

Abstract

Essays in Experimental Jurisprudence

Kevin P. Tobia
2019

This dissertation studies the concepts of law. Using traditional philosophical analysis alongside experimental methods of cognitive science and experimental philosophy, the dissertation investigates concepts that are central to legal philosophy, including *reasonableness* and *intentional action*. Together, the chapters provide a broader defense of the experimental-philosophical study of legal concepts, “experimental jurisprudence.”

Chapter 1 investigates the concept of legal *reasonableness*. A classic debate concerns whether reasonableness should be understood statistically (e.g., reasonableness is what is common) or prescriptively (e.g., reasonableness is what is good). This chapter elaborates and defends a third possibility. Reasonableness is a partly statistical and partly prescriptive “hybrid,” reflecting both statistical and prescriptive norms. Experiments reveal that people evaluate reasonableness as a hybrid, and the chapter argues that hybrid accounts offer the best theory of legal reasonableness.

Chapter 2 considers the relationship between ordinary and expert legal concepts, with a case study concerning the concept of *intentional action*. Experimental studies reveal divergent judgments of intentional action among ordinary people, law students, and United States judges. I examine whether these differences are better explained by legal experience or some other individual differences. The results support the former explanation. That is, it is not the case that certain people who are skilled at evaluating intentional action disproportionately select to become legal professionals; instead, the data indicate that legal training and experience *teach* a distinctive legal concept of what it is to act intentionally.

Chapter 3 examines the efforts of originalist theories of interpretation to determine the original “public meaning” of legal texts. This chapter evaluates two popular methods used to discover (original) meaning: dictionary-use and corpus linguistics. A series of experimental studies (N = 4,162) reveals systematic divergences among the verdicts delivered by modern concept use, dictionary use, and corpus linguistics use. The same results arise across levels of legal expertise—participants included 230 “elite-university law students” (e.g. at Harvard and Yale) and 98 United States judges. Ultimately, the data suggest that popular originalist methods carry serious risks of error—conservatively estimated, a 20-35% error rate. In some circumstances, even experts’ use carries extremely large error rates—between 80-100%. The chapter argues that these findings shift the argumentative burden onto originalist theories, challenging them to articulate and demonstrate a reliable use of these tools in legal interpretation.

Chapter 4 concludes by reflecting on themes connecting the earlier chapters and the project of “experimental jurisprudence.” The dissertation’s broader lesson is that experimental legal-philosophy, or “experimental jurisprudence,” fosters a unique type of wisdom. There are many reasons to learn how ordinary people understand legal concepts—promoting the clarity and publicity of law, ensuring consistency among judges and juries, and encouraging a democratic law making process. Experimental study also reveals specific conceptual features, enriching philosophical debates and bringing them closer to our world—the real arena of successful philosophical arguments’ practical effects. Empirical study alone cannot resolve philosophical questions about what law should be. But to have the best debate about where law should go, it is important to know from where we start.

Essays in Experimental Jurisprudence

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
In Candidacy for the Degree of
Doctor of Philosophy

by
Kevin P. Tobia

Dissertation Directors: Joshua Knobe and Gideon Yaffe

May 2019

ProQuest Number: 13810230

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 13810230

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

CONTENTS

ACKNOWLEDGEMENTS.....	i
INTRODUCTION.....	1-5
CHAPTER 1: HYBRID THEORIES OF REASONABLENESS	7-91
CHAPTER 2: LEGAL CONCEPTS AND LEGAL EXPERTISE.....	93-150
CHAPTER 3: TESTING ORIGINAL PUBLIC MEANING	151-287
CHAPTER 4: EXPERIMENTAL JURISPRUDENCE.....	289-332
CONCLUSION	333-336

Acknowledgments

This dissertation has been supported by my mentors, friends, and family. Most supporters fall into more than one of these categories.

Special thanks to my dissertation committee: Scott Shapiro, Steve Darwall, and especially co-chairs Joshua Knobe and Gideon Yaffe, for their extraordinary insight and generosity. I am also fortunate to have had several other inspirational and generous academic mentors: Owen Fiss, Abbe Gluck, Guy Kahane, Issa Kohler-Hausmann, Jeff McMahan, Derek Parfit, Regina Rini, Alexander Stremitzer, and especially Stephen Stich.

Thanks to those who have commented on or assisted with work related to these chapters: Mario Attie, Elliott Ash, Stefan Bechtold, Joanna Demaree-Cotton, Jennifer Daigle, John Doris, Brian Earp, Sam Elgin, Antuan Johnson, Dan Kelly, Markus Kneer, Vic Kumar, Arthur Lau, Jamie Macleod, Stefan Magen, Christian Mott, Eddy Nahmias, Laura Niemi, Jonathan Phillips, Juan Piñeros Glasscock, Will Ratoff, Kevin Reuter, Martin Schonger, David Shoemaker, Larry Solan, Larry Solum, Nina Strohminger, Roseanna Sommers, Kate Stanton, Simon Stern, Valerie Tiberius, Nina Varsava, Andrew Verstein, Natalia Washington, Pascale Willemsen, Yuan Yuan, Alex Zhang, and the participants of this experimental research.

Finally, thanks to my friends and family: Roger Aluja, Jason Bell, Christoph Berlin, Eric Branning, Marie Cambe, Nate Diskint, Ariel de Fauconberg, Johanna De Vos and Pascalle Tamis, Joe Fontana, Jon Gaunya, Ben and Katie Gold, Beth Malizsewski, Kevin Miller, Liz Percival, Jared Reichenberg, John Reinhardt, Ali Rizvi, Bill Royce, Jon Sullivan, Zach and Tori Sullivan, Minsern Teh, Trevor Teitel, Kaitlyn Walker, Drew Wagner, Neo Zambas; and the Best, Bilbo, Bollinger, Couzzi, Elliott, Fernstaedt, Klecha, Mallozzi, McCoy, Moore, Tarasuk, Tobia and Waters families. Most of all, thanks to Pat Tobia, Roni Tobia, David Tobia, and Richard Elliott.

© 2019 by Kevin Patrick Tobia

All rights reserved.

INTRODUCTION

On one view, law is a system built upon expert knowledge of technical concepts. Law is replete with specialized concepts: *extradition, disgorgement, habeas corpus, parol evidence, stare decisis, trust, voir dire*. In the United States, students pay hundreds of thousands of dollars to acquire these concepts, and bar examination boards certify who has achieved competence in the governing conceptual rules. Treatises and legal dictionaries articulate legal concepts' features, and legal scholars and jurists expertly debate how these specialized concepts should be applied.

On a radically different view, law is a system built upon ordinary concepts. In the United States, juries of ordinary citizens evaluate legal standards without formal legal training. When negligence cases turn on whether a defendant manifested "reasonable" care, ordinary citizens make this determination equipped only with the ordinary concept of reasonableness. Lay jurors produce similar decisions about a host of central legal concepts: was the act *intentional*, what *caused* the outcome, was the agreement formed with *consent*, what was the person's *motive*, what was in the person's *best interest*? No legal training is required to acquire the ordinary concept of reasonableness, intentional action, or causation. To be sure, law teaches some distinctive concepts that lack ordinary counterparts (e.g. parol evidence), but law's central concepts are ordinary ones.

Call these views the *Expert* and *Ordinary* views of legal concepts. These views oppose each other, but each captures certain aspects of legal practice. Moreover, each view finds further support from certain aspects of legal theory.

The *Expert* view reflects important legal features. For one, law should be *distinctive* from moral and conventional norms. Most legal theories propose that law is a separate sphere from morality or mere convention. Legal norms are not identical to ordinary ones, and ordinary norms often lack legal authority: promise breaking need not imply legal liability, and the state cannot prosecute rudeness. Relatedly, law should be *apt* at solving legal problems. Experts can help discover or craft the legal concepts underlying these distinctive and apt legal norms.

On the other hand, the *Ordinary* view captures the common legal-philosophical proposition that laws are sources of notification, guidance, or deterrence for *ordinary* citizens. Even those without legal training should be able to understand a legal rule that prohibits intentional killing. Such understanding includes, ostensibly, the law's concept of *intentional action*. Plausible accounts of legal legitimacy presume a similar picture about ordinary concepts and law. That law should be *public* suggests that ordinary people's conception of law matches that of the law's drafters. That law should be *consistent* suggests that a judge's or jury's application of a legal standard should not differ with respect to the same facts.

This dissertation is motivated by the tension between these two views of law. A brief reflection upon some examples reveals that neither picture represents a complete account of law. In contrast to the second view, some specialized legal concepts have no ordinary counterpart (e.g. parole evidence) and some terms have a legal meaning that is entirely distinct from an ordinary meaning (e.g. "standing"). Law is not built exclusively from ordinary concepts.

Yet, in contrast to the first view, there are undoubtedly other ordinary concepts that play legal roles. Law relies on ordinary notions of times, dates, and monetary values. If a legal outcome should occur on "the 20th day of January," it is the ordinary notion of this date that applies. The same is true for the concept of *years* used in sentencing guidelines or *dollars* used in any damage award. These do not have a special legal meaning.

However, between these extremes sit a host of legal concepts with more nebulous status. Is legal *reasonableness* a specialized legal concept or an ordinary one? Intriguingly, many of law's most central and outcome-determinative concepts have this more mysterious nature. Is there something distinctive about law's cross-cutting concepts like *causation*, *intention*, *reasonableness*, or *motive*, or does law take the ordinary concepts unaltered? What about civil procedure's concept of *plausibility*, or intellectual property law's concept of *obviousness*? What about constitutional concepts like a right to *keep and bear arms*, or a prohibition against *cruel and unusual punishment*, or even the concept of a *person*?

This dissertation aims to clarify the relationship between ordinary and legal concepts. The dissertation approaches these topics both descriptively and normatively, asking questions that include:

- What are the features of ordinary concepts that are central to legal rules and standards, and *should* legal rules and standards have those features?
- Do legal experts and non-experts have different legal concepts, and how *should* a dualistic system of expert judges and ordinary jurors ensure consistency in light of conceptual differences?
- Do common tools of legal interpretation accurately reflect the ordinary meaning of the concepts referenced in legal texts, and how *should* theories of legal interpretation respond to systematic inaccuracy?

The broadest form of the question—what is the relationship between ordinary concepts and legal concepts—is too broad to answer unequivocally in this dissertation. However, the dissertation begins to make progress through three essays, each of which studies in detail a subset of ordinary and legal concepts. The dissertation's guiding theme is that the experimental study of legal concepts raises new and important philosophical questions. Discovering the contours of ordinary and legal concepts provides a richer factual basis for understanding what law is and debating what it should be. When we ask whether a legal concept should reflect the ordinary concept or an expert concept, it is crucial to first understand the relevant facts about the concept.

Chapter 1 begins by examining a fundamental legal concept: legal reasonableness. Across many legal domains, there is a legal standard of what is “reasonable” or what a “reasonable” person would do, believe, expect, or understand. In many jurisdictions, demonstrating that “reasonable” care was taken defeats negligence liability. Law considers a wide range of reasonable domains: a reasonable number of days within which to accept a contract offer, a reasonable rate of attorney’s fees, a reasonable loan interest rate, a reasonable number of weeks’ delay before a criminal trial. The Chapter begins by exploring the ordinary concept of reasonableness. Experimental studies suggest that, across both ordinary and legal contexts, ordinary judgments of reasonableness represent a hybrid judgment. People’s view of what is reasonable reflects a mixture of both statistical considerations (i.e. what is common) and prescriptive considerations (i.e. what is good). The latter half of the chapter argues that this hybrid feature is one that reasonableness standards *should* reflect. In this case, the legal standard (legal reasonableness) should reflect features of the ordinary concept of reasonableness.

Chapter 2 considers differences between an ordinary concept and its legal counterpart. It conducts experiments that illuminate the concept of acting *intentionally*. The study finds a significant difference in the judgments of intentional action among ordinary people, law students, and United States judges. The differences are not best explained by a selection effect (i.e. the effect of people who have a different concept over-selecting into the legal profession). Instead, the differences are best explained by an effect of legal training and experience. That is, the data suggest that legal training and experience modify the ordinary concept of intentional action.

Chapter 3 tests the assumption that common tools of legal interpretation accurately reflect the ordinary meaning of the concepts referenced in legal texts. Popular theories of legal interpretation posit that interpretive tools, namely dictionaries and corpus linguistics, are accurate measures of the ordinary meaning of a legal text. A series of experiments suggest that, in fact, these methods

are often inaccurate. Dictionaries tend to reflect broad and extensive uses of language, while corpus linguistics tends to reflect narrow and prototypical uses of language. If, for instance, we sought to determine the modern (2019) ordinary meaning of “vehicle,” dictionaries and corpus linguistics would suggest very different conclusions. People (including U.S. judges) relying upon dictionaries are more inclined to categorize entities like trucks and golf carts as vehicles, as well as entities like canoes, bicycles, and even entities like skateboards and crutches. However those using corpus linguistics are much less inclined to categorize all those entities as vehicles. The chapter provides some recommendations for improving legal interpretation in light of these results, and it also elaborates critical challenges to theories of interpretation that rely on such methods.

Chapter 4 concludes by considering the themes that connect the earlier chapters: the relationship between ordinary and legal concepts, the role of legal training and expertise, the connection between ordinary cognition of law and legal legitimacy, and the promise of experimental methods in legal philosophy. In doing so it provides a sketch of a broader project of “experimental jurisprudence,” the experimental-philosophical study of legal concepts.

As we turn to the substantive chapters, it may be useful to keep in mind the two broader views about legal concepts, the *Expert* and *Ordinary* views of legal concepts. For each concept that we consider, we can first ask: should this be a special legal concept or an ordinary concept? As the experimental results reveal facts about the concepts, we can ask more sophisticated questions: does the legal concept reflect *this* feature of the ordinary concept, and should it? These investigations carry important practical implications, revealing facts about how ordinary actors and jurors conceptualize legal standards. But the experiments also raise important philosophical questions, providing insight into whether the conditions of law’s authority and legitimacy actually obtain.

CHAPTER 1

HYBRID THEORIES OF REASONABLENESS

A classic debate concerns whether reasonableness should be understood statistically (e.g., reasonableness is what is common) or prescriptively (e.g., reasonableness is what is good). This Article elaborates and defends a third possibility. Reasonableness is a partly statistical and partly prescriptive “hybrid,” reflecting both statistical and prescriptive considerations. Experiments reveal that people apply reasonableness as a hybrid concept, and the Article argues that a hybrid account offers the best general theory of reasonableness.

First, the Article investigates how ordinary people judge what is reasonable. Reasonableness sits at the core of countless legal standards, yet little work has investigated how ordinary people (i.e., potential jurors) actually make reasonableness judgments. Experiments reveal that judgments of reasonableness are systematically intermediate between judgments of the relevant average and ideal across numerous legal domains. For example, participants’ mean judgment of the legally reasonable number of weeks’ delay before a criminal trial (ten weeks) falls between the judged average (seventeen weeks) and ideal (seven weeks). So too for the reasonable number of days to accept a contract offer, the reasonable rate of attorneys’ fees, the reasonable loan interest rate, and the reasonable annual number of loud events on a football field in a residential neighborhood. Judgment of reasonableness is better predicted by both statistical and prescriptive factors than by either factor alone.

This Article uses this experimental discovery to develop a normative view of reasonableness. It elaborates an account of reasonableness as a hybrid standard, arguing that this view offers the best general theory of reasonableness, one that applies correctly across multiple legal domains. Moreover, this hybrid feature is the historical essence of legal reasonableness: the original use of the “reasonable person” and the “man on the Clapham omnibus” aimed to reflect both statistical and prescriptive considerations. Empirically, reasonableness is a hybrid judgment. And normatively, reasonableness should be applied as a hybrid standard.

CONTENTS

INTRODUCTION	9
I. THEORIES OF REASONABLENESS	12
A. <i>Statistical Theories: Reasonableness as What Is Common</i>	13
B. <i>Prescriptive Theories: Reasonableness as What Is Good</i>	16
II. A THIRD OPTION: HYBRID THEORIES	20
A. <i>Reasonableness as a Partly Statistical, Partly Prescriptive Hybrid</i>	21
1. <i>Prior Suggestions of Hybrid Theories</i>	21
2. <i>Hybrid Theories of Reasonableness</i>	23
3. <i>Normality as a Hybrid Concept</i>	28
B. <i>A Taxonomy of Reasonableness Theories</i>	30
III. EXPERIMENTAL STUDY OF ORDINARY REASONABLENESS JUDGMENT	35
A. <i>The Cognitive Science of Hybrid Concepts</i>	36
B. <i>Experimental Studies: Reasonableness as a Hybrid Concept</i>	38
1. <i>Experiment 1</i>	40
2. <i>Experiment 2</i>	42
3. <i>Experiment 3</i>	45
4. <i>Summary</i>	51
IV. A DEFENSE OF HYBRID THEORIES OF REASONABLENESS	51
A. <i>Hybrid Varieties</i>	52
1. <i>A Corrected Ideal</i>	54
2. <i>A Corrected Average</i>	55
3. <i>A Blended Standard</i>	55
B. <i>Origins of the Reasonable Person</i>	57
1. <i>The Man on the Clapham Omnibus</i>	57
2. <i>The Reasonable Person</i>	59
C. <i>The Ordinary Meaning and Use of Reasonable</i>	64
1. <i>The Ordinary Meaning of Reasonable</i>	65
2. <i>The Use of Reasonable</i>	67
D. <i>A General Theory of Reasonableness</i>	68
V. FURTHER IMPLICATIONS AND FUTURE DIRECTIONS	70
A. <i>Practical Implications of the Data</i>	71
1. <i>Legal Claimants</i>	71
2. <i>Legal Representatives</i>	73
3. <i>Jury Instructions</i>	73
B. <i>Individualization</i>	74
C. <i>The Increasing Use of Reasonableness and Reasonable Person</i>	79
CONCLUSION	85
APPENDIX A: FULL EXPERIMENTAL MATERIALS	86
APPENDIX B: FULL EXPERIMENTAL RESULTS	91

INTRODUCTION

Across countless legal domains, judicial outcomes turn on ordinary people’s determinations of “reasonableness.”¹ Reasonableness is a central legal concept and its history and uses have been well studied.² Nevertheless, there persists significant debate about how reasonableness should be understood.³ One of the most fundamental questions concerns whether reasonableness is a statistical notion (e.g., what is average) or a prescriptive one (e.g., what is good).

This chapter defends a third option. Reasonableness is best understood as a hybrid notion that is partly statistical and partly prescriptive. This chapter defends this claim on both an empirical and normative level. Experimental studies of ordinary people’s reasonableness judgments find that—empirically—reasonableness is more like a hybrid notion (e.g., normality) than a purely statistical notion (e.g., averageness) or prescriptive notion (e.g., welfare maximization). Next, this chapter argues that—normatively—reasonableness should be applied as a hybrid standard, rather than as a purely statistical or prescriptive one.

Part I outlines the two predominant groups of reasonableness theories. The first group of theories posits that reasonableness is a statistical notion. This set of views is sometimes associated

1. See, e.g., *Omnicare, Inc. v. Laborers Dist. Council Constr. Pension Fund*, 135 S. Ct. 1318, 1332 (2015) (“Numerous legal rules hinge on what a reasonable person would think or expect.”).

2. See generally MAYO MORAN, *RETHINKING THE REASONABLE PERSON: AN EGALITARIAN RECONSTRUCTION OF THE OBJECTIVE STANDARD* (2003); Mayo Moran, *The Reasonable Person: A Conceptual Biography in Comparative Perspective*, 14 LEWIS & CLARK L. REV. 1233 (2010); Simon Stern, *R. v. Jones (1703)*, in *LANDMARK CASES IN CRIMINAL LAW* 59 (Philip Handler, Henry Mares & Ian Williams eds., 2017).

3. See Christopher Jackson, *Reasonable Persons, Reasonable Circumstances*, 50 SAN DIEGO L. REV. 651, 652 (2013); Matt King, *Against Personifying the Reasonable Person*, 11 CRIM. L. & PHIL. 725, 726 (2017); Alan D. Miller & Ronen Perry, *The Reasonable Person*, 87 N.Y.U. L. REV. 323, 325 (2012); Christina Carmody Tilley, *Tort Law Inside Out*, 126 YALE L.J. 1320, 1327 (2017); Peter Westen, *Individualizing the Reasonable Person in Criminal Law*, 2 CRIM. L. & PHILOS. 137, 138–40 (2008); Benjamin C. Zipursky, *Reasonableness in and out of Negligence Law*, 163 U. PA. L. REV. 2131, 2132–35 (2015); John Gardner, *The Many Faces of the Reasonable Person*, N.Y.U. SCH. L. 3–4 (2015), http://www.law.nyu.edu/sites/default/files/upload_documents/The%20Many%20Faces%20of%20the%20Reasonable%20Person.pdf (last visited Oct. 6, 2018).

with Oliver Wendell Holmes and the idea that reasonableness is averageness.⁴ The second group of theories posits that reasonableness is a prescriptive notion. For these theories, reasonableness does not reflect something statistical like averageness, but instead reflects something purely prescriptive, such as welfare maximization, justification, virtue, or rightness.⁵

Part II identifies a third possibility. Reasonableness is neither purely statistical nor purely prescriptive. Instead, it is a partly statistical and partly prescriptive “hybrid” notion. Part II motivates and explicates this new possibility. It also provides a taxonomy of reasonableness theories, distinguishing between statistical, prescriptive, hybrid, conventionalist, and nonconventionalist theories of reasonableness.

Part III investigates a strikingly underexplored topic: how do ordinary people (i.e., potential jurors) actually make reasonableness judgments? It presents three original experiments that reveal that judgments of reasonableness are systematically intermediate between judgments of the relevant average and ideal. For example, participants’ mean judgment of the legally reasonable number of weeks for a product to be refundable (five weeks) falls between mean judgments of the average (four weeks) and ideal (six weeks).⁶ This effect arises across numerous legal domains: the reasonable number of days to accept a contract offer, the reasonable number of weeks of construction delay, the reasonable interest rate, the reasonable annual number of loud events on a football field in a residential neighborhood, and so on.⁷

This pattern of intermediacy between the relevant average and ideal is precisely the pattern of judgment that is characteristic of hybrid (partly statistical, partly prescriptive) concepts. The experiments indicate that reasonableness judgments are best predicted by the relevant average and ideal together, rather than by either the average or ideal alone. As such, reasonableness judgment

4. OLIVER WENDELL HOLMES, JR., *THE COMMON LAW* 108 (1881).

5. *See, e.g.*, Miller & Perry, *supra* note 3, at 328–35; Gardner, *supra* note 3, at 4–9.

6. *See infra* Parts II–III.

7. *See infra* Subpart III.B.

is better understood as reflecting a hybrid judgment than as reflecting either a purely statistical one or a purely prescriptive one.

Part IV uses this experimental discovery to support a normative account of reasonableness—an account of how reasonableness standards *should* be applied. It elaborates an account of reasonableness as a hybrid standard, advancing arguments for the distinction between statistical notions like averageness, prescriptive notions like welfare maximization, and hybrid notions like normality. Subpart IV.A distinguishes between three plausible hybrid accounts: reasonableness as a corrected ideal, reasonableness as a corrected average, and reasonableness as a hybrid concept. The next three Subparts offer three arguments for theorizing reasonableness as a hybrid standard. The original use of the “reasonable person” and its companion notion, the “man on the Clapham omnibus,”⁸ aimed to reflect judgment of a hybrid concept. Modern ordinary judgments reflect the same consideration. This conclusion supports two normative arguments—one from history and one from ordinary meaning and use—for hybrid theories of reasonableness. Finally, Part IV argues that a hybrid view is the best *general* theory of reasonableness, one that applies correctly across varied legal domains.

Part V begins by exploring some further implications of the hybrid view of reasonableness. Beyond theoretical implications, Part III’s empirical results have significant implications for legal practice. Understanding how ordinary people tend to make reasonableness judgments provides critical information for prospective legal claimants, legal representatives and decision makers, and drafters of jury instructions. Part V then turns to the well-known “individualization problem,”

8. See, e.g., *Healthcare at Home Ltd. v. Common Servs. Agency* [2014] UKSC 49, [1], [2014] 4 All ER 210 (appeal taken from Scot.).

the question of whether reasonableness should take account of factors like ability, age, culture, gender, mental illness, race, sexuality, or combinations of these. This is a difficult problem for any view—statistical, prescriptive, or hybrid—but the hybrid view reveals new possibilities for understanding and approaching this question. Part V concludes by considering future research on reasonableness. The core account of reasonableness is appropriate across a variety of legal contexts, but a hybrid standard may be a less appropriate one in some other contexts citing reasonableness. As such, the account calls for a cautionary restraint—or at least clarification—of the use of terms like *reasonable* and *reasonable person* in some legal domains.

I. THEORIES OF REASONABLENESS

Reasonableness sits at the core of various legal standards. The most well-known example is in the law of tort negligence. To determine whether a defendant is liable for negligently causing an injury, a jury might be asked to evaluate whether the defendant acted with “reasonable care” or the care of a reasonable person.⁹ Other examples abound. Reasonableness might settle whether someone has caused a public nuisance,¹⁰ whether a contract offer remains open,¹¹ whether a product may be returned within a certain timeframe after purchase,¹² whether a criminal trial is improperly prolonged,¹³ whether attorneys charge inappropriate fees,¹⁴ and many other legal issues.¹⁵

A fundamental debate about reasonableness concerns whether it is a statistical or prescriptive

9. RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 7 (AM. LAW INST. 2009).

10. RESTATEMENT (SECOND) OF TORTS § 821B (AM. LAW INST. 1979).

11. See, e.g., *Coleman v. Davies*, 235 P.2d 199, 203 (Wash. 1951) (“[I]n the absence of an acceptance of an offer . . . within a reasonable time (where no time limit is specified), there is no contract.”); *Sherrod ex rel. Cantone v. Kidd*, 155 P.3d 976, 977 (Wash. Ct. App. 2007) (citing *Minneapolis & St. Louis Ry. v. Columbus Rolling Mill*, 119 U.S. 149, 151 (1886)); RESTATEMENT (SECOND) OF CONTRACTS § 41 (AM. LAW INST. 1981); 1 JOSEPH M. PERILLO, *CORBIN ON CONTRACTS* § 2.16, at 203–04 (rev. ed. 1993).

12. E.g., *Greenwich Indus., L.P. v. Leggett & Platt, Inc.*, No. 07C6550, 2009 WL 1657441, at *4 (N.D. Ill. June 11, 2009).

13. E.g., 32 C.F.R. § 151.7(n) (2018).

14. See generally Samuel R. Berger, *Court Awarded Attorneys’ Fees: What Is “Reasonable”?*, 126 U. PA. L. REV. 281 (1977).

15. See, e.g., Zipursky, *supra* note 3, at 2137 (citing reasonable rates, 16 U.S.C. § 2621(18)(B) (2012), reasonably fair value for capital investment, NEB. REV. STAT. § 71-20, 108(5) (2014); reasonably clear liability, MASS. GEN. LAWS ch. 176D, § 3(9)(f) (2014), and reasonably detailed ballot proposals, R.I. GEN. LAWS § 17-8-10(b)(3) (2010)).

standard. Is “reasonable caution” the caution that an *average* person takes, or the caution that an *ideal* person ought to take? Scholars have defended two primary positions. Some argue that reasonableness is a purely statistical notion (e.g., averageness), while others argue that it is a purely prescriptive one (e.g., welfare maximization, community values, virtue, respect for freedom or rights, an ethic of care, or justification). Subpart I.A outlines the statistical case, and Subpart I.B outlines prescriptive views.

A. Statistical Theories: Reasonableness as What Is Common

Within the first group of statistical theories, most views characterize reasonableness as averageness.¹⁶ This is an interpretation sometimes associated with Holmes, who remarked that a “certain average of conduct . . . is necessary to the general welfare.”¹⁷

Holmes’s account focuses on defining the appropriate standard of precautions to be taken (i.e., it focuses on the negligence context). The theory characterizes the relevant standard as one looking to the way in which actual people “are in the habit of acting.”¹⁸ These statements have led

16. See, e.g., HOLMES, *supra* note 4, at 108.

17. *Id.* at 108–11 (“The standards of the law are standards of general application. . . [W]hen men live in society, a certain average of conduct, a sacrifice of individual peculiarities going beyond a certain point, is necessary to the general welfare. . . . The law considers, in other words, what would be blameworthy in the average man, the man of ordinary intelligence and prudence, and determines liability by that. If we fall below the level in those gifts, it is our misfortune; so much as that we must have at our peril, for the reasons just given. . . . Again, any legal standard must, in theory, be one which would apply to all men, not specially excepted, under the same circumstances. . . . The theory or intention of the law is not that the feeling of approbation or blame which a particular twelve may entertain should be the criterion. They are supposed to leave their idiosyncrasies on one side, and to represent *the feeling of the community*. The *ideal average prudent man*, whose equivalent the jury is taken to be in many cases, and whose culpability or innocence is the supposed test, is a constant, and his conduct under given circumstances is theoretically always the same.” (emphasis added)).

18. *Id.* at 112 (“From the time of Alfred to the present day, statutes and decisions have busied themselves with defining the precautions to be taken in certain familiar cases; that is, with substituting for the vague test of the care exercised by a prudent man, a precise one of specific acts or omissions. The fundamental thought is still the same, that the way prescribed is that in which prudent men are in the habit of acting, or else is one laid down for cases where prudent men might otherwise be in doubt.”). As some readers may notice, it is possible to read these passages to support different views. My own reading is that Holmes is not endorsing a strict average. He refers not to the actual average, but instead to the average of the *prudent* person’s actions. As such, I understand the “hy-

a number of scholars to identify Holmes with the statistical set of views.¹⁹

There are more modern accounts that also characterize reasonableness as a purely statistical notion. For example, the “reasonable person standard . . . considers conduct from the perspective of the hypothetical average person.”²⁰ Here again, the most typical characterization of this view is reasonableness as statistical averageness.

While this view has some modern defenders, it has more modern critics. As Peter Westen argues,

“reasonableness” is not an empirical or statistical measure of how average members of the public think, feel, or behave. Average is not the same as right or appropriate. Regrettably, average persons have been known to think, feel, and behave very differently from the way that the polity to which they are duty-bound believes they should, and when they do, they are answerable to the polity for their failings. Rather, reasonableness is a normative measure of ways in which it is *right* for persons to think, feel or behave—or, at the very least, ways in which it is *not wrong* for them to do so.²¹

For others, the idea that reasonableness is averageness borders on the absurd. Consider a statement from Justice Breyer:

[T]he “reasonable person” standard does not require a court to pretend that [the seventeen-year-old] was a 35-year-old with aging parents whose middle-aged children do what their parents ask only out of respect. Nor does it say that a court should pretend that [he] was the statistically determined “average person”—a working, married, 35-year-old white female with a high school degree.²²

This critique of statistical views appears in the U.S. Supreme Court, as well as the high courts of other countries. A recent decision from the United Kingdom articulates a similar argument. Lord Reed, writing for the majority, asserts that reasonableness (and its British companion, the

brid view,” see *infra* Part IV, as a Neo-Holmesian view. Despite the conventional wisdom, Holmes did not characterize reasonableness as averageness. He characterized it as something close to normality.

19. E.g., Miller & Perry, *supra* note 3, at 370; Robert L. Rabin, *The Historical Development of the Fault Principle: A Reinterpretation*, 15 GA. L. REV. 925, 931 (1981).

20. Deborah Zalesne, *The Intersection of Socioeconomic Class and Gender in Hostile Housing Environment Claims Under Title VIII: Who is the Reasonable Person?*, 38 B.C. L. REV. 861, 863 n.15 (1997); see also Joshua Dressler, *When “Heterosexual” Men Kill “Homosexual” Men: Reflections on Provocation Law, Sexual Advances, and the “Reasonable Man” Standard*, 85 J. CRIM. L. & CRIMINOLOGY 726, 745–49 (1995).

21. Westen, *supra* note 3, at 138 (citations omitted); see also Miller & Perry, *supra* note 3, at 371.

22. *Yarborough v. Alvarado*, 541 U.S. 652, 673–74 (2004) (Breyer, J., dissenting). *But see* *J.D.B. v. North Carolina*, 564 U.S. 261, 275–76 (2011).

man on the Clapham omnibus) is a purely prescriptive notion of justice:

The Clapham omnibus has many passengers. The most venerable is the reasonable man [I]ts most famous passenger, and the others I have mentioned, are legal fictions. They belong to an intellectual tradition of defining a legal standard by reference to a hypothetical person, which stretches back to the creation by Roman jurists of the figure of the *bonus paterfamilias*. As Lord Radcliffe observed . . . “[t]he spokesman of the fair and reasonable man, who represents after all no more than the anthropomorphic conception of justice, is and must be the court itself.”²³

Lord Reed elaborates a strong antistatistical perspective, claiming that statistical facts are entirely irrelevant to determinations of legal reasonableness, and judgments of legal reasonableness should not consider statistical facts about actual persons:

[I]t would [be] misconceived for a party to seek to lead evidence from actual passengers on the Clapham omnibus as to how they would have acted in a given situation or what they would have foreseen, in order to establish how the reasonable man would have acted or what he would have foreseen. Even if the party offered to prove that his witnesses were reasonable men, the evidence would be beside the point. The behaviour of the reasonable man is not established by the evidence of witnesses, but by the application of a legal standard by the court.²⁴

There are two more standard critiques of statistical views. These are the problems of “average accidents” and “reasonable racism.” Both critiques point to a problem with grounding reasonableness in statistical facts about average or prevalent behaviors or beliefs. If reasonableness is averageness, does that not (incorrectly) excuse average accidents, unintentional injuries that occur because of carelessness that is typical? Similarly, if reasonableness is averageness, does that not (incorrectly) excuse racist outcomes that occur because of common racist beliefs and behaviors?

A well-known example of the average accidents challenge comes in *The T.J. Hooper* tugboat

23. *Healthcare at Home Ltd. v. Common Servs. Agency* [2014] UKSC 49, [1]–[2], [2014] 4 All ER 210 (appeal taken from Scot.) (citing *Davis Contractors Ltd. v. Farcham Urban Dist. Council* [1956] AC 696, 728 (appeal taken from Eng.)).

24. *Id.* at [3].

case.²⁵ In that negligence action, the defendant argued that the reasonableness of the decision not to install a reliable radio safety device should be determined by common practice. Judge Hand held that “[c]ourts must in the end say what is required; there are precautions so imperative that even their universal disregard will not excuse their omission.”²⁶ If so, reasonableness cannot be a purely statistical notion.²⁷

The reasonable racism problem has a similar structure. Racist acts cannot be excused as legally reasonable ones simply because they are common ones. As Armour put it, “[t]he role of the courts, from this [statistical] perspective, is to *observe* rather than *define* the attributes of the reasonable man.”²⁸ Critics of statistical views take the job of the court to be the articulation of a judgment of reasonableness, not simply a discovery of it in typical or statistically average practice. That statistical view (discovering reasonableness in typical practice) is particularly vulnerable when typical practices are racist ones.

Although the statistical view has some modern defenders,²⁹ it has more modern critics.³⁰ In response to critiques of the view that reasonableness is a purely statistical notion (like average-ness), theorists have elaborated various thoughtful accounts on which reasonableness is a purely prescriptive notion. The next Subpart outlines those accounts.

B. *Prescriptive Theories: Reasonableness as What Is Good*

In modern legal theory, this second group of prescriptive theories is far more prevalent. Cor-

25. The T.J. Hooper, 60 F.2d 737 (2d Cir. 1932).

26. *Id.* at 740.

27. Note, however, that Judge Hand also remarks that, in “most cases reasonable prudence is in fact common prudence; but strictly it is never its measure; a whole calling may have unduly lagged in the adoption of new and available devices.” *Id.* His account may be better understood as referencing both statistical and prescriptive considerations, while holding that neither is decisive in itself. If so, this gestures towards a hybrid view.

28. JODY DAVID ARMOUR, *NEGROPHOBIA AND REASONABLE RACISM: THE HIDDEN COSTS OF BEING BLACK IN AMERICA* 32 (1997).

29. See, e.g., Dressler, *supra* note 20, at 752–53; Zalesne, *supra* note 20, at 863 n.15.

30. See, e.g., *Yarborough v. Alvarado*, 541 U.S. 652, 673–74 (2004) (Breyer, J., dissenting); *Healthcare at Home Ltd. v. Common Servs. Agency* [2014] UKSC 49, [2014] 4 All ER 210 (appeal taken from Scot.); *Miller & Perry*, *supra* note 3, at 371; *Westen*, *supra* note 3, at 138.

respondingly, there is more diversity within this group of theories. While statistical views theorize reasonableness as some kind of average, prescriptive views offer more numerous theoretical variations. Reasonableness is an economic cost–benefit analysis,³¹ or it is grounded in community “values,”³² or it is some other prescriptive notion (e.g., normative “justification”).³³

This Subpart outlines three distinct groups of prescriptive views. The first group contains the very popular objective (cost–benefit) welfare maximization views. On these views, reasonableness is a standard of efficiency or welfare maximization. The second group characterizes reasonableness as a product of something more subjective, like conventional values—the values of a community. On these views, reasonableness is grounded in community values or a community contractualist moral understanding. The final group contains other objectivist views. On these views, reasonableness is not an objective standard of efficiency but is instead defined by some other objective, prescriptive notion. For example, reasonableness is best understood through virtue ethics, or as the promotion of freedom, or as justification.

The first group of prescriptive theories contains welfare maximization theories. Broadly speaking, these views theorize reasonableness standards as cost–benefit maximization tests. This view is often associated with the tort negligence context.³⁴ To determine whether someone acted reasonably, we do not consider typical or statistically average behavior. Instead, we look to the relevant injury and its probability and burden of avoidance. On one well-known version of this view, liability depends upon whether the expected (dis)value of the injury is greater than the bur-

31. RICHARD POSNER, *ECONOMIC ANALYSIS OF LAW* 3–4 (8th ed. 2011).

32. See Tilley, *supra* note 3, at 1324.

33. John Gardner, *The Mysterious Case of the Reasonable Person*, 51 U. TORONTO L.J. 273, 273 (2001); Gardner, *supra* note 3, at 4–9.

34. See generally *United States v. Carroll Towing Co.*, 159 F.2d 169 (2d Cir. 1947); see also WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF TORT LAW* 85–88 (1987); POSNER, *supra* note 31, at 316–17.

den of avoidance.³⁵

The second group of prescriptivist views understands reasonableness as community values. These are “conventionalist” prescriptivist views. For example, Gregory Keating rejects prescriptivism’s first group of efficiency views, opting for a community contractualist view: “Social contract theory rejects the economic conception of reasonable care as the level of precaution that maximizes wealth. Instead, social contract theory views reasonable care as the level of care that *fairly* reconciles the conflicting liberties of injurers and victims.”³⁶ Although this second group rejects one prescriptive notion (objective cost–benefit welfare maximization), it still contains purely prescriptive views—ones that theorize reasonableness in terms of other purely prescriptive communitarian notions like fairness or local (prescriptive) values.

Importantly, this second group of conventionalist (or “community-values”) prescriptive views has several different varieties.³⁷ For example, reasonableness might be understood as (1) a community’s historical moral standards or (2) a community’s modern moral standards. Theorizing reasonableness as a notion grounded in community values does not imply that it is conventionalist in the sense of being tied to historical values. Reasonableness might be understood as a vehicle of purely modern community values. This would be a modern conventionalist interpretation. Alternatively, reasonableness might be understood as a mixture of historical and modern values. This would be a mixed conventionalist interpretation.³⁸

35. *Carroll Towing Co.*, 159 F.2d at 173 (“Possibly it serves to bring this notion into relief to state it in algebraic terms: if the probability be called P; the injury, L; and the burden, B; liability depends upon whether B is less than L multiplied by P: i.e., whether $B < PL$.”); see also Guido Calabresi, *Some Thoughts on Risk Distribution and the Law of Torts*, 70 YALE L.J. 499, 528–34 (1961).

36. Gregory C. Keating, *Reasonableness and Rationality in Negligence Theory*, 48 STAN. L. REV. 311, 349 (1996) (emphasis added); see also ARTHUR RIPSTEIN, *EQUALITY, RESPONSIBILITY, AND THE LAW* 7 (1999) (“The familiar common-law idea of the reasonable person gives expression to this idea of a fair balance between liberty and security. . . . The reasonable person is neither the typical nor the average person. Nor is the reasonable person to be confused with the rational person, who acts effectively in pursuit of his or her ends. Instead, the reasonable person needs to be understood as the expression of an idea of fair terms of social cooperation.”).

37. For discussion of hybrid conventionalist accounts, see *infra* Subpart II.B.

38. The next Part discusses hybrid views—ones in which reasonableness is partly statistical and partly prescriptive. Conventionalism is consistent with statistical, prescriptive, and hybrid views. As should be clear, theorizing reasonableness as a notion grounded in community conventions does not necessarily imply that reasonableness is a statistical notion, or a hybrid one. The prescriptivist set of views includes only those that theorize reasonableness in a purely prescriptive way.

For an example of a view of (tort) reasonableness as conventional community *values* in a purely prescriptive sense, consider Christina Carmody Tilley's recent article on tort law. She argues that "tort doctrine's reliance on community as the source of norms . . . encourages decision makers to toggle between traditional and modern values—between morality and efficiency."³⁹ The view's degree of conventionalism is "mixed." The view balances traditional and modern values. However, the view's degree of prescriptivism is not at all diluted. This is a doubly prescriptive view, as morality and efficiency are both prescriptive considerations. This is an important demonstration that conventionalism (of any degree) need not imply a statistical view (of any degree, including hybridism). A great number of conventionalist views are purely prescriptive ones.

A third group of prescriptive views explains reasonableness in terms of some more objective normative notion. Like the first group, these are "objectivist" theories, in the sense that they define reasonableness in terms of some objective notion rather than some conventionalist notion. Unlike the first group, these views do not explain reasonableness in terms of welfare maximization. Instead, this third group references other objectivist (i.e., nonconventional) prescriptive notions.

For example, these theories might explain reasonableness in terms of virtue ethics,⁴⁰ a Kantian notion of equal freedom,⁴¹ an ethic of care,⁴² or justification.⁴³ This third set is another very large group of views, but for the purposes of this chapter, the important commonality is that these

39. Tilley, *supra* note 3, at 1325.

40. See Heidi Li Feldman, *Prudence, Benevolence, and Negligence: Virtue Ethics and Tort Law*, 74 CHI.-KENT L. REV. 1431, 1431–32 (2000).

41. Miller & Perry, *supra* note 3, at 348–55.

42. *Id.* at 361–66; see also Leslie Bender, *Feminist (Re)Torts: Thoughts on the Liability Crisis, Mass Torts, Power, and Responsibilities*, 1990 DUKE L.J. 848, 901–08 (1990).

43. Gardner, *supra* note 3, at 4–9.

are all prescriptive views. This third group of prescriptive views theorizes reasonableness not as efficiency maximization or community values, but instead as a separate objective normative standard.

Unlike the second group of prescriptive views, this third set of views is not conventionalist. Reasonableness is not defined by the justifications that are intersubjectively agreed upon; instead, reasonableness is theorized objectively, for example by considering what is really normatively justifiable. But unlike the first group of views, this objective prescriptivism is not about cost–benefit efficiency. Instead, reasonableness is defined by other prescriptive considerations, such as those about virtue, freedom, care, or justification.

The key commonality among all three of these groups—reasonableness as cost–benefit welfare maximization, conventional values, or other objective values—is that reasonableness is some purely prescriptive notion (e.g., justifiability, welfare maximization, or rightness).⁴⁴ These views are “Ideal Person” interpretations, “appeal[ing] not to an average person but a *better* person, such as those who are ideally careful and virtuous.”⁴⁵ Across all of these views, reasonableness is not any kind of reflection of statistical facts or statistical commonalities. Instead, it “is an ideal.”⁴⁶

II. A THIRD OPTION: HYBRID THEORIES

The modern debate about reasonableness largely pits (purely) statistical theories against (purely) prescriptive ones.⁴⁷ However, these two options do not exhaust the debate’s conceptual space. Between these two extremes, there is the possibility of an underdeveloped third view. A

44. See generally Miller & Perry, *supra* note 3 (concluding that a prescriptive definition of reasonableness is superior to a statistical definition).

45. Steven P. Scalet, *Fitting the People They Are Meant to Serve: Reasonable Persons in the American Legal System*, 22 L. & PHIL. 75, 81 (2003).

46. Robert B. Mison, *Homophobia in Manslaughter: The Homosexual Advance as Insufficient Provocation*, 80 CALIF. L. REV. 133, 160 (1992).

47. See, e.g., Miller & Perry, *supra* note 3, at 334–35; Westen, *supra* note 3, at 142 n.23. But see Zipursky, *supra* note 3, at 2145–46 (acknowledging a hybrid possibility).

hybrid theory posits that reasonableness is partly statistical and partly prescriptive. Subpart II.A elaborates this view. Subpart II.B provides a taxonomy of different theories of reasonableness, distinguishing between statistical, prescriptive, hybrid, conventionalist, and nonconventionalist views. It also proposes a new vocabulary to promote clarity in discussions of these different views.

A. Reasonableness as a Partly Statistical, Partly Prescriptive Hybrid

A hybrid view rejects the statistical/prescriptive dichotomy. Reasonableness is neither a purely statistical notion (e.g., averageness) nor a purely prescriptive notion (e.g., welfare maximization, rightness, or virtue ethics). Instead, reasonableness should be understood as a judgment that is informed by *both* statistical and prescriptive considerations. On a hybrid view, considerations about what people *actually do* (i.e., statistical considerations) are neither decisive nor irrelevant. So too for prescriptive considerations about what people *should do*. Reasonableness is not determined by statistical or prescriptive considerations alone; instead, both types of considerations inform reasonableness judgments.

1. Prior Suggestions of Hybrid Theories

Hybrid views are strikingly underrepresented in the reasonableness literature. This may be in part because hybrid views are sometimes dismissed by an overextension of a critique of statistical views. Recall the common critiques of statistical views. For example, Westen remarks: “[R]easonableness’ is not an empirical or statistical measure of how average members of the

public think Average is not the same as right or appropriate.”⁴⁸ This is a fair critique of statistical views—reasonableness cannot be determined by only statistical considerations—but to use this critique to dismiss hybrid theories relies on an invalid inference. The overextended critique begins with (1) the observation that reasonableness cannot simply be what is statistically average. It then infers (2) that reasonableness must therefore be a purely prescriptive notion. Such dismissiveness is unwarranted. The fact that reasonableness is not purely statistical does not mean that it is not at all statistical.

One of the only recent explicit statements in support of a hybrid view comes from Benjamin Zipursky. He endorses a hybrid view of reasonableness, claiming it “involves a kind of judgment that is both normative and descriptive.”⁴⁹ While Zipursky is a clear proponent of this third set of views, his treatment of the normative/prescriptive debate occurs in just two paragraphs of his larger article.⁵⁰

Some statements from other theorists might be understood as endorsements of a hybrid theory. For example, Patrick Kelly articulates a conventionalist account of negligence law. His studies note that, in the language of jury instructions, “ordinary” is commonly used to set the standard of care, and this supports a conventionalist account of tort negligence.⁵¹ On this view, reasonableness (in the negligence context) concerns what the relevant community adheres to conventionally. Or, as Robert Post puts it, the reasonable person is, in essence, “the norms of the . . . community.”⁵²

It is possible to read these conventionalist accounts of reasonableness as endorsing the relevance of statistical features. On such an interpretation, these accounts explain reasonableness with

48. Westen, *supra* note 3, at 138.

49. Zipursky, *supra* note 3, at 2150.

50. *Id.* at 2149–50.

51. Patrick J. Kelley & Laurel A. Wendt, *What Judges Tell Juries About Negligence: A Review of Pattern Jury Instructions*, 77 CHL-KENT L. REV. 587, 622–23 (2002).

52. Robert C. Post, *Community and the First Amendment*, 29 ARIZ. ST. L.J. 473, 477 (1997).

reference to community “norms” or “ordinary customs,” understood in a hybrid way.

However, it is not clear that these conventionalist statements must be understood as statements of a hybrid view. As Subpart I.B noted, and as Subpart II.B develops further, conventionalism is orthogonal to the statistical/prescriptive/hybrid distinction. These conventionalist theories might best be understood as articulating a statistical account: reasonableness is grounded in the community’s (purely statistical) customs. Or they might articulate a prescriptivist account: reasonableness is grounded in the community’s (purely prescriptive) values. Or they might articulate a hybrid account: reasonableness is grounded in a community’s hybrid (partly descriptive, partly prescriptive) norms.

2. Hybrid Theories of Reasonableness

On the hybrid view, reasonableness is neither a purely statistical notion, nor is it a purely prescriptive one. Instead, judgment about reasonableness is a hybrid judgment—one that reflects *both* statistical and prescriptive considerations.

Given the modern preference for prescriptive theories over statistical ones, a notable feature of hybrid theories is their endorsement of an important role for statistical considerations. On a hybrid view, consideration of what is statistically typical is central to reasonableness analyses. However, unlike statistical views, hybrid views do not treat these statistical considerations as decisive. Instead, reasonableness judgment is the product of a more complex consideration of *both* statistical and prescriptive factors.

The hybrid view can be motivated by the intuition that, in many examples of reasonableness standards, both statistical and prescriptive considerations seem critical. More specifically, the

judgment of what is reasonable seems to be a combination of those two types of factors.

As a first example, consider a classic case of tort negligence. Recall *The T.J. Hooper* example in which a boat operator failed to include new safety features on his boat, and a storm caused the boat to lose its cargo.⁵³ The cargo owners sued for negligence under a standard of “reasonable prudence.”⁵⁴ In this example, it seems that the typical prudence of boat owners is relevant, as is the prudence that boat owners ought to have. Moreover, the correct reasonableness judgment results from a combination of these considerations. The typically prudent boat operator has not included every new safety feature, and the ideally prudent boat operator would include all of the available safety features. Our judgment about the reasonably prudent boat owner stems from a combination of these considerations.

As a second example, consider criminal law’s affirmative defense of duress. The defense applies to an allegation of criminal conduct where the person “was coerced to [act] by the use of, or a threat to use, unlawful force . . . that a person of *reasonable firmness* in his situation would have been unable to resist.”⁵⁵ In applying this standard, it seems clear that both statistical and prescriptive considerations are crucial. We care about both the firmness most people *would* have in the relevant situation and what firmness someone *should* have in that situation. Moreover, the right reasonableness determination seems like some combination of these considerations. If it turns out that people are generally weak-willed, reasonable firmness is not simply the firmness that most (weak-willed) people have. Nor is it simply the firmness that an ideal person would have. Both of these factors inform our judgment, but neither is decisive in itself.

These two examples illustrate the combinative aspect of the hybrid view. Reasonableness judgments result from the combination of statistical and prescriptive considerations. The previous

53. *The T.J. Hooper*, 60 F.2d 737, 737–38 (2d Cir. 1932).

54. *Id.* at 740.

55. MODEL PENAL CODE § 2.09(1) (AM. LAW INST. 1985) (emphasis added).

two examples involve *qualitative* judgments. The boat owners vary in terms of their prudence, but it is hard to quantify the exact amount of prudence displayed. Similarly, people display different levels of firmness. Even without assigning numbers to these levels, it is possible to conceptualize a combination of statistical and prescriptive considerations about those levels.

The theory applies similarly to *quantitative* examples. For example, consider a contracts case in which a product ordered online has no specified return policy. What amount of time is reasonable for the customer to return the product for a refund? The hybrid view would predict that the reasonable time is informed by both the average and ideal times. When those significantly diverge, the view predicts that the reasonable quantity would be intermediate. For example, perhaps the average time of return is four weeks, and the ideal time for return seems greater, more like six weeks. If so, the hybrid view would posit a reasonable time that is intermediate (e.g., five weeks).

As another quantitative example, consider the reasonable delay of a criminal trial. What seems reasonable for a delay would be a product of consideration of the relevant average and ideal. For example, if the ideal waiting time is two months and the average waiting time is more like four months, the hybrid view posits an intermediate reasonable amount (say, ten to eleven weeks).

In both the qualitative and quantitative cases, the primary insight of the hybrid view is that a combination of statistical and prescriptive features determines reasonableness. The general view says nothing more about the exact relationship (e.g., one-third statistical, two-thirds prescriptive). As such, it does not posit that, for quantitative cases, the reasonable amount is *precisely* intermediate between the relevant average and ideal.

Nevertheless, the quantitative examples provide useful test cases of the hybrid view's predic-

tions. Part III uses these kinds of quantitative cases to test the view. If people understand reasonableness as a hybrid notion, their judgments of reasonableness should be intermediate between the relevant average and ideal (when the relevant average and ideal diverge).

There are several benefits of the hybrid view. One of the greatest virtues of the hybrid view is that it is most representative of the way in which many courts actually understand and apply reasonableness standards. Consider the “reasonable consumer” standard in false advertising actions.⁵⁶ To challenge an advertisement, plaintiffs must show that a reasonable consumer is likely to be deceived or misled. It would be very strange to apply this standard in a purely prescriptivist way, considering only what *should* mislead people and eschewing all consideration of what *actually* misleads consumers. The standard is not meant to address only what ideally ought to mislead people. Consideration of whether people are typically misled seems highly relevant. And so it is in practice. Most courts consider statistical considerations, and some even require a consumer survey or other evidence demonstrating that the advertisement actually tends to mislead consumers.⁵⁷

While it is unintuitive and unrepresentative for reasonableness to eschew all consideration of statistical factors, it is equally problematic for reasonableness to be determined by only statistical averages. Recall the problems of average accidents and reasonable racism. It is well established that reasonableness cannot be a purely statistical notion.⁵⁸ The advantage of the hybrid view is that it acknowledges both statistical and prescriptive considerations as relevant to reasonableness, but it also rejects treating either as determinative. Thus, one core benefit of hybrid accounts of reasonableness is their capture of the relevance of both statistical and prescriptive considerations.

56. See, e.g., *Fink v. Time Warner Cable*, 714 F.3d 739, 741 (2d Cir. 2013); *In re Horizon Organic Milk Plus DHA Omega-3 Mktg. & Sales Practice Litig.*, 955 F. Supp. 2d 1311, 1330 (S.D. Fla. 2013); *Elias v. Hewlett-Packard Co.*, 950 F. Supp. 2d 1123, 1131 (N.D. Cal. 2013).

57. See, e.g., *Haskell v. Time, Inc.*, 965 F. Supp. 1398, 1406–07 (E.D. Cal. 1997).

58. See, e.g., Douglas Husak, *The “But-Everyone-Does-That!” Defense*, 10 PUB. AFF. Q. 307, 311 (1996); Westen, *supra* note 3, at 138–39.

A second significant benefit is the hybrid view's plausibility as a general theory of reasonableness, one that applies correctly across multiple legal domains. While most theories of reasonableness are plausible in some legal domains (e.g., tort negligence or criminal law), few are plausible as general theories—ones that apply well across the many diverse uses of reasonableness standards.

Most statistical and prescriptive theories of reasonableness are plausible in some specific contexts—for example in the tort negligence context—but highly implausible in others. For example, statistical views of reasonableness might seem compelling as grounds of predictable and unwavering standards for some traditional cases of tort negligence, but they are less plausible as grounds for standards of sexual harassment—or in other circumstances in which actual practice departs radically from legal judgments of liability.

Similar problems also arise for prescriptivist views. Many prescriptivist reasonableness theories were developed in the tort negligence context. These views are plausible in that domain but are much less plausible in others. Although cost-benefit theories or conventionalist value theories might apply well to traditional cases of tort negligence, they are less compelling in other areas. Recall the reasonable-consumer standard for misleading advertising. The prescriptivist interpretation—that such standards should not consider facts about how ordinary people are misled—does not fit well with common sense or actual legal practice.

Many prescriptive views are an especially strange fit in the criminal context. For example, consider prescriptive accounts of “reasonable provocation” to kill. To be provoked to kill in any circumstance seems inconsistent with the typical virtues endorsed by virtue theories of reasonableness. Similarly, it inappropriately infringes on the freedom and rights of the person killed and

is inconsistent with any plausible ethics of care. And although killing might be excusable under the circumstances, it is not appropriately understood as prescriptively justifiable. Similarly, it is hard to see how being provoked to kill could be welfare-maximizing or efficient.

On a prescriptive view, “the reasonable person does not kill at all, even under provocation.”⁵⁹ It would be more welfare-maximizing, justified, virtuous, and morally appropriate never to be provoked to kill. What gets reasonable provocation off the ground (as a standard at all) are statistical considerations: in certain circumstances, ordinary people are typically provoked to kill. Moreover, when someone is judged to be *reasonably* provoked, this affects their legal liability (e.g., liable for manslaughter, not murder), but it does not mean that the act was welfare maximizing, efficient, virtuous, respecting of freedom or rights, or consistent with an ethics of care.

At the same time, these criminal standards are not plausibly defined by only statistical considerations. Reasonable provocation to kill is not the average provocation. The hybrid view captures the intuition that this standard should reflect a combination of both considerations: reasonable provocation judgment reflects the combined consideration of what people actually do and what people should do.

3. Normality as a Hybrid Concept

Another reason that this group of hybrid theories has been less developed than statistical and prescriptivist accounts may be because there is no obvious corresponding hybrid ordinary concept. Statistical theories can characterize the reasonable as “the average” and prescriptive theories can characterize it as “the ideal” (or virtuous, justified, etc.), but no similar ordinary concept for hybrid theories has been identified.

One ordinary notion that may be helpful to hybrid theories is the concept of normality. Re-

59. GEORGE P. FLETCHER, *RETHINKING CRIMINAL LAW* 247 (1978).

cent experimental research has found compelling evidence that one's judgment of what is "normal" is characterized by precisely a statistical-prescriptive blend.⁶⁰ Judgment of what is normal is best predicted by considering both the relevant average and the relevant ideal. This concept of normality can play a helpful analogical role to hybrid theories—a role similar to that played by the concept of averageness for statistical theories or ideality for prescriptive theories.

Scholars across disciplines—including psychology, linguistics, philosophy, and behavioral economics—have studied how representations of normality play a significant role in people's cognition and ordinary life.⁶¹ Most relevant to this chapter is a recent study that suggests that judgments of normality are characterized by a hybrid feature. "[P]eople's normality judgments take into account both statistical considerations (e.g., the statistical notion of the average) and more prescriptive [normative] considerations (e.g., what is morally ideal)."⁶² Strikingly, across a large number of categories, people's representation of what is normal falls between the representation of what is average and the representation of what is ideal. For example, the normal number of hours of television watched per day falls between the average and ideal number.⁶³ The same is true for the normal number of calories to consume each day, the normal number of drinks that a "frat brother" has each weekend, and many other categories of normality.⁶⁴

This research has several implications for the study of reasonableness. First, it makes clear

60. See Adam Bear & Joshua Knobe, *Normality: Part Descriptive, Part Prescriptive*, 167 *COGNITION* 25, 25 (2017).

61. See generally *id.* (citing DAVID R. DOWTY, *WORD MEANING AND MONTAGUE GRAMMAR: THE SEMANTICS OF VERBS AND TIMES IN GENERATIVE SEMANTICS AND IN MONTAGUE'S PTQ* (1979); Robert B. Cialdini, Raymond R. Reno & Carl A. Kallgren, *A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places*, 58 *J. PERSONALITY & SOC. PSYCHOL.* 1015 (1990); Alexander Peysakhovich & David G. Rand, *Habits of Virtue: Creating Norms of Cooperation and Defection in the Laboratory*, 62 *MGMT. SCI.* 631 (2015); and Seth Yalcin, *Modalities of Normality*, in *DEONTIC MODALITY* 230 (Nate Charlow & Matthew Chrisman eds., 2016)).

62. Bear & Knobe, *supra* note 60, at 25.

63. *Id.* at 28.

64. *Id.*

that ordinary hybrid concepts exist. Judgments of normality are informed by both statistical and prescriptive considerations. This provides hybrid theorists of reasonableness with a useful folk notion for analogy. Where statistical views understand reasonableness as a purely statistical notion, like averageness, and prescriptive views understand it as a purely prescriptive notion, like welfare maximization, hybrid theorists might understand reasonableness as a hybrid notion like normality.

This cognitive science research on normality also provides a helpful set of materials and hypotheses to test which of these three types of theories—statistical, prescriptive, or hybrid—best represents ordinary judgments of reasonableness.⁶⁵ Do ordinary people (i.e., potential jurors) judge reasonableness statistically, prescriptively, or in a hybrid way?

