition Pumps

Introspection is the most common method judges use for ascertaining ordinary usage and the content of folk concepts the law invokes. After all, throughout their everyday lives judges use and are exposed to many of the words, phrases, and concepts at issue in statutory interpretation cases; by consulting their own ordinary understanding of the word or phrase, they can quickly and easily get a sense of its likely ordinary meaning.¹⁵⁴

Still, judges might not do this as reliably and objectively as one would hope. Like the rest of us, they might engage in “motivated reasoning” or “motivated intuiting,” forming initial intuitions that are biased in favor of their preferred outcome,¹⁵⁵ after which “confirmation bias” and “consensus bias” will lead them to become more and more convinced that their interpretation is normal or ordinary.¹⁵⁶ And even absent such systematic biases, we may worry that a sample size of one would still be inadequate to generate reliable inferences about the linguistic usage and intuitions of the public writ large.

Moving from the completely internal to the slightly more external, judges often “test” their intuitions against hypothetical or real world examples (“intuition pumps”) to see what feels comfortable or intuitive to say.¹⁵⁷ The cases above contained various

153. See HART & HONORÉ, *supra* note 23, at 4; ROBINSON, *supra* note 79, at 343.

154. See LAWRENCE M. SOLAN, *THE LANGUAGE OF JUDGES* 16 (1993).

155. See Farnsworth et al., *supra* note 150.

156. See Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. GEN. PSYCHOL. 175 (1998); Lawrence Solan, Terri Rosenblatt & Daniel Osherson, *False Consensus Bias in Contract Interpretation*, 108 COLUM. L. REV. 1268 (2008). But see Farnsworth et al., *supra* note 150.

157. See Raymond M. Kethledge, *Ambiguities and Agency Cases: Reflections After (Almost) Ten Years on the Bench*, 70 VAND. L. REV. EN BANC 315, 319–22 (2017).

examples (the *Burrage* Court on the cause of a baseball team's winning,¹⁵⁸ the *Miller* court on John's reason for bringing an umbrella,¹⁵⁹ etc.). Using intuition pumps to test ordinary meaning is an "often colorful and fun part of the job of statutory interpretation,"¹⁶⁰ and one that permeates judicial opinions.

Like pure introspection, intuition pumps can be useful, but they pose serious risk of bias, rendering them dubious as a guide or constraint on ordinary meaning interpretation. Due to confirmation bias (or, more cynically, "intuition pump shopping"), judges often discount the probity of intuition pumps that cut in the opposite direction from their initial intuition, focusing only on those that confirm their priors.¹⁶¹ And while intuition pumps have the virtue of specifying the relevant context (e.g., a newspaper sports section reporting on a baseball game,¹⁶² a verbal explanation of why someone brought an umbrella,¹⁶³ etc.), subtle differences between the context described in the intuition pump and the context it is supposed to illuminate may render the intuition pump inapt, sometimes for reasons not apparent at first glance.¹⁶⁴

B. Dictionaries and Corpus Linguistics

Moving further from the internal and subjective to the external and objective, judicial use of dictionaries has become extremely popular¹⁶⁵ amongst judges of all ideological inclinations and interpretive philosophies.¹⁶⁶ As the cases reviewed

158. *Burrage v. United States*, 571 U.S. 204, 211–12 (2014).

159. *United States v. Miller*, 767 F.3d 585, 591 (6th Cir. 2014).

160. Kethledge, *supra* note 157, at 321. As an example, Judge Kethledge describes a recent case in which the Sixth Circuit interpreted a criminal statute's use of the term "inflicted," "testing" their intuitions using six hypothetical examples from ordinary language. *Id.* at 321–22. This way, the judges were able to find "the best objective reading" of the text. *Id.* at 319–20 ("[I]n my experience at least, if one works hard enough, all the other interpretations are eventually revealed as imposters. . . . In my own opinions as a judge, I have never yet had occasion to find a statute ambiguous."); see also Kavanaugh, *supra* note 110 (advocating a similar approach).

161. Compare, e.g., *Burrage*, 571 U.S. at 211, with MOORE, *supra* note 23, at 418.

162. See *Burrage*, 571 U.S. at 211.

163. See *Miller*, 767 F.3d at 591.

164. See James A. Macleod, *Belief States in Criminal Law*, 68 OKLA. L. REV. 497, 544 n.190 (2016).

165. James J. Brudney & Lawrence Baum, *Oasis or Mirage: The Supreme Court's Thirst for Dictionaries in the Rehnquist and Roberts Era*, 55 WM. & MARY L. REV. 483, 495 (2013); Kavanaugh, *supra* note 110, at 2150 n.158 ("Under the 'best reading' inquiry, the question is only how the words would be read by an ordinary user of the English language. That's why textualists rely on dictionaries. Dictionaries may not provide authoritative, binding interpretations of the language of a statute, but they do tell courts something about how the ordinary user of the English language might understand that statutory language."); Jeffrey L. Kirchmeier & Samuel A. Thumma, *Scaling the Lexicon Fortress: The United States Supreme Court's Use of Dictionaries in the Twenty-First Century*, 94 MARQ. L. REV. 77, 79 (2010).

166. Brudney & Baum, *supra* note 165, at 522; Samuel A. Thumma & Jeffrey L. Kirchmeier, *The Lexicon Has Become a Fortress: The United States Supreme Court's Use of Dictionaries*, 47 BUFF. L. REV. 227, 256–60 (1999).

above illustrate, it is now common to resort to dictionaries to ascertain the meaning not just of technical terms, but also the ordinary meaning of “very ordinary words.”¹⁶⁷

There are many scholarly criticisms of judicial dictionary use in statutory interpretation,¹⁶⁸ here we will note three. The first and most decisive criticism is that dictionary definitions fail to sufficiently take context into account.¹⁶⁹ This first criticism thus relates to dictionaries’ *accuracy* in providing the ordinary meaning of a word or phrase as it appears in a given context. The lack-of-context criticism leads to a second problem, which concerns dictionaries’ supposed *objectivity* (and hence, their power to constrain judges and reduce bias): with so many acontextual dictionary definitions to choose from, judges go “dictionary shopping” to find the one that supports their priors when applied in the specific context of the case or statute at hand.¹⁷⁰ A third and final problem concerns dictionary definitions’ *indeterminacy*: they often fail to even address, let alone resolve, the issue of disputed meaning that sent judges to the dictionary in the first place.¹⁷¹

167. Gluck & Bressman, *supra* note 117, at 955 (listing examples); Abbe R. Gluck & Richard A. Posner, *Statutory Interpretation on the Bench: A Survey of Forty-Two Judges on the Federal Courts of Appeals*, 131 HARV. L. REV. 1298, 1318 (2018) (listing additional examples).

168. See, e.g., sources cited *supra* notes 165–67.

169. United States v. Costello, 666 F.3d 1040, 1044 (7th Cir. 2012) (Posner, J.); SLOCUM, *supra* note 10, at 215; Craig Hoffman, *Parse the Sentence First: Curbing the Urge to Resort to the Dictionary when Interpreting Legal Texts*, 6 N.Y.U. J. LEGIS. & PUB. POL’Y 401, 404–06 (2003); A. Raymond Randolph, *Dictionaries, Plain Meaning, and Context in Statutory Interpretation*, 17 HARV. J.L. & PUB. POL’Y 71, 73–74 (1994).

170. See Lawrence M. Solan & Tammy Gales, *Corpus Linguistics as a Tool in Legal Interpretation*, 2017 BYU L. REV. 1311, 1333–34 (2017) (claiming that the ability to pick from different dictionary definitions “negates whatever objectivity there is in using a dictionary”); SLOCUM, *supra* note 10, at 215.

171. For example, in *Gross* and other cases discussed in Part I, the Court majority quoted dictionary definitions of “because,” alleging that they bolstered its conclusion that the phrase implies but-for causation. *Gross v. FBL Fin. Servs., Inc.*, 557 U.S. 167, 176 (2009). But in *Gross*, the dictionaries defined “because of” as “on account of” and “by reason of”—definitions the dissenting Justices understandably considered unilluminating. *Id.* (quoting 1 WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 194 (1966); 1 OXFORD ENGLISH DICTIONARY 746 (1933); THE RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE 132 (1966)). Nor do the definitions of those new terms (“on account of” and “by reason of”) shed any additional light, since the same dictionaries define *them* as “because of,” leaving us back where we started. *On account of*, THE RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE (1967) (defining “on account of” as “*by reason of; because of*” (emphasis added)); *By reason of*, THE RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE (1967) (defining “by reason of” as “*on account of; because of*” (emphasis added)); *On account of*, SHORTER OXFORD ENGLISH DICTIONARY (2002) (defining “on account of” as “*because of*”); *By reason of*, SHORTER OXFORD ENGLISH DICTIONARY (2002) (defining “by reason of” as “on account of”); *On account of*, WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY (1961) (defining “on account of” as “for the sake of: *by reason of; because of*” (emphasis added)). See Brief of the Washington Lawyers Committee for Civil Rights and Urban Affairs, the Employment Justice Center, and Employment Litigators as Amici Curiae in Support of Respondent, at 7–8, *Univ. of Tex. Sw. Med. Ctr. v. Nassar*, 570 U.S. 338 (2013) (No. 12-484).

In light of these criticisms, some scholars have recently advocated using “corpus linguistics” as an alternative method of ascertaining ordinary meaning¹⁷² and predicted that it will “revolutionize” statutory interpretation.¹⁷³ Corpus linguistics involves analysis of large databases of text (“corpora”).¹⁷⁴ These corpora contain “naturally occurring” language—books, newspaper articles, transcriptions of public speeches, and other publicly available texts.¹⁷⁵ Legal scholars drawing on corpus linguistics search corpora for a given word or phrase to ascertain the frequency with which it is used in a given manner.¹⁷⁶ When faced with a dispute over which of two candidate meanings is ordinary, they count how many times the corpus contains each.¹⁷⁷ This frequency data is typically said to show which usage is ordinary and

172. For general introductions, see Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. 788 (2018); Symposium, *Law and Corpus Linguistics*, BYU L. REV. (forthcoming 2019). In a few statutory interpretation cases, the Utah and Michigan Supreme Courts have used corpus linguistics, generating yet greater interest in the field. *E.g.*, *People v. Harris*, 885 N.W.2d 832, 838 (Mich. 2016) (interpreting the meaning of “information”); *State v. Rasabout*, 356 P.3d 1258, 1263 (Utah 2015) (interpreting the meaning of “discharge a firearm”).

173. Amanda Kae Fronk, *Big Lang at BYU*, BYU MAG., Summer 2017 (quoting Lawrence Solum), <https://magazine.byu.edu/article/big-lang-at-byu> [<https://perma.cc/952G-DMMN>].

174. Lee & Mouritsen, *supra* note 172, at 795.

175. *Id.* Some corpora are specialized in subject matter (e.g., corpora containing only legal opinions) or in timeframe of source material (e.g., corpora containing texts published only during the decades surrounding the drafting of the U.S. Constitution). *See, e.g.*, Thomas R. Lee & James C. Phillips, *Data-Driven Originalism*, 167 PA. L. REV. 261 (2019).

176. Lee & Mouritsen, *supra* note 172, at 831.

177. *See, e.g.*, Solan & Gales, *supra* note 109, at 258. For example, Solan and Gales analyzed the meaning of “defraud,” for purposes of interpreting the federal bank fraud statute in *Shaw v. United States*, a case then pending before the Supreme Court. The question in *Shaw* was whether the defendant “defrauded” a bank when he deceived that bank into transferring money from one of its customers to the defendant, where the bank (unlike its customer) ultimately bore no loss of property. Solan and Gales consulted 321 instances of the word “defraud” in the Corpus of Historical American English and determined that, in 98% of the instances of “defraud,” the party being “defrauded” was the one who ultimately bore the loss of the property at issue. The “ordinary meaning” of “defraud,” the authors therefore concluded, did not encompass the defendant’s deception of the bank, since the bank was not the one who lost property in the end. *Id.* at 273.

The Supreme Court instead held, in a unanimous opinion, that the defendant *did* “defraud” the bank, explaining that the bank needn’t “ultimately suffer financial harm” to constitute a “defrauded” party, so long as the defendant’s deception had the effect of at least temporarily depriving the bank of property rights, including, e.g., the property rights a bailee possesses with respect to a bailor’s property. *Shaw v. United States*, 137 S. Ct. 462, 466–68, 470 (2016).

which is non-ordinary,¹⁷⁸ and to do so with greater accuracy, objectivity, and determinacy than dictionaries and other traditional methods.¹⁷⁹

Corpus linguistics may indeed represent an improvement over intuitions and dictionaries in terms of both objectivity and determinacy (though to be sure, it does present its own problems with respect to both).¹⁸⁰ But the bigger challenge for corpus linguistics in legal interpretation concerns the method's *accuracy*—that is, whether it provides data about the sort of thing courts do or should seek to track when they engage in ordinary meaning analysis.

Here is the problem. Ordinary meaning analysis does not and should not concern the *frequency* of a given usage *per se*, nor especially frequency data of the relatively decontextualized sort that corpus linguistics provides. More specifically, proponents of corpus linguistics in legal interpretation typically treat “ordinary meaning” as a question about “the relative frequency of competing senses of a given term” in reported usage.¹⁸¹ But that is not what ordinary meaning analysis typically does or should seek; rather, it seeks the “sense of a word or phrase that is most likely implicated in a given linguistic *context*.”¹⁸² Proponents of corpus linguistics sometimes conflate the two,¹⁸³ but they may be quite different, for at least the

178. Solan & Gales, *supra* note 109, at 263 (“‘Ordinary meaning,’ especially as applied to particular words and phrases, is a distributional fact.”); *see also* Lee & Mouritsen, *supra* note 172, at 829 (“ordinary meaning” concerns “the relative frequency of competing senses of a given term”).

179. *See, e.g.,* Stephen C. Mouritsen, *Hard Cases and Hard Data: Assessing Corpus Linguistics as an Empirical Path to Plain Meaning*, 13 COLUM. SCI. & TECH. L. REV. 156, 174–75, 202, 205 (2011); Lee & Mouritsen, *supra* note 172, at 831.

180. As to *objectivity*, corpus linguistics can provide replicable, falsifiable results, but bias remains a problem throughout the coding that produces those results, as the coder may be faced with difficult decisions regarding which instances of a given word or phrase count as instances of one of the candidate types of usage as opposed to the other. (Imagine, for example, the difficulties one would face in attempting to code instances of the word “because” according to whether the cause to which they referred was a but-for cause or not.) Each discrete coding decision presents potential room for infiltration of bias. *See* Mouritsen, *supra* note 179, at 202.

As for *determinacy*, corpus linguistics can at least get judges arguing about the same thing, rather than talking past each other with dictionary definitions that fail to even bear on the interpretive question at hand. That is because whereas dictionaries were not designed to disambiguate between the relevant senses of the word under dispute, corpus searches can be constructed to address the precise problem at issue. Still, where multiple judges perform analyses of even the same corpora, they may often differ in their interpretation of the results, rendering the method at least somewhat indeterminate. In the first and most extensive example of courts using corpus linguistics, for example, all justices of Utah’s supreme court weighed in on the results of Justice Lee’s corpus linguistics analysis, but they were deeply divided as to the implications of his findings for the statutory meaning at issue and the applicability of the rule of lenity. *State v. Rasabout*, 356 P.3d 1258 (Utah 2015); *see also* *People v. Harris*, 885 N.W.2d 832 (Mich. 2016).

181. Lee & Mouritsen, *supra* note 172, at 829; *see also* Solan & Gales, *supra* note 109, at 263 (“‘Ordinary meaning,’ especially as applied to particular words and phrases, is a distributional fact.”).