For those examples in which the relevant average and ideal differ, the three theories of reasonableness offer different verdicts. Statistical theories posit that reasonableness is and should be best described by the statistical average. Prescriptive theories posit that reasonableness is and should be best described by the prescriptive ideal. And hybrid theories posit that reasonableness is and should be best described by an intermediate hybrid notion (like normality). For example, imagine that people understand the time for a product to be refundable (with no specifying warranty) as four weeks and the ideal time to be six weeks. The statistical view posits that the reasonable time is closer to four weeks. The prescriptivist view posits that the reasonable time is closer to six weeks. And the hybrid view posits that the reasonable time is informed by both considerations and therefore closer to five weeks.

B. A Taxonomy of Reasonableness Theories

Before turning to the next Part, it is worth providing additional clarity regarding the concep-

65. See *infra* Part III.

tual distinction between (1) statistical, prescriptive, and hybrid views and (2) conventionalist vs. nonconventionalist views.⁶⁶ This Subpart proposes a new vocabulary to help distinguish among these theories and some subtle conceptual differences. This Subpart's most important broader point is this: Conventionalism is orthogonal to the debate among statistical, prescriptive, and hybrid theories, so hybrid theories should not be confused with conventionalist ones. Although there are many well-developed conventionalist accounts, most of those are best understood as nonhybrid accounts. This insight helps explain why hybrid views have been overlooked.

Consider Table 1 below, which proposes a vocabulary to track these distinctions. The table's particular terminology is less important than the fact that these divisions track important distinctions about reasonableness. Without such a vocabulary, too often these significant distinctions are blurred.

Each box of the Table proposes a terminology for the relevant object of reasonableness theories. For example, a "Traditional Conventionalist Statistical" theory of reasonableness theorizes reasonableness as a judgment of traditional community customs, while a "nonconventionalist prescriptive" theory theorizes reasonableness as a judgment of objective values.

The first row identifies the distinction between statistical, prescriptive, and hybrid theories. This difference is the primary focus of this chapter. The final entry in the first row ("ambiguous") identifies terms that are ambiguous between the other three (statistical, prescriptive, and hybrid). For example, "objective standards" might be the object of a statistical theory ("objective customs"), a prescriptive one ("objective values"), or a hybrid one ("objective norms").

The first column represents a different theoretical axis. It distinguishes between three types of

66. *See supra* Subpart I.B & Section II.A.1.

conventionalist views (traditional, modern, and mixed), nonconventionalist views, and views that are ambiguous among different conventionalist and nonconventionalist possibilities. Because conventionalism is not this chapter's focus, these possibilities are not developed in greater detail, but the fundamental distinctions should be clear. The objects of traditional conventionalism are traditional standards (customs, values, or norms). The objects of modern conventionalism are modern standards. The objects of mixed conventionalism are both traditional and modern standards. And the objects of nonconventionalism are objective standards.

Table 1. Proposed Vocabulary for Reasonableness Theories

	<i>Statistical</i>	<i>Prescriptive</i>	<i>Hybrid</i>	<i>Ambiguous</i>
Traditional	Traditional	Traditional	Traditional	Traditional
Conventionalist	Community	Community	Community	Community
Modern Conventionalist	Modern	Modern	Modern	Modern
Mixed Conventionalist	Community	Community	Community	Community
Nonconventionalist	Mixed	Mixed	Mixed	Mixed
	Community	Community	Community	Community
	Objective	Objective	Objective	Objective
	Customs	Values	Norms	Standards
Ambiguous	Customs	Values	Norms	Standards

The proposed vocabulary tracks distinctions that are sometimes overlooked or blurred. For example, some scholars assume that conventionalism about reasonableness implies a statistical view.⁶⁷ However, that is an invalid inference. The Table makes clear that conventionalism does not necessarily imply a statistical view, or even a hybrid one. Recall Tilley's conventionalist view of tort law, an example of a conventionalist and prescriptivist view. She argues that "tort doctrine's reliance on community as the source of norms . . . encourages decision makers to toggle

67. This interpretive mistake may be responsible for the common characterization of Holmes as a statistical theorist and not a hybrid one. Emphasis on ordinary community standards is consistent with statistical, prescriptive, and hybrid theories. *See infra* Subpart IV.A.

between traditional and modern values—between morality and efficiency.”⁶⁸ In the proposed terminology, this view is that of a “mixed conventionalist prescriptivist,” one that theorizes reasonableness with concern for both traditional and modern community values.

The proposed stipulative terminology is not necessarily consistent with all prior use in the reasonableness literature or ordinary language. For example, in the previous paragraph’s quotation, Tilley’s language of “norms” is somewhat ambiguous. In ordinary language, “norms” might refer to statistical practices, prescriptive ideals, or hybrid judgments of normality. In the proposed terminology, I use “norms” as the object of hybrid views. This is a stipulative choice of vocabulary, and some other theorists may wish to redefine “norm” to mean a purely statistical or prescriptive notion. The vocabulary’s usefulness is that it distinguishes among notions that are purely statistical (“customs”), purely prescriptive (“values”), and hybrid (“norms”).

I use “standards” as an ambiguous generalist phrase, one that might refer to customs, values, or norms. This approach can help distinguish different debates. For example, scholars might debate the merits of conventionalist and nonconventionalist views, while sidestepping debate about statistical, prescriptive, and hybrid views. Such a debate would concern whether reasonableness is a judgment of objective standards or community standards.

Similarly, I use “customs,” “values,” and “norms” as ambiguous general phrases to refer to (respectively) statistical, prescriptive, and hybrid views. These views might be conventionalist or nonconventionalist. For example, this chapter debates the merits of statistical, prescriptive, and hybrid views, while sidestepping the debate about conventionalism and nonconventionalism. Thus, the present debate can be described as one about customs versus values versus norms.

68. Tilley, *supra* note 3, at 1325.

A final aspect of these distinctions that is worth further explanation is the difference between objective- or nonconventionalist-statistical views and conventionalist ones. It may not be immediately clear how a statistical interpretation of reasonableness can be nonobjective. The key is that people's judgments of statistically typical behavior need not correspond to facts about typical behavior.⁶⁹ Statistical theories have a very important choice about whether to ground reasonableness in people's judgments of common practice or in facts about common practice.

This observation makes clear that the "hybrid" column actually contains more possible views than first meet the eye. Because hybrid views account for both statistical and prescriptive considerations, a hybrid view must make choices about two relevant sets of factors. The focal aspects of the first three hybrid views—traditional community norms, modern community norms, and mixed community norms—are all understood in a purely nonobjective, conventionalist manner. That is, both the statistical and prescriptive components are grounded in community understandings. On a modern conventionalist hybrid view, the focus is modern community norms—determinations that reflect both statistical and prescriptive judgments.

However, on the nonconventionalist (or objective) hybrid view, the focus is objective norms. The most straightforward way to understand this box is as one representing theories that consider both an objective prescriptive factor (e.g., justification) and an objective statistical one (e.g., actual facts about ordinary practice). However, this box also contains views that are only partly objective. For example, a hybrid view of reasonableness might consider an objective prescriptive factor (e.g., moral rightness) and a conventionalist statistical one (e.g., judgments—not facts—about ordinary practice).

This Subpart has outlined significant distinctions between several views of reasonableness and proposed a stipulative terminology to track these distinctions. These distinctions also clarify

69. For thoughtful elaboration of this point, see Roseanna Sommers, *A Psychological Critique of the Reasonable Person Standard 1–8* (2018) (unpublished manuscript) (on file with author).

the chapter's primary focus: the debate about theorizing reasonableness as (purely statistical) customs, (purely prescriptive) values, or (hybrid) norms.

III. EXPERIMENTAL STUDY OF ORDINARY REASONABLENESS JUDGMENT

Debates about legal reasonableness operate on two levels. One level addresses a normative question: how *should* reasonableness be theorized and applied? Is it right for jurors to consider averages, ideals, or both when they make reasonableness decisions? Another level concerns an empirical question about human cognition: how do ordinary people (i.e., potential jurors) *actually* make reasonableness judgments? This Part addresses this second level, investigating how ordinary people actually evaluate reasonableness.

There are several motivations for this approach. For one, reasonableness is a widely used legal standard, and the nature of reasonableness judgments presents a significant legal question. Second, the approach has practical value. Reasonableness is often determined by a jury judgment. Understanding the ordinary mechanism of layperson reasonableness judgments illuminates jury decision making and the actual application of reasonableness standards. Finally, investigation of these ordinary judgments enriches debates in legal theory. Insofar as some of those theories of reasonableness make empirical predictions, experimental study can support or challenge those theories.

This Part conducts the first experimental investigation of ordinary judgments of reasonableness. Three experiments investigate ordinary people's reasonableness and legal reasonableness judgments. To test the predictions, I draw on a paradigm from recent cognitive science research on *normality*, which finds that judgments of normality are best predicted by a complex combina-

tion of judgments of the (statistical) average and the (prescriptive) ideal.⁷⁰ For example, the mean judgment of the normal number of lies told per week is intermediate between mean judgments of the average and ideal.⁷¹ Moreover, those judgments are best predicted by the same complex combination of judgments of a (statistical) average and (prescriptive) ideal, rather than by judgments of either alone.

Although the hybrid view of reasonableness is a less-developed view in legal scholarship, this recent work on the cognitive science of normality judgments involves just this kind of statistical and prescriptive blend. This opens up an exciting possibility to empirically test theories of reasonableness. The results indicate that ordinary reasonableness judgments are neither purely statistical nor prescriptive, but are instead better understood as a hybrid notion.

The experiments test whether reasonableness is a *hybrid* judgment, one that is partly statistical and partly prescriptive. They test the prediction that ordinary reasonableness judgments are systematically intermediate between the relevant average and ideal. Moreover, they test the hypothesis that reasonableness is better predicted by judgments of the relevant average *and* ideal, rather than by either alone.

A. *The Cognitive Science of Hybrid Concepts*

My hypothesis is that ordinary judgments of reasonableness reflect a hybrid judgment, one that is partly statistical and partly prescriptive. To test this hypothesis, I draw on recent work in cognitive science about normality judgments. Adam Bear and Joshua Knobe find that exactly this hybrid feature characterizes judgments of “the normal.”⁷² They find that, across many varied categories, people’s representation of what is normal falls between the statistical representation of

70. Bear & Knobe, *supra* note 60, at 25–26.

71. *Id.* at 28.

72. *Id.* at 25–26.

what is average and the prescriptive representation of what is ideal.⁷³ More generally, Bear and Knobe find that normality judgments are better explained by a model of both average and ideal judgments, compared to a model of only one judgment.⁷⁴

This research provides a useful paradigm within which to test the three views about reasonableness. An experiment can compare participants' mean judgments of reasonable quantities to their mean judgments of average and ideal quantities. To minimize researcher degrees of freedom, the first experiment uses the exact items from Bear and Knobe.⁷⁵

The different theories of reasonableness make different experimental predictions. Statistical views hypothesize that average judgments will best predict reasonableness judgments. Prescriptive views hypothesize that ideal judgments will best predict reasonableness judgments. And hybrid views predict that a more complex model of both judged averages and ideals will best predict reasonableness judgments.

Furthermore, statistical views predict that reasonableness judgments are randomly distributed around the relevant average. Where average and ideal judgments diverge, statistical views predict that some reasonableness judgments will be on the "ideal side" of the average, and others will be on the "nonideal side" of the average. Prescriptive views predict that reasonableness judgments are randomly distributed around the relevant ideal. For prescriptivists, when average and ideal judgments diverge, some reasonableness judgments will be on the "average side" of ideal, and others will be on the "nonaverage side" of ideal. However, hybrid views predict a very different

73. *Id.*

74. *Id.*

75. To minimize researcher degrees of freedom is to minimize the number of design choices a researcher makes that might support the desired hypothesis. For example, rather than choosing my own examples, I use the exact items developed by Bear and Knobe—who had no knowledge of the present hypothesis. *See id.* at 28.

and very specific pattern. Hybrid views predict that reasonableness judgments are *intermediate* between the relevant average and ideal. In other words, where the judged average and ideal diverge, hybrid views predict that reasonableness judgments should fall both on the ideal side of the relevant average and on the average side of the relevant ideal.

B. Experimental Studies: Reasonableness as a Hybrid Concept

Three studies examine ordinary judgments of reasonableness. Because the target population is ordinary people (i.e., potential jurors, not legal experts), participants were recruited from Amazon's Mechanical Turk, an online research platform that is representative and reliable for cognitive science research.⁷⁶

Experiment 1 compares mean judgments of twenty reasonable quantities (e.g., the reasonable number of calories to consume each day and the reasonable amount to cheat on one's taxes) to mean judgments of the relevant average and ideal. Experiment 2 invites participants to make the same judgments about reasonableness while in a legal context. Experiment 3 examines judgments about thirteen legally relevant quantities inspired by real legal-reasonableness standards (e.g., the reasonable number of days to accept a contract and the reasonable interest rate).

All data analyses, including participant exclusion criteria, follow the exact tests used by Bear and Knobe. As the experiments follow the exact methods and statistics, I report the exact list of statistical tests that could support or weaken the case for the hypothesis that reasonableness is judged as a hybrid notion. In each experiment, participants report their judgments of either average, ideal, reasonable, or legally reasonable quantities.

76. Mechanical Turk (MTurk) is an online platform that enables researchers to collect large samples from a population that is more representative than many other typical research samples. See generally Adam J. Berinsky et al., *Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk*, 20 *POL. ANALYSIS* 351 (2012); Gabriele Paolacci et al., *Running Experiments on Amazon Mechanical Turk*, 5 *JUDGMENT & DECISION MAKING* 411 (2010). The service is understood to provide high-quality data. See Michael Buhmester et al., *Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality Data?*, 6 *PERSP. ON PSYCHOL. SCI.* 3, 4–5 (2011).

In each case, I will run the following statistical tests and make the following predictions. First, I will regress participants' mean reasonableness judgments on participants' mean average and mean ideal judgments (complex model), predicting that both average and ideal will significantly predict reasonableness. Next I will regress participants' mean reasonableness judgments on participants' mean average judgments (average model), predicting that the complex model will explain more variance than the average model. Next I will regress participants' mean reasonableness judgments on participants' mean ideal judgments (ideal model), predicting that the complex model will explain more variance than the ideal model.

Additionally, I will compute the Akaike Information Criterion with finite-sample correction (AIC_c) for the complex, average, and ideal models and compute the evidence ratio comparing the models. I predict that the AIC_c will be lower for the complex model than for the average or ideal models, and the evidence ratio will support the complex model.

Lastly, I will compare to chance (.5) the proportion of mean reasonable responses that are on the average side of mean ideal responses and the proportion of mean reasonable responses that are on the ideal side of mean average responses. I predict that the proportion of mean reasonable responses on the average side of ideal will be greater than chance (.5) and the proportion of mean reasonable responses on the ideal side of average will be greater than chance (.5). Finally, I will compare to chance (.33) the proportion of mean reasonable responses that are intermediate between mean average and mean ideal responses, predicting that the proportion of reasonable responses that are intermediate between mean average and mean ideal response will be greater than chance (.33).

All three studies provide strong evidence that reasonableness is best predicted by judgments

of *both* the average and ideal, and that reasonableness judgments are intermediate between judgments of the relevant average and ideal. This result provides evidence for the hybrid view of reasonableness, that reasonableness is partly statistical and partly prescriptive. Moreover, since this represents a similar pattern to judgments of normality, the studies also suggest that ordinary judgments of reasonableness are importantly similar to ordinary judgments of normality.

1. Experiment 1

a. Method

Forty-eight participants were recruited from Amazon's Mechanical Turk.⁷⁷ Each participant rated the reasonable quantity of twenty items. For example, participants are asked, "What do you think is a reasonable number of calories that a person consumes in a day?" and "What do you think is the reasonable number of books that a person reads in a year?" To minimize researcher degrees of freedom, the items were taken directly from Bear and Knobe.⁷⁸ Items were presented in a random order. The full text of the questions can be found in Appendix A.

b. Results

The mean ratings for each item are displayed in Table 2 of Appendix B. Following Bear and Knobe, individual participant responses that were three standard deviations away from the mean answer for a given question were excluded.⁷⁹ These results were compared to the average and ideal ratings for each item reported by Bear and Knobe, and I analyzed the data following their exact statistical methodology.⁸⁰ Tests of the eight a priori hypotheses were conducted using Bonferroni adjusted alpha levels of .006 per test.

77. Of this group, 65% were male, 35% were female, and 0% were nonbinary. The mean age was 35.6.

78. See Bear & Knobe, *supra* note 60, at 28.

79. See *id.* at 27.

80. See *id.* at 27–28.

First, I examined whether reasonableness judgments are predicted from average and ideal judgments. Because the questions asked about varied quantities (e.g., minutes vs. calories), mean responses for each measure were converted to a logarithmic scale.

Reasonableness judgments were regressed on judged averages and judged ideals.⁸¹ Both judged averages⁸² and judged ideals⁸³ significantly predicted reasonableness judgments.

Next, I compared this complex model to two simpler regression models, one in which only the average judgment predicted reasonableness and one in which only the ideal judgment predicted reasonableness. The more complex model explains more variance than the model in which only average judgments predict reasonable judgments⁸⁴ and the model in which only ideal judgments predict reasonable judgments.⁸⁵

Moreover, in addition to the complex model explaining more variance, the Akaike Information Criterion with finite-sample correction (AIC_c) for the complex model⁸⁶ was lower than that for either the model in which only judged averages predict reasonableness judgments⁸⁷ or the model in which only judged ideals predict reasonableness judgments.⁸⁸ This suggests that the more complex model is more appropriate. Quantifying the strength of evidence in favor of the more complex model, by calculating an evidence ratio based on Akaike weights,⁸⁹ indicated a result of over 1,000 for the more complex model compared to the average-only model and a result

81. $F(2, 17) = 127.71, r^2 = .98, p < 0.001$.

82. $\beta = .572, SE = .084, p < 0.001$.

83. $\beta = .480, SE = .073, p < 0.001$.

84. $F(1, 18) = 127.71, r^2 = .88, p < 0.001$.

85. $F(1, 18) = 119.23, r^2 = .87, p < 0.001$.

86. -40.16.

87. -17.46.

88. -16.26.

89. See generally Eric-Jan Wagenmakers & Simon Farrell, *AIC Model Selection Using Akaike Weights*, 11 *PSYCHONOMIC BULL. & REV.* 192 (2004).

of over 1,000 for the more complex model compared to the ideal-only model.⁹⁰ These results very strongly support the more complex model.⁹¹

I also compared the degree to which reasonableness judgments were intermediate between average and ideal ones. I compared both (i) the proportion of the reasonable responses that were on the average side of ideal and (ii) the proportion of the reasonable responses that were on the ideal side of average, to chance (.5). Nineteen out of twenty, 95%, of the items had reasonableness judgments that were on the average side of the ideal,⁹² while eighteen out of twenty, 90%, had reasonableness judgments that were on the ideal side of the average.⁹³ Seventeen out of twenty items, 85%, had reasonableness judgments that were intermediate between average and ideal ratings.⁹⁴

These statistical analyses, individually and collectively, indicate that the results provide evidence that judgment of reasonableness is a hybrid judgment.

2. Experiment 2

Experiment 2 tested whether the same effect arises in judgments of reasonableness when participants are invited to make their judgments in a legal context.

a. Method

Fifty-nine participants were recruited from Amazon's Mechanical Turk.⁹⁵ Participants com-

90. This is an expression of "the evidence ratio as the normalized probability" that the more complex model is preferred over the simpler model, reflecting "an intuitive feeling for how much support [the] evidence ratio provides." *Id.* at 194. In other words, it is a normalized ratio of the evidence weights for each model, reflecting a comparison of the relative strength of the more complex model over the simpler one.

91. *See id.*; *see also* Bear & Knobe, *supra* note 60, at 28 (citing a ratio of 269 as a decisive result).

92. Binomial $p < 0.001$ (compared to null hypothesis of .5).

93. Binomial $p < 0.001$ (compared to null hypothesis of .5).

94. Binomial $p < 0.001$ (compared to null hypothesis of .33). For this binomial test, I compare the rate of intermediacy to a null hypothesis rate of .33. The null hypothesis is that the mean reasonableness judgment is equally likely to fall on the ideal side of average, the average side of ideal, or between average and ideal.

95. Of this group, 40% were male, 60% were female, and 0% nonbinary. The mean age was 38.0.

pleted the same task as in Experiment 1, but they were instructed to consider these items in a legal frame:

In the following screen we ask you to judge the legally reasonable quantity of a number of different things. We ask you to imagine that you are making these judgments in a legal setting for a legal purpose. For example, imagine that you are a jury member in a jury deliberation.

Jurors are often asked to make legal judgments by comparing someone's actual behavior to a hypothetical reasonable one. For example, imagine Mike was painting the outside of his house and left the can of lead-based paint open by his garage for some amount of time. During that time, the neighbor's dog ate some of the paint and was injured. To determine whether Mike is legally liable for the injury to the dog, jurors might be asked to compare Mike's actual behavior to "reasonable" behavior in similar circumstances. If Mike acted with the reasonable amount of care (or more), he is not liable for the dog's injury; if Mike acted with less care than the reasonable amount, he is liable for the dog's injury.

These examples can vary very widely. For example, a contract might specify that employees are entitled to a "reasonable" number of sick days per year. Settling a contract dispute between the employer and employee would involve comparing the number of sick days that the employee actually took to the reasonable number of sick days.

In the next screen we ask you to estimate the reasonable quantity of different things. For some of these things, it will seem very clear how the question of its reasonableness might arise in a legal setting; for others, it will be less obvious. We ask that in all cases, you keep in mind the legal context.

The full text of the questions can be found in Appendix A.

b. Results

The mean ratings for each item are displayed in Table 2 of Appendix B. Again, mean responses for each measure were converted to a logarithmic scale. Again, tests of the eight a priori hypotheses were conducted using Bonferonni adjusted alpha levels of .006 per test.

First, I examined whether legal reasonableness judgments are predicted from average and ideal judgments. Reasonableness judgments were regressed on average and ideal judgments.⁹⁶ The model revealed that both judged averages⁹⁷ and judged ideals⁹⁸ significantly predicted reasonableness judgments.

This more complex model explains more variance than both a model in which only average judgments predict legally reasonable judgments⁹⁹ and a model in which only ideal judgments predict legally reasonable judgments.¹⁰⁰

Moreover, in addition to explaining more variance, the AIC_c for the complex model¹⁰¹ was lower than that for both a model in which only judged averages predict reasonableness judgments¹⁰² and a model in which only judged ideals predict reasonableness judgments,¹⁰³ suggesting that the more complex model is the most appropriate one. Quantifying the strength of evidence in favor of the more complex model, by calculating an evidence ratio based on Akaike weights, indicated a result of over 1,000 for the more complex model compared to the average-only model and a result of over 1,000 for the more complex model compared to the ideal-only model. These results very strongly support the more complex model.

Again, I compared the degree to which reasonableness judgments were intermediate between

96. $F(2, 17) = 339.12, r^2 = .98, p < 0.001$.

97. $\beta = .489, SE = .065, p < 0.001$.

98. $\beta = .493, SE = .057, p < 0.001$.

99. $F(1, 18) = 119.07, r^2 = .87, p < 0.001$.

100. $F(1, 18) = 153.51, r^2 = .90, p < 0.001$.

101. -50.09.

102. -18.95.

103. -23.43.

average and ideal ones. Nineteen out of twenty, 95%, of the items had reasonableness judgments that were on the average side of the ideal,¹⁰⁴ and nineteen out of twenty, 95%, had reasonableness judgments that were on the ideal side of the average.¹⁰⁵ Eighteen out of twenty items, 90%, had reasonableness judgments that were intermediate.¹⁰⁶

These statistical analyses, individually and collectively, indicate that the results provide evidence that judgment of reasonableness is a hybrid judgment.

3. Experiment 3

The previous experiments suggest that reasonableness is a partly statistical and partly prescriptive hybrid concept. Experiments 1 and 2 used the exact twenty items used by Bear and Knobe to minimize researcher degrees of freedom.¹⁰⁷ However, some of those items represent more plausible legal examples (e.g., lies told per week), while other are less-legally-pertinent examples (e.g., servings of vegetables per month). The third Experiment tests reasonableness judgments using more typical legal examples.

a. Method

Two hundred seventeen participants were recruited from Amazon's Mechanical Turk.¹⁰⁸ Following Bear and Knobe, I selected examples in which I expected a significant difference between average and ideal quantity judgments. I aimed to include a representative sample of reasonable-

104. Binomial $p < 0.001$.

105. Binomial $p < 0.001$.

106. Binomial $p < 0.001$.

107. See Bear & Knobe, *supra* note 60, at 28.

108. Of this group, 50% were male, 50% were female, and 0% were nonbinary. The mean age was 36.8.

ness items from various legal domains and of varied specificity. For example, participants were asked what is the reasonable “number of days taken to accept a business contract when no deadline is specified” and what is the reasonable “number of loud events held at a football field close to a quiet neighborhood, per year.” Participants answered one of four sets of questions about the same thirteen items. Three groups were instructed to estimate the average, ideal, or reasonable quantity of each item. A final treatment (the “legally reasonable” condition) received the same “legal” contextual information presented in Experiment 2 and responded to questions about reasonableness. The full text of the questions can be found in Appendix A.

b. Results

The mean ratings for each item are displayed in Table 2 of Appendix B. Tests of the sixteen a priori hypotheses were conducted using Bonferonni adjusted alpha levels of .003 per test.

First, consider the results for the group estimating the reasonable quantity of each item (without additional legal context). I examined whether legal reasonableness judgments are predicted from average and ideal judgments. Again, mean responses for each measure were converted to a logarithmic scale. Reasonableness judgments were regressed on both average and ideal judgments.¹⁰⁹ The model revealed that judged ideals¹¹⁰ significantly predicted reasonableness judgments. The model indicated that judged averages did not predict reasonableness judgments at the level of statistical significance.¹¹¹

This complex model explained more variance than both a model in which only average judgments predict reasonable judgments¹¹² and a model in which only ideal judgments predict

109. $F(2, 12) = 449.60, r^2 = .99, p < 0.001$.

110. $\beta = .806, SE = .080, p < 0.001$.

111. $\beta = .177, SE = .078, p = 0.047$.

112. $F(1, 11) = 78.40, r^2 = .88, p < 0.001$.

reasonable judgments.¹¹³

The AIC_c for the complex model¹¹⁴ was lower than that for a model in which only judged averages predict reasonableness judgments¹¹⁵ and a model in which only judged ideals predict reasonableness judgments,¹¹⁶ suggesting that the more complex model is the most appropriate one. Quantifying the strength of evidence in favor of the more complex model, by calculating an evidence ratio based on Akaike weights, indicated a result of over 1,000 for the more complex model compared to the average-only model and a result of 2.22 for the more complex model compared to the ideal-only model. These support the more complex model.

Again, I compared the degree to which reasonableness judgments were intermediate between average and ideal ones. All thirteen, 100%, of the items had reasonableness judgments that were on the average side of the ideal,¹¹⁷ and eleven out of thirteen, 85%, had reasonableness judgments that were on the ideal side of the average.¹¹⁸ Eleven out of thirteen items, 85%, had reasonableness judgments that were intermediate.¹¹⁹

Second, consider the results for the group estimating the legally reasonable quantity of each item (with additional legal context). I ran a regression model in which both average and ideal judgments predict legally reasonable judgments.¹²⁰ Both judged averages¹²¹ and judged ideals¹²² significantly predicted reasonableness judgments.

113. $F(1, 11) = 648.85, r^2 = .98, p < 0.001$.

114. -41.37.

115. -12.78.

116. -38.80.

117. Binomial $p < 0.001$.

118. Binomial $p = 0.011$.

119. Binomial $p < 0.001$.

120. $F(2, 10) = 590.23, r^2 = .99, p < 0.001$.

121. $\beta = .364, SE = .070, p < 0.001$.

122. $\beta = .650, SE = .072, p < 0.001$.

This complex model explained more variance than both a model in which only average judgments predict reasonable judgments¹²³ and a model in which only ideal judgments predict reasonable judgments.¹²⁴

The AIC_c for the complex model¹²⁵ was lower than that for a model in which only judged averages predict reasonableness judgments¹²⁶ and a model in which only ideal average predict reasonableness judgments,¹²⁷ suggesting that the complex model is the most appropriate.

Quantifying the strength of evidence in favor of the more complex model, by calculating an evidence ratio based on Akaike weights, indicated a result of over 1,000 for the more complex model compared to the average-only model and a result of over 1,000 for the more complex model compared to the ideal-only model. These very strongly support the more complex model.

Twelve out of thirteen, 92%, of the items had reasonableness judgments that were on the average side of the ideal,¹²⁸ and twelve out of thirteen, 92%, had reasonableness judgments that were on the ideal side of the average.¹²⁹ Eleven items, 85%, had intermediate reasonableness judgments.¹³⁰ These statistical analyses, individually and collectively, indicate that the results provide evidence that judgment of reasonableness is a hybrid judgment.

123. $F(1, 11) = 131.24, r^2 = .92, p < 0.001$.

124. $F(1, 11) = 341.40, r^2 = .97, p < 0.001$.

125. -44.70.

126. -18.14.

127. -29.92.

128. Binomial $p = 0.002$.

129. Binomial $p = 0.002$.

130. Binomial $p < 0.001$.

Figure 1. Representation of Intermediacy Results for Reasonableness

(Bars depict the log (base 10) of mean judgments for average, ideal, and reasonable quantities in Experiments 1 and 3.)

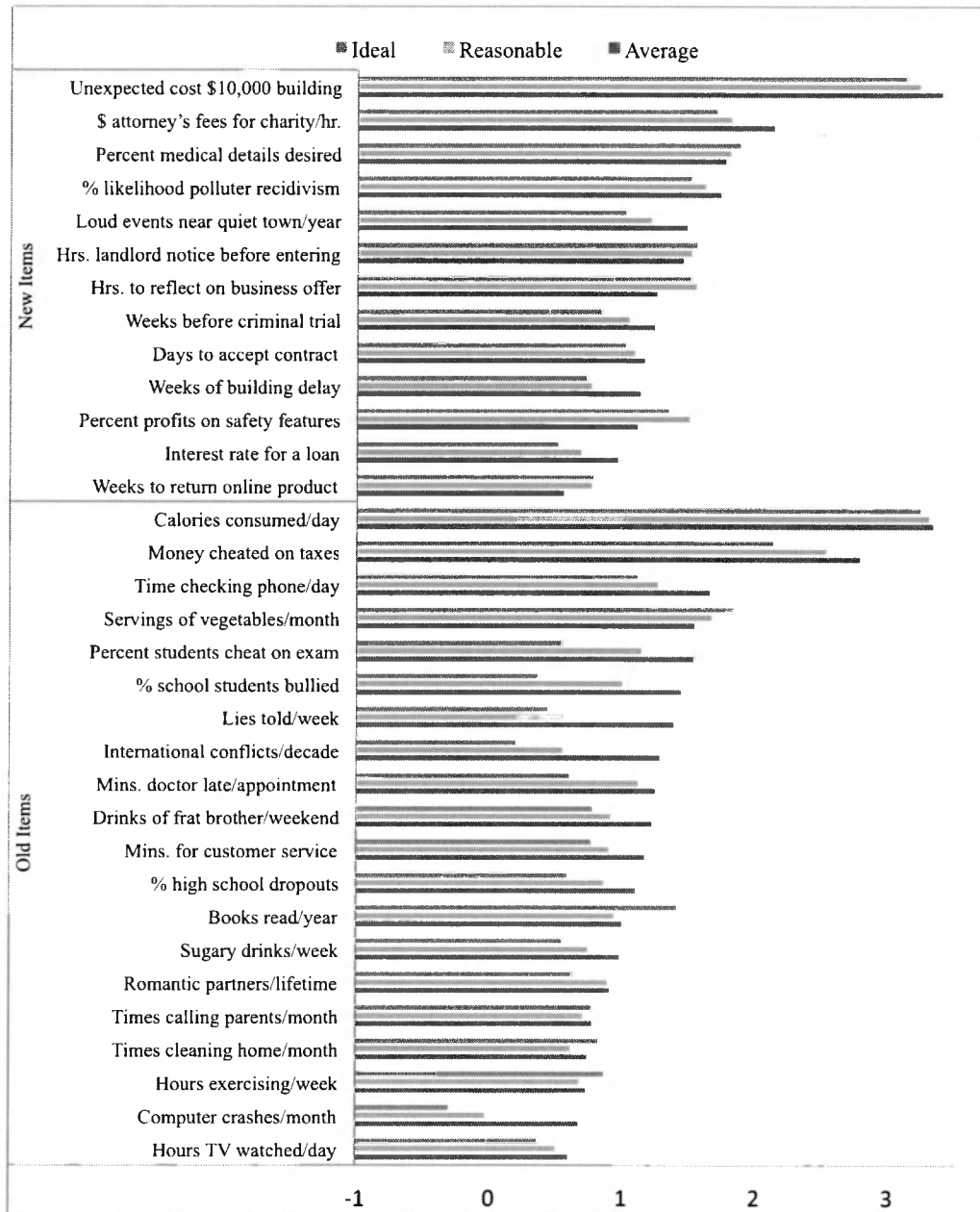
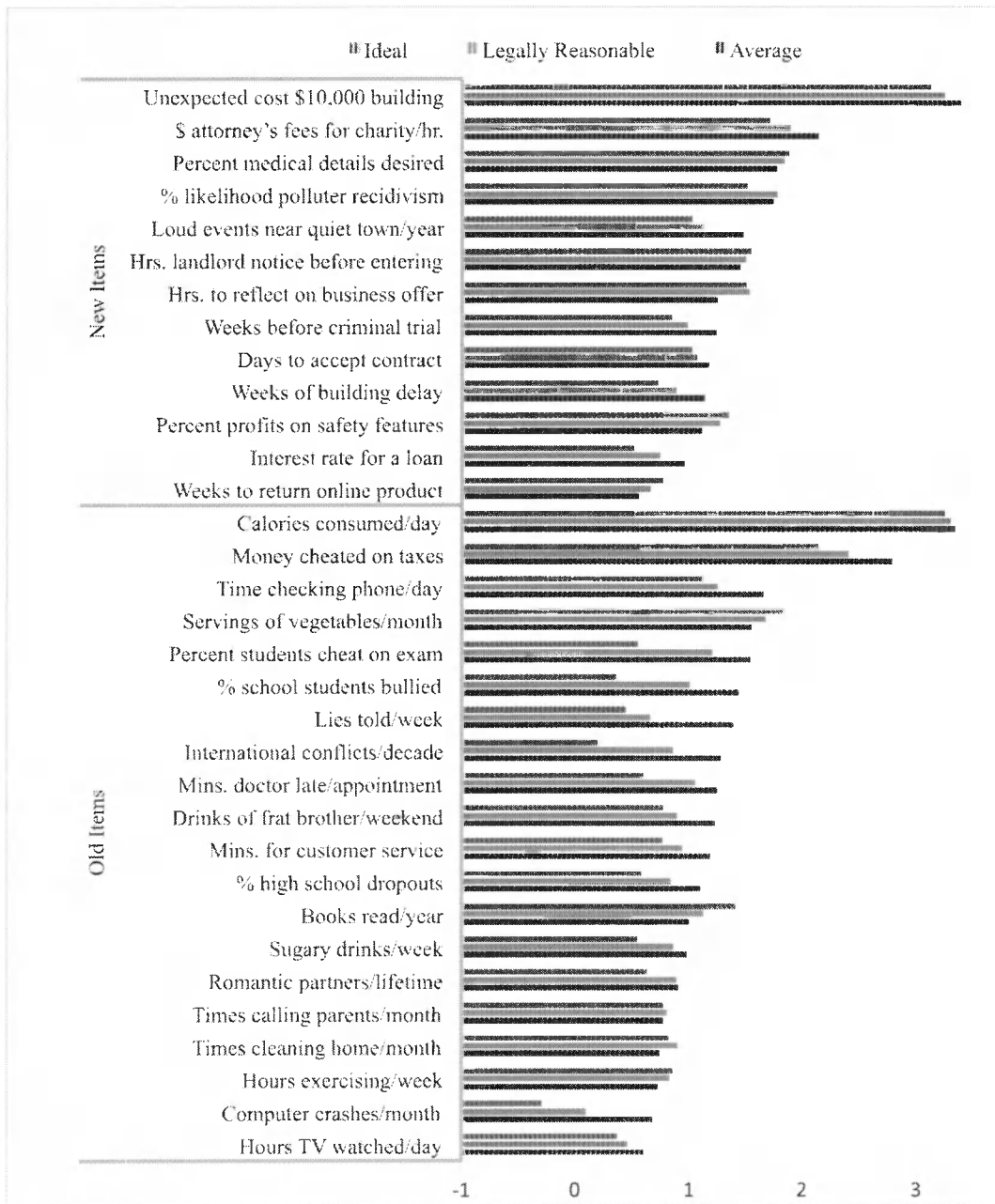


Figure 2. Representation of Intermediacy Results for Legal Reasonableness

(Bars depict the log (base 10) of mean judgments for average, ideal, and legally reasonable quantities in Experiments 2 and 3.)



4. *Summary*

In the three experimental studies in which ordinary people make judgments about reasonableness and legal reasonableness, the results show that a combination of both statistical and prescriptive judgments best predicts reasonableness judgments. Moreover, across various domains, a striking pattern emerged: the reasonable quantity was intermediate between divergent average and ideal quantities. These findings provide strong evidence that reasonableness is a hybrid judgment, reflecting both statistical and prescriptive considerations.

IV. A DEFENSE OF HYBRID THEORIES OF REASONABLENESS

While the previous Part revealed how reasonableness is understood, this Part returns to the normative question of how reasonableness should be theorized. This Part presents arguments in favor of theorizing legal reasonableness as a hybrid standard, rather than as a purely statistical or purely prescriptive one.

Subpart IV.A begins by outlining several different varieties of hybrid theories. Each of these is consistent with Part III's data, and each provides a different theoretical basis for theorizing reasonableness as a hybrid standard. Future empirical and theoretical work might help distinguish among these more specific hybrid views, but the remainder of this Part (and this chapter) does not endorse one of these narrower hybrid variations. Instead, I defend the broader class of hybrid theories that encompasses each of the three variations.

Subpart IV.B presents a historical argument. It begins by considering Britain's man on the Clapham omnibus, the historical predecessor of the reasonable person. Both terms were introduced to capture statistical and prescriptive properties, informing reasonableness analysis with a

hybrid of statistical and prescriptive considerations. Insofar as this concept reflects the original and traditional meaning of reasonableness, this provides a reason to theorize reasonableness in the same way today.

Subpart IV.C focuses on the modern context. It returns to the experimental findings of Part III and recasts them to support a normative argument. Various legal theories—making diverse assumptions—posit that legal reasonableness should reflect either the ordinary meaning or ordinary use of *reasonableness*. On these views, an experimental finding about how reasonableness is applied actually provides a reason for how reasonableness *ought* to be applied.

Subpart IV.D presents a third line of arguments for reasonableness as a hybrid standard. Theorizing reasonableness as a hybrid notion is the best contender for a conception of reasonableness that avoids absurdities and achieves appropriate instantiations of reasonableness judgments across multiple legal domains. A hybrid view offers the best general theory of reasonableness.

A. *Hybrid Varieties*

This Subpart begins the normative defense of hybrid theories of reasonableness by distinguishing among three plausible variations on a hybrid theory. Although I do not endorse one of these variations, this Subpart's work demonstrates the breadth of hybrid views and also illuminates the diverse possible theoretical underpinnings of those views.

I consider three variations. The first is a “corrected ideal” interpretation. Very broadly speaking, on that account, reasonableness is determined by considering the relevant ideal and adjusting it in line with statistical norms. The second view is a “corrected average” view. Again, broadly speaking, on this view reasonableness is determined by considering the relevant statistical norms and adjusting those to take account of some prescriptive considerations. The final view is the “blended standard” view. On that account, reasonableness is best understood as a standard that treats statistical and prescriptive facts as two relevant signals of reasonableness. For example, one

version of that third view might theorize reasonableness as normality.

Before detailing each of these variations, it is worth noting that there may be some plausible nonhybrid interpretations of Part III's data that give rise to nonhybrid theories. For example, perhaps ordinary judgments of reasonableness are driven by a process in which people take the average of plausible ideals. In other words, if there is disagreement about the ideal X, those who judge the reasonable X might consider only ideals but simply take the average of those ideals.

To clarify this "average of ideals" interpretation, consider a simple example. Imagine people disagree about the ideal time that a product should be refundable. One-third of them think it should be four weeks, a third think five weeks, and a third six weeks. On the average-of-ideals interpretation, someone judging the reasonable number of time to return a product would take the average of ideals (five weeks).

Importantly, this is not a hybrid view, but it may be a defensible one. A corresponding normative theory of reasonableness might be motivated by pluralism, the desire to accommodate diverse conceptions of ideal behavior.

However, this account actually would not make sense of the data. If people produced a reasonableness judgment by considering the average of ideals, we would expect reasonableness judgments to fall on either side of the ideal at rates of chance. However, the data show that reasonableness judgments actually fall on the average side of ideal at rates above chance. Although this outcome does not definitely refute the average-of-ideals hypothesis, the data is more consistent with various hybrid accounts.

The average-of-ideals interpretation is only one possible nonhybrid interpretation. I cannot consider every possible competing interpretation of the data. Thus, the chapter's data and conclu-

sions are open to further interpretation and debate. However, the burden to articulate plausible nonhybrid accounts rests with defenders of those interpretations. The pattern of data favors hybrid views over the many plausible nonhybrid ones (e.g., an average-of-ideals interpretation).

1. A Corrected Ideal

The first variation on a hybrid account of reasonableness is a corrected-ideal interpretation. On this account, the formula for determining reasonableness modifies a prescriptive standard (e.g., the ideal, the welfare-maximizing, or the virtuous) by adjusting it in line with statistical considerations (e.g., statistical norms).

The motivation for this account is not hard to see. In various domains—such as criminal law, tort negligence, consumer protection, and privacy—reasonableness seems to capture a *moderate* standard. The reasonable person does not have to do everything that an ideal person would do, but instead must adhere to a (somewhat) less restrictive standard.

As one example of this view, consider Benjamin Zipursky's account of reasonableness in negligence law:

The [reasonable care] standard is not whether the defendant did all she could have done, or what would have been best to do, or what every last precaution would have been. . . . Due care in the normal case, as conventionally understood and as still represented in jury instructions, lies in between “the highest degree of diligence and care” and “gross negligence” Our system elaborates on this sort of ordinariness by saying that the jury is to determine whether the defendant acted as “a reasonably careful person” or “a reasonably prudent person” would have acted.”¹³¹

This view of reasonableness is *probably* a hybrid view, one that recommends setting the standard of reasonable care by adjusting the ideal degree of care to account for facts about actual behavior. I say this is probably a hybrid since there are other ways in which one might adjust an ideal standard. For instance, perhaps reasonableness is whatever standard is 20% less stringent than the

131. Zipursky, *supra* note 3, at 2154.

ideal standard. This kind of “arbitrarily moderating” view would *not* be a hybrid view since its adjustment to the ideal standard is not grounded in statistical considerations.

What makes the corrected-ideal view a variation on a hybrid view is that the adjustment of the ideal is grounded in statistical considerations. Insofar as Zipursky’s view recommends moderating the ideal standard of reasonable care to account for statistical considerations, it is a hybrid view.

2. A Corrected Average

A second variation is a corrected-average view. This is structurally analogous to the corrected-ideal view. On this account, the formula for determining reasonableness modifies a statistical standard (e.g., averageness) by adjusting it in line with prescriptive considerations (e.g., virtuousness or welfare maximization).

There are a number of plausible motivations for this hybrid variation. For example, perhaps reasonableness is largely meant to capture traditional community customs (i.e., statistical practices), insofar as those customs are consistent with the basic tenants of community values (i.e., prescriptive norms). A corrected-average variation of reasonableness might capture this aim. Reasonableness is fundamentally defined by statistical facts, but the standard is adjusted in cases in which those statistical facts represent significant norm violations.

3. A Blended Standard

A third variation defines reasonableness as a hybrid standard *because* both statistical and prescriptive factors are highly significant signals of reasonableness. Compare this to the views out-

lined in the two previous sections. One way to understand those is that they set reasonableness as a hybrid standard because of problems with defining reasonableness in terms of only one type of factor. For example, a “corrected ideal” theory might be motivated by arguments that reasonableness is fundamentally a prescriptive notion, but that theory adjusts that standard because sometimes the ideal is too difficult to attain. Similarly, a “corrected average” theory might be motivated by arguments that reasonableness is fundamentally about statistical norms, but that theory adjusts that standard because sometimes the average seems an inappropriate standard. This third variation is very different. On this view, reasonableness is a hybrid standard because statistical and prescriptive factors are relevant to reasonableness in every case (not just exceptional ones).

Among the distinctive considerations in support of this account is the fact that it allows reasonableness to leverage more information. Both statistical and prescriptive considerations form the signal that is relevant to reasonableness.

The experiments of Part III focused on examples in which judgment of the average and ideal diverge. But in many—and perhaps most—other examples, the average and ideal would not diverge so significantly. On this third conception of a hybrid view, there are two signals of reasonableness: statistical facts and prescriptive facts. In cases in which there is little significant divergence between statistical and prescriptive considerations, judgment of reasonableness has two relevant sources of information.

Despite differences among these three variations (corrected ideal, corrected average, and hybrid concept), the views have a central hybrid essence. Rather than defining reasonableness as a purely statistical or purely prescriptive standard, they each define reasonableness with respect to *both* statistical and prescriptive considerations. The next three Subparts offer arguments in support of such a hybrid theory of reasonableness.

B. Origins of the Reasonable Person

This Subpart defends hybrid accounts of reasonableness by considering the history of the “reasonable person” and its historical colleague, the “man on the Clapham omnibus.” The history of both terms suggests that they were originally introduced and used to reflect a hybrid notion.

1. The Man on the Clapham Omnibus

Scholars and jurists often characterize the man on the Clapham omnibus as the historical predecessor or colleague of the reasonable person.¹³² Although the idea of the reasonable man/person may predate that of the man on the Clapham omnibus,¹³³ it is worth reflecting on the man-on-the-Clapham-omnibus legal fiction, since this terminology shares a long history with reasonableness.

Many cases treat the two terms as synonymous. For example, a 1903 Court of Appeals case regarding whether a comment on a literary work is libel or a “fair comment” notes, “‘Fair,’ therefore, in this collocation certainly does not mean that which the ordinary reasonable man, ‘the man on the Clapham omnibus,’ as Lord Bowen phrased it, the juryman common or special, would think a correct appreciation of the work”¹³⁴ Here we see not just a proximity of the terms *reasonable man* and *man on the Clapham omnibus*, but also a suggestion of their equivalence or coextension.¹³⁵

132. See, e.g., *Healthcare at Home Ltd. v. Common Servs. Agency* [2014] UKSC 49, [2014] 4 All ER 210 (appeal taken from Scot.).

133. See Stern, *supra* note 2, at 59 (tracing the reasonable-person concept to 1703).

134. *McQuire v. W. Morning News Co., Ltd.* [1903] 2 KB 100 (CA), 109.

135. *Id.* Intriguingly, here the “reasonable man” is used to indicate a kind of judgment that is inappropriate for the jury determination. The views of the man on the Clapham omnibus about fairness of criticism ought to be set aside. In this instance, the court holds that such factors are not relevant to the question about the limits of fair comment; this is not the appropriate place for reasona-

An early nonlegal use of the man on the Clapham omnibus may trace to Walter Bagehot's 1873 treatise on the English Constitution:

The middle classes—the ordinary majority of educated men—are in the present day the despotic power in England. “Public opinion,” now-a-days, “is the opinion of the bald-headed man at the back of the omnibus.” It is *not* the opinion of the aristocratical classes as such; or of the most educated or refined classes as such; it is simply the opinion of the ordinary mass of educated, but still commonplace mankind.¹³⁶

Here, the opinion of the man at the back of the omnibus is “simply the opinion of the ordinary mass.” This admits of multiple plausible readings. But some prescriptivist interpretations are particularly unlikely. Bagehot's phraseology does not suggest that the person is simply an embodiment of an ideal normative standard. Nevertheless, it might be understood as referring to the ordinary community values (prescriptivist) or the average community view (statistical) or something mixed (hybrid).

That said, it is notable that the omnibus person is the product of the opinions of the “heavy sensible” masses.¹³⁷ That such a person represents the opinion of a class that is *both* large and sensible is a recurring theme in the story of the man on the Clapham omnibus and the reasonable person. Many of the early descriptions of reasonableness gesture towards a dualistic hybrid feature: reasonableness is mass *and* sensibility, average *and* ideal, ordinary *and* educated.

Additional evidence about the man on the Clapham omnibus comes from considering Clapham itself. At the time of the introduction of the omnibus man, Clapham was a London commuter suburb. More importantly, it was *conceived of* as a London commuter suburb. In a novel of the time, Edmund Yates contrasts two types of persons. The first are those heading for Epsom Downs

bleness analysis. See *Hall v. Brooklands Auto-Racing Club* [1933] 1 KB 205 (CA), 224–26. *But see* Gardner, *supra* note 3, at 18 (suggesting that the man on the Clapham omnibus may play a distinct legal role from that of the reasonable person).

136. WALTER BAGEHOT, *THE ENGLISH CONSTITUTION* 325–26 (new & rev. ed. 1873).

137. *Id.* at 326 (“The English constitution in its palpable form is this — the mass of the people yield obedience to a select few; and when you see this select few, you perceive that though not of the lowest class, nor of an unrespectable class, they are yet of a heavy sensible class — the last people in the world to whom, if they were drawn up in a row, an immense nation would ever give an exclusive preference.”).

(an exciting racecourse) with “a scorn of the respectable conventionalities of society, a freedom of thought and action possessing a peculiar charm of their own.”¹³⁸ The second are those persons “who married and settled, and paid taxes and tradesmen’s bills, and had children, and went to bed before morning, and didn’t smoke clay pipes and sit in their shirt-sleeves . . . the City-clerk going to business on the Clapham omnibus.”¹³⁹

In both Yates’s description of the Clapham-omnibus type and Bagehot’s reference to the “sensible masses,” there is a reflection of both statistical and prescriptive factors. The Clapham type represents a certain kind of typical or common person—one who *also* has a basic respect for law and morality, paying taxes, and abstaining from smoking.

Perhaps the “man on the Clapham omnibus” initially referred to a person who seemed both a statistically average person and a prescriptively praiseworthy person. In other words, in late nineteenth-century Britain it just so happened that there existed a class that seemed large *and* sensible; that this typical-seeming person coincided with this good-seeming person is a contingency.

Close inspection of the history of the man on the Clapham omnibus gives good reasons to suspect that he originally evoked considerations of some hybrid notion. This is a far cry from the view that the reasonable man is an entirely normative “anthropomorphic conception of justice.”¹⁴⁰

2. The Reasonable Person

Legal theorists often trace the reasonable man to *Vaughan v. Menlove* (1837),¹⁴¹ which links

138. EDMUND YATES, *LAND AT LAST: BOOK I — MAKING FOR SHORE* 227 (London, Chapman & Hall 1866).

139. *Id.*

140. *Davis Contractors v. Fareham Urban Dist. Council* [1956] UKHL 3, [1956] AC 696, 728 (appeal taken from Eng.).

141. *Vaughan v. Menlove* (1837) 132 Eng. Rep. 490; 3 Bing (N.C.) 468.

reasonableness with the standard for tort negligence.¹⁴² In a thorough historical analysis, Simon Stern rejects this conventional wisdom, tracing reasonableness to 1703's *R. v. Jones*.¹⁴³ Stern builds a compelling case for a “tension between the normative and statistical functions” that drove the historical development of the reasonableness standard.¹⁴⁴

A closer look at *Vaughan* supports this historical narrative. The court held that “the question for [the jury] to consider, was, whether [a] fire had been occasioned by gross negligence on the part of the Defendant; adding, that he was bound to proceed with such reasonable caution as a prudent man would have exercised under such circumstances.”¹⁴⁵

Some suspect that this mention of the prudent man exercising reasonable caution draws from Adolphe Quetelet's inquiry into human nature,¹⁴⁶ which employed the concept of an abstract “average man” and which was published just two years earlier.¹⁴⁷ However, it is unclear whether Quetelet's project had legal ambitions, and *Vaughan* does not explicitly reference Quetelet's work. Moreover, although Quetelet and *Vaughan* are often credited as the birthplace of the term *reasonable man*, neither actually makes use of the phrase.¹⁴⁸ Both reference the “prudent man,” and Quetelet focuses much attention on the average man—but not a reasonable man.¹⁴⁹ And *Vaughan* only mentions reasonable *caution*, not reasonable men or persons.

Vaughan links the negligence standard to the “reasonable caution” of a “prudent man.”¹⁵⁰ We

142. See Stern, *supra* note 2, at 59.

143. *Id.* Stern traces a precursor of the reasonable man to the 1703 case *R. v. Jones* (1703) 87 Eng. Rep. 863; 6 Mod. 105 (styled as *Anonymous* in the English Reports, but as *R. v. Jones* in all other reporters, see Stern, *supra* note 2, at 59 n.2). That court uses a personified standard to draw the line between civil and criminal harms. See *Jones*, 87 Eng. Rep. at 863–64. Notably, the court describes this personification as the “person of an ordinary capacity.” *Id.* at 864.

144. See Stern, *supra* note 2, at 66.

145. *Vaughan*, 132 Eng. Rep. at 492.

146. See, e.g., Miller & Perry, *supra* note 3, at 370.

147. A. QUETELET, SUR L'HOMME ET LE DÉVELOPPEMENT DE SES FACULTÉS, OU ESSAI DE PHYSIQUE SOCIALE 29 (Paris, Bechalier 1835). For the original English translation of Quetelet's work, see A. QUETELET, A TREATISE ON MAN AND THE DEVELOPMENT OF HIS FACULTIES (R. Knox trans., Edinburgh, William & Robert Chambers 1842) (1835).

148. See generally sources cited *supra* note 147. Quetelet's “*l'homme moyen*” more plausibly translates to “the average man.”

149. See generally sources cited *supra* note 147. For example, “*l'homme raisonnable*” or “*l'homme rationnel*.”

150. *Vaughan*, 132 Eng. Rep. at 492 (“Patteson J. before whom the cause was tried, told the jury that the question for them to consider[] was[] whether the fire had been occasioned by gross negligence on the part of the Defendant; adding, that he was bound to proceed with such reasonable caution as a prudent man would have exercised under such circumstances.”).

can better understand the doctrinal function of reasonableness from some particularities of the decision. First, the standard of “the conduct of a man of ordinary prudence”¹⁵¹ is cited as reaching back to the 1703 *Coggs v. Bernard* decision;¹⁵² it is and has “always been the rule laid down.”¹⁵³ *Coggs* is a decision handed down in the months before 1703’s *R. v. Jones*.¹⁵⁴ Insofar as those decisions reflect “tension between the normative and descriptive functions of the [reasonableness] standard,”¹⁵⁵ this provides a reason to interpret *Vaughan* similarly.

Secondly, it is striking to note the standard that was meant to *oppose* the man-of-ordinary-prudence standard. The opposing standard would have asked whether the “[d]efendant had acted honestly and bonâ fide to the best of his own judgment.”¹⁵⁶ This is a plausible candidate for a purely prescriptive standard, invoking judgment about the defendant’s honesty. The proto-reasonableness standard, on the other hand, seems at least partly concerned with the factual question about the conduct of an actual, ordinary prudent person.

Subsequent cases provide a similar perspective on the reasonable person of tort negligence. If *Vaughan* is the quintessential introduction to the ordinary prudent person of tort law, *Blyth v. Birmingham Waterworks Co.* is the quintessential introduction to the reasonable man of tort law: “Negligence is the omission to do something which a reasonable man, guided upon those considerations which ordinarily regulate the conduct of human affairs, would do, or doing something which a prudent and reasonable man would not do.”¹⁵⁷

151. *Id.* at 493.

152. *Coggs v. Bernard* (1703) 90 Eng. Rep. 905; Holt 13.

153. *Vaughan*, 132 Eng. Rep. at 493.

154. *See Stern*, *supra* note 2, at 64.

155. *Id.* at 66.

156. *Vaughan*, 132 Eng. Rep. at 493.

157. *Blyth v. Birmingham Waterworks Co.* (1856) 156 Eng. Rep. 1047, 1049; 11 Ex. 781, 784.

Again, the text provides evidence for a partly statistical theory of the reasonable man:

A reasonable man would act with reference to the average circumstances of the temperature in ordinary years. The defendants had provided against such frosts as experience would have led men, acting prudently, to provide against; and they are not guilty of negligence, because their precautions proved insufficient against the effects of the extreme severity of the frost¹⁵⁸

Following this line of the reasonable man of ordinary prudence, early U.S. jurisprudence cast the reasonable man as the reasonable, prudent man. Consider, for example, a typical definition of negligence: it is “the failure to do what a reasonable and prudent person would ordinarily have done under the circumstances of the situation.”¹⁵⁹ Here again we see reference to statistical and prescriptive features: prudence and ordinariness. Early interpretations of the reasonable person (extended to areas beyond tort law) suggest the same hybrid interpretation: “[C]ourts should rely upon prevailing social norms for their definition of reasonable behavior.”¹⁶⁰

The prudent and reasonable man has become a classic standard for negligence: “Negligence is the omission to do something which a reasonable man, guided upon those considerations which ordinarily regulate the conduct of human affairs, would do, or doing something which a prudent and reasonable man would not do.”¹⁶¹ In this capacity, the reasonable and prudent man is not a purely prescriptive notion; his existence is grounded in the considerations that ordinarily regulate the conduct of human affairs. Like the man from Clapham, the reasonable man may have been both normal and praiseworthy; as a contingent matter, it might have been the case that the masses were sensible and prudent, following ordinary rules of conduct.

Study of the origins of America’s reasonable man indicates similar concern for considerations of a hybrid notion. The reasonable man did not robustly enter into U.S. Supreme Court jurisprudence until the middle of the nineteenth century.¹⁶² The reasonable man’s role was similar to that played by the man on the Clapham omnibus in England; the reasonable man was characteristical-

158. *Id.*

ly tied to legal rulings concerning negligence.

Holmes, writing at this critical time, characterizes the reasonable man's role, explicitly endorsing the use of a "certain average of conduct."¹⁶³ In *THE COMMON LAW*, Holmes refers to the reasonable man, but more often to the average man or prudent man.¹⁶⁴ The application of the reasonableness standard channels and "represent[s] the feeling of the community," ensuring the law applies generally.¹⁶⁵

Subpart I.A noted that Holmes is sometimes associated with statistical views of reasonableness. But Holmes's view of reasonableness is a matter of debate. He has also been associated with cost-benefit efficiency versions of prescriptivist views. The awareness of hybrid views raises further interpretive possibilities. Perhaps Holmes is best understood as endorsing something like a hybrid theory of reasonableness. Is Holmes's "ideal average prudent man"¹⁶⁶ best understood as a statistical, prescriptive, or hybrid notion? This locution is a strange mixture of statistical and normative properties, like England's "sensible masses." It seems to reference a person who is both (statistically) average *and* (prescriptively) prudent.

Focus on Holmes is useful, as he provides an important example of the jurisprudential view of reasonableness at the time of its increasing use in the United States. But a broader survey of the uses of *reasonable man* and *reasonable person* also indicates the terms' original connection to

159. *R.R. Co. v. Jones*, 95 U.S. 439, 441–42 (1877).

160. Nancy S. Ehrenreich, *Pluralist Myths and Powerless Men: The Ideology of Reasonableness in Sexual Harassment Law*, 99 *YALE L.J.* 1177, 1181 (1990).

161. *Blyth*, 156 Eng. Rep. at 1049.

162. *See infra* Subpart V.C (documenting the Supreme Court's use of "reasonable man").

163. HOLMES, *supra* note 4, at 108.

164. *See id.* at 93–126.

165. *Id.* at 111.

166. *Id.*

consideration of a hybrid notion like normality. Consider an early case in which the reasonable man appears, which concerned a question of negligence liability for damage caused by a nitro-glycerin explosion.¹⁶⁷ The Court looked to the English definition of negligence: “‘Negligence’ has been defined to be ‘the omission to do something which a reasonable man, guided by those considerations which ordinarily regulate the conduct of human affairs, would do, or doing something which a prudent and reasonable man would not do.’”¹⁶⁸

Again, considerations of both statistical and prescriptive features are relevant. The reasonable man is prudent and guided by the considerations that ordinarily regulate human conduct. Rather than reading this as inconstancy or absurdity (a clash of statistical and prescriptive considerations), we might instead understand this as an articulation of a hybrid theory.

A series of other cases from the same time looked to ordinary conduct as a basis for judgments about the reasonable man and negligence liability.¹⁶⁹ Some of these cases employed the reasonable man for different purposes: as a standard for fraud in bankruptcy law and as a counter a defense of equity.¹⁷⁰ Importantly, although reasonableness was not tied to negligence in these two cases, reasonableness was still partly grounded in what was considered ordinary or normal.