182. Lee & Mouritsen, *supra* note 172, at 795 (emphasis added).

183. Of course, relatively de-contextualized frequency of the sort addressed by corpus linguistics may often be probative evidence of ordinary meaning in the sense I describe in this

following reasons: relative frequency in the corpus may simply reflect the *frequency of the underlying phenomenon*¹⁸⁴ or the *newsworthiness of its occurrence*¹⁸⁵ in the time and place generating the sources that form the corpus¹⁸⁶—yet those things should not influence our ordinary meaning determination.¹⁸⁷ If it were to turn out, for example, that but-for causes occur more often or are more frequently reported on than other sorts of causes, we should not conclude that they are the only kind of “cause,” nor even that they are the most likely kind of “cause” implicated in a given statutory or factual context.

In any event, as we saw with respect to the cases discussed in Par I, courts’ ordinary meaning analysis tends to focus on whether or not it feels “natural” or “appropriate” to apply a given term or phrase to a particular example,¹⁸⁸ rather than focusing on the frequency with which situations like those described in the example arise.¹⁸⁹ And for good reasons—namely, those examined in Part II. A wholly frequency-based regime like that proposed in much of the law and corpus linguistics literature, in other words, would risk undermining many of the values that support

Article. It may even be *more* probative than judges’ usual methods, i.e., use of introspection supplemented by intuition pumping examples and dictionary definitions. And indeed, in some statutory disputes (e.g., concerning terms in old statutes the meaning of which has changed significantly over time), relatively decontextualized frequency in sources from the time of enactment may be more probative of historical ordinary meaning *than would a modern survey* (that is, it may be more probative, compared to a modern survey, of the likely results that a hypothetical survey administered at the time of enactment would have yielded).

184. Ethan J. Herenstein, *The Faulty Frequency Hypothesis*, 70 STAN. L. REV. ONLINE 112, 116–17 (2017). *But see* Lee & Mouritsen, *supra* note 172, at 874 (contending that frequency of the underlying phenomenon reveals the word meaning or sense that “seems likely to be the one that first comes to mind when we think of” the word or phrase at issue, and for reasons of fair notice, ordinary meaning interpretation ought to track only that “top-of-mind” sense). On the tendency of corpus linguistic approaches to favor interpretations that do not extend much beyond prototypical usage of disputed terms, see Kevin P. Tobia, *Testing Public Meaning: Are Dictionaries and Corpus Linguistics Reliable Measures of Meaning?* (2019) (unpublished manuscript on file with the *Indiana Law Journal*), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3266082 [<https://perma.cc/8J4Q-R9NW>].

185. *See* Herenstein, *supra* note 184, at 118–19.

186. *See* Solan & Gales, *supra* note 109, at 259.

187. Herenstein, *supra* note 184, at 116. On potential uses of corpus linguistics in legal interpretation that do not fall prey to these frequency-based critiques, see NEAL GOLDFARB, *CORPUS LINGUISTICS IN LEGAL INTERPRETATION: WHEN IS IT (IN)APPROPRIATE?* (2019) (unpublished manuscript on file with the *Indiana Law Journal*), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3333512 [<https://perma.cc/6UBS-BQEN>].

188. *See e.g.*, *Burrage v. United States*, 134 S. Ct. 881, 891 (2014); *Gross v. FBL Fin. Servs., Inc.*, 557 U.S. 167, 176 (2009) (quoting *Safeco Ins. Co. of Am. v. Burr*, 551 U.S. 47, 63–64, n.14 (2007)); *id.* at 180; *Miller*, 767 F.3d 585, 591 (6th Cir. 2014); *see also* *Johnson v. United States*, 529 U.S. 694, 718 (2000) (Scalia, J., dissenting) (“[T]he acid test of whether a word can reasonably bear a particular meaning is whether you could use the word in that sense at a cocktail party without having people look at you funny.”).

189. *See* Carissa Byrne Hessick, *Corpus Linguistics and the Criminal Law*, 2017 BYU L. REV. 1503, 1506 (arguing that corpus linguistics’ exclusive focus on “how frequently a term is used in a particular way . . . represent[s] [a] significant departure[] from current theories of statutory interpretation”).

ordinary meaning interpretation in the first place, especially as they relate to criminal and tort law.¹⁹⁰ As one scholar recently put the point,

Proponents of corpus linguistics believe that the frequency with which words are used in various non-legal publications can tell us how a member of the public would understand those words when they appear in statutes. While there are good reasons for courts to interpret statutes—especially criminal statutes—in a fashion that is consistent with how an ordinary member of the public would understand the statute, the frequency with which a term is used does not give us that information.¹⁹¹

What, then, *would* give us that information?

IV. EXPERIMENTAL EVIDENCE OF ORDINARY MEANING AND CAUSAL ATTRIBUTION

This Part proposes a novel answer: to determine how ordinary speakers would apply a disputed word, phrase, or concept in a given context, ask them to apply it and observe what they do. More specifically, ask a lot of them, and do so via a controlled experiment, so that subtle variations (in precise wording, in factual context, etc.) can be altered as needed to ensure that the results are robust and informative. Before demonstrating the approach in action, let's briefly examine why it holds promise as an improvement over introspection, intuition pumps, dictionaries, and corpus linguistics.

A. Advantages and Limitations of Experimental Statutory Interpretation

Recall that, as a method of ascertaining ordinary meaning, judicial introspection faced problems stemming from its sample size of one—an unreliable basis for drawing inferences about the general population, especially insofar as the judge's own biases and motivated reasoning distort the analysis. Survey experiments, on the other hand, can actually sample the general population, letting idiosyncratic intuitions and biases cancel each other out. As for less idiosyncratic biases, much irrelevant and potentially biasing information—for example, the defendant's skin color—can simply be omitted from the experimental materials, unlike in a bench or jury trial.¹⁹² Other possible influences on interpretation (e.g., anti-drug-dealer sentiment in interpreting the CSA's penalty provisions) can be experimentally manipulated or controlled for to the extent they represent biases.¹⁹³

Intuition pumps, like introspection, similarly fare poorly in terms of objectivity: due to the realities of “intuition pump shopping,” they provide very little constraint and present significant risk of bias. And recall that intuition pumps run into accuracy problems due to the potential for subtle disanalogy in factual or linguistic context (though, to their credit, intuition-pumping examples at least make that surrounding

190. *Id.* (discussing notice and accountability); *see also supra* Sections II.B–C.

191. Hessick, *supra* note 189, at 105–06; *cf.* Baude & Sachs, *supra* note 10, at 1144–45; Fallon, *supra* note 10, at 1246.

192. *Cf.* Bennett Capers, *Evidence Without Rules*, 94 NOTRE DAME L. REV. 867 (2018).

193. *See, e.g., infra* Section IV.B.3.b.

context explicit). Again, survey experiments represent an improvement in both accuracy and objectivity because they can ask participants about the actual situation and wording in question. To the extent one suspects that the survey has presented participants with some subtly disanalogous variation of the linguistic and factual context, one can manipulate those features via controlled experiment to reveal whether they do in fact render the survey results untrustworthy. And of course, to the extent one suspects that the survey creator has engaged in another sort of “shopping”—namely, running multiple surveys and reporting only the results that support her position—replication tests and other devices can help separate the reliable from the unreliable (as they already do in areas like trademark law, where courts routinely consider expert witnesses’ dueling surveys).¹⁹⁴

As for dictionaries, recall that they risked serious inaccuracy, and raised related objectivity concerns, due to their failure to sufficiently account for factual and linguistic *context*, whereas surveys are constructed to include the relevant context. And recall that dictionary definitions often fail to even address the disputed aspect of the word or phrase at issue. Survey experiments, on the other hand, can be designed to disambiguate between the candidate meanings at issue in an actual case. And even when they do not clearly favor one position over another, surveys, unlike dictionaries, can at least reveal the *extent* to which a given term is vague or ambiguous in the context at issue—a relevant consideration in criminal law given rules like lenity and void-for-vagueness.¹⁹⁵

Finally, while corpus linguistics may in many cases roughly equal survey experiments in terms of objectivity and determinacy, survey experiments allow for comparatively greater *accuracy*, because they track actual usage in a very specific and controlled context, rather than frequency of usage across all contexts represented in the corpus (which in any event may reflect irrelevant factors such as the frequency of the underlying event and its newsworthiness).¹⁹⁶

Of course, no method of ascertaining ordinary meaning is perfect, and it is worth examining some limitations on survey experiments before finally observing them in action. First, surveys test *current* usage and concepts, not historical ones; where folk usage and concepts have substantially shifted postenactment, surveys are likely to be less useful.¹⁹⁷ Still, such cases may be relatively unusual.¹⁹⁸ Indeed, courts—like

194. See Ben-Shahar & Strahilevitz, *supra* note 15, at 1807–08; Daniel J. Hemel & Lisa Larrimore Ouellette, *Public Perceptions of Government Speech*, 2017-2 SUP. CT. REV. 33, 59–60 (2017).

195. See *supra* notes 131, 133 and accompanying text; *infra* notes 261–66 and accompanying text.

196. Whereas corpus linguistics’ focus on relatively de-contextualized frequency data may give rise to notice concerns, see notes 184–90 and accompanying text, surveys directly answer the question of notice: if someone contemplating whether to do *x* had looked dispassionately at the relevant statutory provision, would they have understood it to apply to *x*? Indeed, it is hard to imagine a method more suited to addressing notice concerns.

197. Cf. Lawrence B. Solum, *The Fixation Thesis: The Role of Historical Fact in Original Meaning*, 91 NOTRE DAME L. REV. 1, 22 (2015).

198. For especially old statutes, and for disputes over statutory terms that one has reason to suspect have changed considerably over time, corpus linguistics may well provide

philosophers, psychologists, and everyday folk—rightly treat many everyday concepts as stable over fairly broad stretches of time, especially core concepts like “knowledge,”¹⁹⁹ “intent,”²⁰⁰ and “causation,”²⁰¹ and core logical terms like “any,” “if,” and “would.”²⁰² Hence judges’ (and philosophers’, psychologists’, etc.) frequent recourse to their own, present-day intuitions and linguistic usage when determining ordinary meaning even as a historical matter.²⁰³ In any event, the historical drift concern is of little moment for the causation language found in recently enacted legislation such as the criminal sentencing, hate crime, and discrimination statutes considered in Part I and tested below.²⁰⁴

Second, even though surveys have become increasingly cheap to conduct online, they still cost money (a few thousand dollars for a large-scale, nationally representative survey like the one reported below). Granted, this is a drop in the bucket for many litigants.²⁰⁵ Still, for most cases, courts and litigants will continue to deem introspection, intuition pumps, and dictionaries sufficient, and this will often be perfectly sensible in light of cost concerns. Here again, though, investigating core psychological concepts and logical terms avoids much of the worry: the very fact that these terms appear throughout federal and state codes, and that even their appearance in a single place like Title VII can form the basis of so many discrete cases, means that experimental research can have a potentially large and wide-ranging payoff. Insofar as parties and courts lack the resources and incentives not only to carry out experimental research but also to make the results public, academic studies can fill some of the void. And even the mere awareness that such data may someday be forthcoming might usefully temper the temptation (whether on the part of judges, lawyers, legal theorists, or philosophers) to treat arguments from ordinary

significantly more probative evidence of original public meaning than surveys would provide. See *supra* note 188.

199. Cf. Macleod, *supra* note 164.

200. Cf. Julia Kobick, *Discriminatory Intent Reconsidered: Folk Concepts of Intentionality and Equal Protection Jurisprudence*, 45 HARV. C.R.-C.L. L. REV. 517 (2010).

201. Cf. sources cited *supra* notes 36–68; sources cited *infra* notes 311–18.

202. See Gluck & Bressman, *supra* note 117, at 955 (citing additional examples).

203. See *infra* notes 270–77 and accompanying text (discussing the folk concept of “punishment”). For examples of cross-cultural (as opposed to inter-temporal) uniformity in folk conceptual understandings concerning what I’ve labeled “core” folk concepts and logical terms, see sources cited in Knobe & Nichols, *supra* note 25, § 2.1.1.

204. The provision of the CSA at issue in *Burrage* was enacted in 1970. *Burrage v. United States*, 134 S. Ct. 881 (2014). The ADEA, at issue in *Gross*, was enacted in 1967. *Gross v. FBL Fin. Servs.*, 557 U.S. 167, 176 (2009). Title VII, at issue in *Nassar*, was enacted in 1964. Changes to other parts of Title VII that could arguably have altered the meaning or implication of the causal language at issue in *Nassar* took place in 1991. *Univ. of Tex. Sw. Med. Ctr. v. Nassar*, 570 U.S. 338 (2013). The HCPA, at issue in *Miller*, was enacted in 2009. *United States v. Miller*, 767 F.3d 585 (2014).

205. See *supra* notes 18–20 and accompanying text. And in trademark law and several other areas, private parties already routinely pay for and introduce via expert witnesses survey evidence similar to that reported below—indeed, they did so even before today’s (far less expensive) online surveys were possible. Ben-Shahar & Strahilevitz, *supra* note 15, at 1769–70, 1805–08. See generally J. THOMAS MCCARTHY, 4 MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 32:158 (Thomson Reuters 5th ed. 2017).

meaning, plain meaning, and “commonsense” intuitions as if they are immune from rigorous criticism.

* * *

In sum, then, although not every statutory case, term, or concept will warrant experimental examination, the case for surveys is at its strongest for recent legislation, such as the statutory tort and criminal legislation discussed in Part I, and with respect to everyday folk concepts like causation. Here, in other words, is an area where litigants, judges, and (as we’ll see in Part V) even legal theorists and philosophers could benefit from empirical data. Yet despite the plentiful armchair theorizing about causation in law, there is surprisingly little publicly available empirical research on point.²⁰⁶

B. The Experiments

Having argued for the experimental method’s potential as a way of determining ordinary meaning, this Part now offers a demonstration. It describes a proof-of-concept survey based on statutory causation cases discussed in Part I. After explaining who the survey’s participants were, how it was designed, and what materials were used, it reports the results and examines what they reveal about the ordinary meaning of causal language and about courts’ ordinary meaning decisions more broadly.

1. Participants

The survey was administered online to a nationally representative sample recruited by Toluna, a well-regarded survey research firm with an established

206. The only study in the legal literature concerning whether uninstructed laypeople ascribe causality in the absence of but-for causation is from 2001. *See* Spellman & Kincannon, *supra* note 72. The authors tested an overdetermined murder scenario (two shooters simultaneously kill one victim). Participants (an unspecified number of college students) ascribed greater causality and assigned greater punishment in the overdetermined double-murderer scenario than in the alternative scenario where death would *not* have resulted if either shot had missed. *Id.* at 253. In the late 1990s, Erich J. Greene and John Darley tested punishment preferences, but not causality ascription, in a similar two-murderer scenario where but-for causation was present and each murderer’s contribution was either independently sufficient or insufficient to cause the victim’s death. *See* Erich J. Greene and John M. Darley, *Effects of Necessary, Sufficient, and Indirect Causation on Judgments of Criminal Liability*, 22 LAW & HUM. BEHAV. 429 (1998) (finding that, where but-for causation is satisfied, independently sufficient murderers were allotted more punishment than independently insufficient murderers); *see also* ROBINSON, *supra* note 79, at 392 (noting the need for further research on this point); *cf. infra* note 311 (citing sources critical of Hart and Honoré and subsequent work on causal intuitions in law as lacking empirical support beyond the authors’ own intuitions).

panel.²⁰⁷ The sample consisted of 1486 participants,²⁰⁸ and was closely matched with U.S. census data to be representative on gender, age, race, ethnicity, and education.²⁰⁹ Participants were not told that the study had anything to do with law or interpretation.²¹⁰ The study was conducted with the approval of the Columbia University Institutional Review Board.

2. Design and Materials

Each participant was randomly assigned to read a short vignette modeled on one of the three main cases examined in Part I—specifically, *Burrage* (did death “result from” the drug sold by the protagonist?); *Miller* (did the protagonist assault the victim “because of” the victim’s religion?); and *Gross* (did the protagonist terminate the employee “because of” the employee’s age?).²¹¹ (As explained below, a fourth vignette mirrored the *Burrage* vignette but with slight variations that eliminated the impulse to morally blame.²¹²) For ease of reference, call the drug, the religious hatred, and the ageism the “focal causes” in each vignette (as opposed to the “nonfocal causes,” which were, respectively, the other drugs the victim took in *Burrage*, the other motivations the attacker had for assaulting the victim in *Miller*, and the other reasons the employer had for firing the employee in *Gross*).