C. *The Ordinary Meaning and Use of Reasonable*

For another argument supporting the account of reasonableness as a hybrid notion, this Subpart turns from a historical argument to one about language. The argument can be stated in a general form: insofar as the ordinary use of “reasonableness” determines or provides evidence about its legal meaning or appropriate application, the meaning of legal reasonableness is better cap-

167. *The Nitro-Glycerine Case*, 82 U.S. 524 (1872).

168. *Id.* at 536 (quoting *Blyth v. Birmingham Waterworks Co.* (1856) 156 Eng. Rcp. 1047, 1049; 11 Ex. 781, 784).

169. *See, e.g., R.R. Co. v. Lockwood*, 84 U.S. (17 Wall.) 357, 376 (1873); *Duncan v. Jaudon*, 82 U.S. (15 Wall.) 165, 176 (1873) (citing ordinary loan practices as a basis for reasonable conduct).

170. *Upton v. Tribilcock*, 91 U.S. 45, 55 (1875) (as the standard for fraud); *Walbrun v. Babbit*, 83 U.S. (16 Wall.) 577, 583 (1873) (as a counter to the defense of equity).

tured by a hybrid concept than by a purely statistical or purely prescriptive one.

I present this general argument in two formats, which rest on different theoretical assumptions. The first is grounded in assumptions about the relationship between ordinary meaning and legal meaning. To the extent that ordinary meaning determines or supports legal meaning, the ordinary use of reasonableness indicates facts about ordinary meaning that support a hybrid view, as a normative matter.

The second argument rests on very different assumptions. It begins from the observation that law, especially tort law, typically describes and applies reasonableness in a hybrid fashion. Insofar as this indicates appropriate legal use (*not* necessarily ordinary meaning), this provides support for a hybrid view.

1. The Ordinary Meaning of Reasonable

The first version of the argument rests on assumptions about the relationship between law and ordinary meaning.¹⁷¹ The strongest version of the argument posits that the ordinary meaning of *reasonable* actually determines its (appropriate) legal meaning and application. Here, there is only the slightest gap between *is* and *ought*. The experimental discoveries about how *reasonableness* is applied provide evidence about the term's ordinary meaning, which grounds its legal effect. The fact that prescriptive and statistical notions together best predicted participants' understanding of reasonableness suggests that the ordinary meaning of *reasonableness* is hybrid. Inso-

171. See, e.g., ANTONIN SCALIA & BRYAN A. GARNER, *READING LAW: THE INTERPRETATION OF LEGAL TEXTS* 69–77 (2012). For a discussion of ordinary meaning in the tort context, see generally Patrick J. Kelley, *The Carroll Towing Company Case and the Teaching of Tort Law*, 45 ST. LOUIS U. L.J. 731 (2001); Henry T. Terry, *Proximate Consequences in the Law of Torts*, 28 HARV. L. REV. 10 (1914).

far as legal meaning should reflect ordinary meaning, legal reasonableness should reflect this hybrid conception.

Of course, some might object that it is possible that participants in these studies are somehow making systematic or consistent mistakes. If that were the case, the experimental results would reflect noise and confusion, but not the ordinary meaning of the term *reasonable*.

However, in this instance it seems that the burden falls on the objector to make the case for why such widespread performance errors are plausible. Across a large number of participants, items, and contexts, the ordinary meaning of *reasonableness* reflected the same pattern: intermediacy between the relevant average and ideal. An objector could still stipulate that *reasonableness* means averageness, claiming that any person or participant who says otherwise is mistaken. But this approach is inappropriately untethered to facts about meaning. Scholars consistently remark that ordinary or public meaning is an empirical question.¹⁷² Answering an empirical question requires at least some deference to actual empirical facts. And in this case, there is good empirical evidence supporting an ordinary hybrid conception of reasonableness.

Not all theories will endorse a strong determination claim that the ordinary meaning of *reasonable* should determine its legal effect. But the argument can operate on a less rigid level. Even if a theory holds that ordinary meaning does not determine the legal effect of reasonableness, it might hold that ordinary meaning informs the legal effect of legal reasonableness, alongside other factors. Under this pluralistic approach, there are several factors that determine the legal effect of reasonableness—one of which is ordinary meaning. On this kind of view, the experimental data still provides a reason—just not a decisive reason—to theorize reasonableness in a hybrid way.

Under a variety of legal theory views, ordinary meaning provides this kind of reason. And as

172. See, e.g., Randy E. Barnett, *Interpretation and Construction*, 34 HARV. J.L. & PUB. POL'Y 65, 66 (2011) ("It cannot be overstressed that the activity of determining semantic meaning at the time of enactment required by the first proposition is *empirical*, not *normative*." (citing KEITH E. WHITTINGTON, CONSTITUTIONAL INTERPRETATION: TEXTUAL MEANING, ORIGINAL INTENT, AND JUDICIAL REVIEW 6 (1999))).

argued previously, too large a gap between ordinary judgments of reasonableness and the “true” theory of reasonableness moves such a theory in the direction of impracticality and implausibility.

2. The Use of Reasonable

Even for those who are skeptical of assumptions about the legal significance of ordinary or public meaning, there is a different argument that supports a hybrid account with the experimental results—an argument built on very different assumptions. Some scholars note that reasonableness theories should be informed by how the community or law itself describes, treats, or applies reasonableness, particularly in the tort-law context.¹⁷³ This kind of relationship between legal theory and practice is especially compelling in the context of reasonableness for tort law, an area often tied to customs, conventions, norms, or traditions.¹⁷⁴

If the correct theory of reasonableness (in tort or elsewhere) should be determined or informed by its actual use,¹⁷⁵ the experiments also provide evidence for the hybrid view. Insofar as the experiments model the typical cognitive process underlying reasonableness judgments—and how that same process influences jury decision-making—the results suggest facts about typical legal determinations of reasonableness.

In sum, upon diverse theoretical assumptions, the experimental findings about how reasonableness *is* applied provide reasons for how reasonableness *ought* to be applied. That is, the experimental results support a normative theory of reasonableness as a hybrid.

173. See generally, e.g., Tilley, *supra* note 3.

174. See Kelley & Wendt, *supra* note 51, at 621–22; Tilley, *supra* note 3, at 1345–46; see generally, e.g., The T.J. Hooper, 60 F.2d 737 (2d Cir. 1932); Richard A. Epstein, *The Path to The T.J. Hooper: The Theory and History of Custom in the Law of Tort*, 21 J. LEGAL STUD. 1 (1992); Clarence Morris, *Custom and Negligence*, 42 COLUM. L. REV. 1147 (1942).

175. See, e.g., Zipursky, *supra* note 3, at 2153–54 (noting how *no* states instruct jurors to make a cost–benefit analysis in tort negligence determinations, providing a reason against adopting a welfare maximization theory of reasonableness).

D. A General Theory of Reasonableness

In addition to arguments from history and ordinary meaning and use, this Subpart offers a final argument in favor of reasonableness as a hybrid standard: this characterization offers the best general or unified explanation of reasonableness, one that is apt across many of its varied legal uses. The hybrid view does better than its statistical and prescriptive competitors at avoiding absurdities, and it is the most plausible conception of reasonableness as a general legal standard.

By theorizing reasonableness as a hybrid standard, one that reflects *both* statistical and prescriptive considerations, hybrid accounts capture important insights of both statistical and prescriptive views. Across many reasonableness standards, both statistical considerations and prescriptive considerations seem intuitively relevant. Reasonableness is not simply an average, nor is it simply an ideal. Instead, it is a hybrid judgment informed by both types of considerations.

This dualistic feature allows hybrid views to avoid some absurdities generated by pegging reasonableness strictly to one set of considerations. Recall how in some circumstances reasonableness cannot be a simply statistical standard, while in other circumstances reasonableness cannot be simply a prescriptive standard. Defining reasonableness as a purely statistical standard raises problems of average accidents (i.e., accidents resulting from common behavior that should nevertheless carry legal liability) and reasonable racism (i.e., harms stemming from common attitudes or beliefs that should nevertheless carry legal liability). These absurdities arising from statistical theories of reasonableness are well known.¹⁷⁶

There are also absurdities of prescriptivist views. For example, recall the reasonable-consumer standard in false advertising actions, in which plaintiffs must show that a reasonable

176. See, e.g., Husak, *supra* note 58, at 311; Westen, *supra* note 3, at 138–39.

consumer is likely to be deceived or misled.¹⁷⁷ Most purely prescriptive interpretations of this standard are bizarre. The standard is not meant to protect consumers in only those situations in which they *ought* to be misled. Instead, facts about whether people are actually misled are central.¹⁷⁸

The hybrid view's advantage is that it treats both statistical and prescriptive considerations as relevant to reasonableness, while refusing to treat either as determinative. Thus, one core benefit of hybrid accounts of reasonableness is their capture of the importance of both statistical and prescriptive considerations.

An objector might see the hybrid view's flexibility as coming at the cost of unboundedness: since the hybrid view permits both statistical and prescriptive considerations, it is too flexible, hardly a useful standard at all. The hybrid view is certainly more flexible than most statistical and prescriptivist views, but the hybrid view is not unbounded. As Part III's experiments indicate, when statistical and prescriptive considerations diverge, reasonableness is intermediate between the two. So in cases in which statistical and prescriptive considerations support the same judgment, the hybrid view also supports that judgment. And in cases in which statistical and prescriptive considerations support divergent judgments, the hybrid view is not unconstrained, but is instead bounded by those two judgments.

In addition to avoiding absurdities, the hybrid view is the best contender for a general theory of reasonableness. While many views of reasonableness were modeled in the tort negligence con-

177. *E.g.*, *Fink v. Time Warner Cable*, 714 F.3d 739, 741 (2d Cir. 2013); *In re Horizon Organic Milk Plus DHA Omega-3 Mktg. & Sales Practice Litig.*, 955 F. Supp. 2d 1311, 1330 (S.D. Fla. 2013); *Elias v. Hewlett-Packard Co.*, 950 F. Supp. 2d 1123, 1131 (N.D. Cal. 2013).

178. *See, e.g.*, *Haskell v. Time, Inc.*, 965 F. Supp. 1398, 1406–07 (E.D. Cal. 1997).

text, few are plausible candidate theories of reasonableness across all legal domains that use reasonableness standards.

For example, consider prescriptivist interpretations of reasonableness across various legal domains. The interpretation of reasonableness as welfare maximization is surprisingly uncommon in actual legal practice,¹⁷⁹ but it is also unintuitive in many domains outside of tort law. Consider, for example, criminal reasonableness standards. There are numerous plausible justifications of criminal law, but welfare maximization is an especially strange fit. The same is true for reasonableness standards used in procedural protections. Unreasonable trial delay or unreasonable searches are not adequately explicable in terms of welfare maximization.

These considerations might suggest that what is required is simply a prescriptivist analysis that considers a broader spectrum of factors. Perhaps we should conclude that reasonableness is not just about welfare maximization but also other normative considerations like justice and virtue. However, other uses of reasonableness make clear that prescriptive considerations alone do not set the standard. Recall the example of reasonable provocation to kill. Under a prescriptivist view, there should be no reasonable provocation. What makes the standard sensible at all are statistical considerations about actual patterns of human behavior. It is more just and virtuous to never be provoked to kill. The considerations motivating reasonable provocation as a standard are statistical ones: sometimes ordinary people are in fact provoked to kill.

V. FURTHER IMPLICATIONS AND FUTURE DIRECTIONS

This Part elaborates upon some implications of a hybrid view of reasonableness. Subpart V.A considers practical implications of the data. Subpart V.B addresses the individualization problem. The hybrid view of reasonableness provides new insights into this classic and difficult challenge.

179. See Zipursky, *supra* note 3, at 2153–54.

Subpart V.C considers the implications for reasonableness across some legal uses in which the function of reasonableness or the reasonable person is unclear. The earlier experiments investigated reasonableness across many domains, finding that reasonableness *is* a hybrid judgment across tort, contract, criminal, and many other areas of law. Subpart V.C argues that these features should give us pause about uses of reasonableness that do not reflect a hybrid standard.

A. Practical Implications of the Data

This chapter defends both a statistical and normative account of reasonableness as a hybrid, and these empirical findings provide one line of support for the normative view.¹⁸⁰ However, in this Subpart, I bracket those legal theory implications in order to highlight another set of significant implications of the data. Regardless of one's view about how reasonableness *should* be applied, there are very significant implications of the discovery about how reasonableness *is* applied.

This Subpart outlines some of these practical implications of the data. Since reasonableness is often a jury determination, understanding how ordinary people generate reasonableness judgments is of great value to various members of the legal system. These facts provide relevant information for those who are considering pursuing legal claims, for legal representatives and decision makers, and for those drafting or providing jury instructions.

1. Legal Claimants

Insight into the way in which ordinary people judge reasonableness provides potential liti-

180. *See supra* Subpart IV.C.

gants with greater information about the likely success of their claims. As a heuristic for estimating jury intuitions about reasonableness, a litigant might reflect on both statistical and prescriptive factors, as well as the perceived normality of the relevant issue. For example, if a contract dispute or public nuisance claim turns on a reasonableness standard, the potential litigant might consider what people actually *would* do in the situation, not what people *should* do in the situation.

To be sure, there is a plausible gap between the results of the experiments and the reality of an actual jury deliberation. As in any controlled experiment, there are questions of external validity. Nevertheless, in this case there are good reasons to think the experiments have external value, particularly in the jury decision-making context. For one, both the experimental study and jury context involve judgments by ordinary people (i.e., not legal experts). In both cases, the decision makers have relatively little training and are encouraged to use their ordinary judgment to answer a legal question.

Moreover, jury deliberation involves an aggregative feature similar to that of the experimental analysis. Even if it is not the case that each individual's reasonableness judgments are intermediate between their relevant average and ideal judgments, this pattern emerges on average across a group of decision makers. This feature of jury deliberation is replicated in the experimental analysis, which considers aggregate responses rather than individual ones.

Of course, future research could provide evidence that weakens or strengthens the external validity of these studies. Perhaps the small degree of training that jurors receive makes them judge reasonableness in a very different way. Or perhaps the context of a jury deliberation has a significant effect on the pattern of aggregate reasonableness judgment. These are open empirical questions. As initial evidence, the results provide modest support for the claim that modeling jury verdicts of reasonableness would be better predicted by considering *both* the relevant average and ideal rather than either the average or ideal alone.

2. Legal Representatives

Similar recommendations extend to legal representatives. The increase of sophisticated modeling of settlement rates and litigation success rates evinces a demand for evidence about likely trial success.¹⁸¹ The most obvious source of data is empirical study of actual litigation results, but legal representatives might also consider other sources of empirical evidence. Consulting existing experimental data is one promising source.¹⁸² The study here offers evidence about how people would judge the reasonableness of various legal fact patterns.

Running one's own test panels would present another useful perspective. Given the experimental finding about reasonableness, legal representatives might benefit from running panels with facts tailored to their case. Where there is a reasonableness standard, asking ordinary people about reasonableness provides one data point. But this could be supplemented by asking about the relevant average and ideal, providing plausible decision boundaries. Such experimental trials could be particularly useful in estimating success under different possible factual discoveries.

3. Jury Instructions

A final domain in which the results carry practical implications is that of jury instruction.¹⁸³ Insofar as the drafters of jury instructions wish to convey the appropriateness of a hybrid reasonableness judgment, they might elaborate upon the factors contributing to the desired judgment. Jury instructions on reasonableness could acknowledge that in making determinations about

181. See, e.g., Theodore Eisenberg & Charlotte Lanvers, *What Is the Settlement Rate and Why Should We Care?*, 6 J. EMPIRICAL LEGAL STUD. 111, 112–14 (2009).

182. See, e.g., Dennis J. Devine et al., *Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups*, 7 PSYCHOL. PUB. POL'Y & L. 622, 699–701 (2001).

183. See, e.g., Kelley & Wendt, *supra* note 51, at 618–22.

whether something was reasonable, jurors might be helped by considering both statistical and prescriptive factors such as the relevant average and ideal, or what they think people *would* typically do and what they think people *should* do.

Alternatively, if jury instruction drafters seek to convey the inappropriateness of either statistical or prescriptive considerations in a reasonableness determination, the experimental results suggest that they should include explicit instructions to that effect. Since the ordinary judgment about reasonableness is a hybrid one, in both a legal and nonlegal context, it is plausible to assume that this is how many jurors understand a task of judging reasonableness. So if, for example, statistical considerations should be irrelevant to a particular reasonableness analysis, the experimental results indicate a reason to instruct jurors explicitly to ignore such statistical considerations in their decision making. Similarly, if prescriptive considerations should be irrelevant to a particular reasonableness analysis, the results indicate a reason to instruct jurors to ignore prescriptive considerations.

B. Individualization

Thus far, the chapter has been silent on another crucial question about reasonableness: the “individualization problem.”¹⁸⁴ This is the problem of which personal characteristics, such as age,¹⁸⁵ culture,¹⁸⁶ gender,¹⁸⁷ mental illness,¹⁸⁸ race,¹⁸⁹ sexuality,¹⁹⁰ or combinations of these,¹⁹¹

184. See, e.g., MORAN, *supra* note 2, at 5; Westen, *supra* note 3, at 139.

185. See, e.g., David E. Seidelson, *Reasonable Expectations and Subjective Standards in Negligence Law: The Minor, the Mentally Impaired, and the Mentally Incompetent*, 50 GEO. WASH. L. REV. 17, 20–26 (1981).

186. See, e.g., *People v. Wu*, 286 Cal. Rptr. 868, 884 (Ct. App. 1991) (depublished).

187. See *Ellison v. Brady*, 924 F.2d 872, 879–80 (9th Cir. 1991); Jeremy A. Blumenthal, *The Reasonable Woman Standard: A Meta-Analytic Review of Gender Differences in Perceptions of Sexual Harassment*, 22 L. & HUM. BEHAV. 33, 34–35 (1998); Naomi R. Cahn, *The Looseness of Legal Language: The Reasonable Woman Standard in Theory and in Practice*, 77 CORNELL L. REV. 1398, 1406–11 (1992); Barbara A. Gutek & Maureen O'Connor, *The Empirical Basis for the Reasonable Woman Standard*, 51 J. SOC. ISSUES 151, 160–63 (1995). See generally Elizabeth L. Shocnfelt, Allison E. Maue & JoAnn Nelson, *Reasonable Person Versus Reasonable Woman: Does It Matter?*, 10 AM. U. J. GENDER SOC. POL'Y & L. 633 (2002); Jolynn Childers, Note, *Is There a Place for a Reasonable Woman in the Law? A Discussion of Recent Developments in Hostile Environment Sexual Harassment*, 42 DUKE L.J. 854 (1993) (examining the strengths and weaknesses of a reasonable-woman standard).

188. See generally, e.g., Kristin Harlow, Note, *Applying the Reasonable Person Standard to Psychosis: How Tort Law Unfairly Burdens Adults with Mental Illness*, 68 OHIO ST. L.J. 1733 (2007).

should be included in reasonable-person analyses. This problem is important and difficult,¹⁹² and it is largely separate from the statistical, prescriptive, and hybrid debate. These competing views can be combined with various different accounts that address the individualization problem. Nevertheless, given the problem's importance, it is worth remarking briefly on its nature and how the hybrid view enhances the individualization debate.

The individualization problem is theoretically intriguing but also practically impactful. A reasonableness analysis might result in a different determination depending upon whether it includes or excludes certain individual features. For instance, a child's age "would have affected how a reasonable person" would perceive her freedom to leave a police interview.¹⁹³ If age were excluded from the analysis (and the reasonable person were conceptualized as an adult), a child would likely be held to have perceived greater freedom to leave than if age were included.

Consider the responses of statistical and prescriptivist views to the individualization problem. Statistical views might simply assert that no individual features are relevant; reasonableness is judged by reference to a person generated by empirical facts about *all* persons. Intuitively, this view gets things wrong. There may be a number of individual differences that would not manifest as part of the statistical normality analysis, but that are nevertheless relevant to reasonable person analysis (e.g., blindness).

189. See, e.g., andré douglas pond cummings, "Lions and Tigers and Bears, Oh My" or "Redskins and Braves and Indians, Oh Why": *Ruminations on McBride v. Utah State Tax Commission, Political Correctness, and the Reasonable Person*, 36 CAL. W. L. REV. 11, 26–33 (1999). See generally, e.g., Mia Carpinello, *Striking a Sincere Balance: A Reasonable Black Person Standard for "Location Plus Evasion" Terry Stops*, 6 MICH. J. RACE & L. 355 (2001).

190. See, e.g., Dressler, *supra* note 20, at 756–57.

191. See, e.g., Stingley v. Arizona, 796 F. Supp. 424, 428–29 (D. Ariz. 1992); Zalsne, *supra* note 20, at 863–65.

192. It is worth noting that, although the individualization problem might seem a special problem for reasonableness, it is a problem that plagues the application of many standards to particular instances. A court must determine which features of a reasonable person are relevant (how the person should be "individualized"), but also which features of many categories germane to analyses are relevant. For instance, courts often determine which features of an act are relevant (how an act should be individuated).

193. J.D.B. v. North Carolina, 564 U.S. 261, 271–72 (2011) (quoting Stansbury v. California, 511 U.S. 318, 325 (1994)).

To account for such differences, a statistical view would typically refer to systematic patterns of difference. For example, if blind persons were statistically far more likely to behave in a certain way, the view would not necessarily hold blind persons to the reasonableness standard determined by reference to *all* people, but instead to a standard determined by the average blind person.

This suggestion obviously requires some limiting principle. For instance, if *every* distinguishing feature of some person is deemed relevant to reasonable person analysis, the reference class upon which the person is determined will shrink to a very small size. That is, imagine incorporating a person's age, gender, race, ability, but also favorite color, birthday, full name, and so on. If the reasonable person is just equal to the normal person with those features, the reasonable person will just *be* the specific person.

And even if the statistical view only acknowledges features with a statistically significant impact (e.g., perhaps excluding favorite color), we will still have far too many individualized reasonableness standards. The useful generality of reasonableness suffers a death by a thousand cuts.

A similar problem arises for prescriptive views. The view might individualize the reasonable person not based on empirical differences, but instead on prescriptive ones. The blind person is held to a different standard, not because "the average" does not represent him well, but because the standard that is normatively justifiable or welfare maximizing for blind persons is different than the one for nonblind persons. However, this results in a similar concern: can such individualization be limited?

The account of reasonableness as a hybrid offers fresh insight into this difficult problem. To be sure, it does not quickly solve the problem. Individualization remains a hard question for any theory of reasonableness, be it prescriptive, statistical, or hybrid. What the hybrid view does offer is an enriched perspective on the nature of the problem. The individualization problem is typically approached from a prescriptivist perspective. From that view, the stakes of individualization

concern prescriptive (or normative) justifications. Alternatively, adopting a hybrid view makes clear that the stakes of individualization concern *both* prescriptive and statistical questions.

For example, consider a reasonableness standard for sexual harassment and the question of whether we should individualize the standard (the reasonable woman) or not (the reasonable person). On a prescriptivist view, the only types of considerations that determine reasonableness are prescriptive ones. So what is at stake in individualization is something prescriptive. *Should* women understand a series of remarks differently from men?

However, on a hybrid view, both statistical and prescriptive considerations determine reasonableness. The individualization problem is thus a problem about both types of considerations. *Should* women understand a series of remarks differently from men, but also *would* women in fact typically understand those remarks differently?

Again, this feature of the hybrid view does not solve the individualization problem. One could adopt a hybrid view and hold that we should not individualize in any context with respect to merely statistical differences. For example, one might contend that reasonable-woman standards should only be applied where there is a compelling normative difference. Or one could adopt a hybrid view while holding that we should individualize with respect to *either* statistical or prescriptive differences. Or one could adopt a hybrid view combined with the view that we should individualize with respect to only features involving *both* a statistical and prescriptive difference. And so on.

The hybrid view's crucial insight is that individualization choices might also involve considerations about statistical factors. This contrasts sharply with the typical characterization of the individualization problem, which arises from the prescriptivist perspective. Because those views

characterize reasonableness in terms of prescriptive considerations, they tend to treat individualization similarly. From a prescriptivist perspective, individualization—like reasonableness—is not about statistical patterns of behavior; instead, it is about normatively correct behaviors.

Once again, this prescriptivist treatment is most sensible in the tort negligence context. The choice about individualization for tort negligence seems plausibly driven by prescriptive factors. All persons should be held to the same general standard of care, regardless of their demographic idiosyncrasies.

However, for other domains of reasonableness, there seems to be a relevant individualization choice from mere statistical facts alone. Irrespective of whether it would be welfare-maximizing, virtuous, or justifiable for women to understand certain types of remarks differently from men, a fact that they do suggests a substantive individualization question.

Other areas of law share this feature. Individualizing a criminal defendant based on race, age, or ability might be motivated by prescriptive considerations—someone of that race, age, or ability should act differently from others. But it might equally be motivated by statistical ones; whether or not it is morally right or objectively justifiable for those group members to act differently, the mere fact of statistical differences could generate an individualization question. For example, perhaps sixteen-year-olds should act like adults in ways relevant to a criminal reasonableness standard, but on average they do not. Prescriptivist views should not individualize, but on the hybrid view, an individualization question remains.

This approach to individualization pays respect to two seemingly conflicting aims of reasonableness. On the one hand, reasonableness standards are standards of general application.¹⁹⁴ On the other hand, reasonableness is sometimes taken to represent precisely the opposite: the *aptness*

194. See HOLMES, *supra* note 4, at 108.

of law, or its resistance to overgenerality.¹⁹⁵ The individualization problem remains an open question for prescriptivist, statistical, and hybrid views.¹⁹⁶ But for now, the hybrid view offers a new gloss on an old problem: the choice about whether to individualize could be motivated by prescriptive considerations or statistical ones.

C. *The Increasing Use of Reasonableness and Reasonable Person*

A final implication of the hybrid view concerns the appropriateness of the legal use of *reasonableness* and *the reasonable person*. If the appropriate role of reasonableness standards is to reflect a hybrid notion, reasonableness standards that do not serve this role warrant cautious scrutiny. This Subpart suggests that *reasonableness* and *the reasonable person* are used broadly and—to some surprise—increasingly. This is true even in nontraditional uses (e.g., reasonableness is not exclusive to state tort claims). On the account defended in Part IV, reasonableness should reflect a hybrid judgment like that of normality. However, it is sometimes unclear what these various uses of *reasonableness* do or should reflect. As such, the present account advises a cautionary restraining of the use of legal reasonableness, particularly outside of its more traditional applications.

The studies of Part III provide evidence of a systematic process guiding reasonableness

195. See, e.g., Timothy Endicott, *The Subsidiarity of Law and the Obligation to Obey*, 50 AM. J. JURIS. 233, 240 (2005) (“Aristotle also saw the *disadvantage* in the generality of laws. He insisted that laws are to be applied to particular cases with *epieikeia*: reasonableness or aptness. He explains that error can arise in the application of general norms because the lawmaker needs to impose a general control on conduct that may reasonably vary according to the circumstances in which the general norm will apply. So laws are inevitably overgeneral because of the incapacity of lawmakers to tailor a general norm to a variety of circumstances. So citizens and officials need to act with *epieikeia* to avoid applying a general norm in a manner contrary to the lawmaker’s own rationale.” (citing ARISTOTLE, NICOMACHEAN ETHICS V.10 (W.D. Ross trans., Batoche Books 1999), <https://socialsciences.mcmaster.ca/econ/ugcm/3113/aristotle/Ethics.pdf>)).

196. And I suspect that the right view will be highly domain-specific with respect to individualization (e.g., the way in which the right hybrid view should individualize based on statistical considerations, prescriptive considerations, or both will vary from tort to contracts to criminal law).

judgment. Across a number of different domains, mean reasonable quantities thread a narrow gap between mean average and ideal judgments.¹⁹⁷ This finding suggests that an ordinary judgment of reasonableness is not idiosyncratic among different legal domains. Part IV defended the normative appropriateness of this pattern of judgment, supporting a normative theory of reasonableness that should apply generally. Of course, in many legal areas reasonableness standards have evolved to address very specific questions and are accompanied with correspondingly specific jury instructions. Sometimes, reasonable is better understood as meaning something else entirely (e.g., foreseeable).

The previous Parts have focused primarily on theories that posit a unified core account of reasonableness (e.g., as averageness, justification, or normality). However, some are skeptical about this very project. Reasonableness is used in many different legal contexts, and some wonder whether there is any coherency or consistency among these various legal reasonableness standards.¹⁹⁸

This chapter counters this skepticism by defending a general theory of reasonableness as a hybrid notion. This account provides a plausible explanation of reasonableness standards across most domains. However, an implication of the present view is that uses of reasonableness (and the reasonable person) that do not reflect this hybrid judgment are inappropriate. To understand the scope of the challenge, first consider a brief study estimating trends in the frequency of use of reasonableness standards.

Although the most typical use is in state tort claims (e.g., negligence), reasonableness is used much more broadly. To quantify the frequency of important uses that plausibly exceed use in state tort claims, consider a study of the Supreme Court's use of reasonableness standards. Because the term *reasonable* might be used in many contexts (including ones not intended to invoke

197. See *supra* Subpart III.B.

198. See, e.g., Zipursky, *supra* note 3, at 2132–33.

legal reasonableness), I instead study the use of the terms *reasonable man*, *reasonable woman*, and *reasonable person*. While these terms could be used in other contexts, quantifying use of these phrases allows a plausible estimation of trends in citation of legal reasonableness in the relevant sense.

Like *reasonableness*, *the reasonable person's* breadth and frequency are striking. The reasonable person is “perhaps the common law’s most enduring and expounded upon fiction,”¹⁹⁹ found across a number of areas of law—including tort negligence,²⁰⁰ contract,²⁰¹ criminal law,²⁰² and police conduct.²⁰³

As Figure 3 indicates, the explosion of the use of the phrase *the reasonable person* in the U.S. Supreme Court did not begin until around 1975. Supreme Court uses of *the reasonable man* have actually remained level since 1870—and the phrase was only used once before that time.²⁰⁴

This small study challenges several myths about the reasonable person in U.S. law. Most importantly, for present purposes, it debunks the myth that reasonableness is largely a remnant of tort negligence claims. Intriguingly, *reasonable man* and *reasonable person* are not longstanding legal fixtures in Supreme Court jurisprudence. A significant use of the former did not occur before the mid-nineteenth century, and the robust appearance of the latter did not commence before the mid-1970s. Second is the myth that the reasonable person has consistently played a large and

199. MORAN, *supra* note 2, at 18.

200. See Vaughan v. Menlove (1837) 132 Eng. Rep. 490, 493; 3 Bing. (N.C.) 468, 474–75; RESTATEMENT (SECOND) OF TORTS § 283 (AM. LAW INST. 1965).

201. See RESTATEMENT (SECOND) OF CONTRACTS § 2 (AM. LAW INST. 1981).

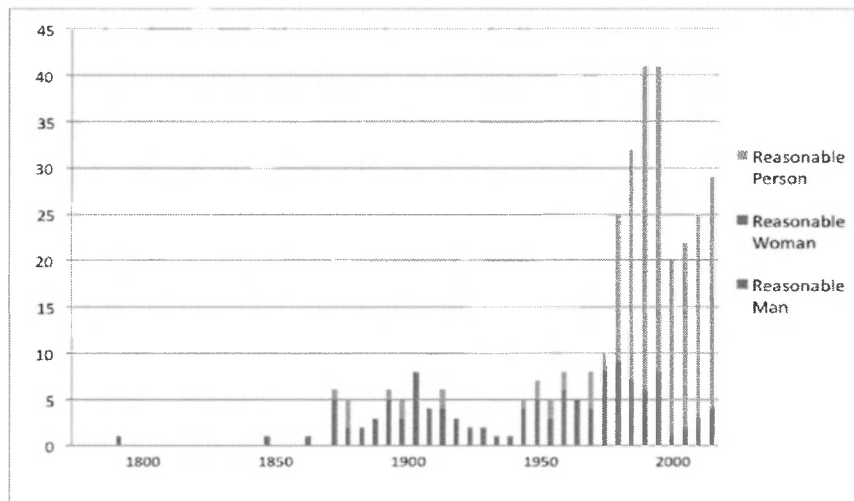
202. See, e.g., MODEL PENAL CODE § 2.02(2)(d) (AM. LAW INST. 1985) (culpability); *id.* § 3.04 (use of force for self-protection); *id.* § 3.05 (use of force for protection of other persons); *id.* § 3.08(4)(b) (use of force by persons with special responsibility for care, discipline or safety of others); *id.* § 210.3(1)(b) (manslaughter); *id.* § 210.4 (negligent homicide).

203. See, e.g., J.D.B. v. North Carolina, 564 U.S. 261, 270–79 (2011); United States v. Mendenhall, 446 U.S. 544, 554 (1980).

204. Chisholm v. Georgia, 2 U.S. (2 Dall.) 419, 431–32 (1793), *superseded by constitutional amendment*, U.S. CONST. amend. XI, as recognized in Va. Office for Prot. & Advocacy v. Stewart, 563 U.S. 247 (2011).

central role across many areas of law. The quantitative uptick in *reasonable person* uses is recent. Part III noted a similarly recent qualitative change: while reasonable-man standards were introduced primarily to deal with cases of negligence (often invoking the “prudent, reasonable” person or person of “ordinary prudence”), the expansion of these standards to other areas of law is more recent.²⁰⁵

Figure 3. *Reasonable Person, Reasonable Woman, and Reasonable Man* in U.S. Supreme Court Decisions, 1789–2015²⁰⁶ (five-year increments)



In addition to its frequent use, *the reasonable person* also has varied use, spanning numerous legal doctrines.²⁰⁷ Moreover, the reasonable man, turned reasonable person, now has several other reasonable companions. The Supreme Court has recently referred to the “reasonable investor”²⁰⁸;

205. The method employed here comes with significant limitations. For one, the term *reasonable person* may appear prior to and exhibit different trends in state courts or statutes, as compared to Supreme Court opinions. The method employed to track U.S. Supreme Court uses is useful because of its precision. Nevertheless, a preliminary review of U.S. statutes suggests some even earlier uses of *reasonable person* than those cited here, but also generally similar trends in the use (and increase) of *reasonable person*.

206. $R^2 = .565$, $p < .0001$ (highly statistically significant), Runs' Statistic = 8 (strong evidence for trend), Durbin-Watson Statistic = .122 (moderate evidence against autocorrelation). I follow the general methodology introduced and articulated by John Calhoun. Note, *Measuring the Fortress: Explaining Trends in Supreme Court and Circuit Court Dictionary Use*, 124 YALE L.J. 484, 493–96. The plain text copies of opinions can be found at <https://bulk.resource.org/courts.gov/c>. Given the recent decline in the Supreme Court's docket, the trend of modern increase of usage is only more extreme when counting uses as a percentage of Supreme Court holdings.

207. See *supra* notes 9–15 and accompanying text.

208. *Omnicare, Inc. v. Laborers Dist. Council Constr. Indus. Pension Fund*, 135 S. Ct. 1318, 1327 (2015).

the “reasonable lawyer” and “reasonable judge”²⁰⁹; and the “reasonable judge or jury.”²¹⁰ This growth in references to reasonable actors is not limited to United States jurisprudence.²¹¹

The theory of reasonableness as a hybrid offers a cautionary implication for some of these expansive and increasing uses of reasonableness. On the present view, reasonableness has a very particular function: it should reflect a hybrid standard informed by statistical and prescriptive considerations. On that view, it is inappropriate for reasonableness standards to reflect something besides that hybrid standard.

Important exceptions to this principle include doctrines that refer to reasonableness, but also have substantial guidelines giving content to the standard.²¹² These uses of *reasonableness* do not reflect the essence of the term, but they are also not inappropriately trading on the currency of reasonableness.

The more problematic class is the broader expansion of reasonableness—e.g., various types of reasonable legal assertions such as those of the form “it is reasonable that *X*.” Of course some

209. *Caperton v. A.T. Massey Coal Co.*, 556 U.S. 868, 896 (2009) (Roberts, C.J., dissenting).

210. *Walker v. Martin*, 562 U.S. 307, 317 n.5 (2011).

211. *See, e.g.*, *Healthcare at Home Ltd. v. Common Servs. Agency* [2014] UKSC 49, [2014] 4 All ER 210 (appeal taken from Scot.). Reasonable person standards can be found across a variety of legal systems, from Canada’s patent law, *Beloit Can. Ltée/Ltd. v. Valmet Oy*, 1986 CarswcllNat 588, para. 17 (Can. Fed. Ct.) (WL), to Australia’s employment law, *Re Sortiros Pandos & Commonwealth of Austl.* [1991] AATA 18, (25 Jan. 1991) (Austl.), <http://www8.austlii.edu.au/cgi-bin/viewdoc/au/cases/cth/AATA/1991/18.html> (“Alternatively, would reason dictate or would a person, whether on the Clapham omnibus or a Bourke Street tram consider that the disciplinary action of the Respondent against the Applicant was ‘reasonable?’”), to Hong Kong’s securities law, *Ng Chiu Mui v. Sec. & Futures Comm’n*, [2017] SFAT App. No. 7/2007 (Sec. & Futures Appeals Tribunal), http://www.sfat.gov.hk/english/determination/AN-7_and_8_and_9-2007-Determination.pdf, to Singapore’s administrative law, *Jeyaretnam Joshua Benjamin v. Lee Kuan Yew* [1992] 2 SLR. 310, SGCA 27 (Sing.) (citing *R. v. Liverpool City Justices, ex parte Topping* [1983] 1 All ER 490 (Q.B.), 494), <http://archive.li/0LJri>.

212. For example, consider the classic case of *Hadley v. Baxendale*, in which a question arose about whether a loss was a “reasonable and natural consequence of [a] breach of contract.” *Hadley v. Baxendale* (1854), 156 Eng. Rep. 145, 151; 9 Ex. 341, 356. As Fuller and Perdue note, the court cites reasonableness, but applies a test of foreseeability. L. L. Fuller & William R. Perdue Jr., *The Reliance Interest in Contract Damages: I*, 46 YALE L.J. 52, 85–86 (1936). Reasonableness is explicated by foreseeability. As a second example, consider the reasonableness of a seizure. *Scott v. Harris* cites reasonableness and articulates its application. 550 U.S. 372, 383–86 (2007). The reasonableness of a seizure is fully explicated by a balancing test. There are countless other examples of this broad doctrinal pattern: reasonableness is referenced, but the test or reasoning involved does not require reference to the reasonable person in the same way as in, e.g., tort negligence, since the content of reasonableness has been explicated in other terms.

of these uses reflect attempts to relevantly individualize the reasonable person, for instance by referring to the reasonable woman.²¹³ But perhaps in some other cases, these represent an effort to fill gaps with the influence of reasonableness.

One possibility is that assertions of “reasonable Xs” or the “reasonableness of Y” reflect a stylistic bad habit. Adding *reasonable* softens assertions, but it does not in itself contribute to or support the legal reasoning. A more concerning possibility is that reasonable assertions of this kind function as a placeholder for undefended intuition or opinion, which confer legal authority because of their reference to the reasonable. These uses carry and convey authority, as reasonableness and the reasonable person have a long and venerable legal history.

Interrogating each of these uses of reasonableness is a project for another paper. However, if these uses are not to be grounded in hybrid considerations, they are not serving the general role of reasonableness. And they may be exchanging reasonableness currency for intuition. This worry amplifies in circumstances in which courts theorize reasonableness prescriptively. Replacing judicial uses of *reasonable* with *in line with Justice* or *morally right* would raise red flags in many of those circumstances.

Regardless of whether one accepts the primary argument that reasonableness should be a hybrid standard, an important question remains about the growing use of reasonableness in varied contexts: if these uses should not be served by a hybrid standard, what role exactly is played by reasonableness—and what role should it be playing?

213. *Harris v. Forklift Sys., Inc.*, 510 U.S. 17, 20 (1993).

CONCLUSION

This chapter reconsiders a classic debate about reasonableness. Is reasonableness a statistical notion, like what is common, or a prescriptive notion, like what is good? The chapter defends a third option: reasonableness is a partly prescriptive and partly statistical hybrid standard.

In concluding, it is worth reflecting briefly on the chapter's methodology. Much literature focuses on how we *should* judge what is reasonable, but this chapter's experimental method sheds light on an equally important question: How *do* people judge what is reasonable?

The experimental study informs this descriptive question: People's judgments of what is reasonable reflect *both* statistical and prescriptive considerations. But—perhaps surprisingly—the experimental study also illuminates an underdeveloped normative possibility. Much debate pits statistical against prescriptivist views. But there is a compelling third possibility that reflects insights from each of the other two: reasonableness is partly statistical and partly prescriptive. Empirically, ordinary judgment of reasonableness is a hybrid judgment. And normatively, reasonableness should be applied as a hybrid standard.

APPENDIX A:**FULL EXPERIMENTAL MATERIALS**Experiment 1*Introduction (condition difference in brackets)*

Below, we ask you to estimate the [average, ideal, reasonable] quantity of a number of different things. Please note that you are not in any way being evaluated on these judgments, and we ask that you do not consult outside sources.

Items (presented in random order):

- [Average, Ideal, Reasonable] number of hours of TV that a person watches in a day;
- [Average, Ideal, Reasonable] number of sugary drinks that a person consumes in a week;
- [Average, Ideal, Reasonable] number of hours that a person spends exercising in a week;
- [Average, Ideal, Reasonable] number of calories that a person consumes in a day;
- [Average, Ideal, Reasonable] number of servings of vegetables that a person consumes in a month;
- [Average, Ideal, Reasonable] number of lies that a person tells in a week;
- [Average, Ideal, Reasonable] number of minutes that a doctor is late to see his/her patients;
- [Average, Ideal, Reasonable] number of books that a person reads in a year;
- [Average, Ideal, Reasonable] number of romantic partners that a person has in their life;
- [Average, Ideal, Reasonable] number of international conflicts that a country has in a decade;
- [Average, Ideal, Reasonable] amount of money (in dollars) that a person cheats on his/her taxes;

- [Average, Ideal, Reasonable] percentage of students who have cheated on an exam in any given high school;
- [Average, Ideal, Reasonable] number of times a person checks his/her phone in a day;
- [Average, Ideal, Reasonable] number of minutes that a person spends waiting on the phone for customer service;
- [Average, Ideal, Reasonable] number of times that a person calls his/her parents in a month;
- [Average, Ideal, Reasonable] number of times that a person cleans his/her home in a month;
- [Average, Ideal, Reasonable] number of times that a computer crashes in a month;
- [Average, Ideal, Reasonable] percentage of high school dropouts there are in any given high school;
- Enter the number 15 to show you are paying attention.
- [Average, Ideal, Reasonable] percentage of kids in any given middle school who are bullied;
- [Average, Ideal, Reasonable] number of drinks that a fraternity brother drinks on a weekend.

Experiment 2

Introduction for legal context condition

In the following screen we ask you to judge the legally reasonable quantity of a number of different things. We ask you to imagine that you are making these judgments in a legal setting for

a legal purpose. For example, imagine that you are a jury member in a jury deliberation.

Jurors are often asked to make legal judgments by comparing someone's actual behavior to a hypothetical reasonable one. For example, imagine Mike was painting the outside of his house and left the can of lead-based paint open by his garage for some amount of time. During that time, the neighbor's dog ate some of the paint and was injured. To determine whether Mike is legally liable for the injury to the dog, jurors might be asked to compare Mike's actual behavior to "reasonable" behavior in similar circumstances. If Mike acted with the reasonable amount of care (or more), he is not liable for the dog's injury; if Mike acted with less care than the reasonable amount, he is liable for the dog's injury.

These examples can vary very widely. For example, a contract might specify that employees are entitled to a "reasonable" number of sick days per year. Settling a contract dispute between the employer and employee would involve comparing the number of sick days that the employee actually took to the reasonable number of sick days.

In the next screen we ask you to estimate the reasonable quantity of different things. For some of these things, it will seem very clear how the question of its reasonableness might arise in a legal setting; for others, it will be less obvious. We ask that in all cases, you keep in mind the legal context.

Items (presented in random order):

[The same items as in Experiment 1; for example:

- Reasonable amount of money (in dollars) that a person cheats on his/her taxes.]

Experiment 3

Introduction

[All participants received the same introduction as in Experiment 1. Participants in the legal context condition also received the introduction from Experiment 2.]

Items (presented in random order):

- [Average, Ideal, Reasonable] number of days taken to accept a business contract when no deadline is specified;
- [Average, Ideal, Reasonable] number of weeks taken to return a product ordered online when the warranty does not specify;
- [Average, Ideal, Reasonable] number of hours taken to reflect on an exciting but risky business proposition;
- [Average, Ideal, Reasonable] amount of unexpected additional costs in a \$10,000 building contract;
- [Average, Ideal, Reasonable] number of weeks that a building construction project is delayed beyond its stated completion date;
- [Average, Ideal, Reasonable] number of loud events held at a football field close to a quiet neighborhood, per year;
- [Average, Ideal, Reasonable] percent of profits that a car manufacturer spends on additional safety features;
- Enter the number 17 to show you are paying attention;
- [Average, Ideal, Reasonable] percent of available medical details that a patient wants to hear from his/her doctor;
- [Average, Ideal, Reasonable] number of weeks that a person has to wait before being tried for a criminal charge;

- [Average, Ideal, Reasonable] number of dollars per hour that a charity pays in attorney's fees for legal work for the charity;
- [Average, Ideal, Reasonable] number of hours of notice that a landlord provides a tenant before entering the unit for maintenance or repairs;
- [Average, Ideal, Reasonable] interest rate for a loan;
- [Average, Ideal, Reasonable] percent likelihood that a company found legally liable for pollution will pollute again in the future.

APPENDIX B:

FULL EXPERIMENTAL RESULTS

Table 2. Mean Average, Ideal, Reasonable, and Legally Reasonable Judgment

Domain	Avg.	Ideal	Reas.	Leg. Reas.
Hours TV watched/day	4.00	2.34	3.19	2.87
Sugary drinks/week	9.67	3.52	5.48	7.26
Hours exercising/week	5.37	7.31	4.78	6.74
Calories consumed/day	2159.26	1757.84	2008.48	1997.07
Servings of vegetables/month	34.81	67.67	46.47	46.53
Lies told/week	24.25	2.75	3.52	4.48
Mins. doctor late/appointment	17.78	3.97	13.08	11.18
Books read/year	10.07	26.15	8.66	13.26
Romantic partners/lifetime	8.04	4.25	7.72	7.78
International conflicts/decade	19.3	1.59	3.48	7.19
Money cheated on taxes	604.56	136.45	335.93	247.17
Percent students cheat on exam	34.64	3.5	13.83	15.79
Time checking phone/day	45.33	13.12	18.17	17.58
Mins. for customer service	15.04	5.78	7.89	8.62
Times calling parents/month	6.04	6.00	5.09	6.40
Times cleaning home/month	5.57	6.75	4.08	7.86
Computer crashes/month	4.78	0.50	0.936	1.23
% high school dropouts	12.64	3.82	7.28	6.81
% school students bullied	27.59	2.31	9.91	9.98
Drinks of frat brother/weekend	16.79	5.91	8.15	7.79
Domain	Avg.	Ideal	Reas.	Leg. Reas.
Days to accept contract	14.68	10.52	12.21	11.59
Weeks to return online product	3.57	5.96	5.78	4.53
Hrs. to reflect on business offer	17.81	32.33	35.46	33.31
Unexpected cost \$10,000 building	2492.86	1332.73	1661.37	1744.88
Weeks of building delay	13.56	5.34	5.75	7.57
Loud events near quiet town/year	29.94	10.48	16.13	13.15
Percent profits on safety features	12.80	22.17	31.55	18.47
Percent medical details desired	58.64	75.40	62.92	68.03
Weeks before criminal trial	17.13	6.92	11.02	9.57
\$ attorney's fees for charity/hr.	134.44	50.89	64.56	76.31
Hrs. landlord notice before entering	28.08	35.47	32.41	30.93
Interest rate for a loan	9.18	3.26	4.82	5.44
% likelihood polluter recidivism	54.43	32.37	40.76	58.45

CHAPTER 2

LEGAL CONCEPTS AND LEGAL EXPERTISE

Experimental studies about ordinary concepts are often taken to have legal implications. For example, studies in cognitive science claim to illuminate the nature of legal concepts by studying ordinary people's judgments about terms like "intentionally," "caused," "consent," "reasonable," or "knowingly." However, the truth of this assumption is itself an empirical question: Are legal concepts identical to their ordinary counterparts?

This chapter evaluates this question with a case study. It tests attributions of what it means to act *intentionally*, across four populations (N = 1081): 322 ordinary people without legal training, 310 "elite-university" law students, 318 "elite-university" non-law students without legal training, and 131 United States judges. Participants considered versions of a scenario in which a person's action led to a harmful side effect. Although the person did not desire the harm, it was a foreseeable side-effect of the action. While participants without legal training tended to ascribe intentionality, those with legal training were less inclined to do so. There were two other ways in which the judgments of those with legal training differed. First, it was sensitive to whether the scenario was framed in a legal or ordinary context. Second, the judgments of those with legal training were not affected by a severity bias; the side-effect was not judged to be more intentional when the harm was more severe.

These experimental results suggest that legal training teaches a distinctive expert-legal concept. Legal concepts that correspond to ordinary terms (like "intentionally") are not always identical to the ordinary concept. As such, experimental findings about ordinary concepts do not necessarily carry straightforward legal implications. The acquisition of legal concepts represents a form of legal expertise.

CONTENTS

INTRODUCTION.....	95
I. ORDINARY CONCEPTS AND LEGAL CONCEPTS.....	99
<i>A. Concepts in Law</i>	100
<i>B. Conceptual Acquisition, Learning, and Expertise</i>	102
II. A CASE STUDY: THE CONCEPT OF INTENTIONAL ACTION.....	106
<i>A. Legal Theories of Intentional Action</i>	107
<i>B. Cognitive Science of Intentional Action</i>	111
<i>C. Empirical Questions</i>	114
III. EXPERIMENTAL STUDIES.....	115
<i>A. Experiment 1</i>	115
1. <i>Intentional Action Attributions Differ Among Judges, Law Students, Non-Law Students, and Ordinary People</i>	120
2. <i>Those With Legal Training Distinguish Between Ordinary and Legal Contexts</i>	125
3. <i>Those With Legal Training Do Not Show a Severity Bias</i>	126
4. <i>Summary</i>	128
<i>B. Experiment 2</i>	129
<i>C. Experimental Interpretation</i>	134
1. <i>Legal Experts Have a Distinctive Legal Concept of Intentional Action</i>	134
2. <i>Legal Experts Acquire a Distinctive Legal Concept Through Training</i>	136
3. <i>Legal Experts Retain the Ordinary Concept, with “Spillover”</i>	136
IV. IMPLICATIONS.....	138
<i>A. Ordinary and Legal Concepts</i>	139
<i>B. The Legal Concept(s) of Intentional Action</i>	141
<i>C. Legal Expertise</i>	146
<i>D. New Challenges: Interpretation, Judges Versus Juries</i>	147
CONCLUSION.....	148

INTRODUCTION

Law uses both distinctive legal concepts and ordinary ones. There is no ordinary concept of *parol evidence*, *disgorgement*, *Article III standing*, *voir dire*, or a *joint tenancy with right of survivorship*. These concepts are distinctive legal ones, acquired by legal training and experience. At the same time, law uses many ordinary concepts. When a constitution notes that a presidential term shall end on *the 20th day of January*, the ordinary concept of dates is relevant.

So, some of law's concepts are obviously unique and others are obviously ordinary. But many of the most important legal concepts have a more mysterious status. Often law evaluates questions like: did the defendant *intentionally* harm; did he *knowingly* harm; did his conduct *cause* the injury; did he fail to act with *reasonable* care? All of these legal concepts have an ordinary counterpart. In various non-legal contexts, we think about people's intentions, what they know and cause, and what seems reasonable. The existence of these ordinary and legal concepts arising from the same term (e.g. "intentionally") raises a simple but difficult question: Are corresponding ordinary and legal concepts identical or dissimilar?

As a normative matter—regarding whether these concepts *should* be identical or dissimilar to ordinary ones—both answers are plausible. There are some good reasons to think that legal concepts should be identical to ordinary ones, but there are also good reasons to think that legal concepts should reflect unique notions. In favor of identity, law that employs ordinary concepts is more accessible to ordinary people, promoting values like publicity and clarity. Legal concepts that are identical to ordinary ones also promote consistency: judges and juries faced with the same rules or standards apply the same (ordinary) concepts, facilitating the same legal decisions.

In favor of dissimilarity are considerations about the aptness and sophistication of law. "Relevant evidence" does not simply mean whatever ordinary language suggests; *relevant evidence* a special legal concept. The same is true of legal concepts that have an ordinary

counterpart like *intention*, *causation*, *knowledge*, and *reasonableness*. Law should apply these concepts *correctly*, not simply however ordinary people happen to (mis)understand them.

However, before turning to any normative inquiry—concerning whether legal concepts should be identical to or unique from ordinary ones—we must address the descriptive empirical question: Are there actually differences between ordinary and legal concepts?

Answering this question is important for addressing legal-philosophical issues, but it is also important in assessing the practical stakes and implications of experimental results. An increasing number of studies in cognitive science—including psychology, experimental philosophy, and experimental jurisprudence—attempt to make progress by studying *ordinary* judgments of legal concepts.¹ Although these studies have a number of aims, often the results are taken to suggest that experimental findings about ordinary people carry significant legal implications. For example, in a recent psychology study of the concept of intentional action, the authors interpret their results as threatening “large-scale inner-systemic incoherence” to “nearly every system of criminal law in the world.”² The fact that legal experts employ a flawed concept of intentional action might warrant such conclusions—if true. However, it remains an open question whether legal concepts are, in fact, identical to those ordinary concepts.

A natural way to assess such differences is to consider the judgments of legal experts. Do those with legal training (e.g. judges) apply ordinary concepts (e.g. the ordinary concept of intentional action) in their legal decision making? Or do they operate with distinctive legal concepts, ones that correspond to a term from ordinary language (e.g. “intentionality”), but depart from the ordinary concepts?

¹ See, e.g., Ivar R. Hannikainen & Raff Donelson, *Fuller and the Folk: The Inner Morality of Law Revisited*, in 3 OXFORD STUDIES IN EXPERIMENTAL PHILOSOPHY (forthcoming) (on the ordinary concept of law); Markus Kneer & Sacha Bourgeois-Gironde, *Mens rea Ascription, Expertise and Outcome Effects: Professional Judges Surveyed*, 169 COGNITION 139 (2017) (on the ordinary concept of intentional action); James Macleod, *Ordinary Causation*, 94 IND. L.J. (2019) (on the ordinary concept of causation); Christian Mott, *Statutes of Limitations and Personal Identity*, in 2 OXFORD STUDIES IN EXPERIMENTAL PHILOSOPHY 243 (2018) (on the ordinary concept of identity); Roseanna Sommers, *Commonsense Consent*, (draft manuscript) (on the ordinary concept of consent); Joshua Knobe & Scott Shapiro, *What Cognitive Science can Teach Us about Proximate Causation* (draft manuscript) (on the ordinary concept of causation); and Kevin P. Tobia, *How People Judge What is Reasonable*, 70 ALA. L. REV. 2915 (2018) (on the ordinary concept of reasonableness).

² Kneer & Bourgeois-Gironde, *supra* note 1.

This paper approaches this broad question through an experimental case study on the concept of acting *intentionally*. It studies a specific type of case, which I call “side-effect” cases. In these cases, an actor is indifferent towards a harmful side-effect of his action; he certainly does not desire that the side-effect occur, but its foreseeable likelihood does not halt his action. The experiments examine judgments about whether in such a case the side-effect was produced *intentionally*. The studies examine four populations that vary in their legal training: United States judges, “elite” law students (e.g. Yale and Harvard law students), “elite” non-law students (Yale non-law students), and ordinary people without legal training.

The results suggest that, at least in the case of intentional action, legal concepts can depart dramatically from their ordinary counterparts. When considering whether an action *intentionally* harmed, where the harm was a foreseeable but motivationally insignificant side-effect, attributions of intentional action decreased markedly with legal training. While those without legal training generally assessed the motivationally insignificant side-effect as intentional, this judgment decreases incrementally with legal training.

The experimental results support an explanation of the effect grounded in legal training, rather than in selection. That is, the pattern of results is more consistent with an explanation involving legal experts’ acquisition of a distinctive legal concept of intentional action as a result of their education and training—and not with an explanation pointing to other differences (e.g. intelligence) between legal experts and non-experts.

These findings carry both theoretical and practical implications. They bear on the opening question concerning ordinary and legal concepts, as well as the nature of legal expertise. They also raise a puzzle for ordinary meaning theories of legal interpretation: Which concept should such theories endorse as the “ordinary meaning” of a term in a legal context: the concept common to ordinary people, or the concept common to those with legal training? A similar puzzle arises for legal practice. If judges (with legal training) and juries (without) associate the same term with

two different concepts, this plausibly leads to divergent outcomes based on judge versus jury selection. How should law respond to such inconsistency?

Part I begins by elaborating the primary question concerning the relationship between ordinary concepts and legal concepts. It also provides some background in the cognitive science of concept acquisition and change.

Part II turns to the case study in intentional action. Section II.A surveys views from legal theory and practice about the legal concept of *intentional action*. An influential view in theory and practice is that the core of intentional action involves *motivational significance*. For an outcome (e.g. a side-effect) to be produced intentionally, it must play some role in motivating the causal action. On this account, if an action has foreseeable side-effect that is not motivationally significant, that side-effect would not be understood as one produced intentionally.

This legal account of intentional action differs from the ordinary concept. Section II.B presents evidence from cognitive science about the ordinary concept of intentional action—evidence that suggests distinctive features of the ordinary concept. In particular, ordinary people are willing to attribute intentionality to a foreseen but motivationally insignificant side effect when that effect represents a *negative* or *bad* side-effect. Section II.C provides a list of empirical questions and hypotheses concerning the relationship between the ordinary and legal concept of intentional action.

Part III presents experimental studies. The results indicate that intentional action attributions differ among judges, law students, non-law students, and ordinary people. They also show that only those with legal training (i.e. judges and law students) distinguish between ordinary and legal contexts. Finally, it shows that those with legal training do not show a severity bias. At the same time, the results suggest some evidence of conceptual “spillover” from legal into ordinary contexts. When legal experts evaluate ordinary cases of intentional action, their judgments differ from those of ordinary people.

Part IV considers various implications for debates about ordinary and legal concepts, legal expertise, interpretation, and judges and juries.

I. ORDINARY CONCEPTS AND LEGAL CONCEPTS

This Part begins with an overview of the essay's guiding question: What is the relationship between ordinary and legal concepts? Section I.A elaborates this question and two responses. The legal concept is either (i) *identical* to the ordinary concept or (ii) *dissimilar* to the ordinary concept in some ways.

It also elaborates two hypotheses about how why legal experts and non experts might make different judgments about the same concept. A difference might be explained by something about the *kinds of people* who enter the legal profession (Selection Hypothesis), or by some aspect of *legal training* (Legal Training Hypothesis). These views make divergent predictions. For example, the Selection Hypothesis predicts that differences between experts and non-experts are driven by something special about the legal experts (e.g. intelligence). That view would expect that non-experts who are similar to legal experts in the relevant way (e.g. similar intelligence) should then have similar judgments. On the other hand, the Legal Training Hypothesis predicts that a distinctive legal concept is produced through legal training and experience. That view would predict that non-experts who are similar to legal experts would not make similar judgments, since they do not have the legal concept.

Section II.B provides background in the cognitive science of concept acquisition, learning, and expertise. To explore the possibility of legal concept acquisition, it explores an analogy with a different domain: the acquisition of scientific concepts.

A. Concepts in Law

What is the relationship between an ordinary and legal concept that correspond to the same common term?³ For example, what is the relationship between the ordinary concept of intentional action and the legal concept of intentional action; or between the ordinary concept of reasonableness and the legal concept of reasonableness; or between the ordinary concept of causation and the legal concept of causation? Consider two possibilities:

Identity Theory: The legal concept is *identical* to the ordinary concept.

Dissimilarity Theory: The legal concept is similar to the ordinary concept in some ways, but is distinctive in other ways.

These theories can be approached as descriptive hypotheses or normative views. For example, we might ask as a descriptive matter whether the legal concept of what is reasonable is identical to the ordinary concept. But whether the legal concept *should* reflect the ordinary concept is a different question.