207. See *About Toluna Group*, TOLUNA, <https://www.toluna-group.com/about-toluna-group/about> [https://perma.cc/EP25-MW2A].

208. From an initial sample of 1722, 236 participants were screened out based on prespecified criteria, leaving 1486. Specifically, participants who finished the survey in less than one-third of the median completion time (as fifty-five participants did) were excluded, and participants who failed a simple attention check (as 181 participants did) were excluded. See *infra* note 214 for the text of the attention check.

209. The mean age of the 1486 respondents was 47, with a range of 18–92 and a standard deviation of 17 years. Females comprised 50% of the sample. For race, 76% of the sample self-identified as White, 12% as Black, 6% as Asian, Native Hawaiian, or Pacific Islander, 1% as Native American or Alaska Native, and 6% as Other. On a separate question, 16% of the sample reported that they are Latino or Hispanic. For education, 9% of the sample had not finished high school, 22% had high school diplomas, 22% had some college experience, 12% had 2-year college degrees, 22% had 4-year college degrees, 10% had master’s degrees, and 3% had doctorates or other professional degrees.

210. One potentially interesting follow-up study would test whether the results change when participants are told that the study concerns legal interpretation, and/or when they are provided even greater contextual enrichment (e.g., when they are told to imagine that they are a juror in a case).

211. Participants receiving one of the motive-as-cause vignettes (i.e., *Miller* and *Gross*) received *both* motive-as-cause vignettes, with their order randomized and with the same necessity/sufficiency condition across both vignettes. Participants receiving any version of the *Burrage* vignette did not receive another vignette. Participants’ responses to the *Miller* and *Gross* vignettes showed no statistically significant order effects. Moreover, in all the results reported below, neither the direction nor the fact of statistical significance would change if one were to first remove from the data all responses to the second vignette a given participant received and then recalculate.

212. See *infra* Section IV.B.3; App. at 1028.

The key experimental manipulation for each vignette concerned the focal cause—specifically, whether the focal cause was *necessary* or *unnecessary* to bring about the state of affairs in question (i.e., the death, assault, or employment termination), and whether the focal cause was independently *sufficient* or independently *insufficient* to bring about that state of affairs. So, after having been randomly assigned to one of the four vignettes (*Burrage*, *Miller*, *Gross*, or the low-blame variation of *Burrage*), participants were randomly assigned one of the four possible necessity/sufficiency conditions for that given vignette—namely, focal cause [1] necessary and sufficient; [2] necessary and *insufficient*; [3] *unnecessary* and sufficient; or [4] *unnecessary* and *insufficient*. In total, then, participants were randomly assigned to one of 16 possible stories (i.e., four vignettes, each of which had four versions).

As an example, consider the vignette based on *Burrage*, which is reproduced below. The full text of the other vignettes is reproduced in the Appendix. Whereas the *Miller* and *Gross* vignettes contain only two potential causes, and are thus simpler in structure, the vignette based on *Burrage* contains three—the focal drug (“Fintene”) and two other drugs—so as to more closely mirror the facts in most of the “mixed drug intoxication” CSA cases cited above, including *Burrage* itself.²¹³

DRUG OVERDOSE (*Burrage*):

Fintene, Rextor, and Tamphen are dangerous and illegal recreational drugs. A typical dose of each makes users feel high. But unusually potent doses are very dangerous, and can even be deadly. Drug dealers sell them in powder form, in small plastic bags. To take them, users mix the powder into a glass of orange juice and drink it. Taken this way, Fintene, Rextor, and Tamphen take about an hour to become absorbed into the body.

On Tuesday morning, Josh met up with three different drug dealers. One sold Josh a dose of Fintene, another sold Josh a dose of Rextor, and the other sold Josh a dose of Tamphen. Josh wanted to have a good time, so he decided that that afternoon, he would take all three of the drugs together.

What Josh didn’t know was that the Fintene, Rextor, and Tamphen he bought were all unusually potent. In fact, . . .

213. See, e.g., cases cited *supra* notes 6–9.

[1] Necessary & Sufficient	[2] Necessary & Insufficient	[3] Unnecessary & Sufficient	[4] Unnecessary & Insufficient
... although neither the Rextor by itself, nor the Tamphen by itself, nor the combination of Rextor and Tamphen together would have killed Josh, taking the Fintene by itself would have killed Josh. And...	... although neither the Fintene by itself nor the combination of the Rextor and Tamphen together would have killed Josh, the Fintene by itself would have killed Josh, the Rextor by itself would have killed Josh, and the Tamphen by itself would have killed Josh. Taking any two of the three drugs together would also have killed Josh. And...	... although neither the Fintene by itself, nor the Rextor by itself, nor the Tamphen by itself would kill Josh, taking any two of the three drugs together would kill Josh. And...

... taking all three drugs together would kill Josh.

That afternoon, Josh went home, combined all the powder from all three bags, mixed it into a glass of orange juice, and drank it. Sure enough, after about an hour, the Fintene, Rextor, and Tamphen combined in Josh’s bloodstream, blocking the flow of blood to Josh’s heart, and Josh died.

After reading whichever vignette they were assigned, participants answered a set of questions concerning: the applicable statutory language (for ease of reference, call this the “statutory causation” question); common doctrinal standards / jury charges interpreting that language (call these the “alternative causation standards” questions); and the moral blameworthiness and punishment-worthiness of the actor who would later become the defendant (call these the “blameworthiness” questions).²¹⁴ So, for

214. An attention check question asked, “Which drug or drugs did Josh take?” Participants who failed to select the correct answer, “All three of Fintene, Rextor, and Tamphen,” were excluded. The attention check question for the other vignettes followed this same structure. Note that, for those who received *Miller* and *Gross* vignettes, responses to questions concerning a given vignette remained in the survey results so long as the participant answered the attention check question *as to that given vignette* correctly, even if he or she answered the attention check question wrong *as to the other vignette he or she received* (i.e., even if his or her answers in response to the other vignette were excluded). To reiterate, *see supra* note 211, this was possible only for participants who received *Miller* and *Gross* vignettes because those were the only participants who received two vignettes and corresponding sets of questions, rather than just one.

example, after reading a version of the *Burrage* vignette a participant would be asked whether Josh's death "resulted from the use of" the Fintene (the statutory causation question²¹⁵); whether Josh would still have died had he not used the Fintene (the but-for question); whether the use of the Fintene was a "substantial factor in bringing about Josh's death," and whether it was a "contributing factor" in bringing about his death (the alternative causation standards questions²¹⁶); and how morally blameworthy the drug dealer who sold Josh the Fintene was, as well as how much prison time he should receive (the blameworthiness questions²¹⁷).

After each of the statutory causation and alternative causation standards questions, participants were also asked how confident they were in their "yes" or "no" response (on a scale from 1 ("Not at all confident") to 7 ("Completely confident")). Finally, participants were asked to predict what percentage of *other* survey participants would give the same "yes" or "no" answer to the statutory causation question.²¹⁸ The order of the questions was as follows: (a) statutory causation question (followed by the confidence question); (b) in randomized order, the but-for causation and alternative causation standards questions (each of which was followed by the confidence question); (c) the blameworthiness questions; (d) the prediction of how many others agree with the participant's statutory causation answer; and (e) the attention check.

3. Results

a. Statutory Causation and the Failure of the But-For Test

Figure 1 shows participants' responses to the statutory causation question in the *Burrage*, *Miller*, and *Gross* case vignettes.

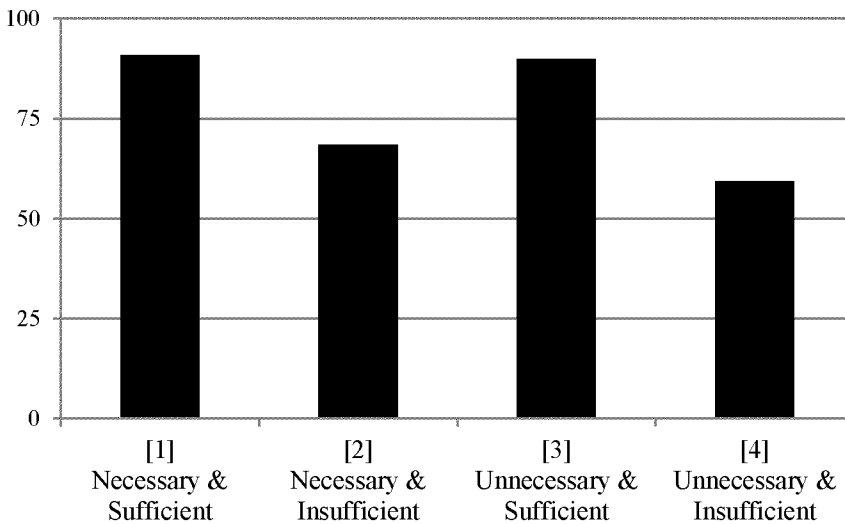
215. Whereas the operative statutory language in *Burrage* was "results from," 134 S. Ct. at 887, in *Miller* and *Gross* it was "because of," hence the statutory causation question for the *Miller* and *Gross* vignettes asked whether the protagonist did something "because of" some motivation, 767 F.3d 585; 557 U.S. at 167.

216. In *Miller* and *Gross*, in line with the most popular jury instruction standards, the substantial factor question concerned whether the focal motive was a "substantial motivating factor." 767 F.3d at 592, 603; 557 U.S. at 171, 183. Since motive-as-cause instructions less frequently use a "contributing factor" instruction, the *Miller* and *Gross* materials did not include that question, whereas the *Burrage* materials did. Since motive-as-cause cases sometimes use a "sole motivating factor" instruction, the *Miller* and *Gross* materials included that question whereas the *Burrage* materials did not.

217. Answers to the "morally blameworthy" question were on a forced seven-point scale, from 1 ("Not at all blameworthy") to 7 ("Extremely blameworthy"). Answers to the prison question were on a forced seven-point scale, from "0–4 years" to "30–34 years." Since prison time is not an option for age discrimination under the ADEA, the *Gross* vignette had no prison question.

218. Specifically, using *Burrage* as an example, "Earlier, when you were asked whether Josh's death 'resulted from' the Fintene, you answered ['Yes'/'No']. 100 other people are taking this survey. Out of the 100, how many do you think will also answer ['Yes'/'No'] to that same question? (Please provide a number between 0 and 100.)."

Figure 1. Percentage Ascribing Causation Using Statutory Language (Burrage, Miller, Gross combined; $N = 1410$)



Two main findings stand out in assessing the courts' but-for test for statutory causation. Importantly, both findings apply not just to all three cases (*Burrage*, *Miller*, and *Gross*) considered *together* (as pictured in Figure 1 above) but also to each of the three cases considered *individually* (as pictured in Appendix Figures A, B, and C). While all three cases are combined here for purposes of exposition, interested readers can consult the Appendix to view the results as to the three cases individually. Now, on to the two findings.

First, where but-for causation was absent (i.e., where the focal cause was unnecessary), a substantial majority (74%) of study participants nonetheless found the statute's causal language to be satisfied (this can be seen by averaging the third and fourth bars in Figure 1). Indeed, even considering only the experimental condition in which the focal cause was *neither necessary nor sufficient*, a clear majority (60%) of study participants found statutory causation satisfied (as represented by the fourth bar in Figure 1).²¹⁹ In short, the "but-for" test appears to be

219. This first main finding cannot be dismissed as resulting from some sort of misunderstanding about whether but-for causation was in fact present. For each vignette, and for each experimental condition, most participants answered the but-for question correctly—including, importantly, the vast majority of those who mistakenly deemed statutory causation satisfied in response to a condition where but-for causation was *unsatisfied*. See App. figs. A, B, C. In other words, the very same people ascribing statutory causation in the absence of but-for causation understood that but-for causation was not satisfied. Indeed, if one removes from the sample of those who mistakenly found but-for causation satisfied in conditions [3] and [4], where it was in fact not satisfied, statutory causation ascription in those conditions drops by only one percent. And the same goes for condition [4] considered by itself—i.e., for the condition where "yes" responses to the statutory causation question were most unexpected. Stated differently: even in the "unnecessary and independently insufficient" condition of each

overly restrictive: most people confidently allege that *y* “resulted from” *x*, and that *y* occurred “because of” *x*, while simultaneously and confidently alleging that *x* was not a but-for cause of *y*.²²⁰ This finding directly contradicts the courts’ pronouncements about ordinary meaning and common understanding surveyed in Section I.A.

Second, the presence or absence of but-for causation appeared to make far less difference in statutory causation ascription than did the presence or absence of independent sufficiency. Specifically, whereas statutory causation ascription increased by 8% where but-for causation was present, it increased by 41% where independent sufficiency was present.²²¹ Again, this is striking given that the cases in

vignette (i.e., condition [4]), one could exclude all responses from those who mistakenly deemed but-for causation satisfied, and it would still clearly remain the case that, of the remaining participants, most deem statutory causation satisfied (for *Burrage*, 58%; *Miller*, 58%; and *Gross*, 60%). Considering all three cases together—and, once again, excluding all those participants who mistakenly ascribed but-for causation—an average of 59% of the remaining participants ascribed statutory causation even in condition [4]. A chi-squared test shows that that 59% average is significantly different from 50%. $\chi^2(1, N=323)=9.365$, $p=0.002$.

A chi-square test is a statistical analysis used to examine the relationship between two discrete variables. The null hypothesis generates expected frequencies that are tested against the observed frequencies. If the frequencies are similar, the null hypothesis is not rejected; if they are sufficiently different, the null hypothesis can be rejected. In what follows, all dichotomous measures (i.e., binary “yes” or “no” responses) were statistically analyzed using chi-square tests.

220. See *supra* note 219; *infra* Section IV.B.3.d.

221. Adding necessity increased statutory causation attribution from 74% to 80%: $\chi^2(1, N=1410)=5.837$, $p=0.015$. Independently, adding sufficiency increased statutory causation attribution from 64% to 90%: $\chi^2(1, N=1410)=139.441$, $p<0.001$.

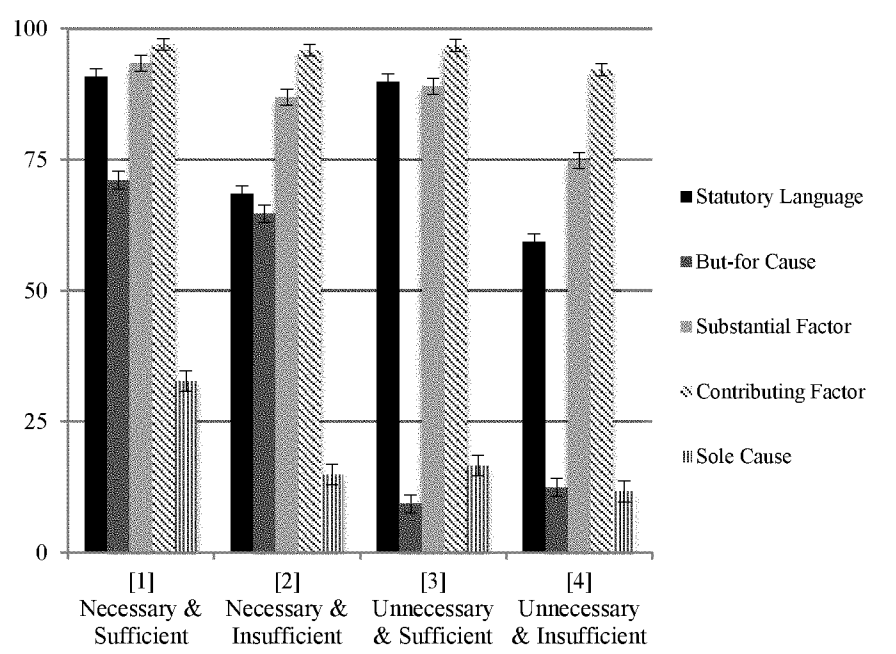
A note of caution is perhaps in order, lest we too hastily conclude that independent sufficiency is truly fundamental to commonsense causal cognition. Consider an alternative explanation for the data regarding the importance of independent sufficiency. It could be that, rather than independent sufficiency *per se* doing the work, independent sufficiency is instead functioning as a mere indicator from which participants *infer* the *absolute amount* of the focal cause that’s present. So, for example, in the *Burrage* vignette participants are not told the quantity of the drug that was sold—it could have been 5 milligrams, 100 milligrams, etc. Perhaps participants who are told that the drug was independently sufficient to bring about the death assume something closer to 100 milligrams, whereas participants who are told that the drug was independently *insufficient* assume something closer to 5 milligrams. And perhaps it is *that assumption* (rather than the presence or absence of independent sufficiency *per se*) that leads participants in the independently sufficient conditions to more readily ascribe causation. I’m skeptical of this alternative explanation, but the data here do not rule it out. And note that it is arguably a less dubious explanation as applied to motive-as-cause cases like *Gross*. So, for example, in the *Gross* vignette, participants were not told the absolute amount of ageism present in the employer’s mind (just as, in *Burrage*, they were not told the absolute amount of the drug that was sold); perhaps when participants learn that the employer’s ageism was independently sufficient to lead him to fire an employee, they *infer* that he is *very* ageist, and again, perhaps it is *that assumption*—an assumption about the absolute amount of the forbidden motive present in the mind of the actor—that leads them to more readily ascribe causation where independent sufficiency is present as opposed to absent. Again, the data here

Section I.A focused exclusively on but-for causation and either neglected to address or actively rejected sufficiency as a factor in commonsense or statutory causal attribution.

b. Alternative Causation Tests

If, contrary to the cases described in Section I.A, a but-for standard so significantly diverges from the most natural interpretation of the relevant statutory language, then the question arises: do any of the three alternative causation tests examined in Section I.B (“contributing factor,” “substantial factor,” and “sole cause”) fare any better?

Figure 2. Percentage Ascribing Causation Using Statutory Language Compared to Alternative Causation Standards (Burrage, Miller, Gross combined; N = 1410)²²²



Neither the “sole cause” nor the “contributing factor” test fares especially well (though the contributing factor test does come closer to tracking statutory causation ascription than does the “but-for” test). More specifically, as shown in Figure 2, the sole cause test is uniformly—and drastically—underinclusive in all four necessity/sufficiency conditions (this can be seen by comparing, in each of the four clusters of bars, the height of the first bar versus the height of the last bar). And the contributing factor test—despite faring slightly better than the but-for test—is

don’t address, let alone rule out, this alternative explanation for the influence of independent sufficiency.

222. Error bars represent .95 confidence intervals.

uniformly over-inclusive (as can be seen by comparing the first bar in each cluster with the second-to-last bar).

The “substantial factor” test (represented by the middle bar in each cluster) comes closer to tracking statutory causation ascription across the four cases than does any of the other tests—including, most importantly, the but-for test (represented by the second bar in each cluster).²²³ And again, although all three of the cases (*Burrage*, *Miller*, and *Gross*) have been collapsed together in Figure 2 for ease of exposition, this finding concerning the substantial factor test’s superior performance also applies to each of the three cases considered individually, as shown in Appendix Figures A, B, and C.

c. Tracking Blame, but Not Merely Blaming

Participants’ moral blame judgments and punishment preferences both tracked statutory causation ascription (and the “substantial factor” test) more closely than they tracked but-for causation.²²⁴ Put another way, moral blame and punishment

223. This can be seen by comparing the first bar in each of the four clusters with, on the one hand, the second bar (representing but-for responses), and on the other hand, the third bar (representing substantial factor responses).

Importantly, the same result appears when one examines individual survey participants’ responses to the three questions: A given participant was more likely to provide the same answer (whether “yes” or “no”) to (a) both the statutory causation question and the *substantial factor question*, than to (b) both the statutory causation question and the *but-for question*. In other words, individual survey participants treated the latter pair of questions as asking about two different things, calling for two different answers; they treated the former pair of questions as asking the same thing, calling for the same answer. This can be seen using McNemar’s test, a form of chi-square that tests the significance of the difference between two responses made by the same participants. Condition [1]: statutory - but-for $\chi^2(1, N=349)=53.149$, $p<0.001$; statutory - substantial factor $\chi^2(1, N=349)=2.207$, $p<0.137$. Condition [2]: statutory - but-for $\chi^2(1, N=337)=0.814$, $p=0.367$; statutory - substantial factor $\chi^2(1, N=337)=32.078$, $p<0.001$. Condition [3]: statutory - but-for $\chi^2(1, N=355)=278.170$, $p<0.001$; statutory - substantial factor $\chi^2(1, N=355)=0.063$, $p=0.801$. Condition [4]: statutory - but-for $\chi^2(1, N=369)=144.310$, $p<0.001$; statutory - substantial factor $\chi^2(1, N=369)=26.353$, $p<0.001$.

As can be seen in Figure 2 and in these McNemar test results, the single exception to the substantial factor test’s superior performance is Condition [2]. There, the but-for test yielded results closer to the statutory causation language than did the substantial factor test—which, surprisingly, yielded *higher* causal attribution than the but-for test here as well as in the other condition in which but-for causation was in fact satisfied. To the extent judges are worried about false positives in cases of necessary and independently insufficient causes (it’s not clear why they would be), in a case where the evidence cannot support a determination that the focal cause was anything other than necessary and independently insufficient, the but-for test could be a better bet, though this is unclear.

224. For example, participants found the protagonists in the vignettes significantly *less* morally blameworthy, and allotted significantly *less* prison time, where the focal cause was [2] a but-for cause but not independently sufficient, compared to where the focal cause was [3] *not* a but-for cause but *was* independently sufficient. Scale measures (i.e., 1 to 7 for moral blameworthiness 0 to 34 years for prison sentences), were analyzed using analyses of variance. An analysis of variance (ANOVA) is a statistical test that examines the relationship between a discrete independent variable and a continuous dependent variable by comparing the

preferences tend to reflect the same basic pattern that statutory causation and substantial factor ascription followed, not the pattern that but-for causation ascription followed.²²⁵

While many theorists would be happy to learn that statutory causation language tracks moral blame and punishment preferences (and unhappy to learn that the courts' but-for test does not),²²⁶ others might worry that moral blame and punishment preferences are somehow distorting the experimental results, rendering them untrustworthy as sources of information about causal intuitions and ordinary meaning (as opposed to merely information about moral blame and punishment preferences).²²⁷

Here are two versions of that type of worry. First, some might worry that, in all four necessity/sufficiency conditions, survey participants' "yes" responses were

variances (averages of the squared deviations from the mean) in order to evaluate the probability of not rejecting the null hypothesis that the predicted experimental effect is absent. Here are the results comparing condition [2] to [3]: Moral blame: $F(1, 690)=13.804$, $p<0.001$; Prison Sentence: $F(1, 469)=14.367$, $p<0.001$. Separating out the three individual cases, these effects were in the same direction in all three but statistically significant only in the *Burrage* vignette. For *Burrage*, where the focal cause was unnecessary and sufficient, the mean blameworthiness assessment was 6.0, where the focal cause was necessary and insufficient, 4.8. This was a highly significant difference. $F(1,246)=29.522$, $p<0.001$. Prison sentences followed the same pattern, with mean sentences of 18.5 years compared to 11.5 years, also highly significant. $F(1,246)=23.521$, $p<0.001$. However, in *Miller* and *Gross*, blameworthiness and punishment preferences were too clustered around the top of the blameworthiness scale to show significant differences (in *Miller*, 6.6 vs. 6.5, $F(1, 223)=0.036$, $p=0.85$; in *Gross*, 5.3 vs. 5.2, $F(1, 217)=0.293$, $p=0.59$), and in the hate crime prison sentences, were too clustered around the bottom of the scale (6.5 years vs. 6 years, $F(1, 223)=0.195$, $p=0.66$). (In *Gross*, which concerns civil employment discrimination, participants were not asked to allot prison time.) Future research might usefully employ a nonlinear scale, or at least a scale with lower prison sentences where less serious crimes are tested, to avoid similar clustering. See, e.g., ROBINSON, *supra* note 79, at 392.

225. This means, in effect, that like in Figure 1, the blameworthiness and punishment ratings are, very roughly speaking: [1] = high; [2] = low; [3] = high; [4] = low. Here, more precisely, are the mean moral blameworthiness ratings in the four necessity/sufficiency conditions, combining all three case vignettes: [1] = 5.66; [2] = 5.48; [3] = 5.96; [4] = 5.27. Here are the mean punishment preference (i.e., prison time) ratings in the four necessity/sufficiency conditions, combining the two criminal case vignettes (*Burrage* and *Miller*, i.e., the only two vignettes for which the question was asked): [1] = 3.39; [2] = 2.77; [3] = 3.47; [4] = 3.04.

226. See *supra* Section II.C.

227. For more information about blame-based motivated reasoning and "blame-first" error theories of causal attribution, see Mark D. Alicke, David Rose & Dori Bloom, *Causation, Norm Violation, and Culpable Control*, 108 J. PHIL. 670, 675 (2011). To be clear, virtually nobody thinks commonsense causal intuitions have nothing to do with picking out factors "responsible" for some occurrence. But responsibility isn't necessarily the same as moral blameworthiness (or punishment deservingness). The error theory under consideration in the text relates to the influence of the desire to morally blame and/or to punish, not the (uncontroversial) influence of responsibility judgments *per se*. Hence this Article's frequent use of "moral blame" as opposed to merely "blame," in case "blame" could be mistaken for mere "responsibility."

“inflated” by their impulse to express moral condemnation. That first worry concerns across-the-board, uniform inflation of the rate of “yes” responses.²²⁸ The second worry concerns the *pattern* of “yes” responses across the four necessity/sufficiency conditions—specifically, whether sufficiency would make less of a difference in statutory causation attribution (compared to necessity) if the impulse to morally blame and punish were absent. For those who feel moral blame and punishment preferences ought not to influence statutory causation ascription, the upshot of these worries—if they were to find empirical support—could be that the judges endorsing the but-for entailment thesis in the cases in Section I.A were not so wrong after all. Indeed, perhaps judges would thus be shown to be experts at setting aside their moral blame and punishment impulses and seeing the ordinary meaning of the relevant language (which, on this theory, really does entail but-for causation), whereas laypeople’s interpretive intuitions are clouded by the moral valence of the particular case (leading them to over-ascribe causation, and to do so especially in cases where independent sufficiency is present). The question, though, is whether either worry is warranted as an empirical matter.

To test these two types of worry, some participants received a version of the *Burrage* vignette that was virtually identical in structure and wording, except that it was tweaked to diminish the impulse to morally blame or punish. Specifically, it substituted the original vignette’s illegal drug for plant food, the blameworthy drug dealer for a blameless gardening store, and the death of a human victim for the death of a houseplant.²²⁹ Again, the goal was to measure whether participants’ statutory causation attributions would look roughly the same—both in across-the-board magnitude (the first worry) and in pattern of responses to necessity and sufficiency (the second worry)—in a context that did not trigger their impulse to punish or express moral condemnation.

The test worked, and the two worries proved largely unfounded. Participants assigned nearly no moral blame,²³⁰ yet across the four necessity/sufficiency conditions the pattern of “yes” responses to the statutory causation question was practically identical to the pattern in the original *Burrage* vignette.²³¹ To be sure, participants ascribed statutory causation (and indeed, all types of causation) at somewhat lower rates across all four conditions, consistent with prior research concerning the role of moral blame in causality ascription.²³² Still, well over half

228. Some might be drawn to an extreme form of this worry—specifically, that participants simply answered “yes” to *all* causation questions, since doing so expressed that the protagonist was blameworthy. But as shown in Figure 2 above, few participants ascribed causation under the “sole cause” test in any of the four conditions, and few ascribed causation under the but-for test in conditions [3] and [4], where but-for causation was in fact absent. So to the extent that blame-based motivated reasoning played a role, see *infra* notes 237–38 and accompanying text, it was clearly not so stark as to result in participants ascribing causation no matter what standard they were asked to apply. See *supra* Figure 2; *infra* note 229 and accompanying text.

229. See App. at 1028.

230. In *Burrage*, the mean blameworthiness was 5.3, in the Plant Food version, it was 2.0, a highly significant difference. $F(1,976)=878.623$, $p<0.001$.

231. See App. figs.A, D.

232. See, e.g., Janice Nadler & Mary-Hunter McDonnell, *Moral Character, Motive, and the Psychology of Blame*, 97 CORNELL L. REV. 255 (2012); see also Thomas F. Icard, Jonathan

(59%) continued to ascribe statutory causation absent but-for causation, and they were just as confident in their answers as were those who ascribed statutory causation in response to the original *Burrage* vignette.²³³

In short, even if we assume that moral valence should be completely irrelevant to these criminal and tort-like statutes' proper interpretation, and if we further assume that moral valence is irrelevant to commonsense causal cognition properly understood, *even then* but-for causation remains just as poor a proxy for the ordinary meaning of the relevant statutory causation language.²³⁴

d. Interpreter Confidence as Noise

Across the board, regardless of vignette or condition, participants tended to express a great deal of confidence in their responses to the statutory causation, but-for causation, and alternative causation standard questions. To be sure, there were differences among conditions and questions, but the overall pattern is of mean confidence ratings that, on a scale from 1 to 7 with 4 labeled "moderately confident" and 7 labeled "completely confident," remained almost entirely between 5.5 and 6.5.

Interpreter confidence, in short, does not appear to be a reliable indication of the degree to which others share the interpreter's intuitions.

The more interesting and decisive proof of that point is found in participants' explicit predictions regarding what percentage of survey participants would agree with their answer to the statutory causation question. These are shown in Table 1.

F. Kominsky & Joshua Knobe, *Normality and Actual Causal Strength*, 161 *COGNITION* 80 (2017).

233. See App. fig.D. On the practical implications of these different causality ascription rates in the low-blame *Burrage* vignette, see *infra* text accompanying notes 245–48.

234. It may even be *worse*, relative to independent sufficiency, here than in the original *Burrage* vignette: in response to the plant food vignette, while slightly less than half of participants (47%) ascribe statutory causation in the unnecessary and insufficient condition, that is a *greater* percentage than the percentage (42%) who ascribe it in the *necessary* and insufficient condition, where but-for causation is satisfied.

Table 1. *Actual Agreement and Predicted Agreement with Statutory Causation Responses* (Burrage, Miller, Gross combined; N=1410)

	<i>Actual Percentage Answering YES</i>	<i>Mean Estimated Percentage Answering YES</i>	<i>Actual Percentage Answering NO</i>	<i>Mean Estimated Percentage Answering NO</i>
[1] Necessary & Sufficient (N=349)	91%	82%	9%	64%
[2] Necessary & Insufficient (N=337)	69%	73%	31%	68%
[3] Unnecessary & Sufficient (N=355)	90%	82%	10%	57%
[4] Unnecessary & Insufficient (N=369)	59%	72%	41%	72%

Note the significant consensus bias exhibited by the minority of survey participants, who found statutory causation lacking (shown in the two bolded far right columns). In each of the four conditions, these participants predicted (inaccurately) that most people would agree with their interpretation of the relevant statutory language—in most instances predicting that more than twice as many people share their interpretive intuitions than was actually the case.²³⁵

C. Summary and Implications

The above results can be briefly summarized in terms of the following six findings. First, across and within all three cases, a clear majority ascribed statutory causation absent but-for causation (indeed, even absent both but-for causation *and* independent sufficiency).²³⁶ Second, independent sufficiency played a far greater

235. As for the majority of participants—who found statutory causation satisfied in a given condition—the results are less clear: on the one hand, where the focal cause was independently sufficient, they *underestimated* agreement; on the other hand, given that over 90% actually did find statutory causation in such cases, this underestimation might simply reflect hesitancy to predict, say, 100% agreement even in the seemingly clearest of cases. And where the focal cause was independently *insufficient*, participants ascribing statutory causation do predict greater agreement than is actually the case, even if the extent of their overestimation is far less than that of the minority of participants who found statutory causation absent.