Various experimental findings about concepts might support different normative arguments. For example, perhaps we find that the legal concept is dissimilar from the ordinary concept. We might argue that the legal concept *should* be reformed to match the ordinary concept.⁴ Alternatively, perhaps we find that the legal concept is identical to the ordinary concept. We might argue that this indicates an inappropriate influence of ordinary cognition on legal judgment. The legal concept should reflect certain legal aims, but instead it reflects ordinary judgment.

Before we turn to these normative questions, it is important to understand the descriptive facts: are legal and ordinary concepts identical or dissimilar? Empirically, the case for identity might be strengthened by demonstrating that in a range of cases, ordinary people and experts

³ There are some kinds of legal concepts that fall outside of the bounds of this discussion. One obvious group consists of concepts that have no ordinary language counterpart: writ of assistance, parole evidence, *res judicata*. Although there are interesting questions about the legal cognition underlying the application of those concepts, it seems likely that little progress would be made by studying ordinary cognition (e.g. lay judgments of *res judicata*). A second group consists of legal concepts that have an ordinary language counterpart, but have an explicit legal definition. The paradigm case is statutorily defined terms. Unless otherwise noted, I bracket these two groups of concepts. The inquiry here concerns concepts that are not otherwise defined or specialized.

⁴ This could occur for reasons specific to the concept. For example, perhaps certain features of the ordinary concept of reasonableness are particularly suited to meeting the aims of legal reasonableness standards. Or it could occur for more general reasons. For instance, if the legal concept corresponded to the ordinary concept, the law might better meet the aims of publicity and clarity.

make the same judgments (or, apply the concept in the same way). The case for dissimilarity might be strengthened by demonstrating the opposite: in a range of cases, ordinary people and experts make different judgments.

If we find support for identity, there may not be much additional mystery about how that legal reality obtains. There are many intriguing questions about how we acquire ordinary concepts, but on the identity account we simply use that ordinary concept in legal decisions. There is nothing special or additional to explain about legal cognition.

However, there are further questions that we might ask if the pattern of judgments differ between ordinary people and legal experts. In particular, we might wonder whether this difference reflects that experts have a special legal concept or whether it reflects that experts are simply better at using the ordinary concept.

Consider two hypotheses:

The Selection Hypothesis: A difference between ordinary and expert judgments corresponding to the same term is explained by factors about the *kinds of people* who enter the legal profession (e.g. the kinds of people who attend law school and become lawyers or judges).

The Legal Training Hypothesis: A difference between the ordinary and expert judgments corresponding to the same term is explained by some aspect of *legal training or experience*.

These two hypotheses posit very different explanations for a divergent pattern of judgments between ordinary people and legal experts.

As an example, imagine that judges make judgments about causation that are different from the judgments made by ordinary people. The Selection account of this difference posits as an explanation something about *the people* who become judges. For example, perhaps people who become judges are naturally more intelligent. Perhaps judges' natural intelligence makes them more reliable users of the concept of causation. So the differences in judgments that we see between judges and ordinary people are a result of ordinary people's performance errors and judges' conceptual expertise (in using the ordinary concept).

Conversely, the Legal Training account of this difference begins by pointing to something that the judges *learn* through legal training, experience, or acculturation. For example, perhaps the difference stems from the fact that judges have three years of legal education, which equips them with a distinctive legal concept of causation. This story posits a very different form of legal expertise. The differences are not a product of judges' innate intelligence, but rather a product of their acquisition of a different concept from the ordinary one. To further elaborate this possibility, the next section presents background on the process of conceptual acquisition, learning, and expertise.

B. Conceptual Acquisition, Learning, and Expertise

There are a number of important empirical questions concerning legal decision making. One set of questions concerns whether ordinary biases affect legal decision making. For example, do judges manifest anchoring biases when they award damages,⁵ and are expert legal-interpretive judgments biased by political affiliation and politically motivated reasoning?⁶ Experimentalists have studied these questions—and for good reason. If political bias or motivated reasoning unduly influenced legal judgments, these judgments would be uncontroversially in need of correction.

A very different set of questions concerns the nature of legal concepts themselves. What is the “legal concept” of causation, intentional action, or reasonableness? Is a legal concept (that corresponds to an ordinary language term) identical or non-identical to the ordinary concept?

To help provide a theory of legal concepts, consider an analogy to a different domain: scientific concepts and learning. A number of important psychological studies have examined the way in which ordinary people and scientific experts acquire and refine their scientific concepts. Although the acquisition of legal concepts may ultimately differ from that of scientific concepts,

⁵ See generally Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, *Blinking on the Bench: How Judges Decide Cases*, 93 CORNELL L. REV. 1, 19-27 (2007).

⁶ See generally Dan Kahan et al., “*Ideology*” or “*Situation Sense*”? An Experimental Investigation of Motivated Reasoning and Professional Judgment, 164 U. PA. L. REV. 349 (2016).

this literature provides a useful starting point in theorizing the nature of legal concepts and expertise.

Science concepts present a familiar and useful example of *conceptual change*.⁷ Susan Carey has documented the way in which development and learning contribute to changes in scientific concepts.⁸ For example, infants and preschoolers have the concept of an *animal*. But research suggests that this concept is embedded in a very different framework from that of ten-year olds. According to Carey, infants conceptualize *animal* in a framework of behaving beings (such as people). But by age ten, children have changed their concept of animal, embedding it within a framework of living things (including plants); *animal* is a subcategory of living things.

Whether or not Carey is right about the concept of an animal, there are undoubtedly many scientific concepts that change over time. There are at least two notable types of change.⁹ One we can call “restructuring,” which reflects changes in the relationship among extant concepts. And one we can call “core changes,” which reflect changes in the features of a particular concept.

The seminal experimental study of scientific expertise and “restructuring” conceptual change was led by Michelene Chi.¹⁰ She asked lay people and physics experts to categorize a set of physics problems. Lay people categorized similar problems on the basis of their superficial features, for example drawing parallels between problems that involved something rotating, or something circular. Experts, on the other hand, categorized problems on the basis of deeper structural features, for example drawing parallels between problems that involved conservation of energy or Newton’s Second Law.

⁷ There is a large philosophical literature that addresses whether conceptual learning is possible. Compare Susan Carey, *Precis of The Origin of Concepts*, 34 BEH. & BRAIN SCI. 113 (2011); Eric Margolis, *How to Acquire a Concept*, 13 MIND & LANGUAGE 347 (1998); Eric Margolis & Stephen Laurence, *Learning Matters: The Role of Learning in Concept Acquisition*, 26 MIND & LANGUAGE 507 (2011); Dylan Sabo, *Where Concepts Come From: Learning Concepts by Description and by Demonstration*, 79 ERKENNTIS 531 (2014); M.J. Cain, *Learning, Concept Acquisition and Psychological Essentialism*, 4 REV. PHIL. PSYCH. 577 (2013); with Jerry Fodor, LOT 2 (2008); Jerry Fodor, THE LANGUAGE OF THOUGHT (1975).

⁸ Susan Carey, *Science Education as Conceptual Change*, 21 J. APPLIED DEVELOPMENTAL PSYCHOLOGY 13 (2000).

⁹ Susan Carey, *Cognitive Science and Science Education*, 41 AM. PSYCHOLOGIST 1123 (1986).

¹⁰ Michelene T.H. Chi, Paul J. Feltovich & Robert Glaser, *Categorization and Representation of Physics Problems by Experts and Novices*, 5 COG. SCI. 121 (1981).

We could imagine something similar characterizing the judgments of ordinary people and legal experts. Where ordinary people might draw connections between different legal cases of *hurting* other people, legal experts might instead distinguish between harms that give rise to tort causes of action and harms that give rise to criminal causes of action. Where ordinary people might group many different types of *discrimination*, legal experts might distinguish between private and public cases (e.g. a putatively discriminatory policy in a public versus private school).

The literature on scientific concepts also highlights changes in core concepts. For example, Carey's research suggests that very young children have one concept of *not alive*, while older children have two distinct concepts: *dead* and *inanimate*.¹¹ Young children might apply the one concept to both a dead dog and a rock, while older children would distinguish between the two. Again, whether or not this specific claim of Carey's is right, the possibility of "strong conceptual differences" represents a worthwhile exploration for legal theory. Do legal experts have different concepts from non-experts?

Yes. Those without legal training do not have the legal concepts of *res judicata*, *disgorgement*, *parol evidence*, or *standing* (in the sense relevant to civil procedure). Legal training teaches these and many other distinctive concepts. However, the more interesting question concerns legal concepts that have an ordinary-language counterpart: *consented*, *caused*, *acted intentionally*, *knowingly*, *reasonable*, and so on. Are these identical to ordinary concepts or has legal training "strongly changed" these concepts?

It is worth noting that both types of change—restructuring and "strong" changes to core concepts—might occur within an individual or across a population over time. For example, Carey suggests a way in which children restructure the concept of an *animal* and learn the concept of *dead*. But societies also restructure and change concepts. The discovery of Newtonian physics or heliocentrism represent changes in population-level structuring; the discovery of new planets or

¹¹ Susan Carey, *Science Education as Conceptual Change*, 21 J. APPLIED DEVELOPMENTAL PSYCHOLOGY 13 (2000).

redefinition/exclusion of existing planets represent changes in concepts between generations. Similar changes likely also occur in law. Incorporation and reverse incorporation of constitutional rights changes the way in which a new generation of people conceptualizes those rights.

This essay cannot explore all of these possibilities of legal conceptual change. It is primarily focused on the possibility of this second type of “strong” change to core concepts, in the context of legal concepts that have an ordinary language counterpart, within individuals (i.e. not across generations). It studies one example, *intentional action*, as an initial test case and proof of concept that legal concepts may differ from ordinary ones.

Compared to the extant work concerning bias in legal decision making, the literature concerning legal concept acquisition, learning, and expertise is relatively thin.¹² But it represents an equally important area of inquiry. In the same way that a legal concept *should not* reflect bias or obviously irrelevant factors, a legal concept *should* reflect either an ordinary notion or a specialized legal one.

However, there is a key difference between these two research programs. In the study of legal biases it is often perfectly clear how experimental findings should be understood in light of legal theory. If for example, experimentalists find that judges’ sentencing determinations are influenced by the race of defendants, it is obvious that this is not a relevant aspect to sentencing, and we should attempt to improve these decision processes. If, on the other hand, experimentalists report that judges’ notion of causation is identical to the ordinary one, the legal-philosophical implications are not immediately clear. To know whether this legal concept should be reformed requires much more reflection on the nature of the ordinary concept and its relationship to law and law’s aims. Of course, just because these questions about concepts represent harder legal-philosophical problems does not mean that they are unimportant.

¹² For some examples, see Leah M. Christensen, *The Paradox of Legal Expertise: A Study of Experts and Novices Reading the Law* 2008 BYU EDUC. & L.J. 53 (2008); Fleurie Nievelstein et al., *Expertise-related Differences in Conceptual and Ontological Knowledge in the Legal Domain* 20 EUR. J. COGNITIVE PSYCHOL. 1043 (2008).

II. A CASE STUDY: THE CONCEPT OF INTENTIONAL ACTION

This essay's question is broad: what is the relationship between ordinary and legal concepts? To make progress on this question requires studying specific examples. With case studies on various concepts (e.g. the concept of *intentional action*, *causation*, *consent*, *reasonableness*, *knowledge*), we can begin to address the broader question about the relationship between (some, most, or all) ordinary and legal concepts.

This Part takes up one such case study. It considers the concept of intentional action. This is a useful case study for several reasons. For one, it is a clear example of a legal concept that corresponds to an ordinary-language term, and some have raised the question of whether the legal concept is an identical, unique, or similar one.¹³

Second, there is a rich empirical literature on interesting and surprising features of the ordinary concept.¹⁴ If the legal concept shares those features, this is some evidence for identity; if it lacks those features, this is some evidence for partial similarity.

Finally, a recent empirical study has created materials to study intentional action that are well-suited to legal questions.¹⁵ That study was conducted without knowledge of this essay or its predictions. Thus, those materials provide ideal test cases, ones that minimize researcher degrees of freedom.

Before turning to the experimental investigation of intentional action, Section II.A provides background on legal theories of intentional action. Section II.B turns to recent work in the cognitive science of intentional action, outlining some of the distinctive features of the ordinary concept. Section II.C concludes by articulating several open empirical questions and the essay's experimental hypotheses.

¹³ E.g. Kimberly Ferzan, *Beyond Intention*, 29 *CARD. L. REV.* 1147 (2008).

¹⁴ E.g. Joshua Knobe, *Intentional Action and Side Effects in Ordinary Language*, 63 *ANALYSIS* 190 (2003).

¹⁵ Markus Kneer & Sacha Bourgeois-Gironde, *supra* note 1.

A. Legal Theories of Intentional Action

This essay focuses on attributions of intentional action in a specific type of case, which we can call “side-effect cases.” This part outlines the legal concept of intentional action, with special focus on such side-effect cases.

When thinking about whether someone *intended* an action’s side-effect, there are a number of considerations that might seem relevant. Broadly speaking, one relevant consideration is whether the person’s action was encouraged by their want or desire for the side-effect to occur. If this condition obtains, the side-effect has “motivational significance” to the actor. Another relevant consideration is the degree of the side-effect’s foreseeability. If the side-effect is completely surprising, it would be not foreseeable. As we the actor’s awareness of the side-effect’s likelihood increases, it eventually reaches the level of being “foreseeable.” And if the actor knew that the effect was practically certain to occur, it could reach the level of being “substantially certain.”

Although there are numerous competing legal theories of intentional action, one influential theory holds that intended outcomes must have motivational significance.¹⁶ If I shoot someone with a BB gun, I have intentionally killed him only if *killing him* was motivationally significant. If I am in no way motivated to kill him, but I shoot because I intended to scare—but not kill—him, many views hold that I did not *intentionally* kill him. I *intentionally shot him* and perhaps *recklessly killed him*, but I did not intentionally kill.

This central view can be found across popular views of intentional action. For example, Duff famously characterized intentions in terms of the test of failure: If the result does not occur, will the action have failed?¹⁷ Another popular view is that intentional acts are ends or means: “things

¹⁶ Ferzan, *supra* note 11, at 1149 n.10 (citing R.A. DUFF, INTENTION, AGENCY AND CRIMINAL LIABILITY: PHILOSOPHY OF ACTION AND THE CRIMINAL LAW 58 (1990); MICHAEL E. BRATMAN, INTENTION, PLANS, AND PRACTICAL REASON 140 (1999); Anthony Kenny, *Intention and Purpose in Law*, in ESSAYS IN LEGAL PHILOSOPHY 146, 148 (Robert S. Summers ed., 2d ed. 1976); John Finnis, *Intention and Side-Effects*, in LIABILITY AND RESPONSIBILITY: ESSAYS IN LAW AND MORALS 32, 36 (R.G. Frey & Christopher W. Morris eds., 1991).

¹⁷ R.A. DUFF, INTENTION, AGENCY AND CRIMINAL LIABILITY: PHILOSOPHY OF ACTION AND THE CRIMINAL LAW 58, 61 (1990).

done as means or ends are intended; those done as side-effects are not.”¹⁸ In a seminal work on legal intention, Kimberly Ferzan argues that each of these different tests “boils down to the same requirement of motivational significance.”¹⁹

Although there are important differences between criminal law and tort law treatments of intentional action, this feature is shared: motivational significance represents a core case of intentional action.

First take the criminal context. In the Model Penal Code “intentionally” or “with intent” means “purposely.”²⁰ Purposeful *mens rea* is elaborated by citing a person’s *conscious object*: “A person acts purposely with respect to a material element of an offense when: (i) if the element involves the nature of his conduct or a result thereof, it is his conscious object to engage in conduct of that nature or cause such a result.”²¹ This is commonly understood as requiring something above mere knowledge of the result; to have the conscious object to cause such a result requires something like the result’s motivational significance.²²

Harder cases are ones involving certain or “inseparable effects.”²³ In circumstances in which an action will clearly lead to one effect that is motivational and also one that is not, some might find it tempting to call the second effect “intentional,” in an ordinary language sense. In general, the law does not concede to this temptation. Outcomes that are intentionally produced are distinguished from actions that are knowingly produced. And in circumstances in which law seeks to impart liability on knowingly produced outcomes, it does so by specifically defining intent as such.

¹⁸ A.P. Simester, *Moral Certainty and the Boundaries of Intention*, 16 OXFORD J. LEGAL STUD. 445, 446 (1996).

¹⁹ Ferzan, *supra*, note 11, at 1154.

²⁰ Model Penal Code 1.13 General Definitions.

²¹ Model Penal Code 2.02 (2)(a).

²² See also Ferzan, *supra*, note 11.

²³ *Id.*

In tort law, “intent to harm” is often defined as a purpose or substantial certainty of producing that harm.²⁴ The Second Restatement reflects this blended definition. “All consequences which the actor desires to bring about are intended... Intent is not, however, limited to consequences that are desired. If the actor knows that the consequences are certain, or substantially certain, to result from his act, and still goes ahead, he is treated by the law as if he had in fact desired to produce the result.”²⁵

The Third Restatement separates this blended definition into a dual definition: “A person acts with the intent to produce a consequence if: (a) the person acts with the purpose of producing that consequence; or (b) the person acts knowing that the consequence is substantially certain to result.”²⁶

The nature of “knowledge of a substantial certainty” is helpfully contrasted with recklessness. The Second Restatement gives an illustration. Imagine that A throws a bomb into B’s office for the purpose of killing B. A has no desire to kill C, but knows that C is in the office. A knows that C’s injury is substantially certain to result.²⁷

The substantial-certainty doctrine is typically construed narrowly, in at least three ways. First, it is limited to acts that an actor knows involve *substantial certainty*. “It is not enough to make the act intentional that the actor realize that it involves any degree of probability of a harmful or offensive contact.”²⁸ In fact, most understand the doctrine to require a level of certainty far beyond mere foreseeability and even beyond high probability.²⁹

Second, it requires *knowledge* of the substantial certainty; certainty without subjective awareness at the level of knowledge is insufficient. Coupling this with the first limitation

²⁴ RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL HARM § 1; *see also* Ellen M. Bublick, *A Restatement (Third) of Torts: Liability for Intentional Harm to Persons—Thoughts* 44 WAKE FOREST L. REV. 1335 (2009).

²⁵ RESTATEMENT (SECOND) OF TORTS: INTENT § 8A.

²⁶ RESTATEMENT (THIRD) OF TORTS: INTENT § 1.

²⁷ *Id.*

²⁸ RESTATEMENT (SECOND) OF TORTS § 18.

²⁹ *See generally* WILLIAM P. STATKSY, ESSENTIALS OF TORTS 19.

produces a stringent standard. Not only must there be awareness of a substantial certainty, there must be *knowledge* of that certainty.

Finally, the doctrine is typically limited to cases in which the defendant has knowledge that the conduct will bring about harm to a *particular victim* within a *particular location* (e.g. an employer's job site).³⁰ A paradigmatic case in which this criterion is satisfied is *Garratt v. Dailey*.³¹ In that case, a five-year old defendant withdrew a chair as his aunt began to sit down, and she fell and sustained an injury. While the child had no desire or purpose to cause injury, he was found to have knowledge of a substantial certainty of the result.

Thus, although the substantial certainty doctrine presents a second definition of intent, "in the vast majority of actual battery cases, the plaintiff in fact proves the defendant's intent by showing the defendant's purpose to cause harm."³²

The preceding analysis does not explore all of the differences in the concept of intentional action between criminal and tort law. For example, the Model Penal Code speaks of consequences that are "practically certain" to occur, rather than ones that are "substantially certain to occur. However, such differences may be of little practical import; "for the purposes of instructing juries, courts can take into account the interchangeability of these various phrases."³³

The core claim of this section is that, across criminal and tort law, there is a central psychological conception of intentional action that requires that the intended outcome is motivationally significant to the actor. In both areas, there are narrow doctrines that allow attributions to be extended to cases of "inseparable effects," or ones that are known to affect a

³⁰ "The applications of the substantial-certainty test should be limited to situations in which the defendant has knowledge to a substantial certainty that the conduct will bring about harm to a particular victim, or to someone with a small class of potential victims within a localized area. The test loses its persuasiveness when the identity of potential victims becomes vaguer and when...the causal sequence connecting conduct and harm becomes more complex." For example, "in many situations a defendant's knowledge of substantially certain harms is entirely consistent with the absence of any liability in tort. For example, an owner of land, arranging for the construction of a high-rise building, can confidently predict that some number of workers will be seriously injured in the course of the construction project... Despite their knowledge, these actors do not intentionally cause the injuries that result."

³¹ 279 P.2d 1091 (Wash. 1955).

³² RESTATEMENT (THIRD) OF TORTS §1.

³³ RESTATEMENT (THIRD) OF TORTS §1.

particular victim with a substantial certainty. But these exceptions aside, intentionality is tightly connected to motivational significance.

If this reflects the core legal concept of intentional action, there is a natural question about the relationship between this concept and the ordinary concept: are they identical or dissimilar?³⁴ As we will see in the next section, work in experimental cognitive science has shown that the ordinary concept of intentional action departs significantly from this motivational significance account. In particular, ordinary people often attribute intentionality to side-effects that are merely foreseen (and motivationally insignificant). If an actor produces a foreseeable (but not substantially certain) bad side-effect that is not motivationally significant, this does not fit the legal schema of intentional action. However, ordinary people would often be inclined to evaluate this bad side-effect as one that was produced intentionally.

B. Cognitive Science of Intentional Action

A well-documented feature of ordinary judgments of intentional action is that those judgments are sensitive to outcome valence. The most famous case concerns an asymmetry between good and bad side-effects.³⁵ An experiment presented participants with two different stories about a business chairman who seeks only to make profits. In the “harm” case, participants consider a chairman who aims to increase profits—a strategy that will foreseeably *harm* the environment. In the “help” case, participants consider an identical chairman, but one whose profit-maximizing strategy will foreseeably *help* the environment. The vast majority of ordinary people judge that the first chairman intentionally harmed the environment but the second did not intentionally help the environment.

³⁴ As Gideon Yaffe puts it: “[I]t is a hypothesis worth exploring that all of the various senses of the legal term “intention” circle around a core notion that has its natural home in ordinary discourse. The law is better when it uses ordinary terms in ordinary ways and employs concepts that bear a close resemblance to those used in everyday life.” Gideon Yaffe, *Criminal Attempts*, 124 *YALE L.J.* 92, 107 (2014).

³⁵ Joshua Knobe, *Intentional Action and Side Effects in Ordinary Language*, 63 *ANALYSIS* 190 (2003).

This effect, known as the Knobe effect or side-effect effect, has been replicated numerous times,³⁶ across various stages of cognitive development,³⁷ and across different cultures.³⁸ There is widespread consensus in the philosophical and psychological literature that this asymmetry is a feature of ordinary judgments of intentional action.

There is, however, significant philosophical debate concerning the normative status of the side-effect effect. Some argue that the effect is a bias,³⁹ while others contend that it reflects ordinary people's conceptual competence.⁴⁰ This essay assumes that the effect reflects a feature of the ordinary concept.

Scholars frequently suggest that this experimental finding raises a problem for law.⁴¹ Broadly speaking, the challenge argues that (i) the ordinary concept of intentional action is sensitive to outcome valence and (ii) legal judgments should not employ a concept of intentional action that has this feature, therefore (iii) legal judgments employ an inappropriate concept.

However, it is not clear how widely the positive-negative asymmetry would manifest in legal practice. Although judges and juries frequently determine intentionality with respect to a negative outcome, it is a less usual legal fact pattern that would require determining whether a positive side-effect was intentional.

³⁶ Joshua Knobe, *Intentional Action and Side Effects in Folk Psychology: An Experimental Investigation*, 16 PHIL. PSYCH. 309 (2003); Joshua Knobe, *Intention, Intentional Action and Moral Considerations*, 64 ANALYSIS 181 (2004); Al Mele & Fiery Cushman, *Intentional Action, Folk Judgments, and Stories: Sorting Things Out*, 31 MIDWEST STUD. PHIL. 184 (2007).

³⁷ Alan M. Leslie, Joshua Knobe, J. & A. Cohen, *Acting Intentionally and the Side-effect Effect Theory of Mind and Moral Judgment*, 17 PSYCH. SCI. 421 (2006).

³⁸ Nikolaus Dalbauer & Andreas Hergovich, *Is What is Morose More Likely?—The Probabilistic Explanation of the Epistemic Side-effect Effect*, 4 REV. PHIL. & PSYCH. 639 (2013); Joshua Knobe & Arudra Burra, A., *The Folk Concepts of Intention and Intentional Action: A Cross-cultural Study*, 6 J. COGNITION & CULTURE, 113 (2006).

³⁹ E.g. Fred Adams & Annie Steadman, *Intentional Action in Ordinary Language: Core Concept or Pragmatic Understanding?*, 64 ANALYSIS 173 (2004); Mark Alicke, *Blaming Badly*, 8 J. COGNITION & CULTURE 179 (2008); Mark Alicke & David Rose, *Culpable control or moral concepts?*, 33 BEHAV. & BRAIN SCI. 330 (2010); James R. Beebe, *A Knobe Effect for Belief Ascriptions*, 4 REV. PHIL. & PSYCH. 235 (2008); Thomas Nadelhoffer, *Blame, Badness, and Intentional Action: A Reply to Knobe and Mendlow*, 24 J. THEORETICAL & PHIL. PSYCH. 259 (2004); Thomas Nadelhoffer, *On Praise, Side Effects, and Folk Ascriptions of Intentionality*, 24 J. THEORETICAL & PHIL. PSYCH., 196 (2004); Hanno Saucier, & Tom Bates, *Chairmen, Cocaine, and Car Crashes: The Knobe Effect as an Attribution Error*, 17 J. ETHICS 30 (2013).

⁴⁰ E.g. Frank Hindriks, *Intentional Action and the Praise-blame Asymmetry*, 58 PHIL. Q. 630 (2008); Joshua Knobe, *Person as Scientist, Person as Moralist*, 33 BEHAV. & BRAIN SCI., 315 (2010); Edouard Machery, *The Folk Concept of Intentional Action: Philosophical and Experimental Issues*, 23 MIND & LANG. 165 (2008); Dean Pettit & Joshua Knobe, *The Pervasive Impact of Moral Judgment*, 24 MIND & LANG 586 (2009); Kevin Uttich & Tania Lombrozo, *Norms Inform Mental State Ascriptions: A Rational Explanation for the Side-effect Effect*, 116 COGNITION 87 (2010).

⁴¹ Matthew Ginther et al., *Decoding Guilty Minds: How Jurors Attribute Knowledge and Guilt*, 71 VAND. L. REV. 241 (2018); Markus Kneer & Sacha Bourgeois-Gironde, *Mens rea Ascription, Expertise and Outcome Effects: Professional Judges Surveyed*, COGNITION 169 (2017); Steven R. Morrison, *Defending Vicarious Felony Murder*, 47 TEX. TECH. L. REV. 129 (2014); Thomas Nadelhoffer, *Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Juror Impartiality*, 9 PHIL. EXPLORATIONS 203 (2006).

A more common legal fact pattern involves assessing the intentionality of a negative/bad outcome, particularly in tort and criminal law. Interestingly, research in cognitive science suggests that a similar “side-effect” effect might arise in these contexts. A recent study reports finding such *severity-sensitivity*. Attributions of intentionality increase when a side-effect is seen as *more* negative. Most strikingly, the study reports that this effect characterizes the judgments of lay persons *and* legal experts.

Kneer & Bourgeois-Gironde conducted two experiments on professional judges and found evidence of outcome severity sensitivity.⁴² Their first study (N=36) replicated the classic positive-negative side-effect effect for the chairman cases. Their second study (N=32) found that judges’ intentional action attributions are sensitive to the severity of a negative outcome.

They presented judges with two versions of a “Beach Town” scenario. One version had a moderately bad side-effect and the other version had a severely bad side-effect:

The mayor of a small beach town is approached by his advisor who says: “We could build a new highway connection. This would make car traffic much more efficient. However, there would be [minor/ severe] adverse effects on the environment. During construction, the animals in the construction zone will [be disturbed/die]. This is [only temporary/not a temporary condition], [everything goes/ things will not go] back to normal once construction is finished.” The mayor responds: “I don’t care at all about the environment. All I care about is making car traffic as efficient as possible. Let’s build the new highway connection.”

They build the new highway connection. The animals in the zone are [temporarily disturbed/die]. [Everything goes/Things do not go] back to normal after construction is finished.

Judges agreed more strongly that the mayor intentionally harmed the environment in the case in which the outcome was severe.⁴³

This is an intriguing finding, but it leaves open a number of questions. For one, is there evidence of this similarity in the ordinary and legal concept across different levels of expertise and in different scenarios? Relatedly, does the severity-sensitivity arise across these levels? Finally, the experimentalists study legal experts’ judgments of intentional action in *an ordinary*

⁴² Kneer & Bourgeois-Gironde, *supra* note 1.

⁴³ M = 5.18 vs. M = 3.33, on a 7-point scale.

context. But might legal experts have a special concept of intentionality that they use specifically in legal decision making?

C. Empirical Questions

Previous work in law and psychology suggests a number of open empirical questions. Most significantly, is the legal concept of intentional action identical to the ordinary concept?

A second question concerns the role of legal context and domain-specific expertise. Do judges' (and law students') attributions of intentional action depend on the legal or non-legal context? To illustrate this possibility of domain-specific expertise and context, recall the analogy to scientific concepts. An electrical engineer might use a very technical concept of *electrocution* at work. Imagine that she returns home and hears her husband warning their child about the toaster: "You just put a fork in there! Did you electrocute yourself?" The child replies, "Yes, it hurt a lot!" Scientifically speaking, this might be a nonsensical exchange: Electrocution is *killing* by electricity. But the ordinary concept of electrocution includes non-lethal shocks. The engineer could understand the ordinary concept operating in this exchange. She has two concepts of electrocution, which she deploys in different scientific or non-scientific contexts.

The same might be true of legal concepts. In a legal context, an expert might understand *intentionally* harming to mean one thing, while in a non-legal context, the legal expert understands it to mean something different.

Although the legal versus non-legal boundary may sometimes be difficult to demarcate, there are some clear examples. First consider some clear legal contexts: A judge in his chambers writing an opinion concerning an intentional infliction of emotional distress; a jury in the jury deliberation room deciding whether a defendant is liable for intentionally injuring the plaintiff's dog. There are also clear non-legal "ordinary" contexts: a teacher decides whether their student intentionally helped others by refraining from taking all of the candy.

A final question concerns severity sensitivity: That is, are the intentionality judgments of those with legal training affected by the degree of the outcome's severity?

III. EXPERIMENTAL STUDIES

The present studies investigate the concept(s) of intentional action among four populations that vary in legal training: a general population sample (MTurk), law students at "elite" universities (e.g. Yale Law), non-law students at an "elite" university (Yale), and U.S. judges.

A. Experiment 1

The first study had three primary aims. First, it sought to test whether attributions of intentional action differed among these populations. Second, the study aimed to test whether attributions of intentional action varied in a legal versus non-legal ("ordinary") contexts. Finally, it sought to test whether judgments of intentionality were affected by the severity of the outcome.

To minimize researcher degrees of freedom, participants were presented with versions of Kneer & Bourgeois-Gironde's "Beach Town" cases.⁴⁴ Those cases are also ones used previously to show that judges (legal experts) make outcome severity sensitive judgments in a non-legal context, supporting an argument that the legal concept of intentional action is similar to the ordinary concept.⁴⁵ As such, these examples represent not only an unbiased type of case (ones developed without knowledge of the present hypothesis), but also a case that is unfriendly to the present hypothesis (one developed with the expectation that legal experts' judgments of intentional action are like ordinary people's, sensitive to outcome severity).

Method

Approximately 1500 professional judges were contacted by email to request voluntary participation in the study. Law students were recruited by emails to administrators and/or student

⁴⁴ *Supra* note 1.

⁴⁵ *Id.*

list-serv emails at fourteen elite law schools. Elite non-law students were recruited by contacting student emails, networks, and list-servs at Yale University.

One hundred and thirty-one United States judges,⁴⁶ 237 law students,⁴⁷ 195 non-law students,⁴⁸ and 211 participants from Amazon's Mechanical Turk⁴⁹ participated in an online experiment.

Judges were recruited from state and federal courts and asked to categorize their and years of experience (e.g. less than 1 year, 1-5 years, 6-10 years, etc.). One-hundred and thirteen judges reported their years of judging experience. Of those, .88% reported less than one year of experience, 15.0% reported 1-5 years of experience, 20.4% reported 6-10 years, 22.1% reported 11-15 years, 16.8% reported 16-20 years, 14.2% reported 21-25 years, and 14.2% reported over 26 years.

Law student participants were from Yale Law School (84.9%), Harvard Law School (7.4%), and Columbia Law School (7.7%), or another law school (7.7%, e.g. Berkley, NYU, Georgetown Law Schools). Most were well training in law, with 11.5% reporting less than one year of legal training, 25.7% reporting one year of legal training, 21.1% reporting two years, 38.3% reporting three years, 1.7% reporting having a JD, and 1.7% reporting "other" (e.g. awarded an LLM).

Non-law student participants were all enrolled in Yale University. None reported being enrolled in Yale Law School or any other law school.

Mechanical Turk ("MTurk") is an online platform that enables researchers to collect large samples from a population that is more representative than many other typical research samples.⁵⁰ Although the sample is not a representative U.S. sample in all respects (e.g. it is not politically

⁴⁶ $M_{age} = 59.4$, 33.3% female, 65.8% male, 0.9% non-binary.

⁴⁷ $M_{age} = 26.2$, 50.0% female, 49.1% male, 0.9% non-binary.

⁴⁸ $M_{age} = 23.7$, 60.5% female, 38.9% male, 0.6% non-binary.

⁴⁹ $M_{age} = 36.4$, 48.8% female, 48.8% male, 0.5% non-binary.

⁵⁰ See Adam J. Berinsky et al., *Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk*, 20 POL. ANALYSIS 351 (2012); Gabriele Paolacci et al., *Running Experiments on Amazon Mechanical Turk*, 5 JUDGMENT & DECISION MAKING 411 (2010). The service is understood to provide high-quality data. See Michael Buhrmester, Tracy Kwang & Samuel D. Gosling, *Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality Data?*, 6 PERS. PSYCHOL. SCI. 3 (2011).

representative), for the purposes of this study concerning legal expertise, the MTurk population offers a useful contrast as a “general population” sample.

All participants were randomly assigned to read one of four versions of a vignette: Ordinary Moderate, Ordinary Severe, Legal Moderate, or Legal Severe.⁵¹ All participants read one scenario that varied in the severity of the bad outcome (moderate or severe) and the context of the decision (ordinary or legal). The Ordinary [Moderate; Severe] vignette read as follows:

The mayor of a small town was approached by his advisor who said: “We could build a new highway connection. This would make car traffic much more efficient. However, there would be [minor; severe] adverse effects on some companion animals, like people’s dogs and cats, close to the construction site. During construction, the animals near the construction zone will experience minor injuries.” The mayor responded: “I don’t care at all about the companion animals. All I care about is making car traffic as efficient as possible. Let’s build the new highway connection.” They built the new highway connection. The animals in the zone [experienced some minor injuries; suffered major injuries, including death].

Please rate whether you agree or disagree with the following statement:

The mayor intentionally injured the companion animals.

The Legal versions were identical, but began with the following text:

Imagine that you are a state judge in the (fictional) U.S. jurisdiction of “East Dakota.” East Dakota has a state statute that imposes liability for intentionally injuring companion animals. The relevant provision reads:

“East Dakota General Statutes § 25-350: Liability For Intentionally Injuring Companion Animals:

Any person who intentionally injures a companion animal, except in defense of such person or another person, shall be liable to the owner of such companion animal for economic damages sustained by such owner.”

You are deciding a case in which a group of citizens are suing their town’s mayor under this statute.

The parties agree to the following facts:

⁵¹ This is a 2x2 between-subjects design.

The remainder of the Legal versions contained the exact text from the relevant Ordinary version (Moderate or Severe).⁵²

The fictional statute “25-350” is modeled closely on existing state statutes, such as the Connecticut General Statute § 22-351a: Liability for intentionally killing or injuring companion animals.

The vignettes are modeled as closely as possible after Kneer and Bourgeois-Gironde.⁵³ Two small changes to the main vignette make the legal version more plausible as a real legal example. First, “animals” are changed to “companion animals” (like dogs) to mirror existing statutes concerning companion animals. Second, because it is unlikely that an animal owner would sue for a “temporary disturbance” to an animal, the moderate outcome (temporary disturbance) was changed to “minor injuries” in order to increase the plausibility of the legal cause of action. Given these minor changes, the experiment’s third aim of testing severity effects should not be understood as a direct replication attempt of Kneer and Bourgeois-Gironde’s severity effect,⁵⁴ but instead as an attempted extension.

Results and Discussion

Ratings of intentional action were analyzed to test for effects of *Population*, whether overall ratings of intentionality differed among judges, law students, other students, and the general population participants; *Context*, whether overall ratings differed between the “ordinary” and “legal” context; and *Severity*, whether overall ratings differed between the mild and severe vignettes.

⁵² Part III.C provides further interpretation of this case. It is worth noting that, although the bad outcome is clearly foreseeable, it likely falls short of tort law’s test of “knowledge of a substantial certainty.” For one, it is not clear that there is a substantial certainty of injury, rather than a likelihood. Second, it is not clear that the chairman has *knowledge* of any substantial certainty; he may have notice or awareness of the possibility, but it is not obvious that he has knowledge. Finally, the injury occurs in a general area to a diffuse population, which differs from the typical substantial certainty case involving a specific area (e.g. workplace) injury to specific people.

⁵³ *Supra* note 1.

⁵⁴ *Id.*

The analysis also tested for the interaction among these factors.⁵⁵ For example, it could detect whether there was a Severity effect within only certain populations (e.g. general population) but not others (e.g. judges); or whether only certain populations (e.g. only law students and judges) were sensitive to the difference between an ordinary versus legal context; whether Severity affected ratings in the ordinary context, but not the legal one; or whether there was a three-way interaction (e.g. whether severity affected intentionality judgments for every population and context, except when judges evaluated a legal context).

The data are explored in greater detail in the next three Sections, but here I provide a very brief summary of the results before turning to those details.

First, overall ratings of intentional action differed among general population participants, non-law students, law students, and judges.⁵⁶ The pattern of results is quite striking: overall attributions of intentional action decrease as years of legal training increase.⁵⁷

Second, only law students and judges distinguished between the ordinary and legal contexts.⁵⁸ The judges and law students agreed more strongly that the action was intentional in the ordinary context. The general population and non-law students showed no differences between the ordinary and legal context ratings.⁵⁹

Finally, there was only a severity effect for one population (in both contexts): the general population sample (MTurk).⁶⁰ Those with legal training showed no severity effect in the legal scenarios, and—contrary to prior research—they showed no severity effect in the ordinary scenarios.⁶¹

Consider the three findings in greater detail.

⁵⁵ I conducted a 4(Population: MTurk, non-law student, law student, judge) x 2(Context: ordinary, legal) x 2(Severity: moderate, severe) ANOVA.

⁵⁶ $F(3, 773) = 24.57, p < .001, \eta_p^2 = .089$.

⁵⁷ See Section III.B.1 *infra*.

⁵⁸ $F(3, 773) = 11.15, p = .02, \eta_p^2 = .013$. There was no main effect of Context, $F(1, 773) = 1.70, p = .192, \eta_p^2 = .002$.

⁵⁹ See Section III.B.2 *infra*.

⁶⁰ There was a significant severity by population interaction, $F(1, 773) = 9.35, p = .041, \eta_p^2 = .011$. There was no main effect of severity, $F(1, 773) = 2.24, p = .135, \eta_p^2 = .003$, and no three-way interaction, $F(3, 773) = 1.67, p = .491, \eta_p^2 = .002$.

⁶¹ See Section III.B.3 *infra*; cf. Kneer & Bourgeois-Gironde, *supra* note 1 (reporting that judges show a severity effect in ordinary cases).

1. Intentional Action Attributions Differ Among Judges, Law Students, Non-Law Students, and Ordinary People

First, consider the overall differences between populations. Mean ratings were highest for the general population,⁶² then the non-law students,⁶³ then the law students,⁶⁴ and finally the judges.⁶⁵

These mean ratings are displayed in Figure 1.

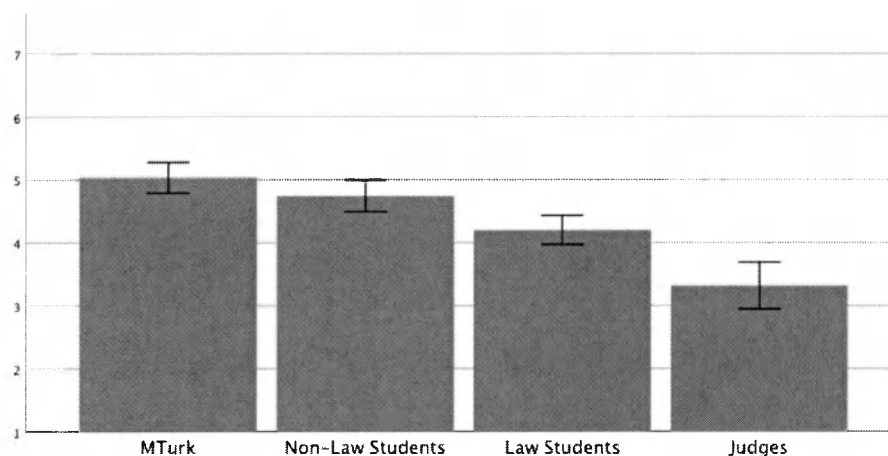


Figure 1. Mean intentionality ratings by Population, collapsing across Context and Severity. Error bars indicate 95% confidence intervals.

One might wonder whether some of these numerically small differences are statistically significant in pairwise comparisons. For example, are law student ratings greater than judge ratings at a level of statistical significance? Further analyses considered pairwise comparisons of the mean ratings among all four populations, corrected for multiple comparisons. Judge ratings were significantly lower than ratings for law students,⁶⁶ non-law students,⁶⁷ and the general population.⁶⁸ Law student ratings were significantly lower than non-law student ratings⁶⁹ and the

⁶² $M = 5.03, SD = 1.81$.

⁶³ $M = 4.74, SD = 1.77$.

⁶⁴ $M = 4.20, SD = 1.80$.

⁶⁵ $M = 3.32, SD = 2.14$.

⁶⁶ $t(366) = 4.20, p < .001, d = .46$ (Bonferonni corrected p value).

⁶⁷ $t(324) = 6.54, p < .001, d = .74$ (Bonferonni corrected p value).

⁶⁸ $t(340) = 7.93, p < .001, d = .89$ (Bonferonni corrected p value).

⁶⁹ $t(430) = 3.13, p = .011, d = .30$ (Bonferonni corrected p value).

general population ratings.⁷⁰ There was no significant difference between the general population and non-law student ratings.⁷¹

We can also consider whether the overall ratings for each population suggest overall agreement or disagreement that the action was intentional. To explore this question, I compared each population mean rating to the scale midpoint (4). Ratings above the midpoint are plausibly interpreted as signifying a population's overall agreement that the action is intentional, while ratings below the midpoint are plausibly interpreted as signifying overall agreement that the action is not intentional. The general population ratings were significantly above the midpoint.⁷² Non-law student ratings were significantly above the midpoint.⁷³ Law student ratings did not significantly differ from the midpoint in either direction.⁷⁴ Judge ratings were significantly below the midpoint.⁷⁵

This primary finding—that attributions of intentional action to a foreseeably bad but motivationally non-significant side-effect decrease with legal training—is evidence against identity theory. While the ordinary concept of intentional action is applied to this scenario's outcome, the legal concept is not.

There are two competing explanations for this main effect of population. These explanations draw from Part I's "Selection" and "Legal Training" theories:

The Selection Explanation: The population difference in intentionality judgments is explained by some factors about the *kinds of people* who attend law school and become judges.

The Legal Training Explanation: The population difference in intentionality judgments is explained by some aspect of *legal training or experience*.

Both explanations are *prima facie* plausible explanations of difference between, on the one hand, non-law students and law students and judges. First consider the Selection explanation. There is variance in the judgments of all populations—although overall judge ratings were lower

⁷⁰ $t(446) = 4.86, p < .001, d = .46$ (Bonferonni corrected p value).

⁷¹ $t(404) = 1.63, p = .623, d = .16$ (Bonferonni corrected p value).

⁷² $t(210) = 8.30, p < .001, d = .57$.

⁷³ $t(194) = 5.88, p < .001, d = .42$.

⁷⁴ $t(236) = 1.73, p = .085, d = .11$.

⁷⁵ $t(130) = -3.63, p < .001, d = .32$.

than the general population ratings, some judges answer “7” and some MTurk participants answer “1.” Perhaps what explains the population difference is that there is some other factor (e.g. intelligence, socioeconomic status, or attention to the survey) that explains why people tend to answer “1.” If the people who become judges are also people with that factor, *that* factor could be the real explanation of the differences.

By contrast, the Legal Training explanation posits that the difference in population means is not explained by differences in (e.g.) the average intelligence of judges versus the general population, but instead by something about legal education.

These two explanations are both plausible accounts of the main pattern of results (e.g. Figure 1), but they make some divergent predictions. For example, the Selection account would predict that people who are very similar to law students, but not in law school, should nevertheless report similar judgments of intentionality. The Legal Training account would make the opposite prediction.

The pattern of results provides some evidence in favor of the Legal Training explanation over the Selection one. Consider that the “elite” law students and “elite” non-law students are similarly situated in a number of respects, but their intentional action attributions differ. The difference cannot be explained by personal characteristics that those populations have in common, but they could be explained by the experience of law school that is unique to the law student population.

Of course, there might be some more elaborate Selection explanations that remain. Although (e.g.) Yale Economics PhD students and Harvard Law students are both likely to share a number of personal characteristics (e.g. education, level of attention in taking surveys), perhaps some other characteristics distinguish one population. The burden seems to rest with such an “Elaborate Selection” account to explain this account in more detail, particularly (1) what the distinctive personal factor is and (2) how it plausibly explains differences in intentionality attributions.

Nevertheless, to provide additional insight into the debate between the Selection and Legal Training accounts I conducted a further post-hoc analysis. Law student participants were asked to report their years of law training: one year or less (“1L”), one-two years (“2L”), two-three years (“3L”), or more or other. Most participants identified as 1L, 2L, or 3L students. The Selection hypothesis would predict little difference between the judgments of these groups: today’s 1Ls are *exactly the same people* as next year’s 2Ls. However, the Legal Training hypothesis would predict a difference among these groups. 1L judgments should be most similar to the judgment of non-law students, while 3L judgments should be most similar to those of judges.

To test these competing hypotheses, I compared the effect of increasing law school training (1L, 2L, 3L) on intentionality ratings. There were sixty-one 1Ls,⁷⁶ fifty-one 2Ls,⁷⁷ and one-hundred 3Ls.⁷⁸ There was a significant effect of law training.⁷⁹ The greatest ratings were from 1Ls,⁸⁰ then 2Ls,⁸¹ then 3Ls.⁸² See Figure 2.

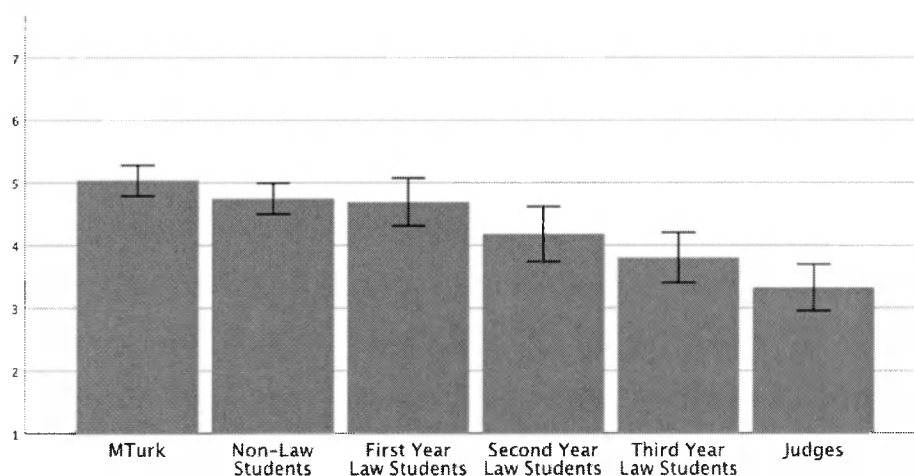


Figure 2. Mean intentionality ratings by Population (law student population divided by years of legal education), collapsing across Context and Severity. Error bars indicate 95% confidence intervals.

⁷⁶ $M_{age} = 25.1$, 54.1% female, 44.3% male, 1.6% non-binary.

⁷⁷ $M_{age} = 26.4$, 51.0% female, 49.0% male, 0.9% non-binary.

⁷⁸ $M_{age} = 26.7$ 48.0% female, 51.0% male, 1.0% non-binary.

⁷⁹ Independent-Samples Kruskal-Wallis test, $p = .016$.

⁸⁰ $M = 4.69$, $SD = 1.49$.

⁸¹ $M = 4.18$, $SD = 1.56$.

⁸² $M = 3.80$, $SD = 2.01$.

I conducted further post-hoc t-tests to compare ratings within law school populations (e.g. 1L vs. 2L), between each law school year group and the non-law students (e.g. Non-Law vs. 1L), and between each law school year group and the judges (e.g. 3L vs. Judges).

1L ratings were greater than 3L ratings,⁸³ but were no different from 2L ratings.⁸⁴ There was no difference between 2L and 3L ratings.⁸⁵

Compared to non-law student ratings (plausibly the most similar non-trained population), ratings of 1Ls⁸⁶ and 2LS⁸⁷ did not differ significantly. However, non-law student ratings were significantly high than those of 3Ls.⁸⁸

Compared to judge ratings, 1L⁸⁹ ratings were significantly higher. There was no difference between judges' ratings and those of 2Ls⁹⁰ or 3Ls.⁹¹

These analyses provide further evidence in favor of the Legal Training Hypothesis over the Selection Hypothesis. There are significant differences among 1L, 2L, and 3L participants consistent with a broader effect of legal training on intentionality attributions. First year law students provide ratings that are no different from that of non-law students, while third year law students provide ratings that are significantly different.

2. Those With Legal Training Distinguish Between Ordinary and Legal Contexts

Next, I explored the relationship between Population (e.g. general population vs. judges) and Context (e.g. ordinary vs. legal scenario). Because the judge population was smaller than the other populations and the primary interest of the study is the effect of legal expertise, I first examined ratings between the Judges and Law Students populations. If there were no significant Severity or Context effects or interactions, it is sensible to treat the law student and judges as a

⁸³ $t(159) = 2.98, p = .027$ (Bonferonni corrected p value).

⁸⁴ $t(110) = 1.77, p = .711$ (Bonferonni corrected p value).

⁸⁵ $t(149) = 1.17, p = 1$ (Bonferonni corrected p value).

⁸⁶ $t(254) = .22, p = 1$ (Bonferonni corrected p value).

⁸⁷ $t(244) = 2.09, p = .342$ (Bonferonni corrected p value).

⁸⁸ $t(293) = 4.14, p < .001$ (Bonferonni corrected p value).

⁸⁹ $t(190) = 4.50, p < .001$ (Bonferonni corrected p value).

⁹⁰ $t(180) = 2.60, p = .090$ (Bonferonni corrected p value).

⁹¹ $t(229) = 1.73, p = .765$ (Bonferonni corrected p value).

single “Expert” population. A 2(Severity: moderate, severe) x 2(Context: ordinary, legal) x 2(Population: law student, judge) ANOVA revealed a main effect of population,⁹² but no significant interactions, all F s < 1. Because the judge sample is smaller than the others and there were no significant context effects or interactions in this analysis, I treated judges and law students as one “expert” population to further explore the Population x Context interaction.

A 2(Context: ordinary, legal) x 3(Expertise: expert (judges and law students), non-law student, MTurk) ANOVA revealed a main effect of Expertise.⁹³ There was no main effect of Context.⁹⁴ There was a Context x Expertise interaction.⁹⁵

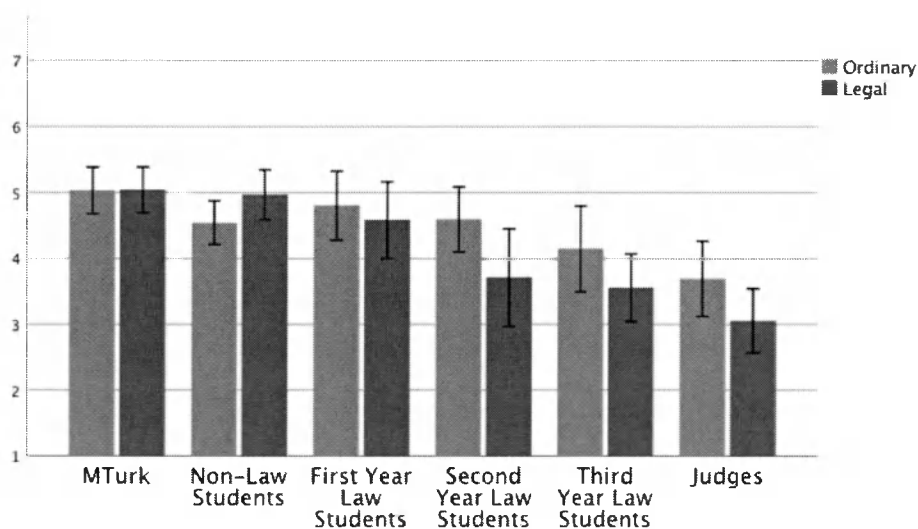


Figure 3. Mean intentionality ratings by Population (law student population divided by years of legal education), collapsing across Severity. Error bars indicate 95% confidence intervals.

To further explore the Context x Expertise interaction, I compared the means between Context conditions for each Expertise level. For the expert population (law students and judges), mean ratings in the ordinary condition⁹⁶ were greater than mean ratings⁹⁷ in the legal condition.⁹⁸

⁹² $F(1, 367) = 16.12, p < .001, \eta_p^2 = .043$.

⁹³ $F(2, 773) = 27.70, p < .001, \eta_p^2 = .068$.

⁹⁴ $F < 1$.

⁹⁵ $F(2, 773) = 5.14, p = .006, \eta_p^2 = .013$.

⁹⁶ $M = 4.22, SD = 1.89$.

⁹⁷ $M = 3.62, SD = 1.97$.

However, for the general population, there was no difference between mean ratings for the ordinary⁹⁹ and legal¹⁰⁰ conditions.¹⁰¹ Similarly, for the non-law student population, there was no significant difference between mean ratings for the ordinary¹⁰² and legal¹⁰³ conditions.¹⁰⁴

The pattern of results suggests further support of the Selection hypothesis. The general population participants and non-law students made no distinction between ordinary and legal contexts. However, those who had legal training made a distinction; ratings for the legal scenario were lower, reflecting the “legal” concept of intentional action.

3. Those With Legal Training Do Not Show A Severity Bias

Next, to analyze the Severity x Population interaction, I conducted a 2(Severity: moderate, severe) x 3(Expertise: expert (judges and law students), non-law student, MTurk) ANOVA. There was a main effect of Expertise,¹⁰⁵ and there was a main effect of Severity.¹⁰⁶ There was also a significant Severity x Expertise interaction.¹⁰⁷

To further explore the Severity x Expertise interaction, I compared the means between Severity conditions for each Expertise level. For the expert population (law students and judges), moderate ratings¹⁰⁸ and severe¹⁰⁹ ratings were not significantly different.¹¹⁰ For the non-law student population, moderate¹¹¹ and severe¹¹² ratings were not significantly different.¹¹³ However,

⁹⁸ $F(1, 367) = 8.55, p = .004, \eta_p^2 = .023$. Moreover, for judges, ordinary ratings were greater ($M = 3.69, SD = 2.12$) than legal ratings ($M = 3.05, SD = 2.13$). For law students, ordinary ratings were also greater ($M = 4.48, SD = 1.72$) than legal ratings ($M = 3.96, SD = 1.84$).

⁹⁹ $M = 5.03, SD = 1.84$.

¹⁰⁰ $M = 5.04, SD = 1.79$.

¹⁰¹ $F < 1$.

¹⁰² $M = 4.54, SD = 1.69$.

¹⁰³ $M = 4.97, SD = 1.83$.

¹⁰⁴ $F(1, 194) = 2.89, p = .091, \eta_p^2 = .015$.

¹⁰⁵ $F(2, 773) = 29.09, p < .001, \eta_p^2 = .070$.

¹⁰⁶ $F(2, 773) = 4.73, p = .03, \eta_p^2 = .006$.

¹⁰⁷ $F(2, 773) = 3.94, p = .020, \eta_p^2 = .010$.

¹⁰⁸ $M = 3.91, SD = 1.99$.

¹⁰⁹ $M = 3.87, SD = 1.96$.

¹¹⁰ $F < 1$. Moreover, for judges, moderate ratings ($M = 3.38, SD = 2.25$) were no different from severe ratings ($M = 3.26, SD = 2.05$). For law students, moderate ratings ($M = 4.22, SD = 1.77$) were no different from severe ratings ($M = 4.19, SD = 1.84$).

¹¹¹ $M = 4.69, SD = 1.71$.

¹¹² $M = 4.80, SD = 1.93$.

¹¹³ $F < 1$.

for the general population (MTurk), moderate ratings¹¹⁴ were lower than severe ratings,¹¹⁵ at a level of statistical significance.¹¹⁶

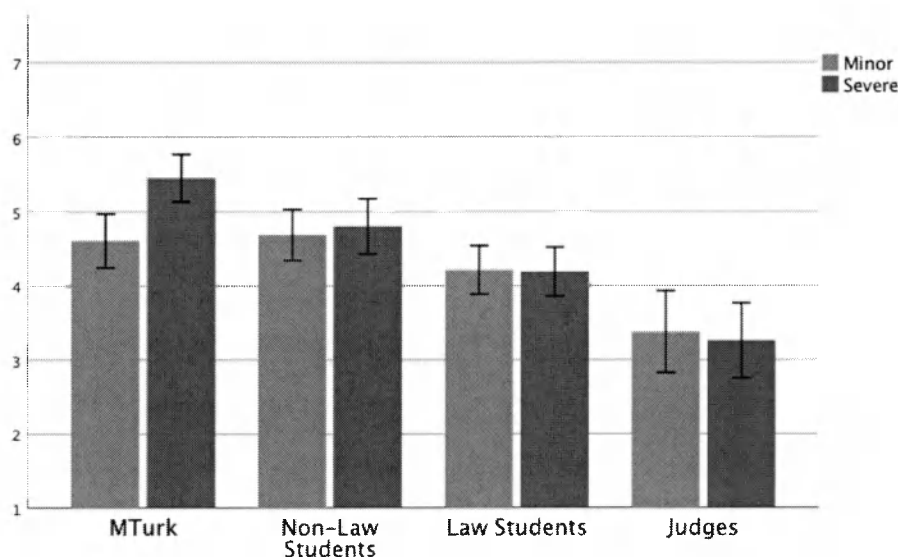


Figure 4. Mean intentionality ratings by Population, collapsing across Context. Error bars indicate 95% confidence intervals.

The results suggest that severity effects may be even less influential than predicted—for both experts and non-experts. For one, legal experts do not show a severity effect in attributions of intentional action. That legal experts do not show severity effects in ordinary contexts is inconsistent with prior research.¹¹⁷ Future research might show that legal experts show severity effects in other cases, but for now the evidence is mixed and confidence in that claim should be more tentative. Even if that were demonstrated, the present study also provides evidence that legal experts do not show severity effects in legal contexts.

Second, the present findings suggest that one non-expert population (non-law students) is not affected by severity. This finding conflicts both with prior research¹¹⁸ as well as the fact that the general population (MTurk participants) did show a severity effect in this study.

¹¹⁴ $M = 4.61, SD = 1.86.$

¹¹⁵ $M = 5.45, SD = 1.65.$

¹¹⁶ $F(1, 773) = 12.08, p < .001, \eta_p^2 = .055.$

¹¹⁷ Kneer & Bourgeois-Gironde, *supra* note 1.

¹¹⁸ *Id.*

One possibility to explain this difference is that online participants may be more likely to show an experimenter demand effect.¹¹⁹ Of course, this hypothesis assumes that at least some MTurkers divined the purpose of this study and that of Kneer and Bourgeois-Gironde. Perhaps that is not so implausible. Might the language of “minor” and “severe” outcomes suggest to participants that the “right” answer is, respectively, “not intentional” and “intentional”? On this hypothesis, severity effects about intentional action present less concern than many think: They arise in psychology studies as a result of experimenter demand, but we have little reason to think they affect people’s intentional action judgment in legal practice (e.g. on a jury panel).

4. Summary

Experiment 1 found evidence that legal training provides a distinctive concept of intentional action, one that manifests more strongly in legal contexts and for those with more legal training and experience. These findings challenge identity theory, the hypothesis that the ordinary and legal concepts of intentional action are identical. Instead, the findings support that there is some dissimilarity between the concepts.