236. Again, this remains true even if one excludes from consideration anybody who mistakenly believed but-for causation to be present when it was in fact not present. *See supra*

role than did but-for causation in altering statutory causation responses. Third, the substantial factor test more closely tracked statutory causation attributions than did the but-for test (or, for that matter, the sole cause test or contributing factor test). Fourth, participants' moral blame and punishment preferences likewise reflected greater concern with independent sufficiency than with but-for causation, and thus tracked the statutory causation language and the substantial factor test more closely than the but-for test. Fifth, and more generally, participants adopting a minority interpretation of statutory causation language were just as confident in their interpretation as those adopting the opposite (majority) interpretation, and consistently predicted that theirs was the majority position even though it was not. Sixth, and most generally, participants were responsive to small differences in causal language, deeming causation present under some causation standards and absent under others (including in high-blame contexts), thereby exhibiting coherent patterns of causal attribution not attributable to mere blame-based motivated reasoning.

These results demonstrate that the courts have been incorrect in claiming that but-for causation tracks the ordinary, plain meaning of the statutory causation language at issue in *Burrage*, *Miller*, *Gross*, and other cases concerning similar statutory language. Whatever degree of obviousness suffices for ordinariness and plainness sufficient to resolve disputed statutory meaning without recourse to other interpretive resources,²³⁷ it cannot be that most people disagree with the purportedly obvious interpretation.²³⁸ This is especially so where alternative interpretations (e.g., the

note 219.

237. Cf. Kavanaugh, *supra* note 110, at 2137–38 (describing different approaches to doctrines like *Chevron* that turn on an initial clarity versus ambiguity determination, and explaining that, “[i]n practice, I probably apply something approaching a 65-35 rule. In other words, if the interpretation is at least 65-35 clear, then I will call it clear and reject reliance on ambiguity-dependent canons. I think a few of my colleagues apply more of a 90-10 rule. . . . By contrast, I have other colleagues who appear to apply a 55-45 rule. If the statute is at least 55-45 clear, that’s good enough to call it clear.”); Richard M. Re, *Clarity Doctrines*, 86 U. CHI. L. REV. (forthcoming 2019) (unpublished manuscript on file with the *Indiana Law Journal*), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3327038 [<https://perma.cc/EH3S-7CQC>].

238. As noted above, this Article does not propose any specific percentage or range of percentages of consensus that would support a determination of “plainness” sufficient to invoke the plain meaning doctrine. As discussed below, in many interpretive contexts such thresholds might shift according to the balance of harms at issue (e.g., the severity of potential punishment, the risk of discriminatory enforcement, etc.), among other things. See Hemel & Ouellette, *supra* note 194, at 65–66; Re, *supra* note 237; Ryan Doerfler, *Going “Clear,”* (Feb. 18, 2019) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3326550 [<https://perma.cc/WRS5-RQWP>]. In any event, there would be two basic approaches to using survey evidence to determine whether any such thresholds are met in a given case. One is to simply decide on some threshold level of consensus in the abstract, without comparison to any set of controls. See, e.g., Hemel & Ouellette, *supra* note 194, at 65 (noting that, in trademark confusion litigation, “[f]igures over 25% will generally support a likelihood-of-confusion finding, though plaintiffs have prevailed on the basis of much weaker evidence as well”). The better approach is probably to compare to a set of control survey responses—i.e., percentage of survey participants that agree in their interpretations when asked to apply language that virtually everybody (all judges, commentators, etc.) agree is

“substantial factor” standard, or even an independent sufficiency requirement) come so much closer to tracking the statutory language’s application by large majorities of interpreters (and, independently, come closer to tracking most people’s intuitions concerning moral blameworthiness and punishment).²³⁹

This Article doesn’t claim to answer the more complex question of how the cases discussed in Part I should come out all things considered. Still, here are a few brief remarks about how cases like those described in Part I would come out if courts were to adhere to their ordinary meaning rhetoric, but to do so in light of this Article’s empirical findings.

As representative examples, consider *Burrage* and *Gross*. In a criminal case like *Burrage*, at least where independent sufficiency is not established, a court concerned exclusively with ordinary meaning (as the *Burrage* majority claimed to be) should most likely—*contra* the actual *Burrage* majority—apply the rule of lenity in light of statutory causation ascription rates of only 57-58% (and perhaps also the 42-47% statutory causation ascription rates in the low-blame version of the case).²⁴⁰ The choice of lenity as opposed to ordinary and plain meaning is no mere technicality; it is consequential for at least the reasons emphasized earlier in this article—namely, courts’ tendency to import ordinary and plain meaning determinations into additional substantive areas and procedural postures.²⁴¹ To return to some earlier examples: courts would not be treating a lenity-based *Burrage* decision as if it dictated the causation standard for unrelated state and federal civil and criminal statutes simply because those statutes also use the phrase “results from”;²⁴² courts would also not be employing jury instructions that lack explicit but-for instructions (and upholding convictions based on them) on the ground that the “results from” instruction entails but-for causation as a matter of ordinary meaning.²⁴³

As for statutory tort cases like *Gross*, a court concerned exclusively with ordinary meaning (as the *Gross* majority claimed to be) would—*contra* the actual *Gross* case—uphold liability determinations absent but-for causation and approve more permissive jury instructions like the substantial factor or contributing factor

completely clear. *See id.*; *see also* Ben-Shahar & Strahilevitz, *supra* note 15, at 1779. The survey experiment here provides the sort of data that such an approach would require. For example, where the focal cause was both necessary and independently sufficient (and all judges and commentators would therefore agree that the statutory language was satisfied), around 90% of respondents agreed that the statutory causation language was satisfied. *See* App. fig. 2. And on the other hand, where survey participants were asked whether a focal cause was the “sole factor” resulting in an outcome, and it clearly was not, fewer than 10% claimed that it was. *See id.* These results suggest that we should not expect 100% agreement even on totally clear interpretive issues (and perhaps should expect closer to 90% agreement in such cases).

239. *See supra* Section IV.B.3.d.

240. *See* App. figs. A, D. Where, on the other hand, independent sufficiency is established beyond a reasonable doubt, the plain meaning of the statutory language would be deemed satisfied and the relevant sentencing enhancement applicable. *See* App. fig. A (showing over 90% statutory causation ascription in cases of independent sufficiency, including where but-for causation is absent).

241. *See supra* text accompanying notes 5–10.

242. *See, e.g.*, cases cited *supra* note 5 and accompanying text.

243. *See, e.g.*, cases cited *supra* notes 6–10 and accompanying text.

standards addressed in *Gross*. After all, it's exceedingly rare that the evidence in motive-as-cause cases like *Gross* would rule out a finding of independent sufficiency. And more importantly, even in a case where independent sufficiency is somehow conclusively shown to be *absent*, there is no rule of lenity in a civil case like *Gross* (as opposed to a criminal case like *Burrage*) to save the defendant from liability where the clear majority interpretation would favor it (here, in response to the ADEA's "because of" language, an approximately two-thirds supermajority).²⁴⁴ Perhaps needless to say, allowing for liability absent but-for causation in *Gross* and other similar statutory torts and civil rights contexts would have vast implications in both federal and state litigation.²⁴⁵

Stepping back, the experimental findings naturally raise the question: why, in these and similar statutory causation cases examined in Part I, have so many judges arrived at the "wrong" result (at least, wrong as assessed according to their own rhetoric)? We'll examine a few possibilities below. But first, note that the question is why judges so often arrived at the "wrong" answer—not why, after arriving at it, they were so confident in its being obviously correct (indeed, so confident that they deemed the plain meaning rule applicable).²⁴⁶ As to that latter issue, the above results concerning overconfidence and consensus bias suggest that judges are doing exactly what most of us do—moving quickly from "x seems to me to be the ordinary meaning" to "x is obviously the ordinary meaning, as dictated by common sense and as understood by most people."²⁴⁷ The oddity that calls out for explanation is why these judges landed on the but-for entailment thesis in the first place, not why they (like the rest of us) exhibited overconfidence and consensus bias with respect to their interpretive intuitions. So, why have so many judges arrived at the "wrong" result, and what might it tell us about interpretation more generally?

One set of possible explanations posits that legal training, perhaps along with related professional socialization and acculturation, leads lawyers and judges to think of causation in the sort of counterfactual terms that the but-for test encapsulates, even though laypeople do not.²⁴⁸ There are many ways this might happen. It could simply be that lawyers and judges are frequently exposed to but-for causation rules in law, and over time, through such repeated exposure, what initially felt like a more

244. See App. fig.C. To be sure, in civil cases with particularly harsh potential penalties such as deportation, void-for-vagueness doctrine applies even though the rule of lenity does not. See *Sessions v. Dimaya*, 138 S. Ct. 1204 (2018).

245. See, e.g., sources cited *supra* notes 36, 43, 47.

246. On the plain meaning rule, see Baude & Doerfler, *supra* note 10, at 541.

247. See Solan et al., *supra* note 156.

248. To be clear, the claim need not be that counterfactual thinking doesn't feature prominently, even centrally, in ordinary folk causal cognition. For example, independent sufficiency might play an important role in folk causality ascription precisely *because* it implicates a close possible world in which the focal cause was a but-for cause (namely, a world where the other, non-focal causes were absent). For discussion of various counterfactual-based models of causal cognition, see Adam Morris, Jonathan Phillips, Thomas Icard, Joshua Knobe, Tobias Gerstenberg & Fiery Cushman, *Judgments of Actual Causation Approximate the Effectiveness of Interventions* (July 24, 2018) (unpublished manuscript) on file with the *Indiana Law Journal*, https://www.researchgate.net/publication/326139192_Judgments_of_actual_causation_approximate_the_effectiveness_of_interventions [perma.cc/8XYL-FDGT].

specialized legal usage begins to feel like the commonsense, ordinary meaning of the term or concept.²⁴⁹ Another variation on the legal training and acculturation theme posits a deeper change in the lawyer or judge's moral psychology compared with that of the average layperson: perhaps lawyers and judges internalize the conceptions of (corrective) justice and (limited) responsibility that many of our legal rules and litigation processes instantiate.²⁵⁰ Having internalized those deeper values, lawyers' and judges' notions of causation and causal responsibility come to reflect them in much the same way that the legal system itself does. Much more could be said on these and similar points, but regardless of whether one prefers a more surface-level explanation about semantic drift or instead a deeper one about drift in values and moral psychology, the basic practical thrust of this first set of explanations remains the same: legally trained interpreters have a somewhat idiosyncratic understanding of causation.

Another set of explanations—with more direct implications for statutory interpretation beyond questions of causation—concerns the influence that traditional extratextual sources of statutory meaning might exert over ordinary and plain meaning determinations, judicial rhetoric to the contrary notwithstanding. These extratextual sources might include considerations of congressional intent or statutory purpose, as well as prior cases construing the same or similar statutory terms on grounds other than ordinary meaning.²⁵¹ If these sorts of sources and considerations heavily influence judges' ordinary meaning determinations, then one important practical upshot is that judges ought to be much more hesitant to import ordinary meaning determinations into new substantive and procedural contexts.²⁵² In any event, if this second set of explanations is correct, then the more general takeaway is that ordinary and plain meaning determinations reflect neither more nor less than all-sources-considered conclusions regarding statutory meaning.

A final set of explanations focuses not on the influence of traditional extratextual sources of meaning, but instead on the influence of judges' substantive preferences concerning case outcomes.²⁵³ Perhaps judges' ordinary meaning determinations are skewed toward their own preferred outcomes, and many judges happen to prefer the outcomes that a but-for causation rule generates over the outcomes that equally clear

249. As we saw in Section I.B.2, though, the but-for test is far from the only causation test at work in legal doctrine. Relatedly, therefore, one cannot easily explain the case outcomes in cases like *Gross* and *Burrage* by claiming that (1) the Justices were seeking a legal, rather than ordinary, meaning of causation, and (2) the “legal meaning” of causation *just is* but-for causation. On the contrary, even if (*contra* Justices' and other judges' indications in those and other statutory causation cases, see *supra* note 10) one were to assume that (1) is true, (2) would still be false (or at the very least, (2) would still have been false at the time those decisions were written). See Section I.B.2; see also Verstein; *supra* note 30; Andrew Verstein, *The Failure of Mixed Motives Jurisprudence*, 86 U. CHI. L. REV. (forthcoming 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3204012 [<https://perma.cc/84VA-EQLA>].

250. Cf. Michael Moore, *For What Must We Pay? Causation and Counterfactual Baselines*, 40 SAN DIEGO L. REV. 1181 (2003).

251. See Adam M. Samaha, *If the Text Is Clear—Lexical Ordering in Statutory Interpretation*, 94 NOTRE DAME L. REV. 155 (2018).

252. See *supra* note 5 and accompanying text.

253. See Farnsworth et al., *Policy Preferences and Legal Interpretation*, *supra* note 150.

and administrable rules like the “contributing factor” test generate. Why might these judges prefer the but-for rule? Return once again to *Burrage* and *Gross* as representative examples. In a criminal case like *Burrage*, the same sorts of fairness concerns undergirding the rule of lenity might lead judges to favor the more restrictive but-for rule. This preference then bleeds into their ordinary meaning determination even as they purport not to apply the rule of lenity—an example of a more general psychological phenomenon that might be called “motivated excusing” (in contrast to the oft-noted “motivated blaming” phenomenon addressed earlier).²⁵⁴ A similar sort of “motivated excusing” might likewise be at work in statutory tort cases like *Gross* where liability carries a similar message of moral disapprobation. In any event, while the present point is not to canvass all the available outcome-preference-based accounts that one could propose, it is worth noting that such explanations at least hold out the possibility of explaining why judges (most prominently majority and dissenting Supreme Court Justices) have so often *disagreed* about the ordinary meaning of statutory causation language, and have done so in ways that map onto traditional ideological lines.²⁵⁵ The prior sets of explanations (focusing on legal training, and on exposure to extra-textual evidence) likely lack the resources to explain that disagreement.

Having surveyed possible explanations for why courts have gotten ordinary and plain meaning wrong, what normative lessons might we draw concerning statutory interpretation methodology more generally? There are once again many possibilities, and I suspect that one’s normative takeaway depends at least in part on which of the above descriptive accounts one finds more convincing. Still, two broad possibilities stand out. One lesson might simply be that when courts and commentators pontificate about ordinary meaning, they should do so more cautiously and accurately, perhaps drawing on experimental evidence similar to this Article’s. A second possible reaction, however, goes further: perhaps now that we’ve seen that courts may be prone to get ordinary and plain meaning wrong, we should prefer that courts give less weight to their ordinary meaning determinations than they currently purport to, and should give more weight to alternative sources or types of meaning (e.g., nonordinary “legal” meaning found in precedent, congressional intent found in legislative history, etc.).

More would need to be said, and more empirical evidence marshaled, before deeming that second reaction warranted. For example, this Article does not speak to the relative merits of other methods of interpretation or sources of statutory meaning,

254. See *supra* note 227 and accompanying text; John Turri & Peter Blouw, *Excuse Validation: A Study in Rule-Breaking*, 172 *PHILOS. STUD.* 615, 615 (2014) (demonstrating that people sometimes claim that a given rule was not broken if they consider it to have been blamelessly broken).

255. This is true not only of the Supreme Court decisions discussed in Part I, but also many of the more ideologically charged state and lower federal court decisions noted there. See *supra* note 72 (discussing conflicting causation determinations in cases where multiple police officers simultaneously injure a victim and each officer’s use of force was independently sufficient to bring about the victim’s injury); *supra* note 103 (discussing conflicting causation determinations in cases where discrimination plaintiffs allege that the defendant possessed multiple unlawful motives, each of which was independently sufficient to motivate the defendant’s discriminatory act).

especially outside criminal and tort law. Nor does this Article's initial study of causation show how often courts are mistaken about ordinary meaning. The Article's more modest lesson is that, to the extent that we rely on claims about public meaning—as we must and should do at least to some extent in criminal and tort settings—we ought to seriously consider asking the public; and where we are unable or unwilling to do so, we ought to treat such claims with caution and skepticism, recognizing that they are often both consequential and falsifiable.

V. FURTHER IMPLICATIONS AND EXTENSIONS

This Part argues that similar methods to those used in Part IV could shed light on further areas of dispute in law and legal theory. Section V.A considers further uses of survey experiments in legal interpretation; Section V.B considers how the methodology may help explain and predict the development of common law doctrine; and Section V.C returns to causation in criminal and tort theory to see where the methodology fits within traditional philosophical analyses of legal concepts.

A. Legal Interpretation Beyond Statutory Causation and Plain Meaning

Until now, the focus has been on how surveys and experiments might inform judges interpreting statutory causation language (and assessing jury instructions based on those statutory terms). One could imagine using a similar experimental methodology in additional contexts—assessing statutory terms outside the criminal and tort context, for example, or analyzing common law terms created and used by judges in nonstatutory contexts (as well as jury instructions based on them).²⁵⁶ And the experimental method need not be limited to ascertaining the meaning of discrete terms and concepts. For example, surveys could test linguistic canons of interpretation to see whether they track common linguistic intuitions.²⁵⁷ Judges could then demote or discard those linguistic canons that fail to reflect ordinary language use²⁵⁸—a practice that might please the many judges and commentators who lament

256. In other words, experiments may illuminate nonstatutory contexts where we are interpreting what a *judge's* words (rather than a legislature's words) meant or will be interpreted to mean in future judicial opinions. Cf. Frederick Schauer, *Opinions as Rules*, 53 U. CHI. L. REV. 682, 686 (1986). The point is also, and more directly, applicable to jurors' interpretations of judges' words, where, as often happens, jury instructions parrot the exact language of appellate opinions that set out the relevant standard. See, e.g., *Boyle v. United States*, 556 U.S. 938, 942 (2009).

257. Linguistic canons, after all, are “guesses about the way actual people actually speak.” Baude & Sachs, *supra* note 10, at 1088; Gluck & Bressman, *supra* note 117, at 925 (collecting sources). If they fail to track ordinary language use, they might lack justification. See SLOCUM, *supra* note 10; Baude & Sachs, *supra* note 10, at 1126; Kavanaugh, *supra* note 110, at 2159–60; Sevilla, *supra* note 15.

258. See, e.g., Kavanaugh, *supra* note 110, at 2159–60 (“[W]e ought to shed semantic canons” that fail to accurately “reflect the meaning that people, including Members of Congress, ordinarily intend to communicate with their choice of words.” (citing Elena Kagan, *The Scalia Lecture: A Dialogue with Justice Kagan on the Reading of Statutes*, HARVARD LAW TODAY, at 35:42 (Nov. 17, 2015) (semantic canons are “just ways of formalizing . . . correct intuitions[] about how people use language” (emphasis added))), <http://today.law.harvard.edu>

the proliferation of conflicting canons, which, due to the possibility of “canon shopping,” fail to constrain judges or to render judicial interpretation more predictable.²⁵⁹

Experiments might also be useful where a legal provision’s ordinary or plain meaning is for practical purposes undisputed, but its degree of vagueness or ambiguity *is* disputed, as often happens especially in criminal law, but also in some statutory and constitutional tort contexts.²⁶⁰ For example, with the benefit of experimental evidence, courts might more clearly articulate the relevant triggers for applying doctrines like void-for-vagueness,²⁶¹ lenity,²⁶² and mistake-of-law defenses in criminal²⁶³ and noncriminal law,²⁶⁴ as well as the standards for what counts as “clearly established law” for purposes of qualified immunity.²⁶⁵ The implications of these experiments need not be as rigid as some might fear: ambiguity or vagueness triggers could be ranges, rather than fixed percentages, and those ranges could shift according to the stakes of the interpretive dispute (e.g., in the lenity context, greater clarity needed for greater punishment) or the policy considerations underlying the interpretive rule (e.g., in the void-for-vagueness context, greater clarity needed where there is greater risk of discriminatory enforcement).²⁶⁶ And in certain contexts, surveys could be administered to groups of specialists, instead of or in addition to laypeople.²⁶⁷ In any event, experimental data would go some way toward promoting

/in-scalia-lecture-kagan-discusses-statutory-interpretation [https://perma.cc/4899-7ARR])).

259. See Karl N. Llewellyn, *Remarks on the Theory of Appellate Decision and the Rules or Canons About How Statutes Are To Be Construed*, 3 VAND. L. REV. 395, 399 (1950).

260. On the common designation of tort suits brought under 42 U.S.C. § 1983 as “constitutional” torts as opposed to “statutory” torts, see Levin & Wells, *supra* note 69, at 2, 13.

261. See *Sessions v. Dimaya*, 138 S. Ct. 1204, 1209 (2018) (“The void-for-vagueness doctrine, as we have called it, guarantees that ordinary people have ‘fair notice’ of the conduct a statute proscribes.”); *Kolender v. Lawson*, 461 U.S. 352, 357 (1983) (“[T]he void-for-vagueness doctrine requires that a penal statute define the criminal offense with sufficient definiteness that ordinary people can understand what conduct is prohibited . . .”).

262. See Gluck & Bressman, *supra* note 117, at 947.

263. See McAdams, *supra* note 129, at 32.

264. See Kenneth W. Simons, *Ignorance and Mistake of Criminal Law, Noncriminal Law, and Fact*, 9 OH. ST. J. CRIM. L. 487 (2012).

265. See Kit Kinports, *The Supreme Court’s Quiet Expansion of Qualified Immunity*, 100 MINN. L. REV. HEADNOTES 62 (2016); see also *Kisela v. Hughes*, 138 S. Ct. 1148, 1152 (2018) (discussing the greater amount of notice required to constitute “fair notice” as applied to police officer defendants as opposed to the general public).

266. Cf. Carissa Byrne Hessick, *Vagueness Principles*, 48 ARIZ. ST. L.J. 1137 (2016).

267. For example, in a dispute over the doctrine of lenity’s application to a technical securities fraud provision, a survey might show how *securities traders* most naturally interpret the statute’s words. Cf. Ben-Shahar & Strahilevitz, *supra* note 15 (discussing surveys of specialized members of trades to determine trade usage in contract interpretation).

more coherent and consistent application of these doctrines, in line with scholarly proposals,²⁶⁸ even if that survey data is not treated as dispositive.²⁶⁹

One final possible extension concerns state and federal constitutional interpretation.²⁷⁰ The idea most naturally applies where modern-day courts and philosophers already analyze a given term or concept found in the constitution according to what they claim modern-day usage entails. Consider, for example, the word “punishment” in the Eighth Amendment to the Federal Constitution.²⁷¹ Justices Scalia and Thomas, as well as Judge Posner, have endorsed the claim that the ordinary concept of “punishment”—“whether we consult the usage of 1791, or 1868, or 1985”—entails an *intent* to punish.²⁷² In other words, there is no such thing as unintentional “punishment”²⁷³—a conceptual and linguistic claim with real consequences for prisoners bringing negligence-based civil rights suits, for

268. See, e.g., Hessick, *supra* note 266; see also Kavanaugh, *supra* note 110, at 2142–43 (contending that “too much of current statutory interpretation revolves around personally instinctive assessments of clarity versus ambiguity And even if judges could make threshold findings of ambiguity in a neutral way, they still would have trouble convincing the public that they were acting impartially. It is all but impossible to communicate clarity versus ambiguity determinations in a reasoned and accountable way This kind of decisionmaking threatens to undermine the stability of the law and the neutrality (actual and perceived) of the judiciary.”).

269. Relatedly, though less tied to criminal and tort law, experimental evidence might aid courts in deciding whether to adopt a “saving construction” that avoids rendering a statute unconstitutional. That is, experimental evidence might assist judges by showing more precisely the degree to which a given saving construction would be “nonordinary” or “unnatural” (an issue that often divides judges disagreeing over whether to adopt a given saving construction). See Adrian Vermeule, *Saving Constructions*, 85 GEO. L.J. 1945 (1997).

270. See, e.g., *District of Columbia v. Heller*, 554 U.S. 570, 576 (2007) (“[T]he Constitution was written to be understood by the voters; its words and phrases were used in their normal and ordinary as distinguished from technical meaning.” (quoting *United States v. Sprague*, 282 U.S. 716, 731 (1931))); cf. Larry Alexander & Lawrence B. Solum, *Popular? Constitutionalism?*, 118 Harv. L. Rev. 1594, 1609, 1621 (2005); Todd E. Pettys, *Popular Constitutionalism and Relaxing the Dead Hand: Can the People Be Trusted?*, 86 WASH. U. L. REV. 313, 321, 354 (2008).

271. U.S. CONST. amend. VIII.

272. *Wilson v. Seiter*, 501 U.S. 294, 300–02 (1991) (“An intent requirement is either implicit in the word ‘punishment’ or is not; it cannot be alternately required and ignored as policy considerations might dictate.” (quoting approvingly *Duckworth v. Franzen*, 780 F.2d 645, 652 (7th Cir. 1985) (Posner, J.) (“[T]he accepted meaning of the word, whether we consult the usage of 1791, or 1868, or 1985 . . . [t]his is what the word means today; it is what it meant in the eighteenth century.”))); accord *Helling v. McKinney*, 509 U.S. 25, 40 (1993) (Thomas, J., dissenting); cf. Raff Donelson, *Cruel and Unusual What? Toward a Unified Definition of Punishment*, 9 WASH. U. JURIS. REV. 1 (2016). Many philosophers of law have likewise purported to provide philosophical analyses of the concept of legal punishment, often claiming that their accounts track the folk concept and usage, but failing to cite empirical support beyond their own individual intuitions. See, e.g., H.L.A. Hart, *Prolegomenon to the Principles of Punishment*, in PUNISHMENT AND RESPONSIBILITY 1–27 (1968); GEORGE P. FLETCHER, *THE GRAMMAR OF CRIMINAL LAW* 227 (2007); Steven Sverdlik, *Punishment*, 7 LAW & PHIL. 179, 190 (1988); DAVID BOONIN, *THE PROBLEM OF PUNISHMENT* ch. 1 (2008).

273. But see Adam J. Kolber, *Unintentional Punishment*, 18 LEGAL THEORY 1 (2012).

example.²⁷⁴ In support of this claim, these judges cite modern dictionaries,²⁷⁵ Founding Era dictionaries,²⁷⁶ and intuition pumps that they themselves have created based on their experience with the concept in ordinary discourse.²⁷⁷ Here, then, is another area where judges and philosophers make competing assertions about a concept, purporting to track lay usage and intuitions, relying on their own intuitions to guide them. It is thus another area where experiments might provide additional relevant evidence of ordinary, public meaning—at least, that is, to the extent that we either (a) care about *current* public meaning *per se*,²⁷⁸ or, perhaps more likely, (b) care about public meaning at the time the Eighth Amendment was ratified, *but agree with these judges and philosophers that the folk concept and ordinary usage of “punishment” has probably not shifted in the relevant respect since then*.²⁷⁹ Again, such limited inquiries need not open the floodgates to popular or populist constitutional interpretation writ large.²⁸⁰

*B. Predicting the Common Law: Legal Realism, the New Doctrinalism, and
Experimental Jurisprudence*

The prior subsection considered how judges *ought* to determine the ordinary meaning of terms or concepts (or at least their degree of vagueness or ambiguity) where judges *already expressly state* that they find ordinary usage relevant. American Legal Realists famously sought to explain and predict judicial decision-making through whatever factors or considerations—whether stated *or unstated*—actually turn out to have the greatest predictive or explanatory power. In their wake, economists, philosophers, and political scientists have taken up this descriptive project with varying degrees of success, explaining judicial decision-making and the evolution of common law doctrine using notions of economic efficiency, theories of justice, ideological preferences, and so forth.

Recently, a movement dubbed the New Doctrinalism has emerged, emphasizing the ways in which law’s doctrinal categories really do channel judges’ intuitions and reasoning, forcing judges to conceptualize legal disputes in categories like “reasonableness” (and “causation”)—influencing and constraining legal decision-making to a degree and in a manner distinct from that which a pure economic or

274. See Donelson, *supra* note 272, at 35–36.

275. E.g., *Wilson*, 501 U.S. at 300–02; *Helling*, 509 U.S. at 40 (Thomas, J., dissenting).

276. E.g., *Wilson*, 501 U.S. at 300–02; *Helling*, 509 U.S. at 38.

277. E.g., *Wilson*, 501 U.S. at 300 (quoting *Duckworth*, 780 F.2d at 652); *Helling*, 509 U.S. at 38. Occasionally, they even cite ordinary language philosophers. See, e.g., *Does #1-5 v. Snyder*, 834 F.3d 696, 701 (6th Cir. 2016) (considering whether the federal Sex Offender Registration Act’s restrictions “meet the general, and widely accepted, definition of punishment offered by legal philosopher H.L.A. Hart”).

278. See Solum, *supra* note 197, at 62–65 (2015) (discussing “the ‘Contemporary Readers’ Meaning Theory” of constitutional interpretation).

279. See *supra* notes 271–77; Solum, *supra* note 197, at 19, 42–50, 57, 73–75 (discussing the concept of “cruelty” in the context of the Eighth Amendment’s prohibition of “cruel and unusual” punishments, along with the Supreme Court’s jurisprudence regarding evolving conceptions of cruelty for Eighth Amendment purposes).

280. See sources cited *supra* notes 272, 277.

positive political science model, for instance, would consider.²⁸¹ The “‘law and’ movement,” some New Doctrinalists suggest, has become too unmoored from how judges, reasoning from the law’s internal point of view, actually decide cases.²⁸² To invoke a phrase popular among the early American Legal Realists: judges’ “situation sense”²⁸³ upon learning the basic facts and issues in a case is filtered through the basic conceptual categories judges take themselves to be bound to apply.²⁸⁴ A realistic positive (as opposed to normative) account of adjudication, the New Doctrinalists usefully remind us, must take account not only of judges’ pre-legal “situation sense” but also legal doctrine.

The positive parts of the Legal Realist and New Doctrinalist research programs stand in need of another, complimentary (and of course, capitalized) movement: Experimental Jurisprudence²⁸⁵—the examination, through controlled experiment, of ordinary concepts, language, and intuitions, combined with analysis of their explanatory role in predicting legal outcomes.²⁸⁶ At this point, though, the New

281. See Cristina Carmody Tilley, *Tort Law Inside Out*, 126 YALE L.J. 1320, 1335–36 (2017) (summarizing New Doctrinalism and collecting sources); Shyamkrishna Balganesh, *The Constraint of Legal Doctrine*, 163 U. PA. L. REV. 1843 (2015); see also Lawrence B. Solum, *The Positive Foundations of Formalism: False Necessity and American Legal Realism*, 127 HARV. L. REV. 2426 (2014). To be clear, most American Legal Realists did not deny that doctrine plays an important role in judicial decision-making, especially in easy cases (where it may fully determine the result), and especially where legal doctrine has been reformed to more accurately reflect the (previously nonlegal) norms that the Realists believed drove judicial decision-making *sub silencio* at the time the Realists were writing. See Brian Leiter, *Legal Realism and Legal Doctrine*, 163 U. PA. L. REV. 1975, 1983 (2015).