However, there is one surprising and unpredicted finding of the first experiment that calls for further investigation. There seemed to be some conceptual “spillover;” in ordinary contexts. In Figure 3, the average ratings in the *ordinary* context also decreased with legal training and experience. The second experiment explores this “spillover” possibility further.

B. Experiment 2

Experiment 2 tests whether the “spillover” effect of Experiment 1 is robust by considering a different ordinary case of intentional action attribution, the classic “side-effect effect.” If we find

¹¹⁹ Adam J. Berinsky, Gregory A. Huber & Gabriel S. Lenz, *Evaluating Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk*, 20 POL. ANALYSIS 351, 366 (2012) (“MTurk respondents . . . may also exhibit experimental demand characteristics to a greater degree than do respondents in other subject pools, divining the experimenter’s intent and behaving accordingly”).

differences among legal experts and non-experts in these ordinary cases, this would provide further evidence that legal training may sometimes spill over into ordinary cognition.

Method

Seventy-three law students,¹²⁰ 123 non-law students,¹²¹ and 111 participants from Amazon's Mechanical Turk¹²² participated in an online experiment. Law students were recruited by emails to students at Yale Law School. Non-law students were recruited by contacting student emails, networks, and list-servs at Yale University. Non-law student participants were all enrolled in Yale University. None reported being enrolled in Yale Law School or any other law school.

All participants were randomly assigned to read one of two versions of a vignette: Harm or Help. After reading the vignette, each participant rated whether the action was intentional and their evaluation of the action (how blameworthy or praiseworthy). The Harm vignette read as follows:

The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment.' The chairman of the board answered, 'I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was harmed.

Please rate whether you agree or disagree with the following statement:

The chairman intentionally harmed the environment.

[next page]

Please rate whether you agree or disagree with the following statement:

The chairman is blameworthy for his action.

The Help vignette read as follows:

The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, and it will also help the environment.' The chairman of the board answered, 'I don't care at all about helping the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was helped.

¹²⁰ $M_{age} = 39.5$, 50.0% female, 49.1% male, 0.9% non-binary.

¹²¹ $M_{age} = 23.4$, 78.6% female, 19.4% male, 1.9% non-binary.

¹²² $M_{age} = 28.0$, 58.0% female, 38.0% male, 4.0% non-binary.

Please rate whether you agree or disagree with the following statement:

The chairman intentionally helped the environment.

[next page]

Please rate whether you agree or disagree with the following statement:

The chairman is praiseworthy for his action.

Results and Discussion

Ratings of intentional action and blame/praise were analyzed to test for effects of *Population*, whether overall ratings differed among law students, non-law students, and the general population (MTurk) participants; and *Valence*, whether overall ratings differed between the “help” and “harm” vignettes.

First consider the results for intentionality ratings. There was a marginally significant main effect of Population; overall ratings of intentional action differed among the general population participants, non-law students, and law students.¹²³ Overall, the ratings decreased with increasing legal experience.¹²⁴ There was also a main effect of Valence.¹²⁵ Ratings of intentionality were lower in the help condition than in the harm condition.¹²⁶ Finally, there was a significant population and valence interaction.¹²⁷ See Figure 5 for the main pattern of results.

¹²³ $F(2, 303) = 3.017, p = .050, \eta_p^2 = .020$.

¹²⁴ MTurk $M = 4.02$ ($SD = 1.77$); Non-Law $M = 3.87$ ($SD = 2.28$); Law Student $M = 3.45$ ($SD = 2.05$).

¹²⁵ $F(2, 303) = 211.0, p < .001, \eta_p^2 = .415$.

¹²⁶ Help $M = 2.30$ ($SD = 1.60$), Harm $M = 5.33$ ($SD = 1.91$).

¹²⁷ $F(2, 303) = 20.0, p = .001, \eta_p^2 = .044$.

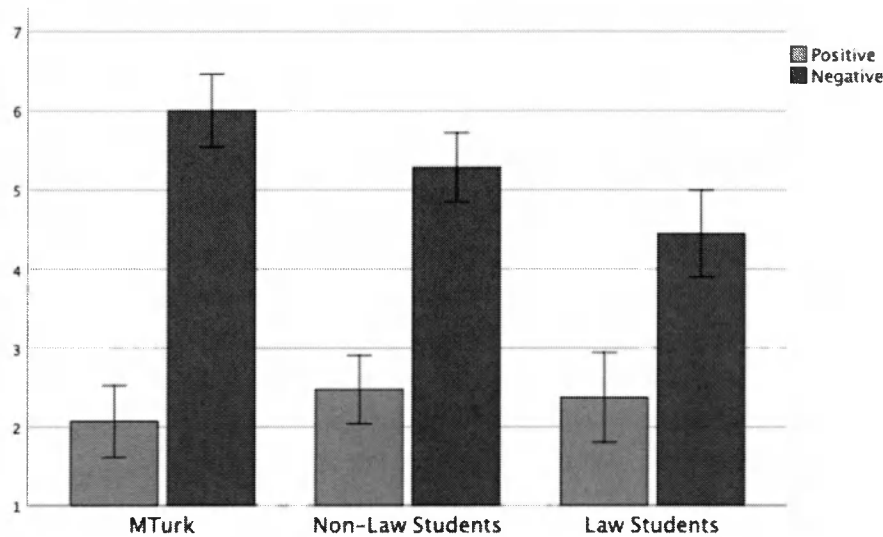


Figure 5. Mean intentionality ratings by Population and Valence. Error bars indicate 95% confidence intervals.

I decomposed the interaction by considering the effect of population within each valence condition (help; harm). Within the help condition, there was no significant effect of population ($F < 1$).

Within the harm condition, there was a significant effect of population.¹²⁸ Overall mean ratings decreased with additional expertise; general population (MTurk) participants had the highest ratings, then non-law students, and then law Students.¹²⁹ There was no significant difference between the mean ratings of the non-law student and general populations.¹³⁰ However, there was a significant difference between law students and the general population,¹³¹ and law students and non-law students.¹³²

¹²⁸ $F(2, 152) = 26.99, p < .001, \eta_p^2 = .098$.

¹²⁹ MTurk $M = 6.00$ ($SD = 1.59$); Non-Law $M = 5.28$ ($SD = 1.85$); Law Student $M = 4.45$ ($SD = 2.09$).

¹³⁰ $t(114) = 1.80, p = .075$.

¹³¹ $t(91) = 3.70, p < .001$.

¹³² $t(97) = 2.15, p = .034$.

Next consider the results for blame/praise ratings. There was no significant effect of Population.¹³³ There was a significant effect of valence.¹³⁴ Overall, participants were more strongly inclined to blame the chairman in Harm than to praise the chairman in Help.¹³⁵ There was no significant interaction between population and valence.¹³⁶

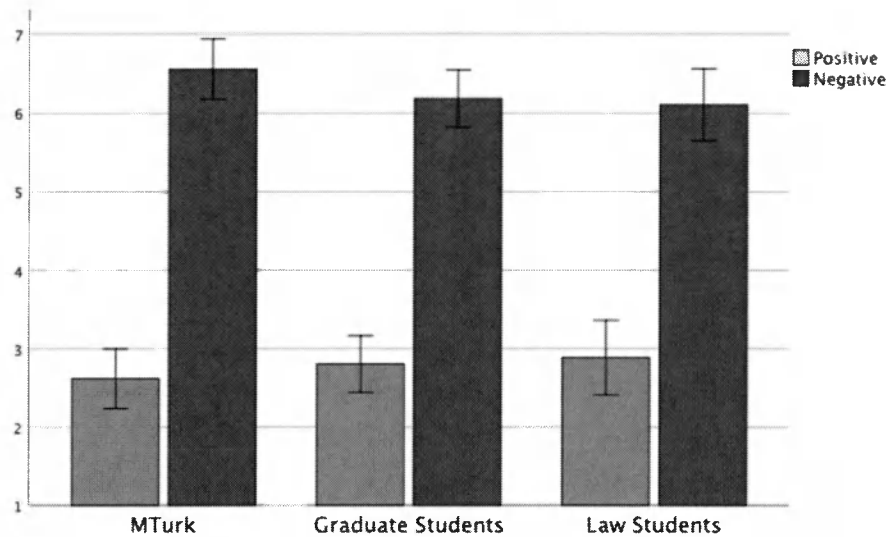


Figure 6. Mean praiseworthiness (positive) and blameworthiness (negative) ratings by Population and Valence. Error bars indicate 95% confidence intervals.

This second experiment provides further support for two claims. First, legal experts seem to have a concept of intentional action that is distinctive in some ways from that of non-experts. This second experiment confirms the unpredicted finding of the first: legal experts apply the “legal” concept not only in legal contexts, but also in ordinary ones. Attributions of intentionality in the harm case were lower for law students than the other populations.

At the same time, legal experts have a concept of intentional action that shares certain features of the ordinary concept. There was still a significant (and large) difference between law

¹³³ $F < 1$.

¹³⁴ $F(1, 303) = 436.51, p < .001, \eta_p^2 = .56$.

¹³⁵ Harm $M = 6.30$ ($SD = 1.13$); Help $M = 2.75$ ($SD = 1.67$).

¹³⁶ $F(2, 303) = 1.702, p = .184, \eta_p^2 = .01$.

student ratings of the help and harm cases. Although the difference was smaller than that of the non-expert populations, the results still suggest that the positive-negative asymmetry that is part of the ordinary concept characterizes the judgments of the legal experts.

C. Experimental Interpretation

The previous sections discuss the pattern of experimental results. This section develops these findings further, offering an interpretation of the results' legal and philosophical significance. There are three key components of the interpretation. First, the results suggest that legal experts have a distinctive legal concept of intentional action. Second, the results support that experts acquire this concept through training. Finally, the results indicate that legal experts retain the ordinary concept of intentional action, but also suggest that some features of the legal concept may “spill over” into judgments made in ordinary contexts.

1. Legal Experts Have a Distinctive Legal Concept of Intentional Action

There are three primary results from the first experiment: (i) *overall* attributions of intentional action in the scenario decreased with legal training; (ii) legal experts distinguished between ordinary and legal contexts, and (iii) legal experts did not show a severity bias. All three of these results support that there is a distinctive legal concept of intentional action, one that is not coextensive with the ordinary concept.

Perhaps the most significant finding is that overall attributions of intentional action decreased with legal training. In ordinary cognition, a bad side-effect that is not motivational but is foreseeable is often understood to be produced intentionally. However, the core legal concept that emphasizes motivational significance, would not obviously apply in this instance. Although it is a closer case, the “substantial certainty” sense of intentional action would also likely not apply. In these vignettes (a) it is not clear that there is a substantial certainty of the side-effect, (b) it is not clear that the actor has knowledge of any such certainty, and (c) the type of harm (unspecified

harm to animals “near” the site) is more remote and disparate than in typical substantial certainty cases (e.g. harm within a specific worksite; the relation between pulling out someone’s chair and them falling)

This distinction is reflected clearly in the data. Both groups of experts (judges and law students) provide reduced attributions of intentionality, compared to two ordinary populations (MTurk and elite university students).

The two additional findings provide further support for the expertise account. Only those with legal training and experience (judges and law students) distinguished between the ordinary and legal contexts. Moreover, they did so in precisely the way that the legal account of expertise would predict. In the legal context, judgments were more in line with the verdict of the putative “legal concept” of intentional action.

Finally, the “severity sensitivity” that characterizes the judgments of ordinary people was reflected in the general (MTurk) population. However, the legal experts showed no severity sensitivity. Surprisingly, even the elite non-law students showed no severity sensitivity. This calls for further experimental investigation, but what is certainly clear is that the legal experts (law students and judges) were not severity sensitive in their attributions of intentional action in these cases.

Together, these three findings support that legal experts produce judgments of intentional action that are different from those of ordinary people. Yet, a question remains about the interpretation of this pattern of results. Does this reflect a special legal concept? Or does it simply reflect that legal experts are simply less confused than ordinary people? On the second competitor account, there is only one concept of intentional action, but ordinary people often make performance errors in applying that concept. Judges and law students are simply more competent in judging intentional action.

This competitor account is unpersuasive in light of the empirical evidence. For one, the account does not provide a compelling explanation of the progressive pattern of results, from early law school students to judges. One might think that additional education helps people refine the one “true” concept of intentional action. But it is striking that it is specifically *legal training* that has this effect; elite university students are no different in their judgments from those participants recruited from MTurk. While it is certainly possible that only legal education teaches the truth about the ordinary concept of intentional action, it is more plausible that legal training teaches something distinctive about law.

More significantly, this competitor account is inconsistent with the finding that legal experts distinguish between ordinary and legal contexts. If judges and law students are simply more competent users of the one concept of intentional action, we should expect that they make better judgments of intentionality (e.g. lower attributions in the scenario) in *both* contexts. That the experts distinguish between the contexts suggests that they are not simply more competent users of one context-independent concept of intentional action.

Finally, this competitor account requires taking a very strong stand against the conceptual significance of a robust pattern of ordinary judgment. It is (in theory) possible that ordinary language users make systematic and repeated errors when judging intentional action. But detaching “competence” from robust patterns of ordinary use threatens theoretical unfalsifiability. If a theory of the concept of intentional action may reject any robust pattern of data as a global “performance error,” one wonders how to make progress on debates about the concept(s). While this essay is not committed to the strongest empirical claim—for example, the claim that empirical patterns of language use are *constitutive* of conceptual competence—it aims to ground its analysis in empirical facts. At the very least, the burden of such an argument rests with those who claim that ordinary patterns of use are irrelevant to the concept that is analyzed.

2. Legal Experts Acquire a Distinctive Legal Concept Through Training

The second important interpretive conclusion is that legal experts acquire a distinctive legal concept through legal training and experience. Recall the two hypotheses about distinctive legal concepts: They might be produced by selection or by training. An unusual form of the selection account might be that people who happen to have the legal concept of intentional action are overrepresented in law; selection into the legal profession explains differences in concepts. The training account posits a very different story; legal concepts are acquired through special legal training and experience.

Perhaps the most striking evidence of this selection interpretation is the progressive pattern of results (Figure 2). Overall attributions of intentional action decrease incrementally with additional legal training and experience. Moreover, when we compare similarly situated groups—like elite law students and elite non-law students—we find that those without legal training are similar to ordinary populations.

These findings count against the selection hypothesis. Elite law students and elite non-law students are similar in a number of respects. However, only the former reflect the kinds of judgments consistent with the legal concept (e.g. distinguishing between context; overall lower ratings). The progressive pattern adds further support to this account. 1L, 2L, and 3L students should be identical in terms of the factors relevant to the selection hypothesis, but these groups differ in precisely the way the training hypothesis would predict.

3. Legal Experts Retain the Ordinary Concept, with “Spillover”

This final interpretive conclusion is offered more cautiously, since it represents a conclusion drawn from an unpredicted finding in the first experiment. The conclusion—offered tentatively—is that, although legal experts retain the ordinary concept, there is some “spillover” of the legal concept into ordinary contexts.

A few results from the studies support this interpretation. First, consider Experiment 1. If legal experts have a legal and an ordinary concept of intentional action, we might expect them to deploy the former in legal contexts and the latter in ordinary ones. The first experiment's results do not support such a simple story. Although legal experts distinguished between ordinary and legal contexts, their average ordinary context ratings were lower than the average ratings of non-expert populations. Moreover, legal experts' attributions of intentional action decreased with training in the legal context *and in the ordinary context*.

Experiment 2 sought to investigate this surprising finding further. It found that, in an ordinary side-effect effect case, legal experts (law students) did show the side-effect effect. However, their ratings of intentionality in the harm case were lower than ratings from general and elite non-law student populations. The first result indicates a way in which the legal experts' ordinary concept is not replaced by a legal concept. But the second result suggests a way in which the legal experts' ordinary concept may nevertheless be altered by legal training.

This first result—that the ordinary concept is not simply replaced by a legal one—finds further support in the cognitive science of scientific concepts. Although experts develop new scientific concepts, the old ordinary concepts are not lost. In fact, scientific experts deploy the “lay” concept more often in cases in which there is time pressure.¹³⁷

Another possibility concerns the difference between Experiment 2's findings for helping (positive) and harming (negative). Perhaps the features of the negative case themselves suggest a “legal” context to some participants. In other words, perhaps the very task of assessing whether someone intentionally caused an injury elicits a common legal schema. Conversely, the task of assessing whether someone intentionally helped does not elicit a common legal schema. This suggests a particular type of “spillover” story. It is not that legal training changes the ordinary

¹³⁷ Andrew Shtulman & Joshua Valcarcel, *Scientific Knowledge Suppresses But Does Not Supplant Earlier Intuitions*, 124 COGNITION 209 (2012); see also Andrew Shtulman & Kate McCallum, *Cognitive Reflection Predicts Science Understanding*, 36 PROCEEDINGS OF THE ANNUAL MEETING OF THE COGNITIVE SCIENCE SOCIETY 2937 (2014); Andrew Shtulman & Kelsey Harrington, *Tensions Between Science and Intuition Across the Lifespan*, 8 TOPICS IN COG. SCI. 118 (2016).

concept, but rather that legal training teaches a distinctive concept that is applied in “legal contexts,” contexts that may arise more broadly than it first seems.

For now, these considerations represent hypotheses rather than interpretive conclusions. Future work might examine whether similar effects characterize legal cognition. What is clear is that legal experts deploy a concept of intentional action in ordinary context with at least some features of the ordinary concept (i.e. the side-effect effect asymmetry).

IV. IMPLICATIONS

This Part considers several implications of the experimental results. First, it considers a broader philosophical question about ordinary and legal concepts, raised by the experimental results. There are legal concepts that are similar to corresponding ordinary concepts, but which nevertheless have important differences. This descriptive conclusion raises the question: Should legal concepts generally reflect ordinary ones?

Next, it considers experimental implications for debates about how law should conceive of intentional action. The experimental results suggest that there is a distinctive legal concept of intentional action that differs from the ordinary concept. But *should* law use this concept?

Next, this Part draws a conclusion about legal expertise. Those with legal training have expertise in unique legal concepts (e.g. *parol evidence*), but also in legal concepts that differ from their ordinary counterparts (e.g. *intentional action*). Moreover, this expertise is developed through legal training, not produced by individual differences like innate intelligence. This defuses some challenges from experimentalists; research about ordinary concepts does not support straightforward conclusions about legal concepts.

Finally, this Part considers how the results raise new challenges for legal theory and practice. One is a puzzle for theories of legal interpretation: What is the ordinary meaning of an ordinary term that evokes different concepts in legal experts and non-experts? Correspondingly, how

should legal practice respond to the inconsistency produced by the fact that “intentional action” evokes a different concept for (legally trained) judges and (legally untrained) juries?

A. Ordinary and Legal Concepts

The experimental findings illuminate the relationship between the ordinary and legal concept of intentional action. They provide initial evidence that, at least for the concept of *intentional action*, the identity hypothesis is false and the dissimilarity hypothesis is true. This section considers some of the broader philosophical questions that this raises about the relationship between ordinary and legal concepts. Most importantly, *should* legal concepts (generally) reflect ordinary ones?

The experiments reveal differences, but an open question remains about whether we should strive to replace the distinctive legal concept with an ordinary one, or whether the legal concept is, in fact, apt. Legal philosophers and jurists disagree on this question, with some supporting using ordinary concepts¹³⁸ and other supporting using specialized concepts.¹³⁹

When empirical evidence suggests that there is a distinction between an ordinary and legal concept, there are two classes of arguments one might consider. One is the class of “specific” arguments, which arise given the specific experimental findings concerning a particular concept. The next section, IV.B, outlined some of these with respect to the concept of intentional action, for example concerning severity-sensitivity.¹⁴⁰

¹³⁸ See generally Nicola Lacey, *A Clear Concept of Intention: Elusive or Illusory*, 56 MOD. L. REV. 621 (1993) (citing R. Buxton, *Some Simple Thoughts on Intention*, CRIM. L. R. 484 (1988)).

¹³⁹ See generally *id.* (citing A. Ashworth, PRINCIPLES OF CRIMINAL LAW (Oxford 1991); E. Griew, *Consistency, Communication and Codification*, in P. R. Glazebrook (ed), RESHAPING THE CRIMINAL LAW.

¹⁴⁰ It is worth emphasizing the distinction between these kinds of arguments (e.g. concerning severity sensitivity) and traditional debates about other experimental-legal findings (e.g. the effect of politically motivated reasoning). While it is uncontroversial that racist attitudes or political beliefs should not affect the application of legal standards, it is an open question whether the legal concept of intentional action should be severity sensitive. This represents an important distinction between traditional work in psychology of law and a more philosophically inspired “experimental jurisprudence” or “experimental legal philosophy.” The discoveries of the latter do not solve debates, but rather enhance them. Learning whether the legal concept of intentional action is severity-sensitive is important, but that knowledge does not come with an easy answer to the normative question: *Should* the legal concept be severity-sensitive?

The other class of arguments one might consider contains “general” arguments. These arguments are ones that might apply across a range of debates about concepts. Rather than addressing something specific about the concept (e.g. should *intentional action* be severity-sensitive?), these arguments point to considerations that apply more broadly (e.g. should legal concepts be ordinary concepts for the sake of publicity or clarity?).

One reason to think that legal concepts *should* reflect ordinary ones is that doing so is consistent with the rule of law and conditions of legal legitimacy. Philosophers including Hart, Dworkin, and Fuller, and “rule of law” theorists often recommend conditions of law including publicity, neutrality, consistency, clarity, prospectivity, coherence, and stability.¹⁴¹ Many of these conditions presuppose that, broadly speaking, legal concepts are *accessible* to ordinary people.

A statute imposing liability on “intentionally” produced outcomes is only public if ordinary people understand what “intentionally” means. One way for this to obtain is for all ordinary people to receive legal training. The other way for this to obtain is for law to use the ordinary concept of intentional action. This “general” argument about publicity likely counts as a reason, in most circumstances, for law to use the ordinary concept.

But not all general arguments support using ordinary concepts. One argument in favor of distinctive legal concepts begins by noting law’s aptness. Law is a system of rules and standards that solve distinctively *legal* problems. What gives law its authority to impose liability for (e.g.) intentionally harming is not something about how most ordinary people (mis)understand legal rules. Instead, law is legitimate for some other reason (e.g. in virtue of being posited by an authoritative source). If most ordinary people understood “murder” to apply only to intentional murders and not to felony murders, that should not change the *legal* concept of murder. So too for the concepts of *causation*, *intention*, *reasonableness*, and so on.

¹⁴¹ See generally John Tasioulas, *The Rule of Law*, in THE CAMBRIDGE COMPANION TO THE PHILOSOPHY OF LAW (Cambridge University Press, 2019); see also R.M. Dworkin, A MATTER OF PRINCIPLE (Harvard University Press, 1985); Lon L. Fuller, THE MORALITY OF LAW (Yale University Press, 1965); Margaret Jane Radin, *Reconsidering the Rule of Law*, BOS. U. L. REV. 69 (1989); Joseph Raz, *The Rule of Law and its Virtue*, in THE AUTHORITY OF LAW (Oxford University Press, 1979).

B. The Legal Concept(s) of Intentional Action

The experimental results suggest that there is a distinctive concept of intentional action that is acquired through legal training and experience. This concept has some different features from the ordinary concept of intentional action. The ordinary concept of intentional action seems particularly connected to judgments about blame (see, e.g., Experiment 2). A foreseeable side-effect becomes more “intentional” as it become more blameworthy (Experiment 2) or severe (Experiment 1). In contrast, the legal concept is not severity sensitive (Experiment 1), nor is it connected tightly to blame judgments (Experiment 2).

So consider these two concepts of intentional action: The ordinary concept, which is responsive to the blameworthiness of the actor and perhaps also the severity of the action, and the legal concept, which is informed primarily by a side-effect’s motivational significance. Which of these concepts *should* the law use as the standard of intentional action?

Before turning to this legal question, it may help to consider some ordinary contexts, in which the ordinary concept is used. For example, consider *moral* decision-making that involves attributions of intentional action. When we come across actors like those in the chairman and beach town cases, we might make moral judgments about whether those persons *intentionally* harmed.

Insofar as these moral decisions aim to allocate blame, the ordinary concept may be very useful. Compare the chairmen in the harm and help cases. Although the side-effect (helping or harming) has the same motivational significance (zero), we might think the chairmen’s attitudes are morally very different. To proceed unflinchingly with an action that has a foreseeably positive side-effect is not very problematic, but to proceed unflinchingly with an action that has a foreseeably negative side-effect is more morally problematic. If the concept of intentional action should tell us something about the type of moral actor we are evaluating, it may be that the features of the ordinary concept are actually apt.

Of course, there is significant debate about this question in the moral domain.¹⁴² But it remains a viable option that some of these surprising features of the ordinary concept are actually apt, given the moral and social contexts in which the concept operates.

Now consider the legal context relevant to the cases here: intentional torts. There is significant debate about the aims of tort law and whether intentional torts carve out a specific set of interests. The aims of intentional torts can also be illuminated by comparing intentional torts to criminal law and negligence law.

First compare intentional torts with criminal law. Historically, there is some conceptual overlap between torts and criminal law. The early common-law courts “approach[ed] the field of tort through the field of crime.”¹⁴³ “Due to the historical influence of the criminal law on tort law, the basic categories of criminal law may have also largely defined the basic categories of tort law.”¹⁴⁴ However, while criminal law is focused on questions of culpability, one may be liable for an intentional tort even without culpability. For example, if I mistakenly take your possession, I am not criminally culpable, but I might be liable in (intentional) tort.

When comparing intentional torts and negligence law, most scholars note a distinction in the underlying motivations.¹⁴⁵ Broadly speaking, negligence law concerns duties owed *to an injured victim*, while intentional torts focus on *aggressive* behavior of the actor. The “duty of care in negligence... [is] a duty not to injure others by conduct that is careless as to them.”¹⁴⁶ Similarly, the “reckless actor ... is not wrongful because he aims to impose his will on another [as in the

¹⁴² See generally Edouard Machery, *The Folk Concept of Intentional Action: Philosophical and Experimental Issues*, 23 MIND & LANG. 165 (2008); K. Uttich & Tania Lombrozo, *Norms inform Mental State Ascriptions: A Rational Explanation for the Side-Effect Effect*, 116 COGNITION 87 (2010); Chandra Sripada, *The Deep Self Model and Asymmetries in Folk Judgments about Intentional Action*, 151 PHIL. STUD. 159 (2009).

¹⁴³ 2 FREDERICK POLLOCK & FREDERIC WILLIAM MAITLAND, *THE HISTORY OF ENGLISH LAW* 530 (2d ed. 1968).

¹⁴⁴ Mark A. Geistfeld, *Conceptualizing the Intentional Torts*, 10 J. TORT L. 159 (2017).

¹⁴⁵ But see Stephen D. Sugarman, *Restating the Tort of Battery* (2017) (arguing that both should fall under a new conception of torts focused on *wrongful action*).

¹⁴⁶ John C.P. Goldberg & Behmain Zipursky, *Torts as Wrongs*, 88 TEX. L. REV. 917, 935 (2010).

case of aggression]. He is wrongful because he imposes his will on the world without regard for the consequences of its imposition.”¹⁴⁷

With these differences in mind, consider the aims of intentional torts. Intentional torts “protect fundamental interests in autonomy, dignity, and security.”¹⁴⁸ On an influential account, the “element of intent has a clear substantive purpose [in intentional torts]: it determines whether or not an interaction is aggressive and properly governed by the intentional torts.”¹⁴⁹

This account, on which intentional torts govern aggressive behavior, makes sense of a number of distinguishing features of intentional torts. For one, “[t]heir shape is more rule-like and more precise than negligence doctrines, which often take the form of flexible and general standards of reasonableness.”¹⁵⁰ Negligence, which traditionally focuses on the *compensation* of an injured person, may be more flexible in determining whether someone acted unreasonably. But intentional torts, which focus on *aggressive behavior*, may be more rule-bound.

On this account, it is not surprising that intentional torts often admit of greater damages. “In the great majority of states, punitive (or “exemplary”) damages may be awarded when the plaintiff has suffered legally recognized harm and the tortfeasor has committed quite serious misconduct with a bad intent or bad state of mind such as malice.”¹⁵¹ This flows from a similar conception of intentional torts. The motivation is not about equitable compensation of an injury, but rather responsiveness to a particular intent or state of mind.

With respect to these traditional aims and features of intentional torts, the distinctive legal concept of intentional action may seem very sensible. The concept should not be flexible (e.g. pliable in light of the severity of the outcome), but rather should reflect a clear indication of what types of actions are proscribed. In the moral context, when we evaluate intentional action, we are

¹⁴⁷ Geistfeld, *supra*, (quoting Anthony J. Sebok, *Purpose, Belief, and Recklessness: Pruning the Restatement (Third)'s Definition of Intent*, 54 VAND. L. REV. 1165, 1179 (2001)).

¹⁴⁸ RESTATEMENT (THIRD) OF TORTS: INTENTIONAL TORTS TO PERSONS, Scope Note to Project (AM. LAW INST. Tent. Draft No. 1, April 8, 2015).

¹⁴⁹ Geistfeld, *supra*, at 15.

¹⁵⁰ RESTATEMENT (THIRD) OF TORTS: INTENTIONAL TORTS TO PERSONS, Scope Note to Project (AM. LAW INST. Tent. Draft No. 1, April 8, 2015).

¹⁵¹ DAN B. DOBBS ET AL., 1 DOBBS' LAW OF TORTS § 483 (2d ed. 2011).

often concerned with blame—did the actor do something wrong? However, the intentional tort question is closer to a question about deliberately aggressive action—did the actor do something bad on purpose?

Similar reasoning might suggest that some features of the ordinary concept are not well suited to the aims of intentional tort law. For example, in an ordinary or moral context, it might be sensible to apply intentionality in a way that is responsive to severity. But in a legal context, severity sensitivity permits a type of moral luck that law might deem unsavory. Across all of the experimental cases, the chairman has the same motivational attitude (disinterest in the additional effects of his actions), but we attribute intentionality to his case in some circumstances in which the outcome happens to be severe. Insofar as motivational significance is the only relevant attitude for ascribing intentional action, severity of the outcome should not change attributions of intentionality.

However, this is not to say that there are no considerations pointing in the other direction. There are also some arguments in favor of incorporating the ordinary concept of intentional action into law.¹⁵² For example, perhaps we might understand intentional torts not as governing *aggressive* behavior, but rather as assigning liability to injury-causing actions that are produced in deliberate violation of *acceptable community standards*. The perceived severity of an outcome may be an indicator of the action's departure from such a standard. A business operation that has some modest negative effects on nearby animals may seem more consistent with our standards than an operation that kills nearby animals. If so, this might provide a way in which severity sensitivity would actually represent an apt part of the legal concept.

¹⁵² See, e.g. Julia Kobick, *Discriminatory Intent Reconsidered: Folk Concepts of Intentionality and Equal Protection Jurisprudence*, 45 HARV. C.R.-C.L. L. REV. 45, 517 (2010) (arguing that law should accommodate the valence-sensitive folk notion of intentionality in equal protection jurisprudence).

Perhaps it is not right (normatively) for the law to fail to hold people accountable in intentional tort for severe foreseeable side-effects. These should not merely be governed by negligence law, one might argue, but also by intentional torts. If there is a moral difference between a person indifferent to possibly harming a little and possibly harming a lot, perhaps there should also be a legal difference.

This kind of reasoning could also inform proposals for tort reform. Consider Sugarman's proposal to eliminate the intentional tort of battery and merge both negligent and intentional imposition of physical harm into a new tort focused on an action's *wrongness*.¹⁵³ The experimental results enrich this debate by providing facts about how the ordinary and legal concepts depart. If, for instance, tort law considered *wrong* actions rather than *intentional* ones, we may find that outcome severity should play a role.

Whether or not tort law *should* be severity sensitive is a difficult question, and the essay cannot resolve all those questions here. But it does suggest one way in which such radical proposals might be implemented more easily: Rather than re-writing tort law, we might instead reconceptualize it, bringing the ordinary concept of intentional action to bear.

Ultimately, this section does not aim to settle this specific question—should law's concept of intentional action reflect the specific features of the ordinary concept? The primary contribution is to call attention to the fact that here there *is* a question—the legal concept of intentional action is dissimilar to the ordinary one, but should it be?—and that experimental findings can enrich these debates by identifying various features of both concepts. In addition to making general arguments about ordinary versus legal concepts (e.g. regarding publicity), we can also ask more detailed normative questions (e.g. about the legal significance of severity sensitivity).

¹⁵³ Sugarman, *supra*.

C. Legal Expertise

While debate about which concept law *should* employ remains open, the essay's results do defuse an experimental challenge to law. Recall the experimental challenge: The ordinary concept of intentional action is severity sensitive, and severity sensitivity is a bias, so we have reason to worry that legal expert decision making is plagued by biased judgments of intentional action.¹⁵⁴ This experimentalist challenge is often couched in broad terms, recommending radical legal implications. As Kneer & Bourgeois-Gironde put it, the severity-sensitive concept of intentionality threatens “large-scale inner-systemic incoherence” to “nearly every system of criminal law in the world.”¹⁵⁵

The findings presented here suggest that this challenge is too quick. Legal experts have a different concept of intentional action, which they use in legal contexts.¹⁵⁶

This supports an important conclusion about the nature of legal expertise. Legal expertise does not consist in merely learning *new* concepts—ones that have no ordinary counterpart (like *parol evidence* or *disgorgement*). Legal expertise also involves *conceptual modification*. Legal experts have a legal concept of intentional action that bears some resemblance to the ordinary concept, but also differs in significant ways.

Importantly, the results suggest that this expertise *develops* with legal training and experience. It is not the case that some smart or well-read people gain access to the legal concept of intentional action. Instead, the concept emerges over time, with exposure to legal training and culture.

¹⁵⁴ E.g. Ginther et al., *supra* note 39; Kneer & Bourgeois-Gironde, *supra* note 1; Morrison *supra* note 39; Nadelhoffer *supra* note 39.

¹⁵⁵ Kneer & Bourgeois-Gironde, *supra* note 1, at 6.

¹⁵⁶ That said, the debate is certainly not over. The experimentalist challenge assumed that the ordinary concept is biased. However, a different question remains: Is the legal concept used by experts actually the *right* concept?

D. New Challenges: Interpretation and Judges Versus Juries

The results also bear on debates about legal interpretation. Some legal concepts are expressed by technical terms (e.g. “ex post facto law,” “parol evidence, “writ of assistance”), and others through explicit definitions (e.g. statutory definitions). But many terms in law are not expressed by unique technical terms or special definitions. Instead, they are expressed by a term that typically expresses an *ordinary* concept. In many cases, legal scholars argue that we should interpret texts through their ordinary meaning. But the results here raise a question: What is the ordinary meaning of “intentionally”?

One tempting response is that the ordinary meaning of “intentionally” should be informed largely by how ordinary people understand the term. But this would suggest that the correct interpretation of the intentional injury to companion animals statute is to impose liability on those who merely foresee negative side-effects. This is inconsistent with the way in which the vast majority of judges and law students understood that statute.

The opposite response is that the ordinary meaning of “intentionally” should be informed largely by what those with legal training think. Problems with this account are immediately apparent. The typical motivations of ordinary meaning theory are inconsistent with this account. If the “ordinary meaning” of intentionality is not accessible to ordinary people, this threatens legal values like publicity, clarity, and democracy.

The puzzle, then, is how best to articulate the relationship between terms in law that have an ordinary language counterpart—terms like “cause,” “intend,” “consent,” and “reasonable.” I maintain that there is no easy answer. Many of the “general” arguments in favor of using ordinary or legal concepts apply in this context. Interpreting ordinary meaning in line with ordinary understanding promotes legal values like democracy, publicity, and clarity. But interpreting ordinary meaning in line with legal understanding protects law from being applied in line with ordinary misunderstandings and respects the sophistication and aptness of law.

A very practical corollary to the interpretation problem concerns divergent judgments between judges and juries. Assume that judges have a legal concept (e.g. of intentional action) that differs from most people's ordinary concept (e.g. of intentional action). We might expect in such a scenario, that different legal outcomes might arise based on whether the case is heard before a judge or jury. Insofar as law should be applied consistently, this presents a problem.

Perhaps, one might reply, this "problem" is already acknowledged by the legal system. Juries composed of different people might reach different verdicts on the same facts. So too might different judges. We aim for law to be applied neutrally, but there inevitable inconsistencies in most legal systems.

However, this response is inadequate. For one, inconsistency concerns are not resolved by pointing to other inconsistencies. Those concerned about legal consistency do not generally commit themselves to all-or-nothing claims. The choice is not between a perfectly consistent legal system or *anything* else. All else equal, a more consistent legal system is preferable.

Second, there is a plausible distinction between random and systematic legal inconsistency. This is a theoretically important distinction; small random inconsistencies may be compatible with legal neutrality, but systematic inconsistencies are not. But the distinction is also of likely practical significance. Insofar as there are systematic patterns of inconsistency, this provides opportunity for comparatively sophisticated actors to realize advantages. Random inconsistency should break even; systematic inconsistency privileges those with the resources to identify and leverage those inconsistencies.

CONCLUSION

This essay considers a broad philosophical question: What is the relationship between ordinary and legal concepts? To address this question, it studies a specific example: the concept of intentional action. Experimental evidence suggests that the legal concept of intentional action is not perfectly identical to the ordinary concept. The legal concept is related to the ordinary

concept, but it is nevertheless distinctive in several significant ways. Importantly, the data suggest that the differences are best explained by legal training, rather than a selection effect. This training teaches expertise in legal concepts that trained experts can deploy in distinctive legal contexts.

The results bear on the relationship between empirical science and legal philosophy. Experimental studies in cognitive science often suggest that discoveries about ordinary concepts have legal implications. Typically, scholars assume that findings about ordinary concepts and cognition generalize to legal decision making contexts involving ordinary language (e.g. “intentionally,” “cause,” “consent,” “reasonable”). The results suggest that, at least sometimes, this assumption is too quick. Legal training facilitates the acquisition of legal concepts that have distinctive features. Contrary to the assumption of some experimentalists, these concepts are not necessary identical to the ordinary ones. As such, experimental findings about ordinary concepts do not necessarily carry straightforward legal implications.

The essay defuses some provocative challenges from experimentalists, but the results also raise new questions and challenges. For one, when the concepts of ordinary people and legal experts diverge, which should be the legal concept? The results raise a related puzzle for theories of interpretation, namely those that claim legal terms should be given their “ordinary meaning.” If the two concepts of intentional action correspond to two separate meanings, which of those is *the* ordinary meaning: what non-legally trained people understand, or what legally-trained people understand? Finally, a practical problem arises for judges and juries. Is it problematic that judges and juries understand “intentionally” differently? The answer is almost certainly yes, and this is likely a difficult problem to resolve.

Ultimately, the essay represents a proof of concept of a distinctive type of legal expertise. It is well known that legal training provides experts with new concepts that lack ordinary counterparts

(e.g. *parol evidence*, *stare decisis*). But legal training produces another type of expertise, expertise in legal concepts that correspond to ordinary terms (like “intentionally”).

The question of whether law *should* use ordinary or legal concepts remains a difficult one. The essay maintains that experimental study into ordinary and legal concepts can help enrich these debates. The most sophisticated debates about whether law should use ordinary or legal concepts will make reference to the features of the relevant concepts, features that experiments can help illuminate.

CHAPTER 3

TESTING ORIGINAL PUBLIC MEANING: ARE DICTIONARIES AND CORPUS LINGUISTICS RELIABLE MEASURES OF MEANING?

Influential theories of interpretation recommend using dictionaries and corpus linguistics to discover a legal text's "original public meaning." This is typically understood as an empirical inquiry, aiming to discover *facts* about how ordinary people comprehended language. However, a central question remains open: Do popular methods of corpus linguistics and dictionary-use accurately reflect original public meaning?

To assess this question, this paper develops a novel method of "experimental jurisprudence." A series of experimental studies (N = 4,162) reveals systematic divergences among the verdicts delivered by modern concept use, dictionary use, and corpus linguistics use. For example, today people apply the concept of a vehicle differently from the way in which they apply modern dictionary definitions or modern corpus linguistics data concerning vehicles. The same results arise across levels of legal expertise—participants included 230 "elite-university" law students (e.g. at Harvard and Yale) and 98 United States judges—and for various terms and phrases, such as "vehicle," "labor," "weapon," "carrying a firearm," and "tangible object." Ultimately, the data suggest that popular methods of dictionary-use and corpus linguistics carry serious risks of error—conservatively estimated, a 20-35% risk on average. And in some circumstances, expert use of these methods carried extremely large error rates—between 80-100%. These findings shift the burden to theorists and practitioners that rely on these tools, to articulate and demonstrate a reliable method of interpretation.

CONTENTS

INTRODUCTION	153
I. PUBLIC MEANING ORIGINALISM	159
<i>A. Original Public Meaning</i>	159
<i>B. Theories of Interpretation</i>	160
II. SOURCES OF INTERPRETIVE EVIDENCE	162
<i>A. Linguistic Intuition</i>	162
<i>B. Dictionaries</i>	164
<i>C. Corpus Linguistics</i>	166
III. PUBLIC MEANING ORIGINALISM’S EMPIRICAL ASSUMPTIONS.....	167
<i>A. Empirical Critiques of Public Meaning Originalism</i>	167
<i>B. Are Dictionaries and Corpus Linguistics Reliable? A Thought Experiment</i>	169
IV. AN EXPERIMENTAL TEST OF DICTIONARIES AND CORPUS LINGUISTICS	172
<i>A. Experiment 1</i>	175
<i>B. Experiment 2: The Process of Using Dictionaries and Corpora</i>	179
<i>C. Experiment 3: Expert Judges</i>	183
<i>D. Experiment 4: Replication Across Ten Examples</i>	185
<i>E. Summary and Interpretation</i>	188
1. Judges and Non-Experts Judge Meaning Similarly.....	189
2. Ordinary Meaning is Often Unclear.....	190
3. Judges and Non-Experts Apply Dictionaries Similarly.....	190
4. Judges and Non-Experts Apply Corpus Linguistics Similarly	191
5. Corpus Linguistics Use Reflects Narrow, Prototypical Uses.....	192
6. Dictionary-Use Can Reflect Extensive Uses.....	193
7. Dictionaries and Corpus Linguistics Provide Divergent Verdicts	194
8. Dictionaries and Corpus Linguistics Diverge from Ordinary Judgment.....	194
9. On Average, Corpus Linguistics and Dictionaries Had 20-35% Error Rates	195
10. In Some Circumstances, Error Rates Reached 80-100%.....	198
V. HOW JURISTS USE DICTIONARIES AND CORPUS LINGUISTICS.....	199
<i>A. Dictionaries Often Supply “Broad” Senses and Corpus Linguistics Provides “Narrow” Ones</i>	200
<i>B. Political Uses of Dictionaries and Corpus Linguistics</i>	205
<i>C. Explaining Debates About Original Public Meaning</i>	211
VI. IMPLICATIONS.....	212
<i>A. Understanding and Improving Originalist Methodology</i>	213
1. Psychological Mechanisms of Use of Dictionaries and Corpora	213
2. Guidelines for Using Originalist Empirical Tools.....	217
<i>B. Challenges to Originalist and Textualist Methodologies</i>	218
1. Fallacies of Corpus Linguistics and Dictionary Use	220
2. The Burden of Demonstrating a Non-Arbitrary and Reliable Methodology	224
CONCLUSION.....	227
APPENDIX A: EXP. 1 ADDITIONAL MATERIALS AND FOLLOW-UP EXPERIMENTS	229
APPENDIX B: EXP. 2 ADDITIONAL MATERIALS AND FOLLOW-UP EXPERIMENTS	237
APPENDIX C: EXP. 3 ADDITIONAL MATERIALS AND FOLLOW-UP EXPERIMENTS	242
APPENDIX D: EXP. 4 ADDITIONAL MATERIALS	251

INTRODUCTION

Theories of constitutional and statutory interpretation often seek to discover a legal text's "original public meaning." There is significant debate about how to elaborate this concept of "original public meaning" and what role it should play in legal decision-making.¹ But broadly speaking, original public meaning is the text's communicative content, or the ordinary meaning of the text to members of the relevant community, at the time it became law.

A diverse range of scholars and jurists increasingly endorse the relevance of public meaning, and many accept that a legal text's original public meaning should constrain legal practice.² Of course, emphasis on original public meaning is not central to all theories of interpretation, and notable detractors question the empirical assumptions required to discover original public meaning: Do judges actually have the ability, insight, or tools to determine the original public meaning of legal texts?³

That critique highlights a crucial insight. Original public meaning inquiries are typically understood as *empirical* ones, which aim to discover descriptive *facts* about meaning.⁴ Theories holding that a legal text must be applied consistently with its original public meaning (hereinafter theories of "Public Meaning Originalism") do not characterize their project as a normative one. Rather than debating how a text should be understood, Public Meaning Originalism asks how a text was *in fact* understood.

¹ See e.g., Victoria Nourse, *Reclaiming the Constitutional Text from Originalism: The Case of Executive Power*, 106 CALIF. L. REV. 1 (2018); Lawrence B. Solum, *The Constraint Principle: Original Meaning and Constitutional Practice* (Apr. 11, 2018). Following Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. 788, 825-826 (2018), I use "original public meaning" to refer to a legal text's communicative content (or "ordinary meaning") at the relevant time. For example, the original public meaning of a 1967 statute is that text's ordinary meaning in 1967.

² See, e.g., WILLIAM N. ESKRIDGE, JR. INTERPRETING LAW: A PRIMER ON HOW TO READ STATUTES AND THE CONSTITUTION 35 (2016) ("[t]here are excellent reasons for the primacy of the ordinary meaning rule"); Lee & Mouritsen, *supra* note 1; Lawrence B. Solum, *Surprising Originalism*, CONLAWNOW (2018); see also Elena Kagan, *The Scalia Lecture: A Dialogue with Justice Kagan on the Reading of Statutes*, HARV. L. TODAY (Nov. 17, 2015), <http://today.law.harvard.edu/in-scalia-lecture-kagan-discusses-statutory-interpretation/> ("[W]e're all textualists now").

³ E.g. Richard H. Fallon, Jr. *The Meaning of Legal "Meaning" and Its Implications for Theories of Legal Interpretation*, 82 U. CHI. L. REV. 1235 (2015); Cass Sunstein, *There is Nothing that Interpretation Just Is*, 30 CONST. COMMENT. 193, 194-195 (2015).

⁴ E.g. Randy E. Barnett, *Interpretation and Construction*, 34 HARV. J.L. & PUB. POL'Y 65 (2011) ("It cannot be overstressed that the activity of determining semantic meaning at the time of enactment required by the first proposition is empirical, not normative." (citing KEITH E. WHITTINGTON, CONSTITUTIONAL INTERPRETATION: TEXTUAL MEANING, ORIGINAL INTENT, AND JUDICIAL REVIEW 5-14 (1999))).

There are several empirical methods that one might use to discover a text's original public meaning, including intuition, dictionaries, and corpus linguistics.⁵ Dictionary-use⁶ and corpus linguistics⁷ are increasingly popular methods. This popularity is not difficult to explain. These methods are relatively easy to use. Moreover, they often seem "objective" and even "scientific."⁸

When Public Meaning Originalism employs dictionaries or corpus linguistics in legal interpretation, they use these as tools of an empirical inquiry. As such, the tools are properly understood as *empirical tools*, which might be inaccurate. Although the use of these tools seems to grow more sophisticated,⁹ the accuracy of dictionaries or corpus linguistics has never been rigorously assessed.¹⁰ That is, there is no test of whether these tools actually track (original) public meaning. Instead, Public Meaning Originalism and other theories relying on these tools typically *assume* that dictionaries and corpus linguistics are accurate measures of original public meaning.¹¹

This paper develops a novel method within "experimental jurisprudence" to test the accuracy of dictionaries and corpus linguistics in interpretation. This provides evidence about the justifiability of core methodologies of theories like Public Meaning Originalism. Insofar as a legal interpretive theory relies upon these tools, the study here also provides evidence about the success

⁵ See Lawrence B. Solum, *Originalist Methodology*, 84 U. CHI. L. REV. 269 (2017).

⁶ John Calhoun, Note, *Measuring the Fortress: Explaining Trends in Supreme Court and Circuit Court Dictionary Use*, 124 YALE L.J. 484 (2014).

⁷ Lee & Mouritsen, *supra* note 1; see also Calhoun *supra*; Lee J. Strang, *How Big Data Can Increase Originalism's Methodological Rigor: Using Corpus Linguistics to Reveal Original Language Conventions*, 50 U.C. DAVIS L. REV. (2017); Evan Zoldan, *Corpus Linguistics and the Impossible Dream of Objectivity* (draft manuscript).

⁸ See, e.g., Lawrence Solan, *Can Corpus Linguistics Help Make Originalism Scientific?* 126 YALE L.J. F. 57 (2016).

⁹ E.g. Lee & Mouritsen, *supra* note 1; Lawrence M. Solan & Tammy Gales, *Corpus Linguistics as a Tool in Legal Interpretation* BYU L. REV. (2018).

¹⁰ There are two published experimental surveys about originalism. Donald L. Drakeman, *What's the Point of Originalism*, 37 HARV. J. L. & PUB. POL'Y 1124; Jamal Green, Nathaniel Persily & Stephen Ansolabehere, *Profiling Originalism*, 111 COLUM. L. REV. 356 (2011). These are fascinating studies, but neither tests the reliability of originalist methodology. Instead, they focus on questions such as why people are originalists. For other important empirical studies of judicial interpretation, see 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 (2013); Lisa Schultz Bressman & Abbe R. Gluck, *Statutory Interpretation from the Inside—An Empirical Study of Congressional Drafting, Delegation, and the Canons: Part II*, 66 STAN. L. REV. 725 (2014).

¹¹ For a demonstration of the use of corpus linguistics see Lee & Mouritsen, *supra* note 1. However, that demonstration is not necessarily a demonstration of a *reliable* method of corpus linguistics.

of such legal theories.¹² Part I outlines the background to these debates and legal theories on which (original) public meaning is significant. Part II surveys the tools that provide interpretive evidence for those theories, including dictionaries and corpus linguistics.

Parts III and IV consider whether these tools are accurate measures of meaning. One reason that the tools of original public meaning have not yet been tested is that such a test may seem impossible—how can we evaluate the accuracy of a 18th century dictionary or corpus linguistics search without facts about the way in which the contested term was actually understood at the time (i.e. without the best data about its original public meaning)?

However, what we can do is evaluate whether *modern* uses of contemporary dictionaries and corpus linguistics reflect terms' *modern* public meanings. If a tool (e.g. dictionary-use) performs poorly in modern interpretation, so long as there are no historically distinguishing factors (i.e. reasons that use of an 18th century dictionary is *more* accurate in historical interpretation than use of a modern dictionary is in modern interpretation), this gives us some evidence that the method is also unreliable in estimating original public meaning.

Part III begins with a philosophical thought experiment: Do originalism's tools accurately reflect *modern* public meaning? If they do not achieve even this task, it seems unlikely that they accurately reflect *historical* public meaning.

Part IV presents an experimental investigation of dictionaries and corpus linguistics, testing their reliability and revealing some of the cognitive processes underlying their uses. Experimental studies (total N = 4,162) show systematic divergences among the verdicts delivered by ordinary concept use, dictionary use, and corpus linguistics use. For example, the way in which people today apply the concept of a vehicle is systematically different from the way in which people apply the modern dictionary definition of a "vehicle" or modern corpus linguistics data

¹² For example, if the study shows that originalist uses of dictionaries and corpus linguistics are not accurate measures of meaning, and Public Meaning Originalism requires the accuracy of either of these methods, the study supports a broader argument against Public Meaning Originalism.

concerning vehicles. This pattern of results arises across samples of ordinary people (N = 3,834), elite law students (N = 230), and United States judges (N = 98).

Section IV.E contains a crucial summary and interpretation of the main experimental findings. There are ten original and important findings. Three concern the surprising similarity between expert and non-expert interpreters (e.g. judges vs. the general population). Judges, law students, and ordinary people were striking similar in their ordinary conceptual judgments (IV.E.1), use of dictionaries (IV.E.3), and use of corpus linguistics (IV.E.4).

Another important finding concerns the modern ordinary meaning of various terms; in *many* cases, there was significant disagreement about ordinary meaning—for judges, law students, and the general population alike (IV.E.2). That is, although there was consistency across these groups, within every group there was disagreement. For example, ordinary people were divided on whether a canoe is a vehicle; the same is true for law students and judges.

Two more findings concern the process of using dictionaries and corpus linguistics. Across the studies, users of corpus linguistics tended to identify prototypical examples (e.g. a car is a vehicle) better than non-prototypical examples (e.g. a moped or airplane is a vehicle) (IV.E.5). Conversely, users of dictionaries sometimes made very extensive judgments (e.g. a pair of rollerskates is a vehicle) (IV.E.6).

The next findings are that the verdicts of both corpus linguistics and dictionary use diverged from ordinary meaning, in a large number of cases (IV.E.7). Moreover, the verdicts of corpus linguistics and dictionary use diverged from each other (IV.E.8).

The final two sub-sections of Section IV.E quantify the error of dictionaries and corpus linguistics. Section IV.E.9 considers the results across a diverse range of plausible assumptions. Across all these different models, each using reasonably conservative estimates, dictionaries and corpus linguistics had between 20-35% error rates. That is, across all the levels of expertise—ordinary people, law students, and judges—the data suggest that relying on a dictionary definition

or corpus linguistics data would lead users to the wrong judgment fairly often: once in every three to five cases.

Section IV.E.10 notes that this 20-35% *average* rate of error may not tell the whole story. The experiments included a number of relatively easy categorizations (e.g. whether a car is a vehicle). Insofar as real legal decisions concern relatively difficult categorizations, it may also be instructive to consider the maximum error rate: What percent of (e.g.) judges using dictionary or corpus linguistics evaluated *the hardest interpretive question* incorrectly? Across all levels of expertise, the data suggest that in some examples, relying on a dictionary definition or corpus linguistics data led 80-100% of users to the incorrect verdict.

Part V considers the ecological validity of the experimental results and interpretation. Do legal uses of dictionaries tend to reflect broad, extensive interpretations, while legal uses of corpus linguistics tend to reflect narrow, prototypical uses? Part V.A finds that caselaw tends to refer to dictionary definitions as “broad” more often than as “narrow,” although dictionary definitions are sometimes narrowed by emphasizing context, which definition is most relevant, or particular features of the definition(s). Corpus linguistics has not yet been used frequently in caselaw, but many uses recommend a narrow meaning.

Part V.B considers whether the nature of dictionary definitions—they generally recommend broad extensive meanings, but ones that can be narrowed—may admit of political decision making. The Part considers two examples from the Bill of Rights, from the Second and Eighth Amendments, each of which contains three terms: “keep and bear arms” and “cruel and unusual punishment.” Republican-appointed justices, at the Supreme Court and Circuit Court level, more frequently cite dictionaries to interpret the former set of terms broadly. Conversely, in Eighth Amendment cases, Republican-appointed justices interpret the broad dictionary definition narrowly. Although Democratic-appointed justices cite dictionaries less frequently, when they do, the pattern is reversed: dictionaries indicate that “cruel and unusual punishment” is broad, but that

“keep and bear arms” is narrow. The dictionary-extensive, corpus-narrow relationship holds for other divisive examples. “Emoluments” seems broad when scholars consider dictionaries, but narrow when they consider corpora. So too for “Commerce.”

Part VI elaborates the normative implications of the experimental results. Constructively, the results provide guidance for originalist and textualist theories of interpretation. The results demonstrate one systematic pattern of judgment; while corpus linguistics tends to track prototypical uses, dictionaries tend to elicit more extensive ones. For example, consider a non-prototypical vehicle, like an airplane. Today, people generally judge that an airplane is a vehicle.¹³ That judgment is also reflected by participants’ use of modern dictionaries, but users of corpus linguistics are hesitant to categorize airplanes as vehicles.¹⁴ A similar pattern arises for other category members; corpus linguistics sometimes provides good evidence about prototypical members, but often fails to provide good evidence about non-prototypical members. Insofar as legal texts indicate the relevance of a very extensive concept application, dictionaries may be more useful. But insofar as legal texts indicate the relevance of only highly prototypical uses, corpus linguistics is more helpful.

Critically, the results support five common fallacies in the use of dictionaries and corpus linguistics. More broadly, the results suggest that dictionaries and corpus linguistics—two central tools of discovering “original public meaning”—are unreliable in interpretation, with error rates plausibly in the range of 20-35% and perhaps as high as 80%. This shifts the argumentative burden to those who use and defend these tools. Public Meaning Originalism must provide a principled account of the use of these tools and a demonstration of how error can be avoided.

¹³ See Part IV, *infra*.

¹⁴ *Id.*

I. PUBLIC MEANING ORIGINALISM

A. *Original Public Meaning*

What is “original public meaning”? As many have noted, there is some irony in the fact that terms like “(original) public meaning,” “plain meaning,” and “ordinary meaning” are notoriously unclear.¹⁵ This paper addresses (original) public meaning, although those interested in other concepts like “plain meaning” may find that the experimental work also has implications for those. Following recent work on corpus linguistics in originalist and textualist interpretation, I treat “ordinary” and “public” meaning coextensively when we consider non-specialized terms and phrases.¹⁶

One of the key features of a text’s “original public meaning” is that it is *not* merely the product of the drafters’ intentions. Public meaning is what the law communicates to its readers, which is not necessarily what the drafters aimed to accomplish in drafting the laws.¹⁷ For example, Public Meaning Originalism “seeks to determine ‘the meaning the words and phrases of the Constitution would have had, in context, to ordinary readers, speakers, and writers, of the English language, reading a document of this type, at the time adopted.’”¹⁸ In most cases, the text’s communicative content is simply its ordinary meaning.¹⁹

Despite continuing debate about the precise contours of “original public meaning,” there is actually remarkable consistency on this point: Original public meaning is a *fact* about how readers of the text would have understood it. Original public meaning is “the likely original

¹⁵ E.g. WILLIAM N. ESKRIDGE, JR. ET AL., *CASES AND MATERIALS ON LEGISLATION: STATUTES AND THE CREATION OF PUBLIC POLICY* 792-93 (4th ed. 2007); Lee & Mouritsen, *supra* note 1; Richard A. Posner, *Statutory Interpretation—in the Classroom and in the Courtroom*, 50 U. CHI. L. REV. 800 (1983).

¹⁶ See Lee & Mouritsen, *supra* note 1.

¹⁷ See generally Solum, *supra* note 1; see also James C. Phillips, Daniel M. Ortner & Thomas R. Lee., *Corpus Linguistics & Original Public Meaning: A New Tool to Make Originalism More Empirical*, 126 YALE L.J. F. 21 (2016).

¹⁸ Phillips, Ortner & Lee, *supra* note 17 (citing Vasana Kesavan & Michael Stokes Paulsen, *The Interpretive Force of the Constitution’s Secret Drafting History*, 91 GEO. L.J. 1113, 1118 (2003)).

¹⁹ Lee & Mouritsen, *supra* note 1, at 792. Lee & Mouritsen argue for this equivalence, except in the case of specialized legal language (e.g. “bill of attainder” or “parol evidence”). Part IV’s experiments and this paper’s arguments set aside study of specialized legal language. As such, following Lee & Mouritsen, I treat ordinary meaning to be equivalent to a text’s communicative content or public meaning.

understanding of the text at the time of its adoption by competent speakers of the English language”²⁰ or “what readers of the historically situated text would have understood the constitutional language to express.”²¹

The “original” in original public meaning refers to the time of the text’s passage or ratification. The original public meaning of a text is the public meaning at the time the text became law.

Consider a very well known example. Imagine that a historical statute, written in 1958, contains the provision: “no vehicles in the park.”²² Now today there are three legal disputes about whether entities that have entered the park have violated the statute: a car, a bicycle, and a horse-drawn carriage. Those debating the legal effect of the statute might argue that its modern legal effect is constrained by the original meaning of “vehicle” in 1958. That is, the ordinary or public meaning of “vehicle” in 1958 fixes the meaning of the statute and thereby constrains how it applies today.

B. Theories of Interpretation

Thirty years ago, Justice Scalia introduced the interpretive theory of “new textualism,”²³ noting that, “[e]very issue of law I resolve as federal judge involves interpretation of text—the text of a regulation, or of a statute, or of the Constitution.”²⁴ The rise of “new textualism”—broadly speaking, the theory that plain and clear text is decisive of legal effect—comes alongside the rise of “new originalism”—broadly speaking, the idea that original public meaning (and not drafters’ intentions) constrains interpretation. These views have fused into a modern thesis of

²⁰ KURT T. LASH, *THE FOURTEENTH AMENDMENT AND THE PRIVILEGES AND IMMUNITIES OF AMERICAN CITIZENSHIP* 277 (2014)

²¹ Christopher R. Green, *The Original Sense of the Equal Protection Clause: Pre-Enactment History*, 19 *GEO. MASON U. C.R. L.J.* 1, 12 (2008).