282. See Tilley, *supra* note 281, at 1336.

283. See KARL N. LLEWELLYN, *THE COMMON LAW TRADITION: DECIDING APPEALS* 59–61, 121–57, 206–08 (1960); Joseph C. Hutcheson, Jr., *The Judgment Intuitive: The Function of the “Hunch” in Judicial Decision*, 14 CORNELL L.Q. 274, 285 (1929); Dan M. Kahan, David Hoffman, Daniel Evans, Neal Devins, Eugene Lucci & Katherine Cheng, “Ideology” or “Situation Sense”? An Experimental Investigation of Motivated Reasoning and Professional Judgment, 164 U. PA. L. REV. 349 (2016).

284. See Balganesh, *supra* note 281, at 1857. To be clear, American Legal Realists maintained that judges’ “situation sense” was a function of nonlegal norms, not legal doctrine (at least as the doctrine existed at the time they were writing). See Brian Leiter, *Rethinking Legal Realism: Toward a Naturalized Jurisprudence*, 76 TEX. L. REV. 267, 277 (1997); Leiter, *supra* note 281, at 1975. For more precise, albeit differing, accounts of judicial “situation sense” as the notion figured in the thinking of various American Realists, see Kahan et al., *supra* note 283, at 373 (adopting Karl Llewellyn’s later characterization of situation sense as a “perceptive faculty”); Leiter, *supra*, at 277 (criticizing this view as “cryptic” and not in keeping with other Realists’ more “naturalistically respectable” accounts of judicial “situation sense”).

285. The label is taken from Solum, *supra* note 281, at 2464–65 (noting “the nascent emergence of experimental jurisprudence (or ‘X-Jur’),” and explaining that the label “is modeled on the analogous phrase ‘experimental philosophy’” (citing EXPERIMENTAL PHILOSOPHY (Joshua Knobe & Shaun Nichols, eds., 2008))).

286. The normative projects contained in of each of these schools of thought would likewise benefit from Experimental Jurisprudence: whether one views folk concepts and ordinary usage as things to be embraced or instead to be manipulated and overcome, implementation of one’s preferred normative theory requires understanding the ordinary

Doctrinalist might wonder: Why “ordinary” concepts, rather than “legal” concepts? Why “ordinary” usage and “ordinary” intuitions rather than lawyers’ and judges’ usage and intuitions? After all, judges exercise significant discretion in giving terms and concepts a *legal* meaning that diverges from *ordinary* meaning.²⁸⁷ Indeed, at their most extreme, concepts seemingly familiar to the layperson can become so infused with legal policy considerations that they become nearly unrecognizable in their application (does a “reasonable” person really do *x*? How can someone not have a “duty” to do *y*? Etc.).

The answer is multifaceted, but we have already seen many of the facets. The very doctrinal rules that judges internalize instruct them to consider ordinary usage, concepts, and intuitions.²⁸⁸ Procedural rules further require them to consider what a reasonable jury could conclude regarding a given concept, including many normatively infused concepts like “reasonableness,” “substantiality,” “knowledge,” “intent,” and “causation.”²⁸⁹ And indeed, judges *say* that they consult and rely on ordinary usage, concepts, and intuitions, often justifying that reliance not only through recourse to positive legal sources (other judicial decisions and procedural rules) that direct them to do so, but also by direct recourse to the various underlying values, examined in Part II, undergirding reliance on ordinary meaning and common understanding in the first place. Across the mine-run of cases and doctrines, through the workaday rulings of the many overburdened judges throughout state and federal court systems, in the vast majority of cases in which they have no particular preference for a given outcome, common law concepts remain tethered to their ordinary meanings and to the folk concepts they originally embodied.²⁹⁰ Judges, in short, apply common law concepts on ordinary meaning’s leash.²⁹¹

By better understanding how most people interpret and apply a given word or concept, we may therefore be able to more accurately predict and explain how judges will interpret and apply it. Scaling up, we may be able to better predict and explain the development of common law doctrine over time. And this may be true even if, in some contexts, judges do not explicitly *say* that they are relying on ordinary usage or folk concepts. As virtually everybody acknowledges, judges, even in common law

concepts and linguistic usage that the actual decisionmakers—judges and juries—bring to the task. See *infra* note 289 (noting normative efficiency theories of causation-in-fact).

287. See Shyamkrishna Balganesh & Gideon Parchomovsky, *Structure and Value in the Common Law*, 163 U. PA. L. REV. 1241, 1244 (2015).

288. See *supra* notes 136–37 and accompanying text.

289. On normativity and notions of “substantiality,” as well as mental states like knowledge and intent, see Macleod, *supra* note 164; cf. Markus Kneer & Sacha Bourgeois-Gironde, *Mens Rea Ascription, Expertise and Outcome Effects: Professional Judges Surveyed*, 169 COGNITION 139 (2017). On reasonableness and normativity, see Kevin P. Tobia, *How People Judge What Is Reasonable*, 70 ALA. L. REV. 293 (2018); Benjamin C. Zipursky, *Reasonableness In and Out of Negligence Law*, 163 PA. L. REV. 2131 (2015).

290. See Scott Hershovitz, *Harry Potter and the Trouble With Tort Theory*, 63 STAN. L. REV. 67, 107 (2010). But see Jody S. Kraus, *Transparency and Determinacy in Common Law Adjudication: A Philosophical Defense of Explanatory Economic Analysis*, 93 VA. L. REV. 287 (2007).

291. Indeed, where they start to drift especially far from their ordinary meanings, commentators and other judges often for that very reason cry foul. See Balganesh & Parchomovsky, *supra* note 287.

contexts, and even at their most creative, daring, and instrumentalist, are loath to give words a “meaning they cannot bear.”²⁹² What they *can* bear typically depends on the ordinary meaning and the plausible extension of folk concepts—something experiments can help ascertain.

C. Experimental Philosophy of Criminal and Tort Law: Causation as a Case Study

Let’s return more squarely to causation, using it now as a case study in methods of conceptual analysis in philosophy of criminal and tort law. We’ve seen that Hart and Honoré set out to analyze “commonsense” notions of causation at work in the law, explicitly relying on ordinary language and intuitions as their guide.²⁹³ But while their analysis remains influential, it has also often been criticized for its reliance on arguments from ordinary language and intuitions. This criticism is exemplified in the work of Michael Moore, the most prominent philosopher of causation in criminal and tort law since Hart and Honoré. It is worth briefly considering Moore’s alternative approach to analyzing causation in the law and the methodological obstacles it runs into—obstacles that, like Hart and Honoré’s work from decades prior, highlight the need for experimental philosophy of law.

Moore’s goal is to provide a “metaphysically robust” account of “the nature of causation”—to find what causation “actually is,” which, on his view, means finding a natural thing or relation that exists out there in the world independently of how people talk and think about it.²⁹⁴ He sees the goal of his project as being much like a natural scientist’s in revealing that, say, what we long ago decided to call “water” actually is H₂O.²⁹⁵ (But why ought *legal* theorists to be concerned with the “metaphysically robust” notion of causation, rather than simply the “commonsense” one that Hart and Honoré thought was at work in the law? Because, on Moore’s

292. See HENRY M. HART, JR. & ALBERT M. SACKS, *THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW* 1374 (William N. Eskridge, Jr. & Philip P. Frickey eds., 1994); RONALD DWORKIN, *TAKING RIGHTS SERIOUSLY* 109–11, 134 (1977). One prominent example of such a daring account in the causation literature is Steven Shavell’s “probabilistic cause” account. See Steven Shavell, *An Analysis of Causation and the Scope of Liability in the Law of Torts*, 9 J. LEGAL STUD. 463 (1980); see also Richard W. Wright, *Actual Causation vs. Probabilistic Linkage: The Bane of Economic Analysts*, 14 J. LEGAL STUD. 435 (1985); Richard W. Wright, *The Efficiency Theory of Causation and Responsibility: Unscientific Formalism and False Semantics*, 63 CHI.-KENT L. REV. 553 (1987); Richard W. Wright, *The New Old Efficiency Theories of Causation and Liability*, 7 J. TORT LAW 65 (2015).

293. See *supra* notes 23, 137–43 and accompanying text.

294. Moore is a metaphysical realist about causation (and, as relevant here, about morality and about semantic reference). Michael S. Moore, *Moral Reality Revisited*, 90 MICH. L. REV. 2424, 2437 (1992); MOORE, *supra* note 23, at 255–60; Michael S. Moore, *The Semantics of Judging*, 54 S. CAL. L. REV. 151 (1981). In other words, Moore believes the following about judgments of causation, judgments of morality, and judgments about what words like “cause” actually refer to: such judgments are all apt for evaluation in terms of truth or falsity; at least some such judgments are true; and, importantly, such judgments’ truth or falsity is entirely independent of whether they seem to any observer (including an ideal and unbiased one) to be true or false. For further background and critical discussion of Moore’s metaphysical realism, see Brian Bix, *Michael Moore’s Realist Approach to Law*, 140 U. PENN. L. REV. 1293 (1992).

295. See, e.g., Moore, *supra* note 146, at 854–55.

account, the notion of causation that both is and should be at work in criminal law,²⁹⁶ tort law,²⁹⁷ and even in correct interpretation of statutory causation language outside these substantive contexts,²⁹⁸ is whatever that metaphysically accurate notion of causation turns out to be—even if it turns out to differ from our “commonsense” notion of causation.)

So, Moore is looking for what causation “actually is,” not just what our ordinary usage and intuitions would have us think it is. And according to Moore, those two things—metaphysically accurate causation and commonsense causation—could be “considerably different.”²⁹⁹ In short, Hart and Honoré’s “sociology”³⁰⁰ does not show us the underlying metaphysical reality that, whatever it turns out to be, is at work in the law.³⁰¹ What, then, would show us that underlying metaphysical reality? How, in other words, does Moore find out what causation “actually is” as a matter of metaphysics (and hence, what it “actually is” for purposes of assigning moral responsibility, determining liability in criminal and tort law, and interpreting what statutory causation language references)? How does he proceed without relying primarily on appeals to the sort of data—ordinary linguistic, conceptual, and moral intuitions—from which Hart and Honoré derived their account?

As others have pointed out, Moore doesn’t: he in fact relies extensively on appeals to ordinary language and intuitions throughout his work on causation, sometimes as decisive evidence in favor of or against a given position.³⁰² (Though to be sure,

296. MOORE, *supra* note 23, at 8.

297. *Id.*; see also, e.g., Michael S. Moore, *Semantics, Metaphysics, and Objectivity in the Law*, in VAGUENESS AND LAW: PHILOSOPHICAL AND LEGAL PERSPECTIVES 146 (Geert Keil & Ralf Poscher eds., 2016) (judges must “*not* look to policies in deciding on tort and criminal liability. Rather, they must apply the best metaphysics of the causal relation, as best they can ascertain it. That, of course, is a formidable task. Fortunately there are good books to help them.”).

298. In Moore’s view, “causation” is a natural kind, like “gold” or “water,” and when the legislature uses natural kind terms, it invokes whatever the metaphysically accurate account of that concept turns out to be. The basis for this latter assumption is a little unclear, though Moore at times appeals to speaker intent and rule of law reasons. See, e.g., Moore, *supra* note 123, at 291, 341; Kimberly Kessler Ferzan, *A Planet by Any Other Name . . .*, 108 MICH. L. REV. 1011, 1026 (2010); LEITER, NATURALIZING JURISPRUDENCE, *supra* note 26, at 268–69.

299. Moore, *supra* note 146, at 855.

300. E.g., *id.* at 853–54, 868; MOORE, *supra* note 23.

301. E.g., MOORE, *supra* note 23, at 255–60; Moore, *supra* note 23, at 854–55.

302. E.g., MOORE, *supra* note 23, at 29–33, 38–39, 55–58, 66, 362, 438, 443, 445–47, 462, 507; see also Richard W. Wright, *Causation: Metaphysics or Intuition?*, in LEGAL, MORAL, AND METAPHYSICAL TRUTHS: THE PHILOSOPHY OF MICHAEL S. MOORE 171 (Ferzan & Morse eds., 2015) (“Although he purports to disavow Hart and Honoré’s reliance on ordinary language usage as an indicator of actual causal relations, his frequent and often heavy reliance on supposed common intuitions to support his arguments belies that disavowal”); Stapleton, *Choosing What We Mean*, *supra* note 141, at 465; Roger Shiner, *Causation, Responsibility, and Metaphysics*, 31 LAW & PHIL. 601, 613 (2012). Wright, for his part, similarly disavows ordinary usage and intuition while appealing to it at times (though, to be sure, rarely for such crucial points in his argument as those for which Moore appeals to them). See, e.g., Wright, *Causation in Tort Law*, *supra* note 81, at 1788–803; Wright, *supra* note 138, at 165; Richard Wright, *Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts*, 73 IOWA L. REV. 1001, 1003, 1009 (1988); Wright,

Moore does not hesitate in reaching a number of conclusions that go against other philosophers',³⁰³ courts',³⁰⁴ and laypeople's³⁰⁵ intuitions.³⁰⁶) In the end, then, Moore's "'I know it when I see it' primitivist"³⁰⁷ account of causation, whatever its metaphysical ambitions,³⁰⁸ is derived largely from data strikingly similar to that on

Legal Responsibility, *supra* note 81, at 1441; *see also* Stapleton, *Choosing What We Mean*, *supra* note 141, at 472; Sara Bernstein, *Intuitions and the Metaphysics of Causation*, in EXPERIMENTAL METAPHYSICS 75, (David Rose ed., 2017) ("Should intuitions play a role in a theory of causation? Metaphysicians often bristle at the idea, while at the same time utilizing intuitions. . . . [M]ost metaphysical and causal theorizing centrally involves intuitions.").

303. *Compare*, e.g., MOORE, *supra* note 23, at 116–17 (regarding the impossibility of omissions as causes), with Jonathan Schaffer, *Disconnection and Responsibility*, 18 LEGAL THEORY 399 (2012) and Paul Henne et al., *Cause by Omission and Norm: Not Watering Plants*, 95 AUSTRALASIAN J. PHIL. (2017); *see also* MOORE, *supra* note 23, at 414–18 (discussing David Lewis's intuitions concerning causality in certain types of overdetermination cases, and conceding that Lewis's intuitions are "tempting," but concluding that "the California Supreme Court's causal intuitions in this actual case seem persuasive here," and then "buttress[ing] this otherwise rather naked intuition by likening these cases to" other cases that come to similar results); *see also*, e.g., Wright & Puppe, *supra* note 71, at 476 (discussing various philosophers' contrary intuitions concerning multiple contributors); RICHARD W. WRIGHT, THE NESS ACCOUNT OF NATURAL CAUSATION: A RESPONSE TO CRITICISMS 12 (same), https://scholarship.kentlaw.iit.edu/cgi/viewcontent.cgi?article=1715&context=fac_schol. Moore raises similar concerns about the counterintuitive implications of Wright's account of causation, implying that incompatibility with common intuitions is, in his own view, to be avoided where possible in generating a correct theory of causation. *See*, e.g., MOORE, *supra* note 23, at 489; *see also* Wright & Puppe, *supra* note 71, at 501–02.

304. Moore is open about treating case law as at times only a relatively minor constraint, since "[t]he law has mixed too many extraneous elements into what it calls 'causation' for there to be much hope for any metaphysical translation." *See* Jane Stapleton, *Causation in the Law*, in THE OXFORD HANDBOOK OF CAUSATION 761 (Helen Beebe, Christopher Hitchcock, and Peter Menzies eds., 2009) (criticizing this move on Moore's part); *see also* note 160, *supra* (comparing Moore's intuitions with the *Burrage* Court's); *cf.* MOORE, *supra* note 23, at 115 (criticizing as ad hoc Wright's explanations for case outcomes seemingly contrary to Wright's account of causation).

305. *See*, e.g., Schaffer, *supra* note 303; Henne et al., *supra* note 303.