²² H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 *HARV. L. REV.* 593 (1958); see also Lee & Mouritsen, *supra* note 1, at 800.

²³ Antonin Scalia, *Common-Law Courts in a Civil-Law System: The Role of United States Federal Courts in Interpreting the Constitution and Laws*, *TANNER LECTURES* 88 (1995).

²⁴ *Id.*

Public Meaning Originalism, a thesis inspiring originalist and textualist views in both constitutional and statutory interpretation.²⁵

Unsurprisingly, textualist and originalist theories place great significance on the (original) public meaning of the text. On the most popular version of these theories, the original public meaning of legal text constrains legal practice.²⁶

However, (original) public meaning is a significant fact to a range of other legal-interpretive theories. For example, pluralist theories might take original public meaning to be one of several relevant considerations in interpretation.²⁷ Even if original public meaning does not necessarily constrain legal effect—for example, it might be overridden by historical, structural, and/or moral considerations—it still plays a role as an important consideration in legal decision making.

The politicization of originalism and textualism can obscure widespread agreement about this point. On many plausible theories of legal interpretation, how the text was understood when it became law is *one* relevant consideration in determining the text's legal effect. So while the stakes of the present project are highest for theories like Public Meaning Originalism, the project also is also relevant to any theory of legal interpretation that places any significance on original ordinary or public meaning.²⁸ Similarly, although the paper focuses on constitutional and statutory law, the findings are also relevant for other areas in which interpretation considers ordinary meaning, such as contractual interpretation.

Of course, some theories of interpretation do not place any interpretive significance on either *original* public meaning or ordinary meaning (at any time). For example, a strong purposivist theory might hold that what determines a text's legal effect is simply the motivating purpose of the text. For example, imagine that we found good evidence that the purpose of the rule “no

²⁵ See, e.g., Solum *supra* note 5; Lee & Mouritsen *supra* note 1.

²⁶ E.g. Solum, *supra* note 5.

²⁷ E.g. Stephen M. Griffin, *Pluralism in Constitutional Interpretation*, 72 TEX. L. REV. 1753 (1994)

²⁸ Of course, on many versions of Public Meaning Originalism, factors beyond public meaning are also relevant when there is no clear ordinary meaning. The most popular version of Public Meaning Originalism focuses on public meaning first. Only if the public meaning is ambiguous, vague, or otherwise underdetermined should we look to other considerations (e.g. intent or purpose).

vehicles in the park” was to keep out noisy vehicles that serve no other useful function in the park. If that purpose determines legal effect, the strong purposivist might hold that the rule allows in an ambulance for an emergency, or a riding lawn-mower for cutting the park’s grass (even if those are both ordinary language “vehicles”). Conversely, the rule might keep out a noisy drone, even if it is not clear that a drone is an ordinary-language “vehicle.”

II. SOURCES OF INTERPRETIVE EVIDENCE

Recall that the search for original public meaning is typically understood as an *empirical* inquiry. Original public meaning is a fact about how the text would have been understood by the public at the time it became law.

However, the fact that many important texts became law far in the past raises an epistemic problem. How do we now reliably determine how the public would have understood the text in historical times? As Lawrence Solum puts it: “the communicative content of the constitutional text is a *fact*. In some cases, there may be epistemic difficulties with discovering the communicative content. Whether we can fully recover the communicative meaning of a particular clause is contingent on our access to the relevant evidence.”²⁹

This Part considers popular sources of originalist interpretive evidence: intuition, dictionaries, and corpus linguistics.³⁰

A. Linguistic Intuition

A common source of interpretive evidence is linguistic intuition.³¹ What does a historical text *seem* to have meant at the time it became law?

In interpretation that seeks to determine the public meaning of a *modern* legal text, linguistic data provides very compelling evidence. People’s collective judgments about the meaning of a

²⁹ Solum, *supra* note 5, at 278.

³⁰ See generally *id.*

³¹ See generally Solum, *supra* note 5, at 281-85.

modern text are extremely closely connected to, if not constitutive of, its modern public meaning.³² Of course, one person's linguistic intuition will not necessarily track these global facts. In fact, research suggests that people often are subject to a false consensus bias, thinking (falsely) that they are good measures of the population's consensus.³³

But in circumstances in which there is very little disagreement, linguistic intuition should be a helpful source in identifying modern public meaning. Imagine again that a local ordinance prohibits all "vehicles" from entering a park.³⁴ Most people today would understand that the ordinance would not prohibit someone from bringing their baby in a baby shoulder-carrier into the park *because* we understand that a baby shoulder-carrier is not a vehicle. Even if there is some disagreement about some entities (e.g. is a skateboard a vehicle?), linguistic intuition provides straightforward guidance in many other cases (e.g. baby shoulder-carriers may be brought into the park, cars must be kept out).

However, linguistic intuitions are not nearly as promising a guide in the search for *original* public meaning. As originalists have cautioned, in the search for original public meaning, "linguistic intuitions formed by immersion in modern linguistic practices can be misleading."³⁵ For a clear example, consider Lawrence Solum's astute observation about the Seventh Amendment's "Twenty Dollars Clause." Most modern readers would assume that this clause refers to the modern Federal Reserve note.³⁶ But, writes Solum, "the word 'dollar' almost certainly referred to the Spanish silver dollar. . . . The 'greenback,' a precursor to the modern note, was not created until much later and was the subject of much controversy."³⁷

³² One natural way to reject the constitution claim begins by noting that individuals sometimes make performance errors. Of course, the claim that most of the population makes a consistent performance error is a much more radical claim. That radical claim is required to reject a connection between ordinary judgments and modern public meaning.

³³ See Solan & Gales, *supra* note 9, at 1333; Lawrence Solan, Terri Rosenblatt & Daniel Osherson, *False Consensus Bias in Contract Interpretation*, 108 COLUM. L. REV. 1268 (2008).

³⁴ This is discussed in much greater detail as the leading example in Parts III-V.

³⁵ Solum, *supra* note 5, at 281.

³⁶ This assumption also arises in some legal scholarship. See, e.g., Note, *The Twenty Dollars Clause*, 118 HARV. L. REV. 1665 (2005).

³⁷ Solum, *supra* note 5, at 282.

Individual intuition is recognized—by originalists and non-originalists alike—as an imperfect source of evidence in modern interpretation and a highly dubious source of evidence in historical interpretation. It is also a source of evidence whose errors are likely hard to identify in practice. Linguistic intuition often *feels* very compelling: Surely “dollar” means dollar. Sometimes, historical research shakes linguistic intuitions of this misplaced confidence. But in legal interpretation, it is a dangerous strategy to rely on unreliable linguistic intuitions *until* and *unless* they are proven erroneous.

There is also an important question of whether intuition-use is even a true *method* of interpretation. While we can generate principles to guide dictionary and corpus use, it is less clear how judges should practice “intuition.” Given the errors that intuition produces in historical interpretation, a method of intuition requires some guiding principles. In the project of determining a text’s original public meaning, it is only defensible to rely on *some methodology*, rather than on some *instinct*—however compelling that instinct feels.

B. Dictionaries

Dictionaries have *prima facie* plausibility as useful sources in interpretation. After all, if the aim is to discover the original public meaning of a term, why not look at how the relevant community defined that term? Recall our example of “no vehicles in the park.” One might seek evidence about the meaning of “vehicle” in 1958 by considering a dictionary definition from that time. As Lee & Mouritsen note, the Webster’s Third New International Dictionary (1961) defines vehicle as a “carrier” or “agent of transmission.”³⁸ Insofar as this definition provides evidence about the public meaning of “vehicle” in 1958, it could help in our hypothetical legal disputes about the car, bicycle, and horse-drawn carriage that entered the park.

³⁸ Lee & Mouritsen, *supra* note 1, at 800.

Part IV presents experimental work that provides evidence about the reliability of dictionaries in interpretation. But there are other aspects of dictionary-use that call for analysis and critique.³⁹ For one, dictionary definitions may be *normative*. That is, while original public meaning theory seeks to determine how some term was—in fact—understood, a dictionary may instead report how some term *should have been* understood. If so, in at least some cases, dictionary definitions are tracking the wrong thing. Rather than tracking public meaning, they might be tracking the dictionary drafter’s conception of desirable meaning. Often, constitutional and statutory interpretation disputes turn on questions about subtle shades of meaning, so such a difference could be consequential.

Another important limitation is that historical dictionaries are less frequent than modern ones. Two important English language dictionaries are published in 1755 and 1828.⁴⁰ Although language change is usually gradual, there are obvious questions about the limits of these dictionaries in interpreting the constitution and early amendments. How reliable is a 1755 dictionary in reflecting the meaning of a provision drafted in 1789?

Finally, historical dictionaries were often the product of a single person’s efforts.⁴¹ Here, too, this raises obvious questions about the reliability of these sources. We might cross-check historical dictionaries to illuminate idiosyncrasies, but the limited number of historical dictionaries severely limits the usefulness of this effort.

Despite all of these concerns, dictionaries are an increasingly popular source of interpretive evidence.⁴² And their use is not without some initial plausibility: It is reasonable to suspect that the public meaning of a term is often reflected well by its definition. Whether this suspicion is true is an open empirical question.

³⁹ See James Brudney & Lawrence Baum, *Oasis or Mirage: The Supreme Court’s Thirst for Dictionaries in the Rehnquist and Roberts Eras*, 55 WM. & MARY L. REV. 483 (2013); see also Solan & Gales, *supra* note 9, at 1334.

⁴⁰ Solum, *supra* note 5, at 283.

⁴¹ *Id.*

⁴² Calhoun, *supra* note 6.

C. Corpus Linguistics

A final source of interpretive evidence is corpus linguistics. Corpora are sets of language data, containing text from books, newspaper articles, online publications, and other sources.⁴³ In recent years, originalist corpus linguistics has evolved from smaller searching to a “big data” approach.⁴⁴

The most prominent defense of this “new corpus linguistics” approach is Lee & Mouritsen’s *Judging Ordinary Meaning*.⁴⁵ That paper advocates a promising account of corpus linguistics use for Public Meaning Originalism. Corpus linguistics can provide an objective, scientific, data-driven approach to constitutional and statutory interpretation.

The core of Lee and Mouritsen’s analysis contains two types of corpus searches: “collocation” and “keywords in context.” Collocation searches in a corpus show the words that are most likely to appear in the same context as the search term. A “keywords in context” search presents the user of corpus linguistics with example of the term in context.

Take the “no vehicles in the park” example. One might seek evidence about the meaning of “vehicle” in 1958 by considering data from the corpus at the time. What are the common collocates of vehicle; with what other words is “vehicle” typically used in the corpus?

Lee and Mouritsen provide a corpus analysis on this exact question. They maintain that collocation provides “a snapshot of the semantic environment in which *vehicle* appears and the kinds of vehicles that tend to appear in that environment.”⁴⁶ For example, in a modern search, the top collocates of “vehicle” include “electric,” “motor,” “gas,” “autonomous,” and so on. As Lee

⁴³ See generally DOUGLAS BIBER, SUSAN CONRAD & RANDI REPPEN, *CORPUS LINGUISTICS: INVESTIGATING LANGUAGE STRUCTURE AND USE* (1998).

⁴⁴ See, e.g., Stefan Th. Gries & Brian G. Slocum, *Ordinary Meaning and Corpus Linguistics*, 6 B.Y.U. L. REV. (2017); Neil Goldfarb, *A Lawyer’s Introduction to Meaning in the Framework of Corpus Linguistics*, 6 B.Y.U. L. REV. (2017); Carissa Byrne Hessick, *Corpus Linguistics and the Criminal Law*, 4 B.Y.U. L. REV. (2018); Solan & Gales, *supra* note 9.

⁴⁵ *Supra* note 1.

⁴⁶ Lee & Mouritsen, *supra* note 1.

and Mouritsen infer, “the collocates of *vehicles* . . . strongly indicate *automobiles* as a likely candidate for the most common use of the term.”⁴⁷

Next they conduct a keywords in context search. This returns examples of the use of “vehicle” in context. For example, “the driver . . . apparently lost control of the vehicle because he was traveling too fast for the wet road conditions.”⁴⁸

It is important to note that this popularized use of corpus linguistics—focused on collocation and examples in context—is very different from versions of corpus linguistics that attempt to build more complex statistical or computation models of meaning, or those that use algorithmic processes to analyze word embeddings in a multi-dimensional vector space. The reason for this choice is likely motivated by practicality concerns. This simpler corpus method is one that many legal interpreters (e.g. judges) can employ cheaply and swiftly. A corpus linguistics “revolution” imagines judges, without much additional technical training, running searches like collocation and keywords in context to assess the frequency of usage in corpora.

III. PUBLIC MEANING ORIGINALISM’S EMPIRICAL ASSUMPTIONS

A. Empirical Critiques of Public Meaning Originalism

The table below outlines different types of critiques of Public Meaning Originalism and examples of the critique leveled at one source of originalist evidence (e.g. intuition, dictionary-use, or corpus linguistics). Most of these are “internal” empirical critiques. That is, they do not criticize originalism from an external perspective, asking whether we *should* be originalists. Instead, they critique originalism on its own terms, asking whether we *can* be originalists: Do originalism’s *own* empirical assumptions withstand empirical scrutiny?

⁴⁷ *Id.*

⁴⁸ *Id.*

<i>Critique</i>	<i>Example of Claim</i>
Accuracy: The originalist method reaches a false verdict concerning original public meaning	Some method (e.g. dictionary-use) recommends an interpretive result that is not consistent with the actual original public meaning.
Inconsistency among interpreters: Different people using the originalist method reach different verdicts	Judges have different linguistic intuitions about the original public meaning.
Inconsistency within interpreters: The same person using the originalist method reaches different verdicts	A judge's use of corpus linguistics at one time recommends an interpretive result that is inconsistent with the result recommended by the same judge's use of corpus linguistics at some other time
Inconsistency with other methods: One originalist method reaches a different verdict from others	An interpreter's linguistic intuition conflicts with the recommendation generated by their use of dictionaries or corpus linguistics.
Inconsistency within a method: The originalist method provides evidence for divergent verdicts	Definitions from two dictionaries provide divergent recommendations about original public meaning; two different plausible search criteria of corpus linguistics provide divergent recommendations about original public meaning
Arbitrariness in Practice: The originalist method's use is plagued by arbitrary decisions	There is no principled application of corpus linguistics, as it is currently used in originalist interpretation.
Arbitrariness in Theory: The originalist method's use cannot escape arbitrary decisions	There is no principled decision among conflicting dictionaries or alternate definitions.
Interpretive Underspecificity: Use of the originalist method does not provide determinate outcomes	For corpus linguistics, what frequency of use constitutes a "public meaning"?
Interpretive Vagueness: Assumptions required to use the originalist method admit of problematic vagueness or implausible cutoffs	A theory holds that if a particular use of a term is reflected in less than 5% of the corpus, then that use is not part of the original public meaning of the term. But, why not 3%; why not 10%?
Bias: Use of the originalist method enables politics or bias to influence interpretation	"Intuition" is subtly or unconsciously influenced by politically motivated reasoning.
Impracticality: The originalist method is too complicated, expensive, or otherwise impractical to use	Some forms of corpus linguistics require technical training that judges do not have.

Table 1. Empirical Critiques of Originalist Methods

These critiques can each be posed for different methods: intuition, dictionary-use, and corpus linguistics. This paper focuses primarily on the Accuracy critique for dictionary-use and corpus linguistics. However, it is important to note that there is a range of critiques to be considered.

Along the way we will consider some of these other critiques. For example, are there inconsistencies between corpus linguistics and dictionaries, and are their inconsistencies between judges?

B. Are Dictionaries and Corpus Linguistics Reliable? A Thought Experiment

With the increasing use of both dictionaries and corpus linguistics in originalist interpretation, a crucial question looms: Are these originalist methods actually achieving their aims? Originalists often assume that dictionary-use and corpus linguistics reflect facts about original public meaning. But this is an open empirical question.

The question has remained open, perhaps, because it seems untestable. To know whether these methods are reliable, we need some verification source, a Rosetta Stone of truths about original public meaning. To know whether an 18th century dictionary reflects public meaning, we need true facts about 18th century public meaning. Because we lack access to any such verification source, we need to use other methods in interpretation (like dictionaries and corpus linguistics) to estimate original public meaning.

But perhaps a Rosetta Stone of original public meaning is not the only option. Although we lack precise verification about historical (e.g. 1787) public meaning, we are much more confident in modern (i.e. 2019) public meaning. Our familiarity with modern public meaning can help assess sources of historical interpretive evidence. Consider a thought experiment:

Modern Amendment: Imagine that a modern Amendment stated that “vehicles” must be registered with the federal government. Two-hundred years later, in 2219, a legal dispute erupts concerning the original (2019) public meaning of “vehicle.” Would consulting 2019 dictionaries and corpus linguistics provide precise and reliable evidence about the Amendment’s original public meaning?

Following Lee & Mouritsen's recent defense of originalist methodology, consider a typical dictionary definition of a vehicle: "an agent of transmission; a carrier."⁴⁹ An interpreter in 2219 who uses this definition might think that roller-skates, or zip-lines, or even baby-shoulder carriers are vehicles. But people today generally judge that these entities aren't vehicles.⁵⁰ The Amendment's 2019 public meaning is not that roller-skates should be registered. But relying on the dictionary could suggest precisely the opposite.

Originalist corpus linguistics may fare no better. As Lee & Mouritsen note, the written word "vehicle" almost always refers to a car. And it most often appears near words associated with cars, like "electric" and "motor."⁵¹ This reflects one common use of "vehicle," but it neglects other acceptable uses. We do not often write today about horse-drawn carriages as "vehicles," and they aren't described as having "motors" or "electric" power. But we understand that they are vehicles. Corpus linguistics indicated that airplanes and helicopters are not "vehicles." But it is far from obvious that the public meaning of the Modern Amendment excludes those entities.

The central empirical assumption of views like Public Meaning Originalism is that its tools (e.g. dictionary-use and corpus linguistics) reflect original public meaning. This assumption remains surprisingly underexplored,⁵² and the thought experiment suggests that it may not be true. If people's modern judgments are not reflected by a method's modern use, we can argue by a historical inference that this also provides evidence that the method is unreliable in historical (i.e. originalist) interpretation:

1. Empirical Claim: The modern use of a method (i.e. use of dictionaries or corpus linguistics) does not accurately reflect people's ordinary judgments.

⁴⁹ Lee & Mouritsen, *supra* note 1, at 800.

⁵⁰ For empirical evidence of this claim see Part IV.

⁵¹ Lee & Mouritsen, *supra* note 1, at 847.

⁵² There has been some prior empirical research on originalism. However, these studies address different questions from those considered here. For example, in an important study Frank Cross suggests that originalism does not, in fact, effectively restrain willful judging. FRANK B. CROSS, *THE FAILED PROMISE OF ORIGINALISM* (2013). See also sources cited *supra* note 10.

2. Reliability Premise: A method that does not accurately reflect people's judgments is not a reliable method of determining public meaning.
3. Intermediate Conclusion: The empirical result provides evidence that the method is unreliable in modern interpretation.
4. Historical Inference: In the absence of historically distinguishing factors, evidence of a method's unreliability in modern interpretation also serves as evidence about that method's unreliability in historical interpretation.
5. Conclusion: The results provide evidence that the method is unreliable in historical interpretation.

An important piece of this argument to unpack is the historical inference. This premise holds that in the absence of historically distinguishing factors, evidence of a method's modern unreliability is also evidence of that method's historical unreliability. A "historically distinguishing factor" would be a compelling reason to think that use of a method is more reliable in historical interpretation.

In the case of dictionaries and corpus linguistics, most of the factors pull in the opposite direction. Modern dictionaries are larger and more frequently revised. Modern corpora are vastly larger and far more easily searchable than historical corpora. Finally, *our* use of a modern tool is presumably at least as accurate in reflecting modern public meaning than is *our* use of a historical tool in reflecting historical public meaning.

The historical inference is a crucial premise of the argument, but it is a highly plausible one. If the method that judges and jurors use in historical interpretation is flawed in modern interpretation, we have good reason to think that it is also flawed in historical interpretation. If our use of a 2019 dictionary does not reflect the modern public meaning of "vehicle," we have good reason to think that our use of a 1960s or 1860s dictionary is at least as poor in generating the historical public meaning. Of course, those who contest the historical inference are welcome

to identify and defend ways in which our use of (e.g.) a 1860s dictionary is more reliable in tracking 1860s meaning than our use of a 2019 dictionary is in tracking 2019 meaning.

The other premise one might resist is the empirical claim. Does our thought experiment really show that originalist methods do not track modern public meaning? The next Part addresses the empirical claim head-on. It investigates whether Public Meaning Originalism's core assumption is true, finding that dictionary-use and corpus linguistics lead to unreliable and also divergent results.

IV. AN EXPERIMENTAL TEST OF DICTIONARIES AND CORPUS LINGUISTICS

This Part turns to an experimental test of the reliability of dictionaries and corpus linguistics in interpretation. To explore the reliability of these methods, this Section presents a series of experiments. There are two broad aims of the present studies. One is an aim that existed at the outset of the studies: to assess whether people's use of corpus linguistic and dictionaries reflect public meaning. The second is an aim that arose during the course of running the studies. After the first experiment revealed that dictionaries and corpus linguistics sometimes provide very different verdicts about public meaning, a natural question arose about what cognitive processes explain the differences.

First, I consider whether the modern public meaning of "vehicle" is reflected in people's application of data from modern dictionaries and corpora. The results suggest that corpus linguistics is particularly unreliable, often providing verdicts that diverge from those delivered by both dictionary definitions and ordinary concept judgments.

Next, I investigate the second question: Why do corpus linguistics judgments differ from dictionary judgments (and why do both sometimes differ from the verdicts supplied by ordinary concept use)? I draw on prototype theory to test the hypothesis that dictionaries tend to generate more extensive uses, while corpus linguistics data is associated with more prototypical examples.

Prototype theory holds that concepts are associated with certain features, and category members that have more of those features are regarded as more central than category members that have less of those features.⁵³ For example, a car is a prototypical vehicle but an airplane is not (although it is still a vehicle).

One possibility is that corpus linguistics data might be seen as supplying the most useful information about *prototypical* category members. Corpus data provides details about the most frequent uses of a term and the most common words associated with the term. For example, corpus linguistics data about vehicles indicates that certain words often appear near “vehicle,” such as “motor” and “electric.” In other words, perhaps corpus data is really supplying the most helpful information about *prototypical* category members. I hypothesized that corpus linguistics is more helpful in identifying prototypical category members, while dictionaries tend to generate a more extensive sense of category membership. This is precisely what the experiments find.

The third series of experiments test an “expertise” objection to the findings about the unreliability of corpora and dictionaries. Are “elite-university” law students (e.g. at Harvard and Yale) or judges experts in the use of dictionaries or corpus data, such that those methods are more reliable in their hands? And does the process underlying the use of these tools differ among these groups? For these studies I used a larger range of questions, in light of the previous experimental results. The data are strikingly similar to the earlier studies on non-experts, suggesting that whatever expertise judges and law students have, it does not provide more reliable use of these interpretive methods. Moreover, similar processes seem to guide all participants’ uses of dictionaries and corpus linguistics.

The final experiment tests whether the discovery of the error of dictionaries and corpus linguistics is limited to the example about vehicles. It attempts to replicate the earlier finding, using ten different examples: vehicle, carry, interpreter, labor, weapon, tangible object, animal,

⁵³ See Eleanor Rosch, *Cognitive Representation of Semantic Categories*, 104 J. EXPERIMENTAL PSYCHOLOGY 192 (1975).

clothing food, and furniture. As a robustness check, the final experiment also varies some of the earlier experimental choices. For example, it uses a different corpus (the Corpus of Contemporary English, rather than the News on the Web Corpus); and it uses the first full definition of the term from Merriam Webster 2019. The data indicate that the results are robust: use of corpus linguistics and dictionary use carries significant error across *all* of these examples.

This paper includes a large number of experimental studies. To avoid redundancy and improve clarity, only the most significant are discussed in the main text. The remainder can be found in Appendices, A, B, C, and D.

Section IV.E is a crucial summary and interpretation of the main experimental findings. Four concern the surprising similarity between expert and non-expert interpreters (e.g. judges vs. the general population). Judges, law students, and ordinary people were striking similar in their ordinary conceptual judgments (IV.E.1; IV.E.2), use of dictionaries (IV.E.3), and use of corpus linguistics (IV.E.4).

Two more findings concern the process of using dictionaries and corpus linguistics. Across the studies, users of corpus linguistics tended to identify prototypical examples (e.g. a car is a vehicle) better than non-prototypical examples (e.g. a moped or airplane is a vehicle) (IV.E.5). Conversely, users of dictionaries sometimes made very extensive judgments (e.g. a pair of rollerskates is a vehicle) (IV.E.6).

Finally, the verdicts of both corpus linguistics and dictionary use diverged from ordinary meaning, in a large number of cases (IV.E.7). Moreover, the verdicts of corpus linguistics and dictionary use diverged from each other (IV.E.8). Section IV.E.9 and IV.E.10 considers these results across a diverse range of plausible assumptions. Across all these different models, each using reasonably conservative estimates, dictionaries and corpus linguistics had between a 20-35% error rate on average, and an 80-100% error rate for some of the hardest examples.

A. Experiment 1

The first experiment tests the verdicts delivered by dictionary and corpus use, as compared to ordinary judgments. To minimize researcher degrees of freedom, I used the first test case mentioned by Lee & Mouritsen's recent article endorsing the use of originalist corpus linguistics in legal interpretation.⁵⁴ That case is the well known "no vehicles in the park" example. I also used the exact corpus method used in that article, and the dictionary definition of "vehicle" that the article recommends.

The experiment divided participants into three groups: corpus, dictionary, and "ordinary concept" participants. The corpus and dictionary participants received corpus or dictionary information for the term vehicle, while concept participants received no information so that they would rely on their ordinary understanding. Each participant answered whether each of a series of ten entities was a vehicle.

Method

Participants. Two-hundred and six participants were recruited from Amazon's Mechanical Turk (52% female, 48% male, 0% non-binary, mean age = 35.8). Mechanical Turk ("MTurk") is an online platform that enables researchers to collect large samples from a population that is more representative than many other typical research samples.⁵⁵

Materials and Procedure. Participants were randomly divided into one of three conditions: Dictionary, Corpus, or Concept. In each condition, participants received some information about a term. Afterwards, they rated whether ten items fell under the category.

⁵⁴ Lee & Mouritsen, *supra* note 1.

⁵⁵ See Adam J. Berinsky et al., *Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk*, 20 POL. ANALYSIS 351 (2012); Gabriele Paolacci et al., *Running Experiments on Amazon Mechanical Turk*, 5 JUDGMENT & DECISION MAKING 411 (2010). The service is understood to provide high-quality data. See Michael Buhrmester, Tracy Kwang & Samuel D. Gosling, *Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality Data?*, 6 PERS. PSYCHOL. SCI. 3 (2011).

In the Concept condition, participants were simply asked to consider the noun “vehicle.” Then they were asked to categorize ten entities. For example, they were asked: “Is an automobile a vehicle?” [Yes / No]; “Is a car a vehicle?” [Yes/No]; and so on.

In the Dictionary condition, participants were given a dictionary definition of a vehicle:⁵⁶

1) a means of carrying or transporting something

2) an agent of transmission : carrier

However, participants were not told to which term that definition applied. Instead, they were told that the definition applied to a fake term, an “ailac” (“Consider this dictionary definition of “ailac:” (noun):”). This fake term guaranteed that any associations with the term “vehicle” would not interfere with participants’ use of the dictionary.⁵⁷ To see the necessity of this design, imagine that dictionary participants evaluated “vehicles” (not “ailacs”). There would be no way to assess whether any success in dictionary-use was attributable to use of the *definition* or people’s conceptual competence concerning vehicles. This methodology ensures that each condition reflects only the use of one method of analysis (ordinary conceptual competence, dictionary definition, or corpus data).⁵⁸

Corpus participants first saw this information:

Consider the noun, “ailac.” To help understand this term, consider some information about the use of “ailac.”

First, consider the top common words used in connection with “ailac.” These words might appear before or after ailac, or sometimes close to ailac, e.g. “electric ailac;” “ailac charging;” “drove the ailac;” etc.

⁵⁶ Note that this dictionary definition mirrors the one suggested by Lee & Mouritsen, *supra* note 1, at 800 (“One attested sense of *vehicle* is the notion of a ‘carrier’ or ‘agent of transmission’” (citing WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 2539 (1961))). See also <https://www.merriam-webster.com/dictionary/vehicle>.

⁵⁷ This is a common method in linguistics. See, e.g. Jean Berko, *The Child’s Learning of English Morphology* 14 WORD 150 (1958) (describing the “wug test”). Thanks to Larry Solan and Tammy Gales for this suggestion.

⁵⁸ All participants received the following introduction to the experiment:

In the following screen you will see some information about a term. The term might be a real term that you know (e.g. a “painter”) or one that is made up (e.g. a “krob”). If the term is one that is made up, the “information” about the term will also be fictional. After you see the information, we will ask some questions about the term.

Top common words: *electric, motor, plug-in, unmanned, armored, connected, cars, aerial, charging, pure, launch, owners, hybrid, traffic, fuel, driving, gas, autonomous, struck, operating, road, safety, accidents, battery, ownership, emergency, batteries, emissions, seat, advanced, driver, primary, demand, commandeered, fuel-efficient, automakers, demonstrators, excluding, lunar, passenger, fleet, gasoline, luxury, drove, parking, retirement, infrastructure.*

Next Corpus participants saw further examples of the term in context.⁵⁹ This corpus data is *precisely* what the advocates of corpus linguistics recommend.⁶⁰ Afterwards, participants in the Dictionary and Corpus conditions categorized ten entities. They were asked: “Is a car an ailac?” [Yes / No]”, and so on.

Results

As predicted, there were significant differences among Dictionary, Corpus, and Concept conditions. Figure 1 indicates the proportion responding “yes” for each entity in each condition.

⁵⁹ “Next, consider some further examples of “ailac” in context:

- 1) ...the driver, Bhaskar Jha, apparently lost control of the **ailac** because he was traveling too fast for the wet road conditions....
- 2) ...of the troopers. Parrott says the suspects in the **ailac** began showing aggression and shots rang out. Corporal Shane...
- 3) ... injury and leaving a child under 12 unsupervised in a motor **ailac** but released on a written promise to appear.) Risk ...
- 4) ... Hybrid electric ailacs use regenerative braking (when the **ailac** captures energy that would be otherwise lost from braking) and
- 5) ... pushed onto the property because of the speed of which these **ailacs** collide,” said Dr. Tom Lawrence , of Clinical Nutrition...
- 6) ..., 2009. That day the two officers saw an **ailac** connected to a domestic violence case in which shots had been...
- 7) ... say automakers would be better. Wakefield says autonomous **ailacs** could erode the image of certain brands more than others. Brands...
- 8) ... biogas, and Daimler, which supplies a number of experimental **ailacs** designed to run on natural gas. The German Federal Ministry of ...
- 9) ... is that they aren’t kept on file with the Motor **Ailacs** Division or any other entity. By contrast, beneficiary...”.

⁶⁰ Lee & Mouritsen, *supra* note 1.

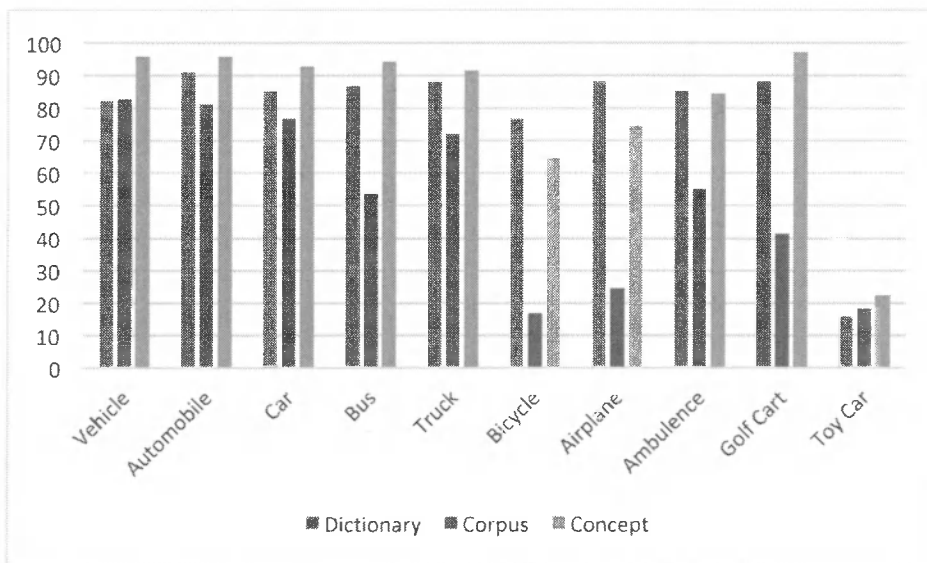


Figure 1. Percentage Responding Yes (the entity is a vehicle) by Dictionary, Corpus, Concept conditions

For each entity (e.g. car), I conducted a chi-square test to compare the proportion of yes responses across Dictionary, Corpus, and Concept conditions. See Figure A1 in the Appendix.

Notably, for each method, the verdicts are different. Moreover, there is a systematic pattern: corpus fails to include a number of entities that are vehicles in the ordinary sense of the term. For example, consider entities like bicycle, airplane, and golf cart. Although there is some disagreement, most people classify these as vehicles. Yet users of corpus systematically hold that these are not vehicles.

As can be seen in Figure A1 (see Appendix), all differences are significant except for toy car. Conventionally, V of .1 is a “small effect,” .3 is a “medium effect,” and .5 is a “large effect.” By effect size conventions, there is a large effect for bicycle, airplane, and golf cart; a medium effect for bus and ambulance; a small effect for vehicle, automobile, car, truck; and no significant effect for toy car. The next natural question is what exactly explains each of these differences.

To address that question, I conducted pairwise chi-square tests to compare the proportion of “yes” responses for each item between conditions (Dictionary v. Corpus, Dictionary v. Concept, Corpus v. Concept). Notably, Dictionary and Corpus methods delivered different verdicts for five

entities (bus, bicycle, airplane, ambulance, golf cart), and Corpus and Concept also diverged for those five.

Discussion

This first experiment represents a small test of the accuracy of corpus linguistics and dictionaries in reflecting public meaning. Broadly speaking, dictionary-use was fairly consistent with people's ordinary judgments: cars, busses, and trucks are vehicles, but a toy car is definitely not.

However, corpus linguistics did not perform nearly as well. A bus is seemingly within our modern conception of a vehicle, but *only half* of the users of corpus linguistics made that categorization. The divergence was not limited to that example. For five of the ten entities, corpus was underinclusive.

Thus, this first study provides initial evidence that dictionaries are reliable in determining ordinary meaning but corpus linguistics is somewhat unreliable.

B. Experiment 2: The Process of Using Dictionaries and Corpora

Consider the patterns of judgment revealed by the previous experiment. Certain entities elicit dramatic differences between Corpus and Dictionary participants. For example, the majority of Dictionary participants judge bicycles, airplanes, and golf carts as vehicles. Yet Corpus participants judge these entities as not vehicles.

What explains these differences? One hypothesis is inspired by research in philosophy and cognitive science about prototypes. According to prototype theory, people associate concepts with certain features, and "prototypical" category members are those that have most or all of those features. For example, both a robin and a penguin are birds, but a robin is a prototypical bird. Studies in cognitive science have shown that people are faster in categorizing prototypical category members than non-prototypical ones. For example, people will categorize a robin as a

bird more quickly than they categorize a penguin as a bird. Moreover, when people are asked to name examples of category members, the more prototypical members are cited more frequently.⁶¹ For example, if you ask someone to name a type of pet, “dog” would be cited more often than “kangaroo.”

I hypothesized that prototype theory might explain some of the differences between dictionaries, which report a broad definition, and corpus linguistics, which reports data indicative of the most popular uses. I hypothesized that corpus linguistics use helps identify prototypical examples, while dictionary use facilitates a more extensive representation. This experiment tests this hypothesis.

Method

Participants. One hundred and one participants were recruited from Amazon’s Mechanical Turk. Eighty-two passed a comprehension check question (51% female, 48.5% male, 1.2% non-binary, mean age = 35.8).

Materials and Procedure. Participants were trained to understand the difference between prototypical and non-prototypical category members (see Appendix A, Part III). Participants were then instructed to “Consider the noun “vehicle.” They were then presented with ten sets of statements, in a random order. For example a participant might first rate two statements appearing like this:

An airplane is a prototypical vehicle. 1 (strongly disagree) to 7 (strongly agree)

An airplane is technically a vehicle. 1 (strongly disagree) to 7 (strongly agree)

⁶¹ See generally Eleanor Rosch, *Principles of Categorization*, in CONCEPTS: CORE READINGS 189 (1999).

Results

Eighty-two participants answered all six check questions correctly and were included in the analysis. As predicted, there were significant differences between the prototypically and technically judgments across the ten entities, see Figure 2. Comparing these results to Experiment 1's results for Corpus and Dictionary participants reveals a striking similarity.

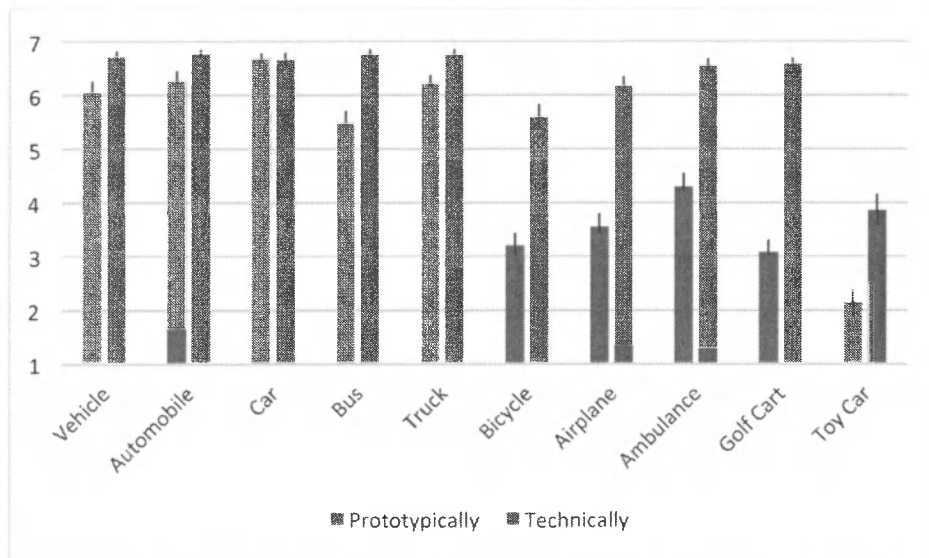


Figure 2. Mean ratings for “prototypical” and “technically” for ten entities. Error bars indicate standard error.

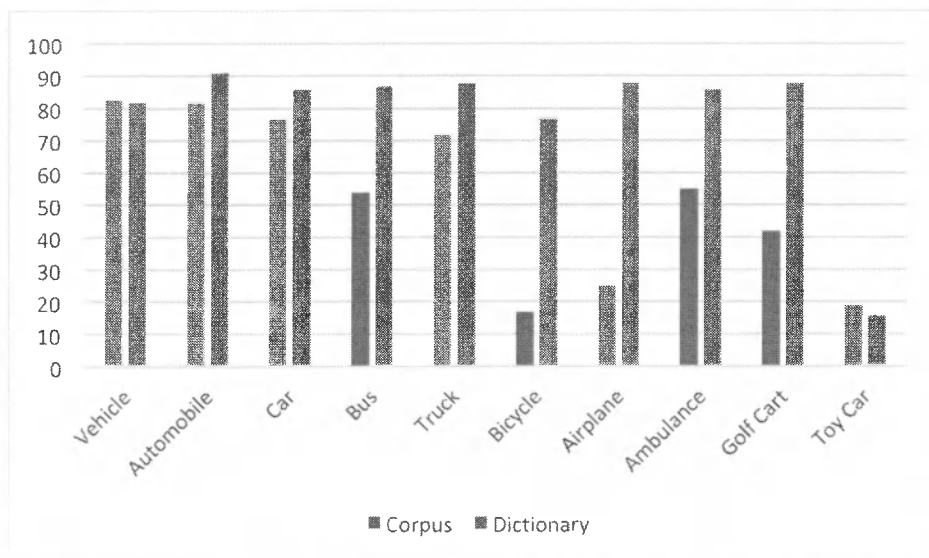


Figure 3. Percentage responding “yes” (entity is a vehicle) for ten entities by Corpus and Dictionary condition (Experiment 1).

To test the statistical significance of this relationship between Corpus-Prototypically and Dictionary-Technically, I conducted two tests for differences between correlations between (i) Corpus and Prototypically and Corpus and Technically, and (ii) Dictionary and Prototypically and Dictionary and Technically.⁶² In the first comparison, I considered the percentage of participants that rated each entity as a vehicle using the corpus, and correlated that with the ratings for prototypically and technically. Prototypically was significantly more correlated with Corpus, $z = 1.841$, $p = .0328$ (one-tailed). Technically was significantly more correlated with Dictionary, $z = 3.489$, $p = .0002$ (one-tailed).⁶³

	Corpus	Dictionary	Concept	Prototypically	Technically
Corpus	1				
Dictionary	0.51	1			
Concept	0.73	0.92	1		
Prototypically	0.93	0.58	0.72	1	
Technically	0.72	0.95	0.99	0.74	1

Figure 4. Correlation Matrix. Boxes indicate the relevant comparisons (Corpus is more correlated with Prototypically, and Dictionary is more correlated with Technically).

Discussion

Section VI.A discusses the legal-philosophical implications of this finding in more detail. Insofar as corpus linguistics elicits prototypical uses of a term but dictionaries elicit a more extensive sense, the former is more appropriate in legal contexts calling for a prototypical uses and the latter is more appropriate in contexts calling for more extensivist ones.

⁶² See Ihno A. Lec & Kristopher J. Preacher, *Calculation for the test of the difference between two dependent correlations with one variable in common* (2013), <http://quantpsy.org>.

⁶³ Because the hypothesis was that prototypically was more correlated with corpus, and technically with dictionary, one-tailed tests were used rather than two-tailed tests. Two-tailed tests indicate similar results. Technically is significantly more correlated with dictionary, $z = 3.489$, $p = .0005$ (two-tailed). Prototypically more correlated with corpus at a level short of the traditional cutoff for statistical significance, $z = 1.841$, $p = .0656$ (two-tailed). To match the previous experiments, this experiment uses only ten items (e.g. car, bus, bicycle, etc.). This limits the power of correlation analyses.

For example, in the context of a rule “any and all vehicles are prohibited from the park,” one might reasonably think that the public meaning of the rule bans even non-prototypical vehicles. But in the context of a rule “only cars, trucks, and other vehicles are prohibited from the park,” one might argue more persuasively that the rule bans only prototypical vehicles. If so, dictionaries would be better guides in the first case, but corpora would be better guides in the second.

C. Experiment 3: Expert Judges

The preceding experiments have studied ordinary, non-expert populations. Judgments of ordinary people provide good evidence about the current public meaning of these terms (e.g. “vehicle”). But some might doubt whether this population contains the best users of dictionaries and corpus linguistics in legal interpretation. To appropriately test the reliability of corpus linguistics and dictionaries, one might argue, we should test legal experts who have the relevant background in interpretation.

This objection is plausible, but it should not be taken to dismiss any significance of the results. After all, even if the previous results do not provide a strong inference into judges’ cognition, they do provide good evidence about juror’s cognition. And jurors, too, are statutory interpreters.⁶⁴

Nevertheless, this section addresses the “expertise” objection head-on. I tested a population of law students from Harvard, Yale, and Columbia and United States judges. There are a few reasons to think such experts might perform differently. For one, it may be that they have some training or expertise that enables them to use dictionaries or corpus linguistics in some expert way. Second, even if they do not have special expertise in corpus linguistics per se, they might be more reliable survey-takers, more likely to devote sufficient attention and produce thoughtful responses.

⁶⁴ Lawrence Solan, *Jurors as Statutory Interpreters*, 78 CHI-KENT L. REV. 1281 (2003).

To more comprehensively test the reliability of dictionaries and corpus, this experiment featured an expanded range of twenty-five entities. In the first three experiments, most entities were “vehicles” in ordinary language and the dictionary categorized these as vehicles. The previous experiment suggested that the dictionary generates an extensive condition of category membership. So to better test dictionaries, this experiment also includes some entities that I predicted are likely *not* vehicles in ordinary language, but which may nevertheless fall under a very extensive sense of a vehicle. These are entities including crutches, a baby should-carrier, and a zip-line.

To avoid redundancy, I present only the judge data in the main text. The law student data (which is strikingly similar) is presented in Appendix A.

Method

Participants. Approximately 1500 professional judges were contacted by email to request voluntary participation in the study. Ninety-eight United States judges participated in an online experiment.⁶⁵ Judges were recruited from state and federal courts and asked to categorize their years of experience (e.g. less than 1 year, 1-5 years, 6-10 years, etc.). Seventy-three judges reported their years of judging experience. Of those, 1.4% reported less than one year of experience, 17.8% reported 1-5 years of experience, 17.8% reported 6-10 years, 24.7% reported 11-15 years, 17.8% reported 16-20 years, 11.0% reported 21-25 years, and 8.22% reported over 26 years (1.37% reported “other”).

Materials and Procedure. As in the previous experiments, participants were randomly assigned to either the Concept, Corpus, or Dictionary condition. In this experiment, participants evaluated the first set of entities (presented in a randomized order): a vehicle, automobile, car, bus, truck, bicycle, airplane, ambulance, golf car, toy car. Participants immediately considered

⁶⁵ $M_{age} = 59.3$, 33.8% female, 66.2% male, 0.0% non-binary.

another set (presented in a randomized order): drone, skateboard, pair of rollerskates, “a non-functioning commemorative truck (e.g. a World War II Truck that has been decorated as a WWII monument)”, baby stroller, electric wheelchair, horse-drawn carriage, wooden canoe, helicopter, moped, pair of crutches, pogo stick, baby shoulder-carrier, liferaft, and zip-line.

Results

First, I conducted chi-square tests for differences among the three conditions. For twenty items, there was a significant difference. To further analyze these differences, I conducted follow-up pairwise chi-square tests. Again, there were a number of significant differences (see Appendix A).

Discussion

The results are strikingly similar to the results of non-experts. United States judges’ use of corpus linguistics and dictionary methods did not consistently track their ordinary judgments about category membership. For the full results and statistics, see Appendix A.

For many entities, the corpus linguistics judgment did not reflect that of judges’ ordinary conceptual competence: consider bus, truck, airplane, ambulance, golf cart, horse-drawn carriage. For many others, the dictionary use did not reflect ordinary judgment: consider skateboard, roller-skates, WWII Truck, baby stroller-carrier, canoe, helicopter, baby-shoulder carrier.

For a very large number of entities, the corpus and dictionary delivered divergent judgments: truck, bicycle, airplane, golf cart, skateboard, roller-skates, baby stroller, wheelchair, horse-drawn carriage, canoe, helicopter, baby shoulder-carrier, liferaft, and zip-line.

D. Experiment 4: Replication Across Ten Examples

Experiment 3 indicates that the findings regarding the inaccuracy of dictionaries and corpus linguistics replicate across levels of legal expertise. Using the example of a vehicle, it found that

the verdicts delivered by dictionary use and corpus linguistics use often depart dramatically from each other and from the verdict indicated by ordinary cognition.

This final experiment sought to test whether these findings replicate across different examples. To examine this question, this experiment tested “vehicle,” as well as nine other terms. More broadly, the final experiment aimed to serve as a “robustness check” of the earlier findings. It altered various parameters from the first experiment: the relevant term, the corpus data used, and the dictionary definition used.

First, the experiment assessed ten terms. Of the ten, the first three were drawn from examples cited by corpus linguistics proponents: “vehicle,” “carry,” and “interpreter.”⁶⁶ The next three are inspired by important interpretation terms: “labor,” “tangible object,” and “weapon” (a modern version of “arms”). The final four are common examples of superordinate categories, which admit of a range of category members: “clothing,” “furniture,” “food,” and “animal.” For each term, the experiment asked about twenty-five entities. For further detail, see Appendix D.

Moreover, while the earlier experiments used the News on the Web Corpus, this experiment used instead the Corpus of Contemporary English.

Finally, while the earlier experiments also used a representative dictionary definition cited by proponents of originalist interpretation,⁶⁷ this experiment simply used the first full definition of the relevant term, from Merriam-Webster 2019 Online.⁶⁸ In some cases, these definitions supplied some examples alongside the definition. For example, “vehicle” is defined as “a means of carrying or transporting something // planes, trains, and other vehicles : such as : a: motor vehicle b : a piece of mechanized equipment.” This fourth experiment included two dictionary conditions. The first “full dictionary” condition included the entire first definition of the relevant term. The

⁶⁶ See Lee & Mouritsen, *supra* note 1; see also Appendix D *infra*.

⁶⁷ See Lee & Mouritsen.

⁶⁸ Merriam-Webster, <https://www.merriam-webster.com/>.

second “bare dictionary” condition included the definition, without examples. The “bare” definition for “vehicle” was “a means of carrying or transporting something.”

The experiment also used a different fake term, “krob” rather than “ailac,” for the corpus and dictionary conditions.

Method

Participants. Two-thousand eight-hundred and thirty-five “general population” participants from the United States were recruited from Amazon’s Mechanical Turk ($M_{age} = 37.88$, 46.1% female, 53.6% male, .03% non-binary).

Materials and Procedure. The procedure was similar to the first three experiments. Participants were randomly assigned to one of four methods (ordinary concept, corpus, full dictionary, bare dictionary) and one of ten examples (vehicle, carry, interpreter, labor, tangible object, weapon, animal, clothing, food, furniture). See Appendix D for full materials.

Results and Discussion

As a first analysis, I conducted chi-square tests to assess differences in categorizations among the ordinary concept, corpus, and full dictionary conditions. There were ten entities, each with twenty-five entities (and twenty-five comparisons). As such, I analyzed each example (e.g. vehicle; weapon) by assessing whether there was at least one significant difference, at a level corrected for twenty-five multiple comparisons.⁶⁹ If there was one significant difference at this level, this would suggest some unreliability among the methods for that example.

The results provide fairly strong evidence of a large degree of unreliability. For *all* ten examples (e.g. vehicle; tangible object), there were at least ten significant differences (out of twenty-five). For vehicle, there were eleven; for carry, there were twelve; for labor, fifteen; for

⁶⁹ I compared the chi-square results to a Bonferonni-corrected p value of .002.

interpreter, twenty; for tangible object, twenty-two; for weapon, fifteen; for furniture, eighteen; for animal, fourteen; for food, twenty; and for clothing, fourteen. See Appendix D for full results.

This is consistent with the earlier experimental results. For example, in Experiment 3, there were eighteen significant differences for judges (concerning vehicle) and twenty for law students (concerning vehicle).

As a second analysis, I conducted chi-square tests to assess differences in categorizations among the ordinary concept, corpus, and bare dictionary conditions. There were ten entities, each with twenty-five entities (and twenty-five comparisons). As such, I analyzed each example by assessing whether there was at least one significant difference, at a level corrected for twenty-five multiple comparisons.⁷⁰ If there was one significant difference at this level, this would suggest some unreliability among the methods for that example.

The results again provide strong evidence of a large degree of unreliability. For *all* ten examples, there were at least ten significant differences (out of twenty-five). For vehicle, there were thirteen; for carry, there were eleven; for labor, twenty-two; for interpreter, thirteen; for tangible object, twenty-two; for weapon, fourteen, for furniture, seventeen; for animal, fifteen; for food, twenty; and for clothing, eleven. See Appendix D for full results.

E. Summary and Interpretation

Before turning to the next Part, it is worth providing some summary considerations and graphics. The experiments suggest that judges and non-experts are similar in (i) their ordinary judgments of common concepts (e.g. vehicle), (ii) how they apply dictionary definitions, and (iii) how they apply corpus linguistics.

Moreover, the results indicate that, perhaps surprisingly, “ordinary meaning” is not as clear as one might think. For a number of entity categorizations, participants were very divided. For

⁷⁰ I compared the chi-square results to a Bonferonni-corrected p value of .002.

example, people are generally divided (about 50%-50%) on whether a canoe is a vehicle. This is true across ordinary people, law students, and judges.

The pattern of results also indicates that dictionaries tend to be more inclusive than corpus linguistics. Corpus linguistics categorizations are correlated with judgments of prototypicality, while dictionary categorizations are more extensive. Unsurprisingly, this implies that dictionaries and corpus linguistics often provide dramatically different verdicts from each other. Moreover, they often provide different recommendations about meaning from what is reflected in ordinary judgments.

1. Judges and Non-Experts Judge Meaning Similarly

First, consider the percentage of participants within each population responding that each entity is a vehicle.

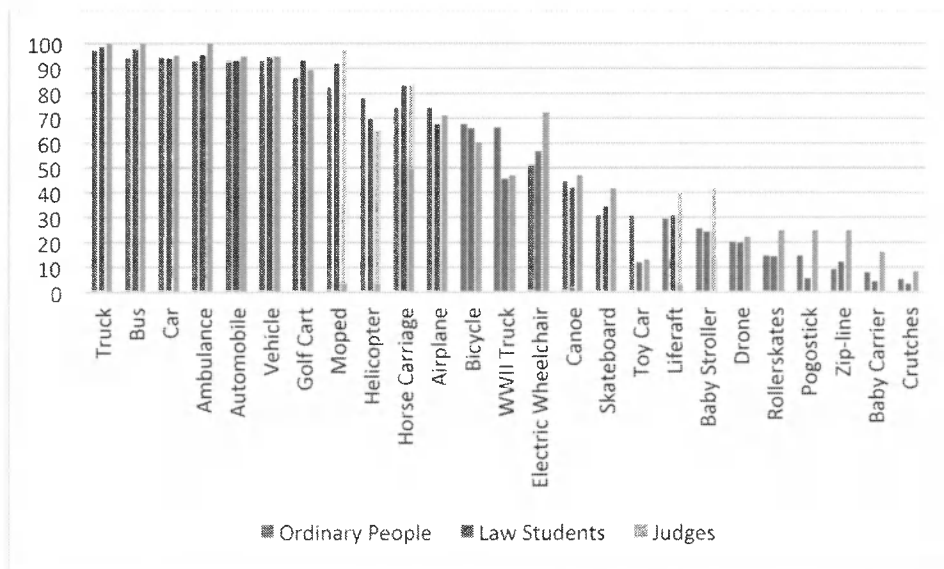


Figure 5. Percentage of participants responding “Yes” to “Is [entity] a vehicle?”

There is a striking similarity in the ordinary concept of a vehicle among those with very different legal and educational backgrounds. Whatever legal experience might provide, it does not seem to dramatically change cognition about ordinary concepts like vehicles.

2. Ordinary Meaning is Often Unclear

Figure 5 indicates a second striking fact about ordinary concepts. Although the results are similar among the three populations, there is notable disagreement among people about which entities are category members. Among judges, law students, and those untrained in law, there is substantial disagreement about whether canoes and skateboards are vehicles.

Although originalists and textualists seek to determine *facts* about ordinary meaning, this result suggests that in many cases those facts may be unclear or indeterminate. Taken at face value, the results suggest that there is no clear fact of the matter concerning whether the modern public meaning of “vehicle” includes a canoe. Moreover, this disagreement is not mitigated by judicial or legal expertise. Disagreement persists (in the same degree) across people with various levels of legal training and experience.

3. Judges and Non-Experts Apply Dictionaries Similarly

Next consider the percentage of participants within each population responding that each entity is a vehicle, according to the dictionary.

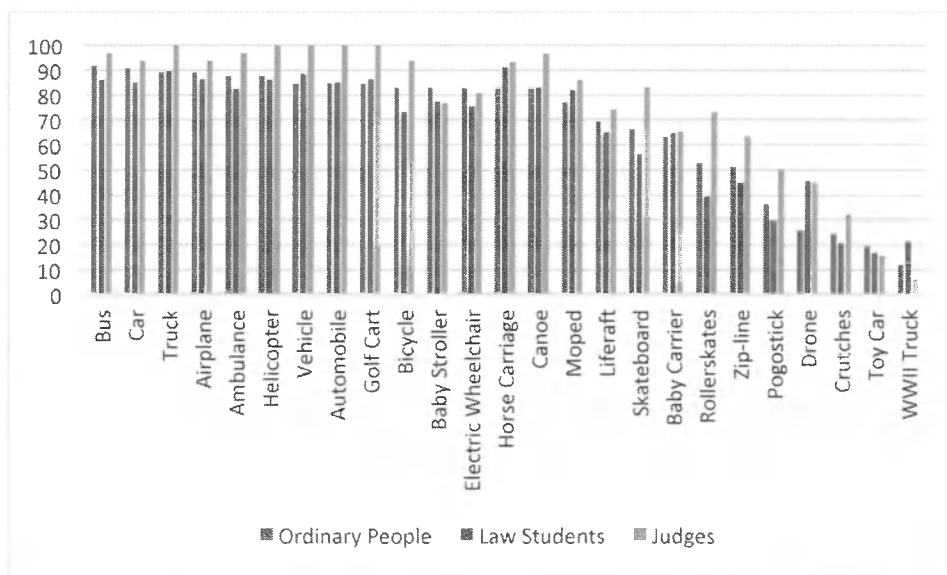


Figure 6. Percentage of participants affirmatively categorizing the entity according to the dictionary definition

The results suggest that legal expertise does not dramatically change the way in which people apply a basic dictionary definition. Judges were modestly more inclined to categorize some entities as vehicles, but overall the pattern of results is fairly consistent among the three populations.

Again, it is also worth noting that although there is remarkable agreement among the populations—judges, law students, and MTurk participants do not disagree *as groups* about how to apply dictionaries—there is striking disagreement within groups for some entities. Consider examples like the zip-line, pogostick, and drone. A substantial proportion of participants in every group categorized these as vehicles, while a substantial proportion did not.

4. Judges and Non-Experts Apply Corpus Linguistics Similarly

Finally, consider the percentage of participants within each population responding that each entity is a vehicle, according to corpus linguistics.

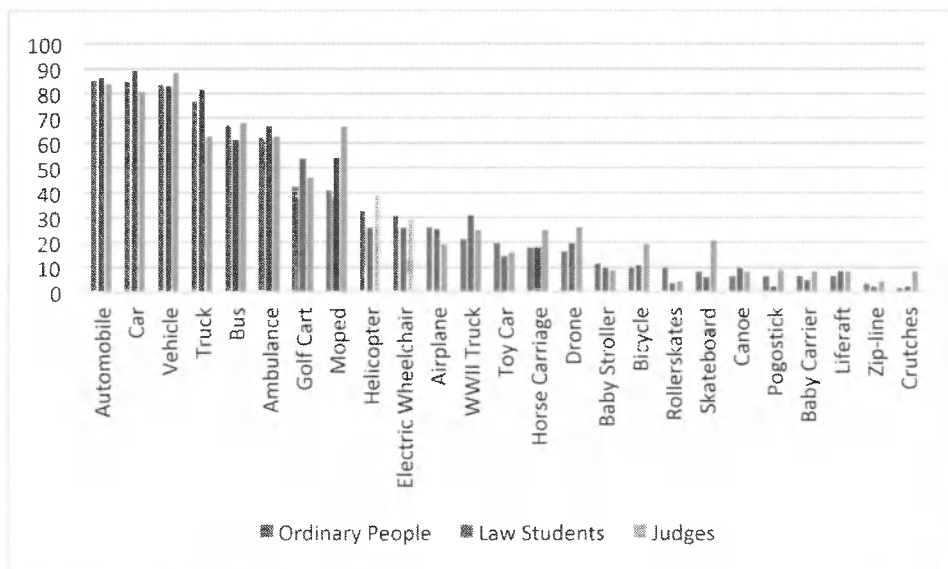


Figure 7. Percentage of participants affirmatively categorizing the entity according to the corpus linguistics data

The results suggest that legal expertise does not dramatically change the way in which people interpret and apply the corpus linguistics data.

Here again, we should note that although there is impressive agreement among the populations—judges, law students, and MTurk participants are not very different *as groups* in their applications of corpus linguistics— but there is striking disagreement *within* groups for some entities. Consider examples like the golf cart, moped, and helicopter. Within each group, a substantial proportion of participants categorized these as vehicles, while a substantial proportion did not.