306. These aspects of Moore's theory might be more easily overlooked if his account were aimed at generating a test for causation that judges or jurors could use, but that is neither an aim nor a result of Moore's staunchly metaphysical project. *See* MOORE, *supra* note 23, at xi. Wright's formalization of Hart and Honore's test for factual causation is likewise widely acknowledged to be too complicated to be applied by juries. *See*, e.g., Johnson, *supra* note 79, at 1762 (calling Wright's test "far too complex to be of any real utility to lay fact finders."); Johnson, *supra* note 80, at 102 (same). Wright's aim, like Moore's, is not, or at least not primarily, to provide a usable test; it is to provide an accurate analysis of "the basic concept of causation, which we all intuitively employ." Wright, *supra* note 81, at 1441; *see also* Wright, *supra* note 81, at 1802 (claiming that his "necessary element of a sufficient set" analysis of causation "is not just a test for causation, but is itself the meaning of causation").

307. Wright, *Causation: Metaphysics or Intuition?*, *supra* note 302, at 171; *see* MOORE, *supra* note 23, at 506 (calling himself a "reluctant primitivist" about the nature of causation); Richard Fumerton & Ken Kress, *Causation and the Law: Preemption, Lawful Sufficiency, and Causal Sufficiency*, 64 L. & CONTEMP. PROBS. 83, 83 (2001).

308. *See* Schaffer, *supra* note 303, at 400.

which Hart and Honoré relied: armchair speculation about intuitive responses to cases.³⁰⁹

An alternative approach to understanding causation in criminal and tort law is to return to something like Hart and Honoré's project,³¹⁰ but with greater empirical rigor.³¹¹ Hart and Honoré, after all, had a sample size of two: Hart and Honoré. An alternative approach, in other words, is to do experimental philosophy of law³¹²—to “demand,” in Brian Leiter's words, “that any account of those features of law which depends on ‘folk intuitions’ as data points in theory-construction ought to answer to empirically sound methods for ascertaining that data. . . . [T]his kind of ‘experimental philosophy’ is likely to set much of the agenda for naturalistic legal philosophy at the dawn of the 21st century.”³¹³

CONCLUSION

There is much to be said for ordinary meaning interpretation, at least in many criminal and torts contexts. As one scholar recently put it, “[o]rdinary meaning requires the interpreter to put herself in the shoes of a nonlegal audience; it has a built-in form of impartiality, not to mention democratic appeal.”³¹⁴ Yet if ordinary meaning is premised on little more than judges' intuitions and biased guesswork about “common sense,” the method loses much of its claim to impartiality. Likewise, the method loses its “democratic appeal” if it allows judicial *ipse dixit* to replace reasoned analysis, and judicial preferences to replace popular ones. Heavy reliance on untested intuitions about ordinary and plain meaning can (perhaps ironically) render statutory interpretation decisions mystifying and misleading to legally trained and lay audiences alike.

Judges, legal philosophers, and the rest of us can be mistaken about whether our own intuitions match the general public's “common understanding.” In light of the empirical data reported in this Article, courts' recent statutory causation decisions appear to present a stark and consequential example of just such a mismatch. By increasing the sample size—by asking the general public itself—judges, philosophers, and the rest of us can more accurately, and more impartially, find ordinary meaning and common understanding in statutory interpretation and beyond. At the very least, by asking from time to time, we can be reminded of the risk we run in failing to do so, and of the skepticism with which we should approach untested

309. Cf. Fumerton & Kress, *supra* note 307, at 83.

310. See Ian P. Farrell, *H.L.A. Hart and the Methodology of Jurisprudence*, 84 TEX. L. REV. 983 (2006); see also LUCY, *supra* note 141, at 184.

311. See, e.g., Stapleton, *supra* note 304, at 759 (“Hart and Honoré's linguistic ‘analysis’ was not based on a rigorous empirical survey . . . patterns of usage were merely asserted and illustrated.”); Stapleton, *Unpacking Causation*, *supra* note 141, at 148–55; David Howarth, *O Madness of Discourse, That Cause Sets Up With and Against Itself*, 96 YALE L.J. 1389, 1402–04 (1987); LUCY, *supra* note 141, at 200.

312. For an introduction and overview of the field of experimental philosophy, see Knobe & Nichols, *supra* note 285.

313. LEITER, NATURALIZING JURISPRUDENCE, *supra* note 26, at 4; Leiter, *Three Approaches*, *supra* note 26, at 9.

314. Nourse, *supra* note 10, at 1004.

but in principle (and often in practice) falsifiable claims concerning common sense and ordinary understanding.

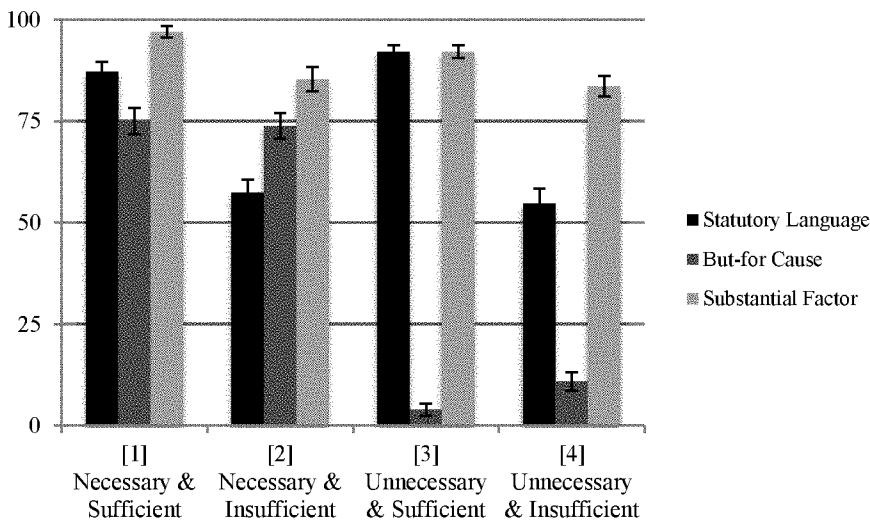
APPENDIX

This Appendix contains the text of each vignette followed by a Figure showing statutory, but-for, and substantial factor causation ascription in response to the vignette in each of its four necessity/sufficiency conditions.³¹⁵

DRUG OVERDOSE (*Burrage*):

For vignette text, see *supra* pages 996 to 997.

Figure A. Drug Overdose (*Burrage*): Statutory, But-For, and Substantial Factor Causation Ascription (N = 508)



315. Figures A through D (unlike Figure 2 *supra*) omit the “contributing factor” and “sole” cause questions. This is simply for ease of visualization, given that those two tests are the least important to this article’s analysis.

HATE CRIME (*Miller*³¹⁶):

Sam worked at a financial company. Sam couldn't stand his new co-worker, Harry, because Harry was a member of a religion that Sam hated, and because Harry got a higher salary than Sam despite being new at the company.

After a few weeks of simmering hatred, Sam decided to attack Harry. So one evening, while Harry was walking home from work, Sam, wearing a ski mask to disguise his identity, ran up to Harry and punched him in the face several times, causing very serious injuries, and then ran off.

Later that night, Sam's best friend, Joe, came to Sam's apartment to talk to him. Sam and Joe were always totally honest with each other. Sam told Joe that he had just attacked Harry. Joe asked Sam why he did it, and Sam told him about how he hated Harry's religion and how he hated that Harry got a higher salary than him.

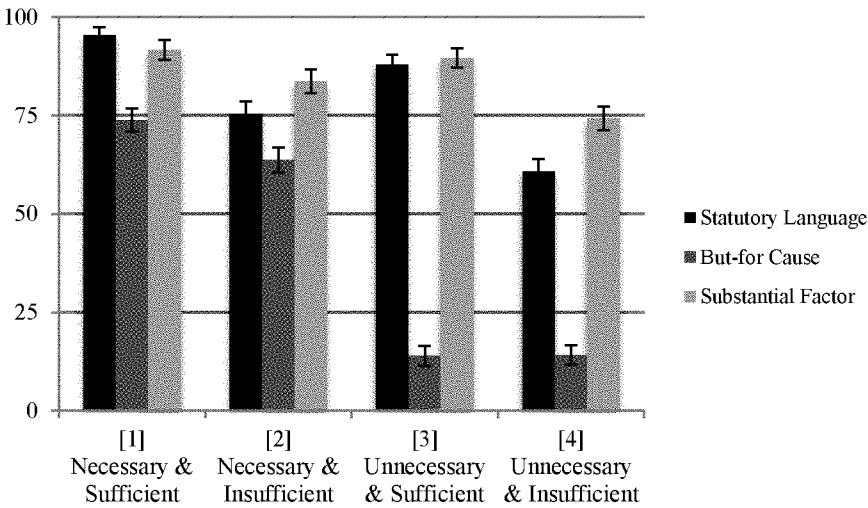
Joe wanted to know more about Sam's motivations for attacking Harry, so he asked Sam a few more questions. Sam was completely honest with him. Sam explained that he...

[1] Necessary & Sufficient	[2] Necessary & Insufficient	[3] Unnecessary & Sufficient	[4] Unnecessary & Insufficient
... still would have attacked Harry even if he had only known Harry's religion (and hadn't known Harry's salary). And Sam explained that he would not have attacked Harry if he had only known Harry's salary (and hadn't known Harry's religion).	... would not have attacked Harry if he had only known Harry's religion (and hadn't known Harry's salary). And Sam explained that he would not have attacked Harry if he had only known Harry's salary (and hadn't known Harry's religion).	... would have attacked Harry even if he had only known Harry's religion (and hadn't known Harry's salary). And Sam explained that he still would have attacked Harry even if he had only known Harry's salary (and hadn't known Harry's religion).	... would not have attacked Harry if he had only known Harry's religion (and hadn't known Harry's salary). And Sam explained that he still would have attacked Harry even if he had only known Harry's salary (and hadn't known Harry's religion).

316. This vignette is based in part on *Miller* and in part on the facts of *United States v. Hill*, 182 F. Supp. 3d 546, 555 (E.D. Va. 2016) (“[I]f the conduct underlying a[n] HCPA] prosecution has any sort of commercial motivation—if, for example, Hill had punched C.T. because of both C.T.’s sexual orientation and C.T. having taken Hill’s job as a packer—then the government could never obtain a conviction.” (citing *United States v. Miller*, 767 F.3d 585, 594 (2014))).

Still, Sam explained, as it actually happened, both things motivated him to attack Harry. Joe knew Sam was telling the truth.

Figure B. Hate Crime (Miller): Statutory, But-For, and Substantial Factor Causation Ascription (N = 452)



EMPLOYMENT DISCRIMINATION (*Gross*):

Steve works as a manager at a large financial company. The company needs to downsize. Steve has never met Jack, one of the company’s employees. But Steve decided to fire Jack because Steve thinks employees Jack’s age aren’t capable of working hard, and because Jack’s position is no longer necessary.

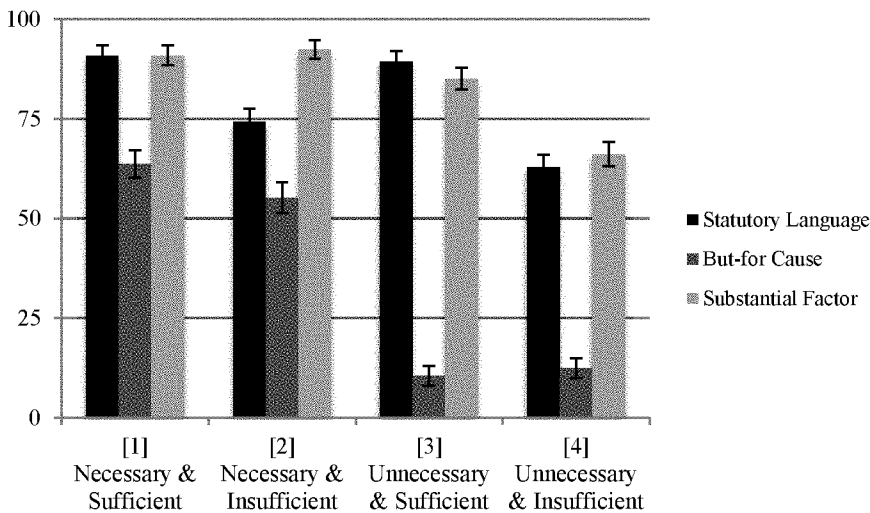
Last Tuesday, Steve fired Jack. Later that night, Steve’s best friend from business school, Robert, came to Steve’s apartment to talk to him. Steve and Robert were always totally honest with each other. Steve told Robert that he had fired Jack. Robert asked Steve why he did it, and Steve told him about how he thinks employees Jack’s age aren’t capable of working hard, and how Jack’s position is no longer necessary.

Robert wanted to know more about Steve’s motivations for firing Jack, so he asked Steve a few more questions. Steve was completely honest with him. Steve explained that he...

[1] Necessary & Sufficient	[2] Necessary & Insufficient	[3] Unnecessary & Sufficient	[4] Unnecessary & Insufficient
...still would have fired Jack even if he had only known Jack’s age (and hadn’t known that Jack’s position was unnecessary). And Steve explained that he would not have fired Jack if he had only known that Jack’s position was unnecessary (and hadn’t known Jack’s age).	...would not have fired Jack if he had only known Jack’s age (and hadn’t known that Jack’s position was unnecessary). And Steve explained that he would not have fired Jack if he had only known that Jack’s position was unnecessary (and hadn’t known Jack’s age).	...would have fired Jack even if he had only known Jack’s age (and hadn’t known that Jack’s position was unnecessary). And Steve explained that he still would have fired Jack even if he had only known that Jack’s position was unnecessary (and hadn’t known Jack’s age).	...would not have fired Jack if he had only known Jack’s age (and hadn’t known that Jack’s position was unnecessary). And Steve explained that he still would have fired Jack even if he had only known that Jack’s position was unnecessary (and hadn’t known Jack’s age).

Still, Steve explained, as it actually happened, both things motivated him to fire Jack. Robert knew Steve was telling the truth.

Figure C. Employment Discrimination (Gross): Statutory, But-For, and Substantial Factor Causation Ascription (N = 450)



PLANT FOOD (Low-Blame Version of *Burrage*):

Fintene, Rextor, and Tamphen are plant foods. A typical dose of each helps houseplants grow. But unusually potent doses are very harmful to houseplants, and can even kill them. Stores sell Fintene, Rextor, and Tamphen in powder form, in small plastic bags. To use them, plant owners mix the powder into a glass of water and pour it into their plant’s soil. Used this way, Fintene, Rextor, and Tamphen take a few hours to become absorbed into the plant.

On Tuesday morning, Josh went to three local stores to buy plant food. One sold Josh a packet of Fintene, another sold Josh a packet of Rextor, and the other sold Josh a packet of Tamphen. Josh wanted his houseplant to grow, so he decided that that afternoon, he would feed his plant all three of the foods together.

What Josh didn’t know was that the Fintene, Rextor, and Tamphen he bought were all unusually potent. (The salesperson at each store had explained this to him, and had told him not to use an entire packet all at once, but he hadn’t listened to what they were saying.) In fact,...

[1] Necessary & Sufficient	[2] Necessary & Insufficient	[3] Unnecessary & Sufficient	[4] Unnecessary & Insufficient
...although neither the Rextor by itself, nor the Tamphen by itself, nor the combination of Rextor and Tamphen together would have killed the plant, the Fintene by itself would have killed the plant. And...	...although neither the Fintene by itself nor the combination of the Rextor and Tamphen together would have killed the plant,...	...the Fintene by itself would have killed the plant, the Rextor by itself would have killed the plant, and the Tamphen by itself would have killed the plant. Any two of the three foods together would also have killed the plant. And...	...although neither the Fintene by itself, nor the Rextor by itself, nor the Tamphen by itself would have killed the plant, any two of the three foods together would have killed the plant. And...

[A]ll three foods together would kill the plant.

That afternoon, Josh went home, combined all the powder from all three bags, mixed it into a glass of water, and poured it into the plant’s soil. Sure enough, after a few hours, the Fintene, Rextor, and Tamphen combined in the plant’s roots, blocking the flow of nutrients to the plant, and the plant died.

Figure D. Plant Food Overdose Vignette (Low-Blame Version of Burrage): Statutory, But-For, and Substantial Factor Causation Ascription (N = 470)