5. Corpus Linguistics Use Reflects Narrow, Prototypical Uses

Next consider the corpus and concept results for judges.

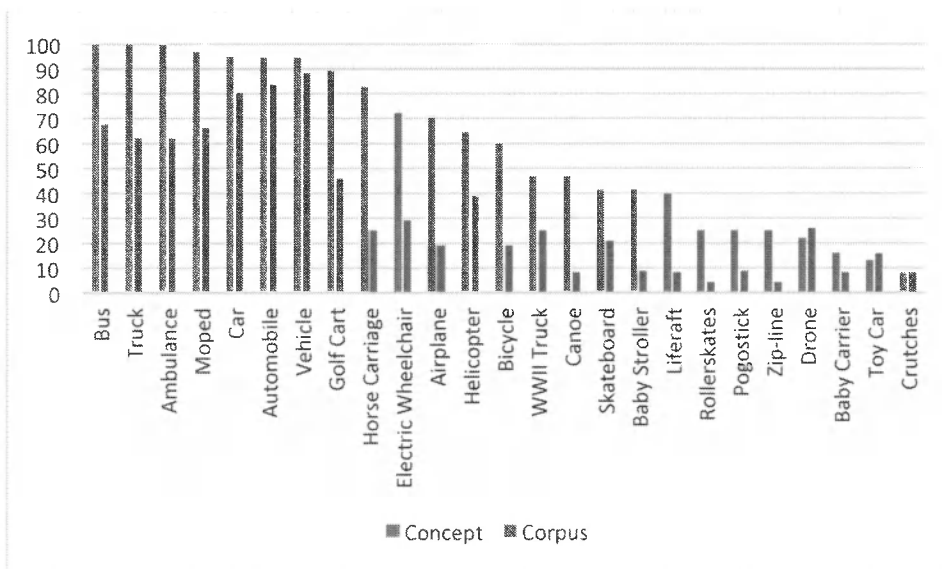


Figure 8. Comparison of percentage of judges responding “Yes” to “Is [entity] a vehicle?” to percentage of judges categorizing the entity as a vehicle according to corpus linguistics.

Although corpus linguistics use is not entirely unrelated to judges’ application of the ordinary concept, in many cases the corpus is underinclusive. For example, a truck is unanimously understood as a vehicle in ordinary language, but use of corpus linguistics returned only a moderate endorsement of trucks as vehicles. Similarly, entities like a horse-drawn carriage, golf cart, airplane, helicopter, and bicycle are largely understood by judges as vehicles in ordinary language, but they are not classified as vehicle by judges using corpus linguistics.

Moreover, when we compare these proportions of judge’s categorizations to the rated prototypicality of the entities (as in Experiment 2), we find the same significant relationship between corpus linguistics and prototypicality (see Appendix A).

6. Dictionary-Use Can Reflect Extensive Uses

Next, consider the concept and dictionary results for judges. There are some large divergences between ordinary judgments and dictionary verdicts. For example, most using the dictionary evaluate a baby-shoulder carrier as a vehicle; however, we generally understand that they are not vehicles. For other controversial entities (e.g. canoe), dictionary-use reflects that these are uncontroversially vehicles.

Moreover, when we compare these proportions of judge’s categorizations to the ratings of whether the entity is a “prototypical” vehicle or “technically” a vehicle (as in Experiment 2), we find the same significant relationship between dictionaries and the extensive question (see Appendix A). This suggests that unlike corpus linguistics, which tracks prototypical uses, dictionaries track a more extensive sense of meaning.

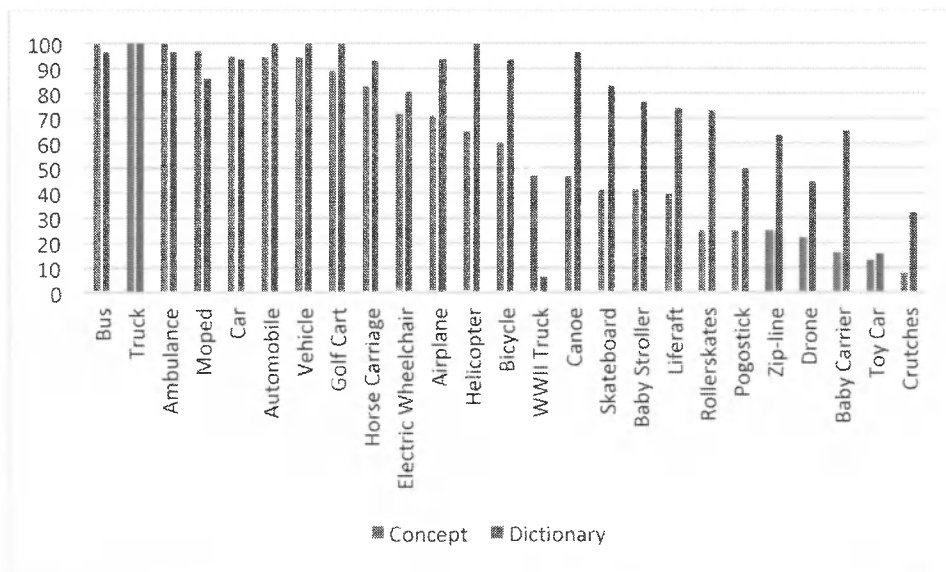


Figure 9. Comparison of percentage of judges responding “Yes” to “Is [entity] a vehicle?” to percentage of judges categorizing the entity as a vehicle according to dictionary use.

7. Dictionaries and Corpus Linguistics Provide Divergent Verdicts

It is worth considering the dictionary and corpus results together. These results reflect some dramatic differences between the results suggested by dictionaries and corpus linguistics.

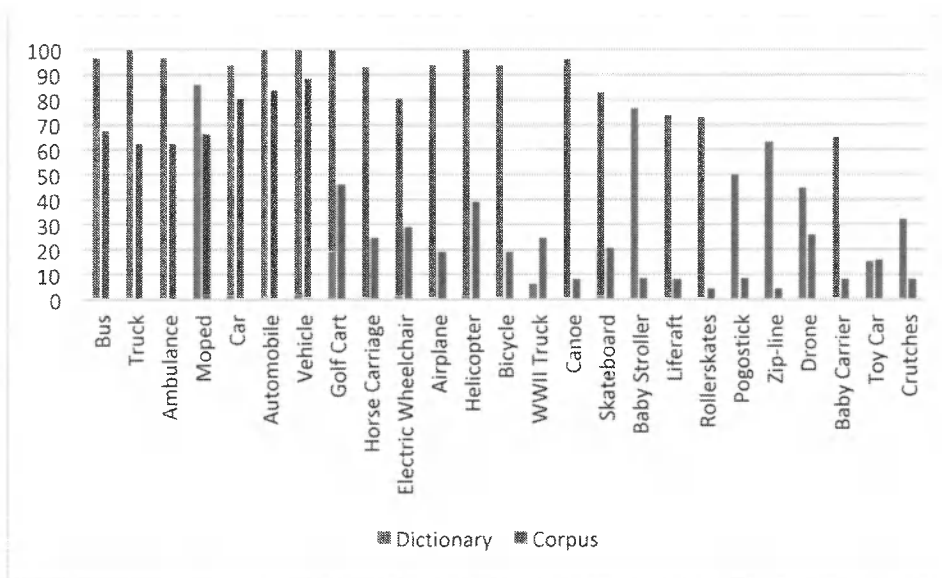


Figure 10. Comparison of percentage of judges categorizing the entity as a vehicle according to dictionary use and corpus linguistics use.

In many cases, these results reflect extreme differences between dictionaries and corpus linguistics. Nearly every judge using a dictionary assessed entities like canoes, bicycles, and airplanes as vehicles; while nearly every judge using corpus linguistics assessed those as *not* vehicles.

8. Dictionary and Corpus Linguistics Verdicts Diverge from Ordinary Judgment

Figures 9 and 10 indicate that, for many examples, use of dictionaries and corpus linguistics did not reflect ordinary judgments.⁷¹ The most straightforward interpretation of this pattern of results is that dictionaries and corpus linguistics were not always accurate measures of modern public meaning.

⁷¹ See also Appendix C *infra*; Appendix D *infra*.

Considering just the results of Experiment 3, in some cases, corpus linguistics indicated that clear vehicles were not, in fact, vehicles. For example, busses, trucks, and ambulances were unanimously understood to be vehicles. Yet over one-third of those using corpus linguistics evaluated these as not vehicles. Conversely, in some cases the dictionary use indicated that clear non-vehicles were, in fact, vehicles. For example, dictionary-using judges overrated rollerskates and baby-shoulder carriers as vehicles, compared to judges' ordinary evaluation of those entities.

9. On Average, Corpus Linguistics and Dictionaries Had 20-35% Error Rates

This Section considers the error rates for both dictionary and corpus linguistics methods, across Experiments 3 and 4. That is, it considers how often one relying upon a dictionary definition or corpus linguistics data would reach the *wrong* verdict about ordinary meaning.

To assess the error rates, we must make some assumptions about what percentage of agreement in the ordinary concept condition indicates that the use of part of the ordinary meaning. For example, should we assume that the ordinary meaning of "vehicle" includes a car if at least 50% of people agree; or must some higher threshold, like 75% or 90%, be met?

This part considers these three cutoffs, 50%, 75%, and 90%, as a representative range of plausible options. As such, this analysis does not require us to take a stand on this hard question about ordinary meaning (i.e. we need not commit that 50% is the right cutoff to determine ordinary meaning from ordinary judgments). Instead, this analysis allows us to consider the error rates across a range of plausible options. As we will see, there is some similarity in the error estimated across these options. This allows us to conclude that, under many plausible assumptions, there is a certain range of error for dictionaries and corpus linguistics.

To give a sense of how this computation works, consider a 50% cutoff. That is, assume that if over 50% participants (in the ordinary concept condition) categorized something as a vehicle, then it counts as a vehicle. To take one example, 100% of judges assessed a bus to be a vehicle.

Because 100 is greater than 50, we treat this as a vehicle. Only 68% of judges using corpus linguistics made the same judgment. So, 32% of corpus users made a judgment (i.e. that a bus is not a vehicle) that is incorrect on these assumptions. On this assumption, there is a 32% error rate for the bus item for judges using corpus linguistics. Repeating this process for all items (i.e. all 25 items in Experiment 3; and all 250 items in Experiment 4), we can compute an average error rate.

I performed these computations, using 50%, 75%, and 90% cutoffs, for the corpus and dictionary results from Experiment 3 (judges, law students, and general population evaluations of vehicles) and Experiment 4 (general population evaluations of ten examples). The results are displayed in the first three columns of Tables 2 and 3.

A final method of assessing error is to consider the “Difference Between Percents.” On this method, we consider the absolute value of the difference between the percentages of affirmative judgments in the ordinary concept condition and one of the corpus or dictionary conditions. For example, for judges, 32.3% of dictionary users categorized crutches as a vehicle, and 8.3% of ordinary concept condition participants made the same categorization. So, the “Difference Between Percents” error rate for dictionaries for this item is 24.0% (32.3% minus 8.3%). As should be clear, this calculation of error is *very generous* to corpus linguistics and dictionaries. The most natural interpretation of the crutches data is that it is *not* a vehicle in the ordinary sense; 8.3% of participants in the concept condition were wrong; and 32.3% of judges using corpus linguistics made an incorrect categorization. In this case, the “Difference Between Percents” method computes a dictionary error rate that is 8.3% lower. These results for corpus linguistics and dictionaries are displayed in the final column of Table 2 and 3, respectively.

Average Corpus Linguistics Error				
	Ordinary Meaning = 50%	Ordinary Meaning = 75%	Ordinary Meaning = 90%	Difference Between Percents
Judges (Vehicle)	30.3%	22.8%	20.5%	22.4%
Law Students (Vehicle)	27.8%	21.0%	22.6%	25.0%
General Pop. (Vehicle)	32.5%	21.1%	18.4%	24.1%
General Population (Ten)	41.9%	39.1%	39.5%	25.1%

Table 2. Average error for corpus linguistic judgments in Experiments 3 and 4, under different theoretical assumptions.

Average Dictionary Error				
	Ordinary Meaning = 50%	Ordinary Meaning = 75%	Ordinary Meaning = 90%	Difference Between Percents
Judges (Vehicle)	29.9%	43.3%	50.79%	22.4%
Law Students (Vehicle)	28.6%	35.9%	45.96%	16.7%
General Pop. (Vehicle)	33.8%	41.9%	49.9%	21.7%
General Pop. (Ten, "Full")	34.2%	36.4%	41.3%	18.4%
General Pop. (Ten, "Bare")	35.1%	46.9%	47.6%	20.8%

Table 3. Average error for dictionary judgments in Experiments 3 and 4, under different theoretical assumptions.

The important takeaway from these tables is that the error rates for dictionaries and corpus linguistics are certainly not trivial. For judges, law, students, and the general population (across many examples), the error rate for both tended to fall between 20% to 35%. Sometimes it was higher (e.g. 50.8% for judges' use of dictionaries, with a 90% cutoff); and sometimes it was lower (e.g. 18.4% for the general population in Experiment 4, with a 90% cutoff). But the results overwhelmingly indicate that these methods carry real risks of error. The range (of 20-35%) error rates suggest that one relying on dictionaries or corpus linguistics would reach the wrong verdict *in every three to five cases.*

10. In Some Circumstances, Error Rates Reached 80-100%

Although the notion of an average error rate is helpful, it is also useful to consider the maximum error rates. The experiments included a number of relatively easy categorizations (e.g. whether a car is a vehicle; whether factory working is labor; whether a book is a tangible object). Insofar as real legal decisions concern comparatively more difficult categorizations (e.g. whether an airplane a vehicle; whether preaching is labor; whether a fish is a tangible object), it may also be instructive to consider the maximum error rate: What percent of (e.g.) judges using dictionary or corpus linguistics evaluated *the hardest interpretive question* incorrectly?

I conducted a similar analysis to that conducted in Section IV.E.9. But in this analysis, I computed the maximum error rate, under each of the different assumptions. As Tables 4 and 5 indicates, across all levels of expertise, the data suggest that in some examples, relying on a dictionary definition or corpus linguistics data led 80-100% of users to the incorrect verdict.

Maximum Corpus Linguistics Error				
	Ordinary Meaning = 50%	Ordinary Meaning = 75%	Ordinary Meaning = 90%	Difference Between Percents
Judges (Vehicle)	80.8%	75.0%	46.1%	58.3%
Law Students (Vehicle)	92.5%	85.0%	90.0%	76.0%
General Pop. (Vehicle)	90.2%	67.2%	42.6%	57.7%
General Pop. (Ten)	75.3%	73.1%	73.5%	37.1%

Table 4. Maximum error for corpus linguistic judgments in Experiments 3 and 4, under different theoretical assumptions.

Maximum Dictionary Error				
	Ordinary Meaning = 50%	Ordinary Meaning = 75%	Ordinary Meaning = 90%	Difference Between Percents
Judges (Vehicle)	96.6%	100%	100%	49.3%
Law Students (Vehicle)	87.2%	87.2%	94.4%	66.1%
General Pop. (Vehicle)	87.9%	89.4%	89.4%	57.7%
General Pop. (Ten, "Full")	85.1%	87.8%	93.9%	73.2%
General Pop. (Ten, "Bare")	86.7%	86.1%	88.6%	76.5%

Table 5. Maximum error for dictionary judgments in Experiments 3 and 4, under different theoretical assumptions.

These results indicate the potential gravity of the risk of error in relying on dictionaries and corpus linguistics in interpretation. In a number of interpretive tasks, the percentage of judges, law students, and ordinary people reaching incorrect verdicts on the basis of corpus linguistics and dictionaries reached 50%, 75%, and even 100%.

V. HOW JURISTS USE DICTIONARIES AND CORPUS LINGUISTICS

Part V considers the ecological validity of the experimental results and interpretation. Do legal uses of dictionaries tend to reflect broad, extensive interpretations, while legal uses of corpus linguistics tend to reflect narrow, prototypical uses? While corpus linguistics is relatively new in legal decision making, dictionaries are frequently cited. Part V.A surveys the pattern of citation and finds that caselaw tends to refer to dictionary definitions as "broad" significantly more often than as "narrow." Moreover, while dictionaries are often understood to be broad, their definitions are sometimes narrowed by considering contextual features or which of multiple definitions is most relevant.

Part V.B considers whether this property of dictionary definitions—they are generally broad but can be narrowed by some interpretive choices—may admit of political decision making. The Part considers two examples from the Bill of Rights, each of which contains three terms: the Second Amendment’s “keep and bear arms” and the Eighth Amendment’s “cruel and unusual punishment.” Republican-appointed justices, at the Supreme Court and Circuit Court level, more frequently cite Founding Era dictionaries to interpret terms like “keep” “bear” and “arms” broadly. Conversely, when Republican-appointed justices cite dictionaries in Eighth Amendment cases, the broad dictionary definitions are interpreted narrowly. Although Democratic-appointed justices cite dictionaries less frequently, when they do, the pattern is reversed: dictionaries support that “cruel and unusual punishment” is broad, but “keep and bear arms” is narrow.

Part V.C suggests that the dictionary-extensive, corpus-narrow relationship holds for other divisive examples. “Emoluments” seems broad when scholars survey Founding Era dictionaries, but narrow when scholars consider usage in historical corpora. So too for “Commerce.” The experimental insight about originalist uses of dictionaries and corpus linguistics sheds light on these and other debates about original public meaning.

A. Dictionaries Often Supply “Broad” Senses and Corpus Linguistics Supplies “Narrow” Ones

The experimental results indicate that for many (but not all) examples, dictionary definitions tended to reflect a broad sense of category membership. Those using the dictionary were inclined to include far more entities as category members, compared to those using corpus linguistics. And those using dictionaries were even inclined to categorize some entities as category members that are not judged to be category members in ordinary language. For example, dictionary users evaluated baby-shoulder carriers and rollerskates as vehicles, but most people do not consider those entities to be vehicles.

This result may seem less surprising if we reflect on the nature of a dictionary. Dictionaries often present brief definitions that aim to comprehensively reflect a broad range of permissible uses. A “vehicle” is defined as an “agent of transmission” or a “carrier.” This definition is broader than the definition of the most prototypical vehicles. For example, a car might be defined as an entity with four wheels that drives on roads. But using that as the definition of a vehicle would (inappropriately) exclude airplanes.

If this interpretation is right, we might expect courts’ usage of dictionaries reflect a similar sense that dictionaries provide broad definitions. As one approximation, consider courts’ usage of the terms “broad” and “narrow” in the context of discussing dictionaries. Figure 11 shows citations to the terms “broad” and “narrow” within the same sentence (“/s”) or paragraph (“/p”) in the Supreme Court, and a sample from Lower Federal Courts and Circuit Courts.⁷²

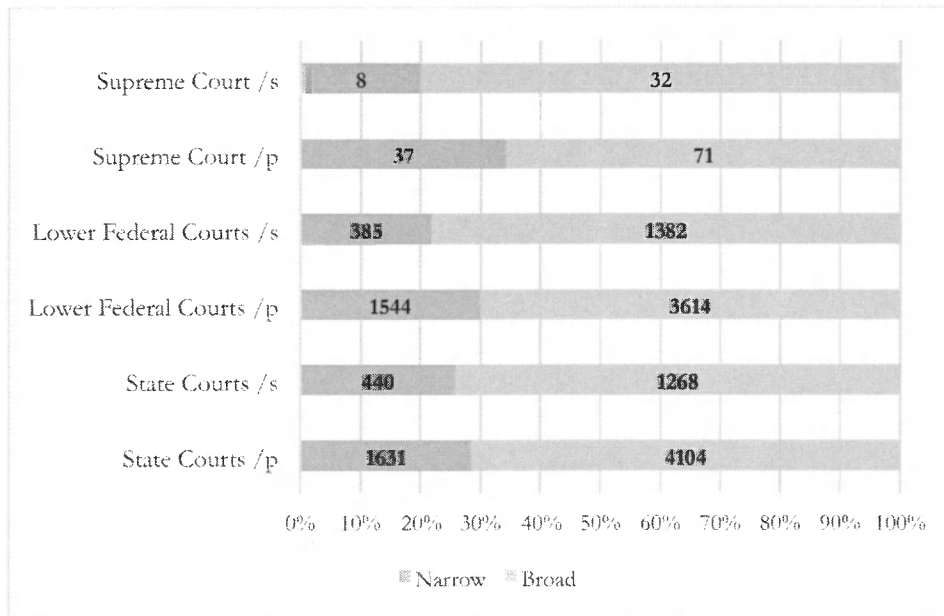


Figure 11. Court citations of “broad” and “narrow” in the same sentence (/s) or paragraph (/p) as “dictionary”⁷³

⁷² For further detail, see notes 73-85 *infra*.

⁷³ Searches conducted on Westlaw. “/s” indicates that the terms are within the same sentence; “/p” indicates that the terms are within the same paragraph.

This suggests that more often courts describe dictionaries as broad.⁷⁴ About 70% of the dictionary citations are near “broad” rather than “narrow.” Compared to an estimation that citations would appear randomly—50% near “broad” and 50% near “narrow”—this represents a statistically significant effect at all levels, for Supreme Court within sentence uses,⁷⁵ Supreme Court within paragraph uses,⁷⁶ lower federal court within sentence uses,⁷⁷ lower federal court within paragraph uses,⁷⁸ state court within sentence uses,⁷⁹ and state court within paragraph uses.⁸⁰

The same pattern of results holds true when taking into account the general frequency with which courts use “broad” and “narrow.” In one comparison, the effect was not statistically significant: Supreme Court within paragraph uses.⁸¹ However for all other comparison, the same pattern held: for Supreme Court within sentence uses,⁸² lower federal court within sentence uses,⁸³ lower federal court within paragraph uses,⁸⁴ state court within sentence uses,⁸⁵ and state court within paragraph uses.⁸⁶ Overall, this pattern of results strongly suggests that dictionaries are more often cited in the context of “broad” than “narrow.”

To further evaluate the significance of this pattern, consider some of the Supreme Court “broad” dictionary examples:

⁷⁴ Note, some might wonder whether this corpus linguistics-style analysis can consistently be relied upon given the earlier critique of corpus linguistics. Importantly, this inquiry is very different from using corpus linguistics to establish public meaning. Corpus linguistics has a number of tremendously useful possibilities. The earlier critique is leveled at the claim that corpus linguistics reflects public meaning. That is independent from the this claim that corpus linguistics provides evidence about whether dictionaries are typically described as broad or narrow.

⁷⁵ Binomial $p = .0099$.

⁷⁶ Binomial $p = .0275$.

⁷⁷ Binomial $p < .0001$.

⁷⁸ Binomial $p < .0001$.

⁷⁹ Binomial $p < .0001$.

⁸⁰ Binomial $p < .0001$.

⁸¹ Binomial $p = .2478$.

⁸² Binomial $p = .0147$.

⁸³ Binomial $p < .0001$. To provide a test case, I took the “broad” and “narrow” citations in the lower federal courts for three years, 2015, 2016, and 2017, and selected the highest ratio of broad/narrow uses, 63.9%. This selection provides a stringent test for the hypothesis.

⁸⁴ Binomial $p < .0001$.

⁸⁵ Binomial $p < .0001$. To provide a test case, I took the “broad” and “narrow” citations in state courts for January to June 2017 and July to December 2017 and selected the highest ratio of broad/narrow uses, 67.1%. This selection provides a stringent test for the hypothesis.

⁸⁶ Binomial $p < .0326$.

- “That a dictionary definition is broad enough to encompass one sense of a word does not establish, however, that the word is ordinarily understood in that sense.”⁸⁷
- “Just as the context of Rule 16 supports giving ‘tangible object’ a meaning as broad as its dictionary definition, the context of § 1519 tugs strongly in favor of a narrower reading”⁸⁸
- “...the dictionary definitions of that word are very broad.”⁸⁹
- “Broad definitions of the term in modern and older dictionaries are unhelpful”⁹⁰

One striking feature of these uses is that, while most suggest dictionary definitions are broad, many cite this as a reason that legal outcome should *not* be constrained by the meaning recommended by the dictionary.

Of course, dictionaries are not uniformly understood to provide broad definitions. About 20-30% of the time, they are referred to near “narrow.” In some of those examples, dictionaries are taken to provide a narrow definition:

- “To our knowledge *all* English dictionaries provided the narrow definition of ‘modify’ [connoting only moderate, and not fundamental, change]”⁹¹

However, many of the dictionary references near “narrow” actually suggest that dictionaries are understood to be broad:

- “Some [law dictionaries] define ‘firm’ [narrowly] But other dictionaries, while recognizing that narrow definition, also state that the word has a broader meaning...”⁹²
- “[C]ommon usage at the time of the National Bank Act prevents the conclusion that the Comptroller’s refusal to give the word ‘rate’ the narrow meaning petitioner demands is

⁸⁷ *Taniguchi v. Kan Pacific Saipan, Ltd.*, 566 U.S. 560 (2012).

⁸⁸ *Yates v. U.S.*, 135 S.Ct. 1074, 1083 (2015).

⁸⁹ *Smith v. U.S.*, 508 U.S. 223 (1993).

⁹⁰ *Bullock v. BankChampaign, N.A.* 569 U.S. 267 (2013).

⁹¹ *MCI Telecommunications Corp. v. American Tel. & Tel. Co.*, 512 U.S. 218 (1994).

⁹² *U.S. v. Cook*, 384 U.S. 257, n.5. (1966).

unreasonable. The 1849 edition of Webster's gives as one of the definitions of 'rate' the '[p]rice or amount stated or fixed on any thing.'⁹³

- “When we have stated that sovereignty is a political question, we have referred not to sovereignty in the general, colloquial sense, meaning the exercise of dominion or power, see Webster's New International Dictionary 2406 (2d ed. 1934) ('sovereignty,' definition 3), but sovereignty in the narrow legal sense of the term, meaning a claim of right”⁹⁴

These empirical results are consistent with the experimental findings: Dictionary definitions often (but not always) supply a broad, extensive sense of meaning. Importantly, they sometimes provide narrow meanings. Of course, there are some ways in which broad dictionary definitions might be narrowed. Most notably, one might narrow a broad definition by (a) considering linguistic context, (b) selecting a relatively more narrow definition, where there are multiple definitions, or (c) emphasizing the necessity of particular features of the definition.

Corpus linguistics is relatively new and has yet to appear in a range of court decisions. So, inevitably, the conclusions that we can draw from caselaw practice are much more limited. However, in the few cases that explicitly cite corpus linguistics, the results tend to narrow the contested sense of meaning.

For example, in the first opinion using corpus linguistics, *In re the Adoption of Baby E.Z.*,⁹⁵ Justice Lee analyzed the phrase “custody determination.” He considered 500 sample sentences from the Corpus of Contemporary American English, and reported that the most common use of “custody” was in the context of divorce, rather than in the context of adoption. He concluded that “the custody proceedings covered by the Act are limited to proceedings resulting in the modifiable custody orders of a divorce,” rather than in a broader range of custody proceedings.⁹⁶

⁹³ *Smiley v. Citibank (South Dakota)* 517 U.S. 735, 746 (1996).

⁹⁴ *Boumediene v. Bush*, 553 U.S. 723, 754, 128 S. Ct. 2229 (2008).

⁹⁵ 266 P.3d (Utah 2011).

⁹⁶ *Id.*

This is consistent with the experimental findings, in which corpus linguistics often suggests that ordinary meaning is limited to prototypical uses.

Similarly, in *State v. Rasabout*, corpus linguistics suggested that “discharge” was largely used to refer to a single shot of a firearm, rather than emptying the entire magazine.⁹⁷ This, too, is a narrowing interpretation, limiting the ordinary meaning of “discharge” to the most common and prototypical use.

Finally in 2018 Justice Thomas made the first explicit reference to corpus linguistics in the Supreme Court. In a dissent regarding the meaning of “expectations of privacy,” Thomas notes that “[t]he phrase ‘expectation(s) of privacy’ does not appear in . . . the papers of prominent Founders, early congressional documents and debates, collection of early American English texts, or early American newspapers.”⁹⁸ This reflects a broadly similar use of corpus linguistics: the relative infrequency of a use from the corpus (in this case, the absence of a use) is taken to suggest that the use is not part of public meaning.

To be sure, corpus linguistics need not always provide a narrowing/exclusive recommendation concerning meaning. But the early judicial uses of corpus linguistics are suggestive of such a trend.

B. Political Uses of Dictionaries and Corpus Linguistics

If dictionaries often provide broad, extensive senses of meaning, we should expect that jurists that cite dictionaries should reach inclusive or exclusive conclusions when dictionaries are cited at equal rates across similar types of cases. In other words, if dictionaries often reflect a broader sense of a term (say seventy or eighty percent of the time), we might expect that citations of dictionary definitions lead to an inclusive interpretation at similar rates (e.g. seventy or eighty percent of the time).

⁹⁷ 356 P.3d 1258 (Utah 2015) (Lee, J., concurring).

⁹⁸ *Carpenter v. United States*, 585 U.S. ___ (2018) (Thomas, dissenting) (citing corpus linguistics data).

However, one might wonder whether dictionaries are sometimes used politically. If so, we might expect that jurists who cite dictionaries reach narrow/exclusive interpretations of the definition at surprisingly high rates when that narrow interpretation is consistent with the outcome associated with their political affiliation.

As one example, consider the contrast between two important clauses from the Bill of Rights: the right to “keep and bear arms” and the protection against “cruel and unusual punishments.” Broadly speaking, modern Republicans would prefer the former right interpreted broadly and the latter protection narrowly, while modern Democrats would prefer that the former right is interpreted narrowly and the latter protection broadly. But what do judges actually do?

Consider how Republican-appointed and Democratic-appointed federal jurists interpret dictionaries to support broad interpretations equally in Second and Eighth Amendment cases. First take “keep and bear arms.” The only Supreme Court case in which dictionaries are used to interpret these Second Amendment terms is *District of Columbia v. Heller*.⁹⁹ But this is a rich case. The majority cites dictionaries to interpret all three terms, “keep,” “bear,” and “arms.” And the dissent also cites a dictionary to interpret “bear arms.”¹⁰⁰

First take the majority holding, authored by Republican-appointed Justice Scalia. He cites dictionary definitions of “keep,” “bear,” and “arms.” For “arms,” Scalia cites Samuel Johnson’s 1773 dictionary, which “defined ‘arms’ as ‘[w]eapons of offence, or armour of defense.’”¹⁰¹ He also cites Timothy Cunningham’s 1771 legal dictionary, which “defined ‘arms’ as ‘any thing that a man wears for his defense, or takes into his hands, or useth in wrath to cast at or strike another.’”¹⁰² Scalia also cited, but does not print, Noah Webster’s 1828 definition.

⁹⁹ 554 U.S. 570 (2008).

¹⁰⁰ *Id.*

¹⁰¹ *Id.* at 581.

¹⁰² *Id.*

Scalia also cites dictionary definitions of “keep.” He cites Johnson for the claim that keep meant “most relevantly, ‘[t]o retain; not to lose,’ and ‘[t]o have in custody.’”¹⁰³ Moreover, “Webster defined it as ‘[t]o hold; to retain in one’s power or possession.’”¹⁰⁴ Thus, Scalia concludes, “the most natural reading of ‘keep Arms’ in the Second Amendment is to ‘have weapons.’”¹⁰⁵

Finally, Scalia cites Johnson, Webster, Sheridan, and the Oxford English Dictionary for the claim that “bear” meant “carry.”¹⁰⁶ For all three terms, the dictionary definition is understood to convey a broad sense of meaning, one that is inclusive in the context of *Heller*.

Conversely, the *Heller* dissent, signed by all the Democratically-appointed Justices, cites “bear arms,” finding that its dictionary meaning is “to serve as a soldier, do military service, fight.”¹⁰⁷ It also cites the very same Johnson Dictionary definition that Scalia cites—“weapons of offence, or armour of defence”—but understands it to apply narrowly, exclusive of the use contested in *Heller*.

This same pattern emerges in circuit courts.¹⁰⁸ Of all cases citing dictionaries in the same sentence as “keep,” “bear,” or “arms,” one of those defining a Second Amendment term was authored by a Democratic-appointed judge, and six were authored by Republican-appointed judges.

In the one case in which a Democratic-appointed judge used a dictionary, it supported interpreting “bear arms” narrowly.¹⁰⁹ In the other six, a Republican-appointed judge used a dictionary to interpret keep and bear arms broadly.¹¹⁰ Thus, at both the Supreme Court and circuit

¹⁰³ *Id.* at 285.

¹⁰⁴ *Id.*

¹⁰⁵ *Id.*

¹⁰⁶ *Id.* at 584

¹⁰⁷ *Id.* at 646.

¹⁰⁸ I considered all Westlaw-listed cases heard in Federal Courts of Appeal that cite the Second Amendment. I searched within those for uses of “dictionary” within the same sentence as “keep,” “bear,” or “arms.” Seven cases were returned, six of which used a dictionary to define a Second Amendment term.

¹⁰⁹ *Silveira v. Lockyer*, 328 F.3d 567 (2003).

¹¹⁰ *Kolbe v. Hogan*, 813 F.3d 160 (2016); *Parker v. District of Columbia*, 478 F.3d 370 (2007); *Nordyke v. Kind*, 319 F.3d 1185 (2003); *Silveira v. Lockyer*, 312 F.3d 1052 (2002).

court level, a judge's dictionary use in Second Amendment cases matched the outcome predicted by the political affiliation of the appointing president.

Contrast this with the use of dictionaries in Eighth Amendment cases. In *Furman v. Georgia*, Justice White refers to the broad dictionary sense of "cruel": "The imposition and execution of the death penalty are obviously cruel in the dictionary sense."¹¹¹

However, more recent conservative-authored opinions use dictionaries to construe the Eighth Amendment's protection narrowly. Consider Thomas and Scalia in *Baze v. Rees*, arguing that lethal injections for executions are constitutional:¹¹²

Embellishments upon the death penalty designed to inflict pain for pain's sake also would have fallen comfortably within the ordinary meaning of the word "cruel." See 1 S. Johnson, *A Dictionary of the English Language* 459 (1773) (defining "cruel" to mean "[p]leased with hurting others; inhuman; hard-hearted; void of pity; wanting compassion; savage; barbarous; unrelenting"); 1 N. Webster, *An American Dictionary of the English Language* 52 (1828) (defining "cruel" as "[d]isposed to give pain to others, in body or mind; willing or pleased to torment, vex or afflict; inhuman; destitute of pity, compassion or kindness").

It is worth considering the full definition of cruel cited in these dictionaries. First, take "cruel" in Johnson's 1773 dictionary:

1. Pleased with hurting others; inhuman; heard hearted; barbarous. *Dryden*.
2. [Of things.] Bloody; mischievous; destructive. *Psalms*.

Now consider "cruel" in Webster's 1828:

1. Disposed to give pain to others, in body or mind; willing or pleased to torment, vex or afflict; inhuman; destitute of pity, compassion or kindness; fierce; ferocious; savage; barbarous; hardhearted; applied to persons or their dispositions.

They are *cruel* and have no mercy. Jeremiah 6:23.

2. Inhuman; barbarous; savage; causing pain, grief or distress; exerted in tormenting, vexing or afflicting.

Cursed be their wrath, for it was *cruel*. Genesis 44:1

The tender mercies of the wicked are *cruel* Proverbs 12:10

Others had trials of *cruel* mockings. Hebrews 11:36

¹¹¹ 408 U.S. 238, 312 (1972).

¹¹² 553 U.S. 35, 97 (2008).

It is striking that Scalia and Thomas use the definitions that are applied to persons, rather than the definition applied to “things” (like the Eighth Amendment’s “punishment”). Although the definitions relevant to persons appear first, it would seem that the definitions relevant to punishment (a thing) may be more apt. Understanding “cruel” punishment as ones that are “destructive,” or “causing pain, grief or distress” suggests a much broader ordinary meaning.

Republican-appointed Supreme Court justices apply dictionary definitions similarly (i.e. exclusively/narrowly) when defining “unusual.” For example, in *Harmelin*, the court considered whether the imposition of mandatory sentences of life in prison without the possibility of parole, without any consideration of mitigating factors, constituted cruel and unusual punishment. Scalia writes for the majority, concluding that such punishment is not “unusual.” Unusual means, according to Scalia “such as [does not] occu[r] in ordinary practice, Webster’s American Dictionary (1828), “[s]uch as is [not] in common use,” Webster’s Second International Dictionary 2807 (1954).¹¹³

Finally, in *Farmer* and *Helling*,¹¹⁴ the Republican-appointed opinions indicate a narrow dictionary construal of “punishment.” It does not include an attack on a prisoner, and punishment is only the penalty for the commission of a crime, not jail conditions.

This same pattern is consistent with the limited evidence from circuit courts.¹¹⁵ Of all cases citing dictionaries in the same sentence as “cruel,” “unusual,” or “punishment,” only one of those defined an Eighth Amendment term. This is *Duckworth v. Franzen*, 780 F.2d 645 (1985), in which a Republican-appointed judge cites Johnson’s dictionary to support that “punishment” does not include injuries sustained when a bus to which prisoners were chained caught fire. He argues

¹¹³ *Harmelin v. Michigan*, 501 U.S. 957 (1991).

¹¹⁴ *Farmer v. Brennan*, 511 U.S. 825 (1994); *Helling v. McKinney*, 509 U.S. 25 (1993).

¹¹⁵ I considered all Westlaw-listed cases heard in Federal Courts of Appeal that cite the Eighth Amendment. I searched within those for uses of “dictionary” within the same sentence as “cruel,” “unusual,” or “punishment.” Seven cases were returned, one of which used a dictionary to define an Eighth Amendment term.

that the dictionary definition requires that punishment be deliberate or reckless in the criminal law sense, or a “strong sense.”

It is helpful to consider the broader pattern that emerges in these uses of dictionaries to determine whether legal texts are exclusive (narrow construal) or inclusive (broad construal). Recall Figure 11, which suggests that, cross all levels of the judiciary dictionaries tend to admit of “broad” interpretations about 70% of the time and “narrow” interpretations about 30% of the time. *If* we expect broad interpretations to imply inclusive legal determinations and narrow interpretations to imply exclusive legal determinations, we should find similar proportions across issues and political ideologies.

However, what we have found in Second and Eighth Amendment caselaw at the Supreme Court and Circuit Court level does not reflect this pattern. Instead, Republican-appointees tend to construe dictionary definitions broadly when interpreting the terms “keep,” “bear,” and “arms,” but narrowly when interpreting the terms “cruel,” “unusual,” and “punishment.” Democratic-appointees use dictionaries much less often, but when they do the pattern reverses: dictionaries indicate that Second Amendment terms are narrow, but Eighth Amendment terms are broad.

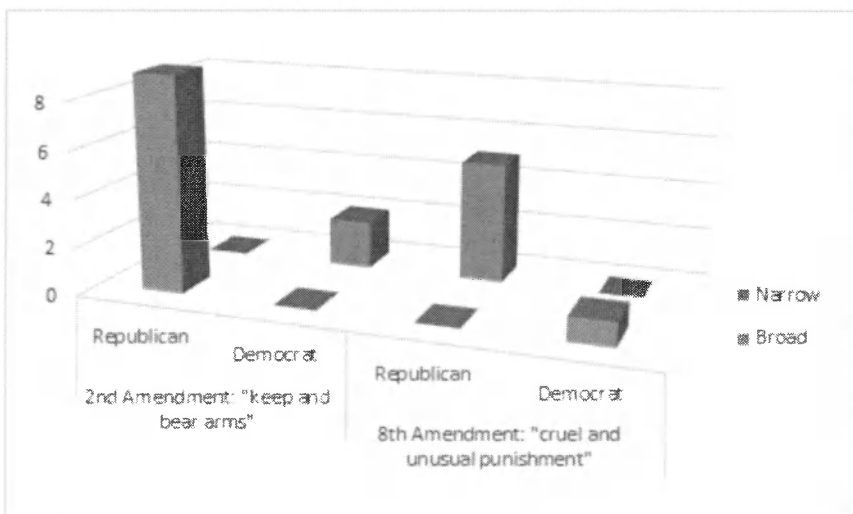


Figure 12. Exclusive (“Narrow”) and Inclusive (“Broad”) Uses of Dictionaries by Republican and Democrat appointed Justices and Judges, in interpreting “keep and bear arms” and “cruel and unusual punishment”

To be sure, this section considers a relatively small sample. Future work may provide further insight into the question of whether and how dictionaries are used politically. The modest empirical analysis here suggests that it is a worthwhile question.

Given the novelty of corpus linguistics, it is difficult to assess its political uses. However, judges may be wise to keep watch of emerging patterns. If current originalist uses of corpus linguistics tend to produce narrow interpretations, does this manifest equally across different areas of law?

C. Explaining Debates about Original Public Meaning

This pattern of results—dictionaries tend to generate broader senses of meaning and corpus linguistics tend to generate narrower senses of meaning, and the sense generated by each may be different—also helps explain some divisive debates about particular terms.

As one example, consider the recent debate about the original public meaning of “emolument” in the Constitution. One putative, “narrow,” sense of the meaning is something like “a profit arising from office or employ.” But another putative, “broad,” sense need not involve “office” or “employment.”

An impressive analysis of Founding Era dictionaries finds support for the broad interpretation.¹¹⁶ Across a sample of over one-hundred dictionaries, the broad meaning gains support.

Conversely, a corpus linguistics analysis finds support for a less extensive meaning.¹¹⁷ The study’s authors report that the broad sense of “emolument” was more common than the narrow sense in an ordinary language corpus (20% more common), but the narrow sense was more common in “elite” and “legal” corpora (35% and 43% more common, respectively). The paper

¹¹⁶ John Mikhail, *The Definition of ‘Emolument’ in English Language and Legal Dictionaries, 1523-1806*.

¹¹⁷ James Cleith Phillips & Sara White, *The Meaning of the Three Emoluments Clauses in the U.S. Constitution: A Corpus Linguistic Analysis of American English, 1760-1799*, 59 S. TEX. L. REV. (2018).

concludes that the Congressional and Presidential Emoluments Clauses would have been understood to contain a narrow sense of emolument, while the Foreign Emoluments Clause is more ambiguous.

A similar debate characterizes analysis of dictionaries and corpus linguistics concerning the original public meaning of “commerce” in the Commerce Clause. Does “commerce” mean something broad like “any gainful activity” or “intercourse,” or something narrower like “the trade and exchange of goods and transportation for this purpose”? Samuel Johnson’s dictionary defines “commerce” broadly.¹¹⁸ However, a thorough corpus linguistics style examination of “every appearance of the word ‘commerce’ [in several Founding Era sources]... finds no surviving example of this term being used in this broader sense.”¹¹⁹

Again, these results are offered primarily as evidence suggestive of the ecological validity of the experimental results. Dictionary-use often reflects broad senses of meaning, while corpus linguistics use tends to reflect narrower senses. The next Part explores this mechanism in more detail and advances normative implications for theories of legal interpretation.

VI. IMPLICATIONS

This Part turns to the experimental findings’ normative implications for theories of interpretation. Part VI.A considers how the results contribute to refining the methodologies of Public Meaning Originalism, as well as other interpretive theories that seek to discover original public meaning.

Part VI.B advances more critical arguments. It identifies several fallacies of the use of corpus linguistics and dictionaries that are supported by the experimental data. Next, it defends a broader burden-shifting argument. The experiments provide evidence that dictionaries and corpus linguistics are unreliable methods in discovering original public meaning. Given the experimental

¹¹⁸ Jack Balkin, *Commerce*, 109 MICH. L. REV. 1 (2010); see also *Gibbons v. Ogden*, 22 U.S. (9 Wheat.) 1 (1824).

¹¹⁹ Randy E. Barnett, *The Original Meaning of the Commerce Clause*, 68 U. CHI. L. REV. 101 (201).

results, Public Meaning Originalists have the argumentative burden of elaborating a non-arbitrary and demonstrably reliable use of corpus linguistics and dictionaries in interpretation.

A. Understanding and Improving Originalist Methodology

1. Psychological Mechanisms of Use of Dictionaries and Corpus

Recall the experimental results that shed light on the psychological processes underlying use of dictionaries and corpora in interpretation. Verdicts from dictionaries were more strongly correlated with a term's extensivist uses than its prototypical ones. And verdicts from corpus linguistics were more strongly correlated with a term's prototypical uses than its extensivist ones.

For example, consider that a car is a prototypical vehicle but an airplane is not (although most people today still judge that an airplane is a vehicle).¹²⁰ Most participants using the dictionary were inclined to classify both entities as vehicles, but most participants using the corpus data only classified the (prototypical) car as a vehicle.¹²¹ Similarly, when participants considered the meaning of "carrying" a firearm, those using the dictionary provided more extensive judgments. Those using corpus data largely categorized prototypical examples as carrying (e.g. taking a gun to a gang fight), but more often excluded non-prototypical examples (e.g. driving to a drug deal with a gun in the rear of the car).¹²²

These results suggest something about the cognitive mechanisms underlying the use of dictionaries and corpus, respectively. While dictionary definitions help identify more extensivist uses of the term, corpus linguistics data tends to help identify prototypical uses. There is something of a puzzle about this interpretation. Corpus data is far more extensive than a brief dictionary definition, yet it is the latter that reflects a more extensive sense of meaning. Somewhat counterintuitively, copious corpus data produce relatively narrow (prototypical) judgments about meaning.

¹²⁰ See Section IV.B *supra*.

¹²¹ *Id.*

¹²² *Id.*

An important objection may be raised here: Although this pattern holds for two examples in this paper, how can we infer that this represents a broader pattern of judgment for many terms and phrases? To answer this challenge, first recall that the examples here were not chosen arbitrarily. The two examples—vehicles and carrying a firearm—are two historically famous cases of statutory interpretation problems.¹²³ They were chosen for this reason and because they are the first two examples used by the recent manifesto on corpus linguistics.¹²⁴ That is, these are examples selected by other researchers who were unaware of the present hypotheses. Moreover, that paper is the leading *defense* of corpus linguistics; insofar as the results here challenge corpus linguistics, I have used cases that should have been selected by its proponents, which should be favorable.¹²⁵

But maybe the original paper from which these examples are drawn just happened to select two unusual examples. That remains an open empirical question. The Corpus-Prototypical and Dictionary-Extensivist relationship is certainly somewhat limited, insofar as not all categories have prototypical and non-prototypical members.¹²⁶ Thus, the response to this objection is not a flat refutation. Instead, it is an acknowledgement that additional data on other examples may very well enrich our understanding of dictionaries and corpus linguistics further.

Nevertheless, there are some further theoretical reasons to expect that this pattern of results would extend to other terms that admit of prototypical and more extensive uses. First, consider the type of evidence supplied by corpus linguistics. Advocates of using corpus linguistics recommend data from collocation and keywords in context searches. Collocation searches reflect the words that appear most frequently with the relevant search term. There is good reason to think that many of these most common collocates are also representative of the features that we

¹²³ See e.g., Lee & Mouritsen, *supra* note 1; Solan & Gales, *supra* note 9.

¹²⁴ Lee & Mouritsen, *supra* note 1.

¹²⁵ *Id.*

¹²⁶ However, even well-defined terms like “even,” “odd,” “female,” and “plane geometry figure” elicit judgments that features similar to judgments of prototype concepts. For example, four judged as a better example of an even number than thirty-four; mother is a better example of a female than an actress; and a circle is a better example of a plane geometry figure than an ellipse. Sharon Lee Armstrong, Lila R. Gleitman & Henry Gleitman, *What Some Concepts Might Not Be*, 13 COGNITION 265 (1983).

commonly attribute to the entity. For example, “vehicle” often appears near “electric,” “gas,” and “motor.” We might also represent those as core features associated with vehicles. According to prototype theory, prototypical members are the ones with most of all of those features. So, insofar as the statistically common collocates also reflect our core associations with the concept, there is good reason to think that collocation data is especially useful in identifying prototypical category members.

The second type of search, keywords in context, might also be especially useful in identifying prototypical category members. Keywords in context searches return example sentences from the corpus. We might think that such a search would return many types of uses—prototypical and non-prototypical. However, in practice, pragmatic considerations might limit the number of non-prototypical uses that we find.¹²⁷

Consider these example sentences:

- (1) Did you see any fish in the ocean?
- (2) Look at that bird!
- (3) The painter will finish painting Mike’s fence tomorrow.

We might expect to find a sentence like (1) that refers to prototypical fish like trout or carp. It would not only be uncommon, but seemingly *inappropriate* to say (1) if we meant to refer to sharks—instead we would ask “did you see any sharks in the ocean?” It would be even more unusual to say (1) if we meant to refer to stingrays. Of course, this pragmatic fact in no way undermines that people understand that sharks and stingrays are fish.

Similarly, (2) might occur when someone describes a prototypical bird, like a robin or sparrow. We would expect to see examples like that in our modern corpus. But it would be a strange way to call attention to a penguin—even though penguins are birds. And we would probably not find many of these kinds of examples referring to penguins in the corpus.

¹²⁷ See generally H.P. GRICE, LOGIC AND CONVERSATION (1975).

In the same way, we would expect to find a sentence like (3) that refers to a prototypical painter (e.g. an adult who works as a painter). Of course, if Mike's twelve-year old niece enjoys painting and will paint his house, (3) could refer to her. But it would be strange, even inappropriate, to say (3) in that context. Instead, we would probably say something like, "Mike's niece will finish painting Mike's fence tomorrow." Mike's niece is still a painter, and anyone familiar with that fact would agree that she is a "painter" in the ordinary sense of the term. Nevertheless, pragmatically, we would not usually say something like (3) if we meant to convey that Mike's niece will paint.

Now consider some of our legal examples:

(4) Asaf said we have to renew the vehicle registration.

A similar phenomenon operates here. It is possible that (4) could refer to an airplane registration, but it is more likely that we take the sentence to indicate a car registration. (4) would be a strange, if not inappropriate, way to describe an airplane. Consider a final example:

(5) How did Jasmin get all those books to school? She carried them there.

This would be an appropriate way to express that Jasmin hand-carried books to school. Of course, it could also express the fact that Jasmin loaded books into a wheelbarrow and towed them to school. But to express that, we would probably say something more specific than (5). Nevertheless, it is still true (in the ordinary sense of the term "carry") that Jasmin carried the books to school.

These examples suggest an interesting phenomenon. Often, it is pragmatically inappropriate to refer to non-prototypical category members by the broader category description. If you want to point out sharks in the water, you don't say "look at those big *fish!*"¹²⁸ This pattern of usage is

¹²⁸ See, e.g., Martin L. Jönsson & James A. Hampton, *On Prototypes as Defaults* COGNITION (2007). But see Andrew C. Connolly, Jerry A. Fodor, Lila R. Gleitman & Henry Gleitman, *Why Stereotypes Don't Even Make Good Defaults*, 103 COGNITION 1 (2007). This phenomenon finds some support in the cognitive science of default interpretations. Terms and propositions often have default,

perfectly consistent with the fact that sharks are understood to be fish (they are part of the ordinary meaning).

Given this phenomenon, we should expect that keywords in context searches can often deliver an incomplete picture of a term's ordinary meaning. Because corpus linguistics does not fully reflect pragmatic effects, there are a number of uses that are entirely consistent with ordinary meaning that nevertheless *should* not appear frequently in the corpus.

2. Guidelines for Using Originalist Empirical Tools

Assuming that dictionaries and corpus linguistics tend to identify extensivist and prototypical members of a category, respectively, there is a straightforward recommendation for the use of these tools in interpretation. Where the interpretive context calls for a broad, extensivist reading of the term or phrase, dictionaries may be better guides than corpora. But where the interpretive context calls for a prototypical reading of the term or phrase, corpora may be better guides than dictionaries.

For example, consider the “no vehicles in the park” ordinance. Although that ordinance does not provide much context that implies the appropriateness of an extensivist or prototypical sense of vehicle, a modified version might provide that information. “*Any and all* vehicles are prohibited from the park” might suggest that “vehicles” should be construed rather extensively. Alternatively, “*Only* cars, trucks, and other vehicles are prohibited from the park” might suggest a more prototypical sense of “vehicles.”

In practice, using both dictionaries and corpora is likely better than relying on either alone. For example, imagine that the interpretive context calls for a broad, extensivist reading of the term or phrase. While dictionaries are a comparatively better source for generating this extensivist sense, we might also cross-check the corpus for relatively rare uses. Although the absence of a

presumptive, or preferred interpretations. These default interpretations are often more easily available and are generated in a shorter time than non-default interpretations. See generally S.C. LEVINSON, PRESUMPTIVE MEANINGS: THE THEORY OF GENERALIZED CONVERSATIONAL IMPLICATURE (2000).

use from a corpus cannot guarantee that such a use is outside of the public meaning, the presence of a use from a corpus can support that a use is within (at least a non-prototypical) sense of the public meaning.¹²⁹

In some other cases, use of both dictionaries and corpus would be necessary. For example, consider criminal contexts in which the rule of lenity applies.¹³⁰ Insofar as dictionaries and corpora elicit different senses of a term in this context (e.g. a “prototypical sense” and an “extensivist sense”), we would want to compare both senses and apply whichever is more consistent with the rule of lenity. Depending on the context, either the more extensive or more prototypical sense could comport with the rule of lenity.

A final recommendation concerns interpretation of recent statutes. When we seek to interpret modern statutes, we might also consider using experimental tools to help discern modern public meaning. We often look to dictionaries or corpora because that is our best historical evidence of public meaning. But when a modern text is at issue, experimental methods may offer a more direct insight into public meaning.¹³¹

B. Challenges to Originalist and Textualist Methodologies

The previous Section elaborates some implications of the data in a way most helpful to theories that rely on dictionaries and corpus linguistics, like Public Meaning Originalism. It strives to use the empirical results to improve the methodology of interpreting original public meaning. This Section considers the data through a more critical lens. It takes the originalist interpretive project itself as an object of evaluation, asking whether the data shakes or bolsters our confidence in the empirical tools commonly used to generate originalist interpretations.

¹²⁹ See also Part IV.B *infra*.

¹³⁰ I.e. there is a textual ambiguity and neither of the two possible senses is inconsistent with legislative intent).

¹³¹ Cf. Omri Ben-Shahar & Lior Jacob Strahilevitz, *Interpreting Contracts Via Surveys and Experiments*, 92 N.Y.U. L. REV. 1753 (2017) (defending an experimental method of contract interpretation).

There may appear to be some tension between these two Sections. Section VI.A argues that the experiments provide beneficial insights for Public Meaning Originalism, but Section VI.B argues that the very same experiments suggest dictionaries and corpus linguistics are unreliable. To some extent, this reflects a true ambivalence about the results. This is the very first study of the reliability of these tools and future research can further support which conclusions (positive or critical) are most warranted.

However, part of the perceived tension is illusory. Section VI.A discusses ways in which corpus linguistics and dictionaries can help, *when we have independent reason to think that a text calls for a very extensive or very prototypical interpretation*. However, in most circumstances, the text does not provide that insight. Moreover, Public Meaning Originalists are generally hesitant to make such claims and inferences—often, they hold that there is just *one* public or ordinary meaning, not competing prototypical and extensive ones. Section VI.B is concerned with those much more common circumstances.

Section VI.B.1 addresses some fallacies of the use of corpus linguistics and dictionaries. The data here suggest that many popular arguments using these tools are, in fact, fallacious.

Then, Section VI.B.2 addresses broader questions about the reliability of these tools in modern and historical interpretation and the implications for theories like Public Meaning Originalism. While the previous Section noted some ways in which the experiments here might improve these methods, ultimately the results suggest that corpus linguistics and dictionaries are surprisingly unreliable measures of modern public meaning. Even if corpus linguistics might help us determine some prototypical category members and dictionaries might help us determine what is part of a very broad category, neither tool consistently reflects straightforward public judgments about what is (simply) a category member. Insofar as that is the most common interpretive question, these tools are less helpful than commonly thought.

I argue that the experimental findings provide evidence of the unreliability of corpus linguistics and dictionary-use in historical interpretation. Moreover, this shifts the argumentative burden to Public Meaning Originalism. If that theory continues to endorse these tools, it must provide a non-arbitrary and demonstrably reliable method for using them (i.e. it must articulate, defend, and demonstrate a way in which these tools reliably track public meaning). This burden may very well be met. But if the theory cannot provide this explanation and demonstration, we should not be confident that its practice achieves its aims (i.e. determining original public meaning). If so, a question arises about whether Public Meaning Originalism is a defensible theory and practice of legal interpretation.

1. Fallacies Corpus Linguistics and Dictionary Use

I begin with some criticisms of the use of dictionaries and corpus linguistics in legal interpretation. There are many other important critiques that are not discussed here.¹³² This Section identifies new critiques, grounded in the experimental results of this paper.

This Section identifies several fallacies in the use of corpus linguistics and dictionaries. These are argumentative or inferential errors in common uses of dictionaries and corpus linguistics. Individually, these fallacies present significant challenges to originalist interpretations; collectively, they threaten the plausibility of relying upon dictionaries and corpus linguistics in originalist interpretation.

First consider some fallacies of corpus linguistics:

The Non-Appearance Fallacy: The non-appearance of some use in the corpus indicates that this use is outside of the public meaning.

It is tempting to think that any acceptable use must be found in the corpus, and any use that is not reflected in the corpus is not part of the ordinary public meaning.¹³³ Defenders of corpus

¹³² See, e.g., Solan & Gales, *supra* note 9.

linguistics defend this argument in holding that airplanes are not vehicles: “With respect to the use of *vehicle* to reference airplane, the answer is simpler. . . . we were unable to find a single collocation or concordance line that reflected the use of *vehicle* to mean *airplane*. . . . [B]ased on its absence from any of our corpus data, we might ask if *airplane* is even a possible sense of *vehicle*.”¹³⁴

But this argument is fallacious. As the experimental results here indicate, corpus linguistics often neglects non-prototypical uses of a term. A corpus search for “vehicle” returns predominantly uses involving cars. But this does not mean that only prototypical uses reflect the public meaning of “vehicle.” As the experimental results (and common sense) indicate, golf carts, airplanes, and horse-drawn carriages are also within the modern public meaning of “vehicle.”

It is highly significant to recognize this fallacy, as the argument often seems to have great rhetorical strength: “in an *entire* corpus, containing tens of thousands of uses, there were *no* uses reflecting such a meaning.” This argument is fallacious in our modern moment. It is also fallacious in historical interpretation. A historical corpus is only smaller than the modern corpus used in the experiments, presumably containing even fewer non-prototypical uses.

A second fallacy flows from the same set of observations and experimental results.

The Uncommon Use Fallacy: The relative rarity of some use in the corpus indicates that this use is outside of the public meaning.

This follows from the logic identifying the first fallacy. Insofar as corpus linguistics data may not adequately reflect non-prototypical uses, one cannot conclude that the rarity of use implies that such a use is not part of the term’s public meaning.

Consider one final fallacy.

¹³³ See, e.g., *Carpenter v. United States*, 585 U.S. ___ (2018) (Thomas, dissenting) (noting that “[t]he phrase ‘expectation(s) of privacy’ does not appear in . . . the papers of prominent Founders, early congressional documents and debates, collection of early American English texts, or early American newspapers.”)

¹³⁴ Lee & Mouritsen, *supra* note 1, at 844.

The Comparative Use Fallacy: When considering two possible senses, the comparatively greater support for one sense in the corpus indicates that this sense is a better candidate of public meaning.

This fallacy is more complicated than the first two, but it may be even more prevalent. It arises when users of corpus linguistics aim to determine which of two possible senses is the better candidate for public meaning. This may happen, for example, if there is debate over whether a term is ambiguous; if one possible sense is much more often reflected in the corpus, one might conclude that this that sense reflects the *only* plain meaning.

However, this too is a fallacious argument. Recall the experimental findings. For ordinary people, law students, and United States judges, there were several entities that were classified as vehicles in ordinary language, but not with respect to the corpus data (e.g. airplane, bicycle, electric wheelchair). Imagine there was a debate over the meaning of vehicle. Sense-1 is the inclusive sense (car, truck, airplane, bicycle, and electric wheelchair) and Sense-2 is the exclusive sense (only car and truck; and not airplane, bicycle, or electric wheelchair). Users of corpus linguistics might be inclined to argue that Sense-1 is the better candidate, as it has more support from the corpus. However, this is a fallacious inference. As discussed previously, the omission of non-prototypical uses from the exclusive sense does not mean it is a better sense or one that reflects *the* (only) plain meaning of “vehicle.”

As it happens, Justice Scalia recognized the dangers of this fallacy: “the fact that [a term or] phrase was commonly used in a particular context does not show that it is limited to that context.”¹³⁵ In other words, the fact that a term or phrase is more common in one context than another, does not mean that that it is not limited to the former context.

These three fallacies—The Non-Appearance Fallacy, The Uncommon Use Fallacy, and The Comparative Use Fallacy—each present an individual challenge to common interpretive arguments grounded in corpus linguistics data. But we should also note that these three arguments

¹³⁵ See *D.C. v. Heller*.

together threaten much of the usefulness of corpus linguistics. If corpus linguistics cannot reliably exclude omitted or rare uses (from public meaning) or determine which of two possible public meanings is more credible, this undercuts much of the method's promise.

I return to this broader theme in the next Part. But first consider two fallacies of dictionary use:

The “It Fits the Definition” Fallacy: When considering whether a use falls under the public meaning, we should conclude that the use is part of the public meaning if it fits the relevant definition.

In the studies presented here, dictionary users categorized as vehicles several items that were not judged to be vehicles by ordinary language users. Often, dictionary definitions seem to aim to convey a comprehensive set of meaning. In defining vehicle, we must provide a definition that includes cars and trucks, but also airplanes, submarines, and mopeds. “An agent of transmission or carrier” is helpful in achieving this. But such a broad definition also applies to many entities that are not understood as vehicles. For example, participants predominantly reported that roller skates, baby shoulder-carriers, and zip-lines are *not* vehicles. Yet, many dictionary users categorized these as vehicles. Thus, while it might seem that a dictionary definition is tied tightly to public meaning, this assumption is erroneous. The mere fact that a use “fits” the dictionary definition does not imply that the use is consistent with ordinary language.

At the same time, there may be particular features of a dictionary definition that seem to exclude certain uses from public meaning.

The “It Doesn’t Fit the Definition” Fallacy: When considering whether a use falls under the public meaning, we should conclude that the use is not part of the meaning if it does not fit the relevant definition.

Sometimes dictionaries include features that are common, but not necessary, criteria of category membership. For example, perhaps “cruel” punishment is often, *but not necessarily*, characterized by the infliction of pain for pain's sake. Or perhaps, a vehicle is typically, *but not*

necessarily, mobile.¹³⁶ It is a mistake to point to a particular aspect of dictionary definition and argue that any use that does not meet that criteria cannot be part of the ordinary meaning.

2. The Burden of Demonstrating a Non-Arbitrary and Reliable Methodology

This Section builds on the specific critiques of corpus linguistics and dictionaries in interpretation, developing a broader challenge to originalist and textualists theories that rely on these tools. The experimental results of Part IV provide evidence of the unreliability of two of the primary tools of Public Meaning Originalism. Recall the empirical-historical argument:

1. Empirical Claim: The modern use of a method (i.e. use of dictionaries or corpus linguistics) does not accurately reflect people's ordinary judgments.
2. Reliability Premise: A method that does not accurately reflect people's judgments is not a reliable method of determining public meaning.
3. Intermediate Conclusion: The empirical result provides evidence that the method is unreliable in modern interpretation.
4. Historical Inference: In the absence of historically distinguishing factors, evidence of a method's unreliability in modern interpretation also serves as evidence about that method's unreliability in historical interpretation.
5. Conclusion: The results provide evidence that the method is unreliable in historical interpretation.

The experimental results have suggested that originalist uses of both dictionaries and corpus linguistics do not accurately reflect people's judgments. As Section IV.E estimated, the error rate for each method is in the range of at least 20-35%, and in some cases over 80%. The studies

¹³⁶ See Part IV *supra* (ordinary concept participants were divided, about 50-50, that a non-mobile WWII memorial truck was a vehicle; however, dictionary participants overwhelmingly disagreed).

provide evidence that these are not reliable methods of determining modern public meaning. Insofar as there is no good reason to think these tools perform better in historical analysis, the results provide evidence that these methods are unreliable in historical interpretation.

Importantly, the debate does not stop here. This conclusion shifts the argumentative burden to Public Meaning Originalism and other theories that rely upon these tools to elaborate and justify non-arbitrary and demonstrably reliable methodologies. The two key features of this challenge are non-arbitrariness and demonstrable reliability. I consider those in turn.

First consider non-arbitrariness. There are many choices one must make in originalist interpretation, many of which threaten arbitrariness. Some of these choices are outside the scope of this paper. For example, for interpretive theories that advocate using dictionaries, *which* dictionaries should be used? Relying on different dictionaries for different cases invites a charge of arbitrariness. The sources of arbitrariness are even broader for corpus linguistics: Exactly how many searches will be conducted, what precisely will be searched and how is the search string determined, what percent of conforming uses “counts” as an instance of public meaning?

Although this paper sets these questions aside, a defense of Public Meaning Originalism must address these fundamental concerns. But there are also new sources of arbitrariness illuminated by the experimental results. Principally, consider the arbitrariness in choosing to use dictionaries *or* corpus linguistics.

The experimental results suggest that corpus linguistics and dictionaries are not just sometimes divergent; they often provide strongly opposing verdicts about public meaning.¹³⁷ Insofar as a theorist or jurist endorses dictionaries in one instance and corpus linguistics in another—with no further supporting reasons—this raises a new question of arbitrariness. This question becomes more pressing where the choice of methodology seems to match the desired political or legal outcome. We should be suspicious of judges or interpreters who are quick to

¹³⁷ See Section IV.E.7 *supra*.

point out the absence of (non-prototypical) uses in a corpus in one case, but the breadth of a dictionary definition in another.¹³⁸

Thus, there is a burden on Public Meaning Originalists (and others who use or defend dictionary and corpus use in interpretation) to elaborate and defend a non-arbitrary use of their tools. Reporting dictionary definitions and detailed corpus data often conveys an impression of legitimacy and scientific rigor. However, these values are illusory if the method of interpretation is subtly (consciously or unconsciously) altered in each case.

This first burden is relatively easier to satisfy. Originalists must simply commit to a public list of interpretive choices. For example, perhaps the first definition of a term in *X* dictionary is our authoritative source in Founding-Era constitutional interpretation.

The second burden, to articulate a demonstrably reliable use of these tools, is more demanding. If the tools of Public Meaning Originalism are unreliable, it does not matter much that they are applied systematically. We can construct many non-arbitrary methods of interpretation (e.g. even principled dice-rolling is not arbitrary in the relevant sense). But any such method unconvincing until the method is also shown to be reliable.

The burden now rests with Public Meaning Originalism and theories that rely on these tools of discovering original public meaning. We should remain open to the very real possibility that such a challenge might be met. But, for a moment, imagine that Public Meaning Originalism does not adequately meet this burden. How should the theory fare?

Recall the error rates for dictionaries and corpus linguistics. Overall, the error for one relying on each method was between 20-35%. In many cases, the error was larger: 50%, 75%, even 100%. These numbers may seem abstract, but consider what they represent: *The data suggest that*

¹³⁸ Compare *Smith v. U.S.*, 508 U.S. 223 241-244 (1993) (Scalia, J., dissenting) (arguing that a broad dictionary definition of “using” should be rejected and instead relying on prototypical examples of “use” in ordinary language to argue that the public meaning of “using” is narrower than the dictionary definition) with *D.C. v. Heller*, 554 U.S. 570 (2008) (Scalia, J.) (arguing that “arms” should be construed extensively on the basis of dictionary definitions).

judges relying on corpus linguistics and dictionary definitions would arrive at the wrong interpretation once in every three to five cases, and perhaps even more frequently.

It is a severe choice for Public Meaning Originalism to simply accept such a serious possibility of error. In doing so, the theory relies on tools that are now *known* to be unfaithful to public meaning at a non-trivial rate. Knowingly predicating important legal outcomes on unreliable methods is a mistake that all theories of legal interpretation should seek to avoid.

CONCLUSION

This paper has developed a novel “experimental jurisprudence” method of testing two of the fundamental tools of Public Meaning Originalism—dictionaries and corpus linguistics.¹³⁹ A series of experiments examined judgments of ordinary people, “elite-university” law students, and United States judges, providing evidence bearing on the process and reliability of dictionary and corpus use in interpretation.

The experimental studies suggest one systematic pattern of the use of dictionaries and corpora: Corpus linguistics tends to reflect prototypical uses, while dictionaries tend to generate more extensive uses. This discovery supports principles for improving the use of both tools in legal interpretation. In an interpretive context that clearly calls for an “extensive meaning” of a term, the currently popular uses of corpus linguistics data are likely unhelpful. Corpus linguistics often reflects prototypical uses but omits non-prototypical ones. In such circumstances, interpreters should look to other sources of evidence (e.g. dictionaries) that tend to provide more extensive uses.

¹³⁹ For other recent examples of experimental jurisprudence, see Ivar R. Hannikainen & Raff Donelson, *Fuller and the Folk: The Inner Morality of Law Revisited*, in 3 OXFORD STUDIES IN EXPERIMENTAL PHILOSOPHY (forthcoming) (on the ordinary concept of law); James Macleod, *Ordinary Causation*, 94 IND. L.J. (2019) (on the ordinary concept of causation); Christian Mott, *Statutes of Limitations and Personal Identity*, in 2 OXFORD STUDIES IN EXPERIMENTAL PHILOSOPHY 243 (2018) (on the ordinary concept of identity); Roseanna Sommers, *Commonsense Consent*, (draft manuscript) (on the ordinary concept of consent); Joshua Knobe & Scott Shapiro, *What Cognitive Science can Teach Us about Proximate Causation* (draft manuscript) (on the ordinary concept of causation); Kevin P. Tobia, *How People Judge What is Reasonable*, 70 ALA. L. REV. 2915 (2018) (on the ordinary concept of reasonableness); and Kevin P. Tobia, *Legal Concepts and Legal Expertise* (draft manuscript) (on the ordinary and expert-legal concepts of intentional action).

Although the experimental results suggest some reforms for using dictionaries and corpus linguistics in special circumstances (e.g. interpretive contexts in which it is clear that only prototypical uses are relevant), they also reveal that these tools are not reliable ones in more common searches for a text's "ordinary meaning" or "public meaning." For one, participants were often divided about the modern meaning of terms, suggesting that "*the* public meaning" may often be indeterminate. Moreover, the experimental studies revealed that both dictionary-use and corpus linguistics-use diverged from ordinary concept use, across a large number of cases. This suggests that two central methods of locating "public meaning" are unreliable in modern interpretation. I argued that, absent distinguishing historical factors, this also provides evidence of the unreliability of dictionaries and corpus linguistics in historical interpretation.

Importantly, similar results arose across levels of legal expertise—including 230 "elite-university" law students (e.g. at Harvard and Yale) and 98 United States judges—and for various terms and phrases, such as "vehicle," "labor," "weapon," "carrying a firearm," and "tangible object." Ultimately, the data suggest that popular methods of dictionary-use and corpus linguistics carry serious risks of error—conservatively estimated, a 20-35% average error rate. And in some circumstances, expert use of these methods carried extremely large error rates—between 80-100%.

The results and arguments here shift the argumentative burden to theories that rely upon these tools in interpretation. Original public meaning theories must provide a principled account of the use of these tools and an explanation and demonstration of how error can be avoided.

APPENDIX A: EXP. 1 ADDITIONAL MATERIALS AND FOLLOW-UP EXPERIMENTS

1. Additional Experiment 1 Analyses

	X^2 (2, $N = 206$)	p (significance)	V (effect size)
Vehicle	7.02	0.0299 *	0.19
Automobile	7.75	0.0208 *	0.19
Car	6.99	0.0303 *	0.19
Bus	37.56	< 0.0001 **	0.43
Truck	10.81	0.0045 *	0.23
Bicycle	53.69	< 0.0001 **	0.51
Airplane	64.61	< 0.0001 **	0.56
Ambulance	21.13	< 0.0001 **	0.32
Golf Cart	.66.70	< 0.0001 **	0.57
Toy Car	1.01	0.6035	0.07

Figure A1. Chi-square test for dictionary vs. concept vs. corpus. * indicates significance at less than .05; ** indicates significance at less than .005 (corrected for multiple comparisons). V indicates an estimate of the effect size.

	<i>Dictionary v. Corpus</i>	<i>Dictionary v. Concept</i>	<i>Corpus v. Concept</i>
	X^2 (1, $N = 134$), p , V	X^2 (1, $N = 140$), p , V	X^2 (1, $N = 136$), p , V
Vehicle	0.01, .9427, .01	6.34, .0118 *, .21	5.93, .0149 *, .21
Automobile	2.74, .0976, .14	1.16, .2809, .09	7.01, .0081 *, .22
Car	1.65, .2024, .11	2.03, .1542, .12	6.96, .0084 *, .22
Bus	17.78, 0.0001 **, .36	2.28, .1310, .13	29.72, < 0.0001 **, .46
Truck	5.55, .0185 *, .20	0.38, .5353, .05	8.65, .0033 *, .25
Bicycle	48.11, < 0.0001 **, .60	2.44, .1181, .13	31.93, < 0.0001 **, .48
Airplane	55.72, < 0.0001 **, .65	4.38, .0363 *, .18	33.98, < 0.0001 **, .49
Ambulance	14.72, 0.0001 **, .33	0.03, .8684, .02	13.86, .0002 **, .31
Golf Cart	32.61, < 0.0001 **, .49	4.07, .0438 *, .17	50.61, < .0001 **, .60
Toy Car	0.14, .6991, .03	0.098, .3229, .03	0.34, .5572, .05

Figure A2. Pairwise chi-squared tests for dictionary vs. concept vs. corpus. Highlighted boxes indicate significance corrected for ten multiple comparisons, $p < .005$.

2. Experiments 1A and 1B

Experiment 1A: Public Meaning with Rules

Experiment 1 uncovered significant differences between the application of the corpus method and the verdicts of dictionaries and ordinary judgments. However, one might wonder how dictionaries and corpus linguistics perform in assessing the meaning of a term in the context of a *rule*. For instance, one might argue, the meaning of “vehicle” is significantly different in the context of the rule “no vehicles in the park.” Importantly, the Public Meaning Originalist’s view is *not* that this rule should be applied differently because of some presumed purpose about keeping certain things out of the park. Instead, the view would be that the meaning of “vehicle” is different in the context of this legal *rule*.

Before turning to a test of this suggestion, it is important to recall the aims of the relevant theories. Original Public Meaning is concerned with determining an empirical fact about how a text was understood. This public meaning is distinct from “teleological meaning,” what the text was meant to convey, the drafter’s intentions, or the “purpose” of the text.

Adding the context of a rule might provide information relevant to some of these other concerns. For this reason, if adding a rule makes a difference, it is important to discern the *process* underlying people’s judgments. Perhaps a rule adds further context that allows the corpus to perform effectively; this would be a friendly finding for Public Meaning Originalists. But perhaps a rule is not really adding semantic context, but instead information about the rule’s *purpose*. If *that* enhanced the performance of corpus linguistics, it is not a friendly finding for Public Meaning Originalists. In that case, corpus linguistics is not tracking public meaning;

instead, the perceived purpose of the rule is driving judgments that seem (to us) to track public meaning.

I return to this issue in Experiment 1B, but first consider the more straightforward results of Experiment 1A, which presents participants with a rule.

Method

Participants. Two-hundred and four participants were recruited from Amazon's Mechanical Turk (51% female, 48.5% male, 0.5% non-binary, mean age = 37.0).

Materials and Procedure. As in Experiment 1, participants were randomly divided into one of three conditions: Dictionary, Corpus, or Concept. In each condition, participants received the same information about a term (ailac or vehicle). However, in this experiment, participants were then instructed about a town ordinance. In the Dictionary and Corpus conditions, this was: Now imagine that a town passes an ordinance that says "no ailacs in the park." In the Concept condition, this was: Now imagine that a town passes an ordinance that says "no vehicles in the park." All participants rated whether ten entities, presented in a random order, were allowed in the park (e.g. "Is a truck allowed in the park" [Yes/No]).

Results.

In this Experiment, the differences among dictionary, corpus and concept methods were strikingly reduced.

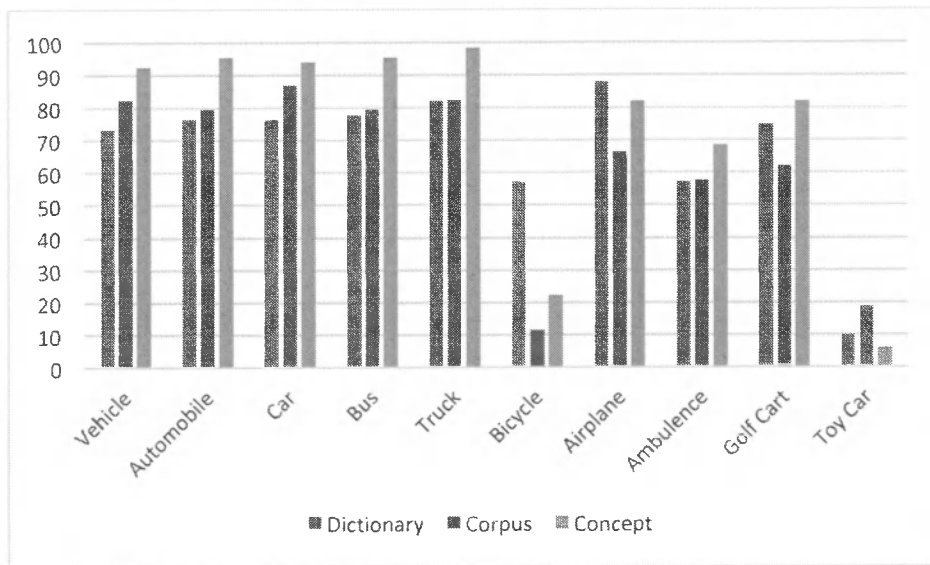


Figure A3. Percentage Responding No (the entity is not allowed in the park) by Dictionary, Corpus, Concept conditions

Chi-squared tests comparing the proportion of “yes” responses revealed smaller effects. Here again, there were significant differences for nearly every entity (except ambulance and toy car).

Pairwise chi-squared tests showed very few significant difference between dictionary and corpus participants. There were some differences between both of those methods and the Concept condition.

	χ^2 (2, $N = 204$)	p (significance)	V (effect size)
Vehicle	8.58	0.0137 *	0.21
Automobile	10.33	0.0057 *	0.22
Car	8.65	0.0132 *	0.21
Bus	9.56	0.0084 *	0.22
Truck	10.75	0.0046 **	0.23
Bicycle	36.92	< 0.0001 **	0.43
Airplane	10.18	0.0062 *	0.22
Ambulance	2.30	0.3166	0.11
Golf Cart	6.95	0.0310 *	0.19
Toy Car	5.64	0.0596	0.17

Figure A4. Chi-squared tests for dictionary vs. concept vs. corpus

*Dictionary v.
Corpus*

*Dictionary v.
Concept*

Corpus v. Concept

	X^2 (1, $N = 137$), p , V	X^2 (1, $N = 135$), p , V	X^2 (1, $N = 136$), p , V
Vehicle	1.65, .1989, .12	8.63, .0033 **, .25	3.06, .0801, .15
Automobile	0.21, .6466, .11	10.13, .0015 **, .27	7.77, .0053 *, .24
Car	2.52, .1121, .14	8.24, .0041 *, .25	1.97, .1607, .12
Bus	.064, .8000, .02	9.02, .0027 **, .26	5.87, .0154 *, .21
Truck	0.02, .9686, .01	10.12, .0017 **, .27	9.34, .0016 **, .26
Bicycle	31.82, < 0.0001 **, .48	17.18, < 0.0001 **, .36	2.82, .0932, .14
Airplane	9.10, .0026 **, .26	1.01, .3149, .09	4.23, .0397 *, .18
Ambulance	0.01, .9416, .01	1.85, .1739, .12	1.67, .1963, .11
Golf Cart	2.56, .1098, .26	1.01, .3160, .09	6.60, .0102 *, .22
Toy Car	2.01, .1566, .12	0.84, .3585, .08	5.15, .0233 *, .20

Figure A5. Pairwise chi-squared tests for dictionary vs. concept vs. corpus.

Discussion

Although there are some differences among corpus, dictionary, and concept methods, those differences are relatively few and relatively small. This suggests the possibility of a redemptive result for dictionary and corpus use. If there is contextual context about the relevant rule, these methods deliver more reliable estimates about original public meaning.

Unfortunately, this redemptive story is not supported by the data. The next experiment suggests that the apparent success of these methods in Experiment 1A is illusory.

Experiment 1B: Public Meaning with Arbitrary Rules

Experiment 1B suggests that the corpus and dictionary methods diverge less when there is a rule. One hypothesis is that this occurs simply whenever there is a rule. Another hypothesis is that this occurs because the rule communicates something about the purpose of the provision. Experiment 3 tests these hypotheses by using an arbitrary rule (i.e. one without a discernable purpose). If the results here look like those of Experiment 2, this suggests that perceived purpose is not the real cause of Experiment 2's results. If the results here look like Experiment 1, this suggests that it is likely the perceived purpose of the rule in Experiment 2 (i.e. the purpose of a rule prohibiting things from a park) that drives corpus and dictionary judgments.

Method

Participants. Two-hundred and one participants were recruited from Amazon's Mechanical Turk (51% female, 48.5% male, 0.5% non-binary, mean age = 35.5).

Materials and Procedure. As in Experiments 1 and 1A, participants were randomly assigned to one of three conditions: Dictionary, Corpus, or Concept. In each condition, participants received the same information about a term (ailac or vehicle). As in Experiment 1A, participants were then instructed about a town ordinance. In the Dictionary and Corpus conditions, this was: Now imagine that a town passes an ordinance that says "all ailacs can display a blue sticker, but everything that is not an ailac cannot display a blue sticker." In the Concept condition, this was: Now imagine that a town passes an ordinance that says "all ailacs can display a blue sticker, but everything that is not an ailac cannot display a blue sticker" All participants rated whether ten items were allowed to display a blue sticker (e.g. Can a bicycle display a blue sticker [Yes/No]).

Results.

The results in this arbitrary rule case are strikingly similar to those of Experiment 1.

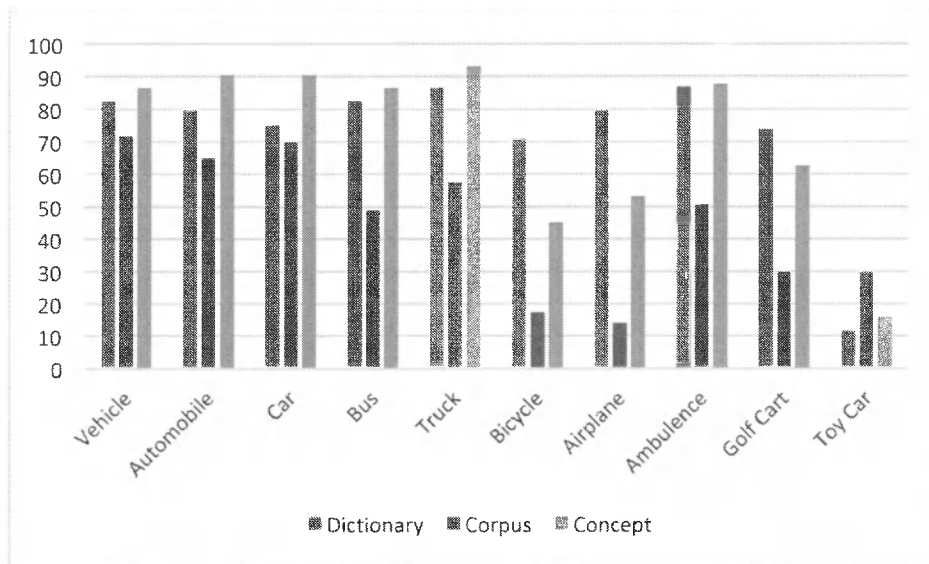


Figure A6. Percentage Responding Yes (the entity can display a blue sticker) by Dictionary, Corpus, Concept conditions

Chi-squared tests indicate significant differences for nine entities, many of which are medium to large effects. Pairwise chi-squared tests indicate significant differences among all three conditions, especially between Dictionary and Corpus and Corpus and Concept.

Discussion

As in Experiment 1, corpus linguistics performed poorly in tracking the ordinary meaning of vehicle. For busses, trucks, bicycles, airplanes, ambulances, and golf carts, the corpus linguistics result was significantly—and often dramatically—different from the verdicts delivered by the dictionary and ordinary concept use. The results from Experiments 1, 1A, and 1B together suggest that whatever success corpus linguistics had in Experiment 1A is not attributable to the fact that there was a rule, but rather to something like a presumed purpose of the rule. Insofar as corpus linguistics is used for textualist purposes, aimed to uncover the *meaning* of the text and not the purpose of its drafting, Experiments 1 and 1B represent the more appropriate test of these methods.

	χ^2 (2, $N = 201$)	p (significance)	V (effect size)
Vehicle	4.74	0.0935	0.15
Automobile	13.23	0.0013 **	0.26
Car	9.54	0.0085 *	0.22
Bus	27.64	< 0.0001 **	0.37
Truck	28.88	< 0.0001 **	0.38
Bicycle	35.94	< 0.0001 **	0.42
Airplane	54.10	< 0.0001 **	0.52
Ambulance	31.06	< 0.0001 **	0.39
Golf Cart	26.24	< 0.0001 **	0.36
Toy Car	7.37	0.0251 *	0.19

Figure A7. Chi-squared test for dictionary vs. concept vs. corpus.

	<i>Dictionary v. Corpus</i>	<i>Dictionary v. Concept</i>	<i>Corpus v. Concept</i>
	χ^2 (1, $N = 126$), p , V	χ^2 (1, $N = 144$), p , V	χ^2 (1, $N = 132$), p , V
Vehicle	2.06, .1513, .20	4.45, .0350 *, .18	0.45, .4989, .06
Automobile	3.47, .0625, .17	3.46, .0627, .16	13.20, .0003 **, .32
Car	0.43, .5139, .06	6.06, .0138 *, .21	9.14, .0025 **, .26
Bus	15.95, .0001 **, .36	0.45, .4989, .06	21.93, < 0.0001 **, .41
Truck	13.64, .0002 **, .33	1.67, .1969, .11	23.73, < 0.0001 **, .42
Bicycle	35.84, < 0.0001 **, .53	9.71, .0018 **, .26	11.25, .0008 **, .29
Airplane	53.85, < 0.0001 **, .65	11.14, .0009, .28	21.61, < 0.0001 **, .41
Ambulance	19.59, 0.0001 **, .39	0.04, .8500, .02	22.12, < 0.0001 **, .41
Golf Cart	24.42, < 0.0001 **, .44	2.09, .1482, .12	13.99, .0002 **, .33
Toy Car	6.52, .0106 *, .23	0.58, .4450, .06	3.61, .0574, .16

Figure A8. Chi-squared tests for dictionary vs. concept vs. corpus

APPENDIX B: EXP. 2 ADDITIONAL MATERIALS

1. Instructions for Experiment 2

Instructions: In the following screen we will ask you some questions about different categories. We will ask whether some things are “prototypical” members of the category. We will also ask whether some things are “technically” part of the category.

For example, consider the term “bird.” A robin is a prototypical bird. A sparrow is another prototypical bird. Other entities are not prototypical birds. An ostrich is not a prototypical bird; neither is a penguin. Nevertheless, robins, sparrows, ostriches, and penguins are all technically birds. Other entities, like whales or chipmunks are not technically birds.

As another example, consider the term “dessert.” An ice cream is a prototypical dessert. A chocolate cake is another prototypical dessert. Other entities are not prototypical desserts. After-dinner cheeses are not a prototypical dessert; neither are candy gummy bears. Nevertheless, ice cream, chocolate cake, after-dinner cheeses, and candy gummy bears are all technically desserts. Other entities, like pizza or salad are not technically desserts.

Check questions:

A robin is a prototypical bird; A penguin is a prototypical bird; A robin is technically a bird;
A penguin is technically a bird; A chipmunk is a prototypical bird; A chipmunk is typically a
bird

The correct answers to these questions were Yes, No, Yes, Yes, No, No.

2. Experiment 2A

Experiment 2A: The Process of Using Dictionaries and Corpora, An Extension

Experiment 2A sought to replicate the result of Experiment 2 using a different example. To minimize researcher degrees of freedom, I selected the second example from Lee & Mouritsen's recent defense of corpus linguistics.¹⁴⁰ Their lead example is "vehicle" and their second example is "carrying" a firearm.

Method

Participants. Two-hundred and six participants participated in an online experiment (47.3% male, 52.7% female, 0.0% non-binary, $M_{\text{age}} = 37.6$).

Materials and Procedure. Participants were randomly assigned to the Concept, Corpus, Dictionary, or Prototypically-Technically Condition. Participants in the Prototypically-Technically condition were presented with the same instructions as in Experiment 4; they then rated whether various actions were prototypically carrying a firearm and technically carrying a firearm. Participants in the Concept condition were told to:

Consider the phrase, "carrying a firearm."

Participants in the Dictionary condition were told to:

Consider this dictionary definition of the phrase "ailacing a firearm":

Ailacing a firearm: Transporting or taking a firearm from one place to another

Participants in the Corpus condition were first told to:

Consider the phrase, "ailacing a firearm." To help understand this phrase, consider some information about the use of "ailacing a firearm."

Corpus participants then received corpus data, beginning with:

¹⁴⁰ *Supra* note 1.

First, consider the top common words used in connection with “ailacing a firearm.” These words might appear before or after ailacing a firearm, or sometimes close to ailacing a firearm, e.g. “while ailacing a firearm;” “guilty of ailacing a firearm;” “ailacing a firearm illegally;” etc.¹⁴¹

Corpus, Dictionary, and Concept participants then answered a series of questions, presented randomly:

Is bringing a gun to a bank robbery carrying/ailacing a firearm?

Is taking a gun to a gang fight carrying/ailacing a firearm?

Is delivering an order of guns to their purchaser by hand carrying/ailacing a firearm?

Is driving to a drug deal with a gun in the rear of the car carrying/ailacing a firearm?

Is shopping in a supermarket with a concealed gun carrying/ailacing a firearm?

Is delivering a bag of crime evidence, including a gun, to a police station carrying/ailacing a firearm?

Is moving a gun into a secure storage locker carrying/ailacing a firearm?

Is making a threat during a drug deal by pointing at a gun on a table carrying/ailacing a firearm?

¹⁴¹ The remainder of the data included:

Top common words: without, license, during, while, using, crime, relation, openly, lawfully, property, under, intoxicated, illegally, possession, permit, drug, weapon, facts, permits, legally, robbery, grounds, guilty, campus, charged, charges, prohibited, trafficking, mall, counts, allegedly, influence, protection, assault, officer, municipally, drug-trafficking, endangerment, self-defense, prohibit, concealed, ammunition, second-degree, punishment, homicide, engaging, citizen, traveling, manner, convicted, violent

Next, consider some further examples of “ailacing a firearm” in context:

- 1) ... Castile had told the officer that he was lawfully **ailacing a firearm** after he was pulled over by Yanez and another officer
- 2) ... the survivor and the witness said Mobley had been **ailacing a firearm** all night. Authorities launched a statewide manhunt for ...
- 3) ... of second-degree burglary, third-degree burglary, **ailacing a firearm**, and drug possession. Houseal was sentenced to 18 months ...
- 4) ... with people engaged in risky behaviors--like **ailacing a firearm** and engaging in criminal activities--increases the
- 5) ... the weapons complaints, and he allegedly admitted to **ailacing a firearm** and engaging in target practice ...
- 6) ... He allegedly was **ailacing a firearm** and heroin and wearing a bulletproof vest at the time ...
- 7) ... by a felon, interference with official acts while **ailacing a firearm** and possession of marijuana ...
- 8) ... by robbery, assault on a federal officer and **ailacing a firearm** during a crime of violence. They also were sentenced ...
- 9) ... Law was decreasing the severity of the punishment for **ailacing a firearm** while intoxicated. Previously, this crime was considered ...”

Is removing a gun from its case and cleaning it carrying/ailacing a firearm?

Is calling someone to arrange selling a gun carrying/ailacing a firearm?

Results

As in Experiment 4, I correlated the proportion of Yes responses per item in each of the Corpus, Dictionary, and Concept conditions with the mean response per item for the prototypically and technically measures. The full correlation matrix is displayed below.

	Corpus	Dictionary	Concept	Prototypically	Technically
Corpus	1				
Dictionary	0.33	1			
Concept	0.73	0.42	1		
Prototypically	0.93	0.29	0.97	1	
Technically	0.68	0.72	0.81	0.74	1

Figure A9. Correlation Matrix. Boxes indicate the relevant comparisons (Corpus is more correlated with Prototypically, and Dictionary is more correlated with Technically).

To test the statistical significance of this relationship between Corpus-Prototypically and Dictionary-Technically, I conducted two tests for differences between correlations between (i) Corpus and Prototypically and Corpus and Technically, and (ii) Dictionary and Prototypically and Dictionary and Technically. Corpus was significantly more correlated with Prototypically than with Technically, $z = 2.067$, $p = .0194$ (one-tailed).¹⁴² Dictionary was significantly more correlated with Technically than with Prototypically, $z = 2.049$, $p = .0202$ (one-tailed).¹⁴³

We can also consider the basic result of whether corpus, concept, and dictionary treatments diverged.

¹⁴² Two-tailed = .0387.

¹⁴³ Two-tailed = .0404.

	χ^2 (2, $N = 206$)	p (significance)	V (effect size)
Robbery	6.69	0.0353	0.18
Fight	9.26	0.0098 **	0.21
Delivery	22.33	< 0.0001 **	0.34
Drug Deal	2.00	0.3679	0.10
Supermarket	39.55	< 0.0001 **	0.44
Evidence	41.41	< 0.0001 **	0.45
Storage	32.01	< 0.0001 **	0.39
Table Threat	54.11	< 0.0001 **	0.51
Cleaning	8.18	0.0167	0.20
Selling	5.01	0.0817	0.16

Figure A10. Chi-squared test for dictionary vs. concept vs. corpus.

	<i>Dictionary v. Corpus</i>	<i>Dictionary v. Concept</i>	<i>Corpus v. Concept</i>
	χ^2 (1, $N = 136$), p , V	χ^2 (1, $N = 139$), p , V	χ^2 (1, $N = 137$), p , V
Robbery	1.44, 0.2301, .13	4.91, 0.0267, .21	0.57, 0.4503, .09
Fight	2.24, 0.1345, .15	7.25, 0.0071**, .25	0.97, 0.3247, .11
Delivery	21.66, <.0001**, .42	9.12, 0.0025**, .27	2.59, 0.1075, .1521
Drug Deal	0.49, 0.4839, .08	0.07, 0.7913, .04	1.48, 0.2238, .12
Supermarket	5.1, 0.0239*, .21	37.84, <.0001**, .54	15.95, <.0001**, .36
Evidence	36.40, <.0001**, .53	21.78, <.0001**, .41	1.97, 0.1604, .14
Storage	30.09, <.0001**, .49	7.57, <.0059**, .25	7.65, 0.0057**, .25
Table Threat	36.01, <.0001**, .53	38.33, <.0001**, .54	0, 1, .01
Cleaning	0.09, 0.7642, .05	5.87, .0154*, .22	3.62, .0571, .18
Selling	2.74, 0.0979, .16	3.67, .0554, .18	0, 1, .02

Figure A11. Chi-squared tests for dictionary vs. concept vs. corpus

Discussion

This study provides further support for the hypothesized mechanism. Dictionaries tend to generate a more extensive sense of a term and corpus linguistics tends to generate a prototypical sense of the term.

APPENDIX C: EXP. 3 ADDITIONAL MATERIALS AND FOLLOW-UP EXPERIMENTS

1. Experiment 3 Additional Analyses

	X^2 (2, $N=$ 98)	p (significance)	V (effect size)
Vehicle	3.78	0.1511	0.20
Automobile	6.57	0.0374	0.26
Car	4.41	0.1103	0.21
Bus	20.57	< 0.0001 *	.46
Truck	29.41	< 0.0001 *	0.56
Bicycle	33.31	< 0.0001 *	0.59
Airplane	36.68	< 0.0001 *	0.62
Ambulance	24.64	< 0.0001 *	0.51
Golf Cart	29.62	< 0.0001 *	0.56
Toy Car	0.13	0.9371	0.04
Drone	4.17	0.1243	0.22
Skateboard	22.50	< 0.0001 *	0.50
Rollerskate	22.94	< 0.0001 *	0.51
WWII Truck	44.25	< 0.0001 *	0.70
Baby Stroller	24.57	< 0.0001 *	0.53
Wheelchair	17.37	0.0002 *	0.44
Horse Carriage	48.98	< 0.0001 *	0.74
Canoe	43.39	< 0.0001 *	0.72
Helicopter	23.76	< 0.0001 *	0.51
Moped	10.80	0.0045	0.35
Crutches	8.50	0.0143	0.31
Pogo Stick	11.26	0.0036 *	0.36
Baby Shoulder-Carrier	26.13	< 0.0001 *	0.54
Liferaft	24.15	< 0.0001 *	0.52
Zip-line	22.20	< 0.0001 *	0.50

Figure A12. Chi-square test for dictionary vs. concept vs. corpus

	<i>Dictionary v. Corpus</i> χ^2 (1, $N = 59$), p , V	<i>Dictionary v. Concept</i> χ^2 (1, $N = 72$), p , V	<i>Corpus v. Concept</i> χ^2 (1, $N = 65$), p , V
Vehicle	1.82, 0.177, 0.23	0.33, 0.566, 0.16	0.2, 0.655, 0.12
Automobile	3.45, 0.063, 0.31	0.36, 0.549, 0.16	1.03, 0.310, 0.18
Car	1.32, 0.251, 0.20	0.10, 0.752, 0.02	2.03, 0.154, 0.23
Bus	6.76, 0.009, 0.39	0.01, 0.920, 0.13	11.49, 0.0007, 0.47
Truck	11.65, 0.0006, 0.51	0, 1, 0	14.14, 0.0002, 0.52
Bicycle	30.25, < 0.0001, 0.76	8.7, 0.0032, 0.39	9.09, 0.0026, 0.41
Airplane	31.05, < 0.0001, 0.76	4.75, 0.029, 0.30	14.57, 0.0001, 0.51
Ambulance	8.83, 0.003, 0.44	0.01, 0.920, 0.13	13.78, 0.0002, .52
Golf Cart	19.32, < 0.0001, .62	1.80, 0.180, 0.22	12.27, 0.0005, 0.47
Toy Car	0.11, 0.740, 0.0	0.0, 1, 0.04	0.0, 1, 0.04
Drone	1.22, 0.269, 0.19	2.79, 0.095, 0.24	0.0, 1, 0.05
Skateboard	18.64, < 0.0001, 0.625	10.22, 0.0014, 0.43	1.95, 0.163, 0.22
Rollerskate	22.49, < 0.0001, 0.69	13.47, 0.0002, 0.48	2.91, 0.088, 0.27
WWII Truck	2.4, 0.121, 0.26	11.70, 0.0006, 0.45	2.14, 0.144, 0.22
Baby Stroller	21.49, < 0.0001, 0.68	6.84, 0.0089, 0.35	5.92, 0.015, 0.36
Wheelchair	12.69, 0.0004, 0.52	0.27, 0.603, 0.10	9.12, 0.003, 0.42
Horse Carriage	23.85, < 0.0001, 0.70	0.74, 0.390, 0.15	18.06, < 0.0001, 0.58
Canoe	38.09, < 0.0001, 0.89	16.11, < 0.0001, 0.53	8.35, 0.004, 0.41
Helicopter	21.78, < 0.0001, 0.68	10.93, 0.0009, 0.44	2.83, 0.092, 0.25
Moped	1.86, 0.17, 0.23	1.41, 0.234, 0.21	8.23, 0.004, 0.42
Crutches	3.24, 0.072, 0.29	4.66, 0.031, 0.30	0.23, 0.63, 0.0
Pogo Stick	8.39, 0.004, 0.44	3.41, 0.065, 0.26	1.50, 0.22, 0.20
Baby Shoulder-Carrier	15.64, < 0.0001, 0.58	14.76, 0.0001, 0.50	0.25, 0.617, 0.11
Liferaft	21.08, < 0.0001, 0.66	6.48, 0.011, 0.34	5.71, 0.0169, 0.35
Zip-line	16.85, < 0.0001, 0.60	8.34, 0.0039, 0.39	2.91, 0.088, 0.27

Figure A13. Pairwise chi-squared tests for dictionary vs. concept vs. corpus.

2. Experiments 3A and 3B

Experiment 3A: Testing Elite Law Students

The preceding experiments have studied ordinary, non-expert populations. Judgments of ordinary people provide good evidence about the current public meaning of these terms (e.g. “vehicle” or “carrying a firearm”). But some might doubt whether this population contains the best users of dictionaries and corpus linguistics in legal interpretation. To appropriately test the reliability of corpus linguistics and dictionaries, one might argue, we should test legal experts who have the relevant background in interpretation.

This objection is plausible, but it should not be taken to dismiss any significance of the results. After all, even if the previous results do not provide a strong inference into judges’ cognition, they do provide good evidence about juror’s cognition. And jurors, too, are statutory interpreters.¹⁴⁴

Nevertheless, this section begins to address the “expertise” objection head-on. I tested “elite-university” law students—law student from the “T-14” law schools—who should have significant legal education or the innate abilities posited by defenders of this expertise objection.

To more comprehensively test the reliability of dictionaries and corpus, this experiment featured an expanded range of twenty-five entities. In the first three experiments, most entities were “vehicles” in ordinary language and the dictionary categorized these as vehicles. Experiments 2 and 2A suggest the dictionary generates an extensive condition of category membership. So to better test dictionaries, this experiment also includes some entities that I predicted are likely *not* vehicles in ordinary language, but which may nevertheless fall under a very extensive sense of a vehicle. These are entities including crutches, a baby should-carrier, and a zip-line.

¹⁴⁴ Lawrence Solan, *Jurors as Statutory Interpreters*, 78 CHI-KENT L. REV. 1281 (2003).

Method

Participants. Two-hundred and thirty participants were recruited from the “T-14” law schools. Solicitation emails were sent to administrators at Berkeley, Columbia, Cornell, Duke, Georgetown, Harvard, New York University, Northwestern, Stanford, the University of Chicago, University of Michigan, University of Pennsylvania, University of Virginia, and Yale. At Columbia, Harvard, and Yale, emails were forwarded directly from a current law student.

Participants were 51.1% female, 47.6% male, 1.3% non-binary, $M_{\text{age}} = 26.5$). Participants were largely from Yale Law School (68.0%), Harvard Law School (12.6%), and Columbia Law School (18.2%). Participants were recruited in May, at the completion of the first year of law school for 27.4%, the second year for 24.8%, and the third year for 43.0%.

	Columbia	Harvard	Yale	Other ¹⁴⁵	Total
J.D. 1L	14		49		63
J.D. 2L	12	2	42	1	57
J.D. 3L	14	26	57	2	99
Post-3L	1	1	1		3
Other¹⁴⁶			8		8
Total	41	29	157	3	230

Figure A14. Participant law school affiliation and academic year.

Materials and Procedure. As in the previous experiments, participants were randomly assigned to either the Concept, Corpus, or Dictionary condition. In this experiment, participants evaluated the first set of entities (presented in a randomized order): a vehicle, automobile, car, bus, truck, bicycle, airplane, ambulance, golf car, toy car. Participants immediately considered another set (presented in a randomized order): drone, skateboard, pair of rollerskates, “a non-functioning commemorative truck (e.g. a World War II Truck that has been decorated as a WWII

¹⁴⁵ There was one response from each of Berkeley, NYU, and Georgetown.

¹⁴⁶ These respondents were enrolled in law school for a non-JD program (e.g. LLM, JSD, PhD in law).

monument)”, baby stroller, electric wheelchair, horse-drawn carriage, wooden canoe, helicopter, moped, pair of crutches, pogo stick, baby shoulder-carrier, liferaft, and zip-line.

Results.

First, I conducted chi-square tests for differences among the three conditions. For twenty items, there was a significant difference. To further analyze these differences, I conducted follow-up pairwise chi-square tests. Again, there were a number of significant differences.

	X^2 (2, $N =$ 261)	p (significance)	V (effect size)
Vehicle	6.27	0.0435	0.155
Automobile	2.99	0.2242	0.11
Car	3.38	0.1845	0.12
Bus	42.42	< 0.0001 *	0.41
Truck	15.06	0.0005 *	0.24
Bicycle	79.11	< 0.0001 *	0.55
Airplane	65.76	< 0.0001 *	0.50
Ambulance	22.24	< 0.0001 *	0.30
Golf Cart	44.35	< 0.0001 *	0.42
Toy Car	0.62	0.7334	0.05
Drone	18.6	< 0.0001 *	0.27
Skateboard	47.38	< 0.0001 *	0.44
Rollerskate	37.01	< 0.0001 *	0.39
WWII Truck	10.91	0.0043	0.21
Baby Stroller	87.54	< 0.0001 *	0.59
Wheelchair	20.02	< 0.0001 *	0.29
Horse Carriage	115.39	< 0.0001 *	0.68
Canoe	90.55	< 0.0001 *	0.60
Helicopter	69.90	< 0.0001 *	0.53
Moped	36.67	< 0.0001 *	0.38
Crutches	24.91	< 0.0001 *	0.32
Pogo Stick	34.23	< 0.0001 *	0.37
Baby Shoulder-Carrier	118.28	< 0.0001 *	0.68
Liferaft	58.58	< 0.0001 *	0.48
Zip-line	49.44	< 0.0001 *	0.45

Figure A15. Chi-square test for dictionary vs. concept vs. corpus

	<i>Dictionary v. Corpus</i> χ^2 (1, $N = 170$), p , V	<i>Dictionary v. Concept</i> χ^2 (1, $N = 180$), p , V	<i>Corpus v. Concept</i> χ^2 (1, $N = 172$), p , V
Vehicle	1.3, 0.254, 0.11	0.34, 0.560, 0.06	3.78, 0.052, 0.17
Automobile	0.03, 0.86, 0.0	1.81, 0.179, 0.12	1.63, 0.202, 0.12
Car	0.05, 0.823, 0.04	2.49, 0.115, 0.14	1.14, 0.286, 0.10
Bus	14.12, < 0.0002, 0.31	5.61, 0.018, 0.20	34.37, < 0.0001, 0.47
Truck	1.76, 0.185, 0.12	5.45, 0.020, 0.20	13.46, 0.0002, 0.30
Bicycle	64.95, < 0.0001, 0.63	0.99, 0.320, 0.09	51.36, < 0.0001, 0.56
Airplane	21.05, < 0.0001, 0.37	7.44, 0.006, 0.22	30.04, < 0.0001, 0.43
Ambulance	5.11, 0.024, 0.19	5.48, 0.019, 0.20	21.05, < 0.0001, 0.38
Golf Cart	19.64, < 0.0001, 0.36	1.81, 0.179, 0.12	33.49, < 0.0001, 0.46
Toy Car	0.00, 1, 0.01	0.31, 0.578, 0.06	0.11, 0.740, 0.04
Drone	11.80, 0.0006, 0.29	11.73, 0.0006, 0.27	0.00, 1, 0.01
Skateboard	45.53, < 0.0001, 0.55	8.16, 0.004, 0.23	18.20, < 0.0001, 0.34
Rollerskate	29.22, < 0.0001, 0.44	7.80, 0.005, 0.23	8.72, 0.003, 0.25
WWII	1.47, 0.225, 0.11	9.59, 0.002, 0.25	2.86, 0.091, 0.14
Truck			
Baby	70.86, < 0.0001, 0.67	45.82, < 0.0001, 0.53	4.97, 0.026, 0.19
Stroller			
Wheelchair	36.79, < 0.0001, 0.49	6.02, 0.014, 0.20	14.37, 0.0002, 0.30
Horse	85.08, < 0.0001, 0.74	2.59, 0.108, 0.14	67.41, < 0.0001, 0.64
Carriage			
Canoe	83.34, < 0.0001, 0.74	30.88, < 0.0001, .43	20.29, < 0.0001, 0.36
Helicopter	60.85, < 0.0001, 0.63	6.83, 0.009, 0.21	13.09, 0.0003, 0.28
Moped	14.31, 0.0002, 0.31	2.55, 0.110, 0.14	29.71, < 0.0001, 0.43
Crutches	13.67, 0.0002, 0.31	11.21, 0.0008, 0.27	0.14, 0.71, 0.07
Pogo Stick	20.24, < 0.0001, 0.37	16.5, < 0.0001, 0.32	0.36, 0.549, 0.08
Baby	61.91, < 0.0001, 0.63	69.68, < 0.0001, 0.65	0.03, 0.862, 0.02
Shoulder-Carrier			
Liferaft	52.95, < 0.0001, 0.59	18.03, < 0.0001, 0.34	13.05, 0.0003, 0.29
Zip-line	36.52, < 0.0001, 0.49	20.62, < 0.0001, 0.36	4.28, 0.039, 0.18

Figure A16. Pairwise chi-squared tests for dictionary vs. concept vs. corpus.

Conventional statistical significance is $p < .05$. The figure above has shaded regions where $p < 0.002$ (corrected for twenty-five comparisons).

Discussion

As the figure indicates, there were striking difference between corpus and dictionary for many items, and for several items either corpus or dictionary was significantly different from the concept condition. The results are consistent with the earlier experiments that studied ordinary

people. For *many* items, use of corpus linguistics did not track people's ordinary categorization judgments.

But these results, which use an expanded range of entities, also suggest the unreliability of dictionaries. In the first experiment, dictionary users categorized most entities as vehicles—and most entities were judged to be vehicles. In this experiment, some ordinary language non-vehicles (e.g. roller-skates, pogo stick, zip-line, baby shoulder-carrier) were categorized as vehicles by dictionary users.

Experiment 3B: The Process of Expert Use of Dictionaries and Corpora

The final experiment uses the data from Experiments 3 and 3A to test whether the Dictionary-Extensive/Corpus-Prototype relationship characterizes the responses of judges and law students.

Method

Participants. One-hundred participants recruited from Amazon's Mechanical Turk participated in an online experiment (52.0% male, 46.0% female, 0.0% non-binary, $M_{age} = 36.4$).

Materials and Procedure. Participants completed the same task as in Experiment 2, except they considered the expanded set of twenty-five entities (including, e.g. liferaft, zip-line, and canoe).

Results

The results from the online experiment were correlated with the percentage ratings for law students and judges, respectively, in Experiments 3 and 3A.

First consider the law student results.

Corpus	Dictionary	Concept	Prototypically	Technically
--------	------------	---------	----------------	-------------

Corpus	1				
Dictionary	0.58	1			
Concept	0.80	0.84	1		
Prototypically	0.94	0.60	0.77	1	
Technically	0.82	0.80	0.96	0.81	1

Figure A17. Law Student Correlation Matrix. Boxes indicate the relevant comparisons (Corpus is more correlated with Prototypically, and Dictionary is more correlated with Technically).

To test the statistical significance of this relationship between Corpus-Prototypically and Dictionary-Technically, I conducted two tests for differences between correlations between (i) Corpus and Prototypically and Corpus and Technically, and (ii) Dictionary and Prototypically and Dictionary and Technically. Corpus was significantly more correlated with Prototypically than with Technically, $z = 2.576$, $p = .0050$ (one-tailed).¹⁴⁷ Dictionary was significantly more correlated with Technically than with Prototypically, $z = 2.390$, $p = .0084$ (one-tailed).¹⁴⁸

Next consider the judges' results.

	Corpus	Dictionary	Concept	Prototypically	Technically
Corpus	1				
Dictionary	0.51	1			
Concept	0.85	0.74	1		
Prototypically	0.91	0.53	0.80	1	
Technically	0.81	0.68	0.95	0.81	1

Figure A18. Judge Correlation Matrix. Boxes indicate the relevant comparisons (Corpus is more correlated with Prototypically, and Dictionary is more correlated with Technically).

To test the statistical significance of this relationship between Corpus-Prototypically and Dictionary-Technically, I conducted two tests for differences between correlations between (i)

¹⁴⁷ Two-tailed = .0100.

¹⁴⁸ Two-tailed = .0168.

Corpus and Prototypically and Corpus and Technically, and (ii) Dictionary and Prototypically and Dictionary and Theory. Corpus was significantly more correlated with Prototypically than with Technically, $z = 1.832$, $p = .0334$ (one-tailed).¹⁴⁹ Dictionary was numerically more correlated with Technically than with Prototypically, but not at a level of traditional statistical significance, $z = 1.51$, $p = .0657$ (one-tailed).¹⁵⁰

Discussion

The results are consistent with those of Experiment 2. The patterns of judgment characterizing the divergent verdicts of dictionaries and corpus linguistics are not entirely random. Rather, dictionary definitions tend to elicit a more extensive meaning while corpus linguistics data tends to elicit a prototypical sense.

It is notable that this pattern arises among three very different populations. The online population (MTurk) may have practice in survey-taking, but they have little practice in using dictionaries and corpus linguistics in interpretation. Nevertheless their applications of those tools were strikingly similar to the applications of law students and judges.

¹⁴⁹ Two-tailed = .0669.

¹⁵⁰ Two-tailed = .1313.

APPENDIX D: EXP. 4 ADDITIONAL MATERIALS

1. Experimental Materials

All participants read the following introduction to the experiment:

In the following screen you will see some information about a term. The term might be a real term that you know (e.g. a “painter”) or one that is made up (e.g. an “ailac”). If the term is one that is made up, the “information” about the term will also be fictional.

After you see the information, we will ask twenty-five short questions about the term. There are no right or wrong answers to these questions; we are simply interested in what *you* think about the questions.

Participants then evaluated twenty-five items (using the fake term “krob,” if in the corpus or dictionary conditions).

The remainder of this appendix section contains the materials used for the corpus linguistics, full dictionary, and bare dictionary conditions. There were ten categories, each with twenty-five items: vehicle, carry, interpreter, labor, tangible object, weapon, furniture, food, animal, clothing. The first three were chosen as the examples endorsed by Lee & Mouritsen 2018. The fourth and final example from that paper was “harbor.” However, the vast majority of COCA uses of harbor referred to Pearl Harbor or “harboring feelings.” Both of these are distinct from the relevant sense of harboring an alien. As such, I chose to exclude “harbor,” since including it may have been unfair to proponents of originalist corpus linguistics. The next three were inspired by additional “famous” interpretation examples: labor, tangible object, and arms. Because “arms” is not commonly used today, I used “weapon” as a suitable modern substitute. The COCA uses of the phrase “tangible object” overwhelmingly came from discussion of *Yates*, the case in which the

ordinary meaning of “tangible object” was at issue. As such, I conducted a corpus search on “tangible,” presented participants with that data, and asked whether entities were tangible (e.g. “is a fish a krob object?”). The final four are other superordinate categories (like vehicle).

A. Corpus linguistics materials

Corpus linguistics data was obtained from the Corpus of Contemporary American English. For each word, the top 50 collocates were used. If the category term appeared within the top 50 words, it was omitted and replaced. For example, if “animal” appeared in the top 50 collocates of “animal,” the word was omitted from the materials and the 51st collocate was included at the end of the list.

Keywords in context searches were run using a frequency of at least 3. Ten random sentences were chosen from the keywords in context results.

[Vehicle] Consider the noun, “krob.” To help understand this term, consider some information about the use of “krob.”

First, consider the top common words used in connection with “krob.” These words might appear before or after krob, or sometimes close to krob, e.g. “electric krob;” “drove the krob;” etc.

Top common words: motor, st, stolen, per, utility, electric, driver, block, oct, ave, krob, armored, sept, sport, driving, fuel, launch, rd, sport-utility, speed, emissions, traffic, passenger, unmanned, parked, theft, all-terrain, recreational, struck, driven, pl, bradley, ford, crashes, drove, accidents, engine, registration, hybrid, off-road, fee, license, four-wheel-drive, description, lane, crash, rear, maintenance, travelled

Next, consider some further examples of “krob” in context:

marijuana. Her 1-year-old daughter was in the car with her. After searching the krob , police found 50 stamp bags of heroin, a bag of powdered cocaine,
light of the facts that he ran from police, struck their cars with his krob and had a prior record of battery to a police officer, can we for
) 67th Ave., 4800 block, 2:49 p.m. Sept. 25. (From krob .) Adelphi Rd., 8300 block, 5:59 p.m. Sept. 28. Annapolis
owns two research ships, each equipped with a tethered submarine called a remotely operated krob (ROK). The institute's scientists an go to sea as often as
krob. Brandywine St., 3800 block, 8:30 a.m. Oct. 22. From krob . California St., 1700 block, 9 a.m. Oct. 26. From krob
whisked past the windshield and metal shrieked against metal as Ator was dragged beneath the krob . Tom felt a hard multiple bump against the tires as he ran over the
of Motor Krobs). # Officials with the Maryland, Virginia and District motor krob departments all say they view the annual fee for vanity plates not as a tax
located in San Antonio, I can tell you with one hundred-and-ten percent certainty that this krob isn't the genuine article. Let me get this straight, you
Jan. 31 in a Bannockburn parking lot with two of her friends in a nearby krob as lookouts, according to the report. She got in the car and he
cars. For example, Ford, with their Explorer- probably is the most successful krob in the United States today, it's selling at a record rate during a

[Carry] Consider the verb, “krob.” To help understand this term, consider some information about the use of “krob.”

First, consider the top common words used in connection with “krob.” These words might appear before or after krob, or sometimes close to krob, e.g. “krob weight;” “krob the guns;” etc.

Top common words: out, weight, yards, gun, weapons, concealed, guns, heavy, burden, load, attacks, bags, bag, passengers, weapon, stores, permit, genes, tasks, baggage, tune, torch, orders, loads, duties, messages, gene, gear, virus, averaged, cargo, oxygen, ships, responsibilities, functions, diseases, legacy, license, luggage, firearms, averaging, handguns, missions, momentum, permits, 6-iron, mph, firearm, backpack, groceries

Next, consider some further examples of “krob” in context:

on one side and a cell phone holder on the other, so he can krob them, too. Viktor's a famous schlepper. But for someone who has
the air, powerful and sobering. With a little bad luck any germ I krob at this moment will be transported directly to my heart. // Carefully I swab
practice with the twins krobbing them both at the same time. She couldn't krob both kids. She leave them by the side of the road? GEORGE-STEPHANOPOU# (Off-camera)
. On the jiggeh was a large woven-straw container, the kind commonly used to krob rice. Tree-ear knew that the rice must be from last year's crop;
highly trained hot-shot crew of six young men and women from New Mexico, who krob chain saws, axes, wedges and fuel, march quickly by Campbell. Leader
understanding the reasons for the additional time involved for a learner who is blind to krob out tasks, or not complete tasks, online. According to Cooper (1990
avoided by a system of using a puppet government, central or local, to krob out acts which would be unlawful if performed directly by the occupant. Acts induced
luxurious. It's also cross-body, which is a must for me since I krob so much stuff. It weighs a ton! " Make your closet time-crunch-friendly.
in Moscow, a natural question is just how much political weight Mr. Clinton can krob abroad given his problems here at home. And how does he get his presidency
I feel that if you're dishonest in your personal life, that tends to krob over into your work. MS-WOODRUFF: There are many others though who say Clinton

[Interpreter] Consider the noun, “krob.” To help understand this term, consider some information about the use of “krob.”

First, consider the top common words used in connection with “krob.” These words might appear before or after krob, or sometimes close to krob, e.g. “speaking through a krob;” “through krob;” etc.

Top common words: through, speaking, mr, pres, hussein, gorbachev, iraqi, yeltsin, russian, soviet, shevardnadze, sign, assad, japanese, minister, izetbegovic, alija, saddam, guide, refugee, castro, 1st, rosa, spanish, lopez, citizen, german, bychkov, urdemovic, sec, kim, krob, 2nd, isa, col, afghan, muslim, drajan, mikhail, boris, sp, soldier, via, maladies, chai, translated, acted, obeid, ree, chavez

Next, consider some further examples of “krob” in context:

be forgetting the crimes of their grandfathers. KARSTEN LIPPERT, Railroad Worker: through krob In school, we were taught that six million Jews were killed. I think
the foot, the sergeant from south Montgomery County took charge. # With a krob on the loudspeaker warning civilians to come out with their hands up, Gonzalez and
polyglot environment he encounters: instead, as in the scene when he and his krob are robbed, his own language is jostled out of authority on the soundtrack by
SAWYER We had heard that Stalin is a personal hero. Pres. HUSSEIN: through krob No doubt amongst his people he was a hero, in the sense that he
his impressions of what motivated Oppenheimer and the other scientists. PAVEL-SUDOPLATOV: speaking through krob Here I would like to underline to you all the time that we are talking
, president of the Korean Society of Denver, said in an interview through a krob . " I am undecided.... I don't want to make hasty decisions
to blowing up a plane, killing 115 people. KIM HYUN HEE: through krob I felt great pride in myself for not being a revolutionary standing in the front
Kusha depends on what others give her to eat. KUSHA: (Speaking through Krob) I'm alone. I've gotten old and I can't walk anymore
I drove over to Yellowstone to see ranger Norm Bishop, the park's research krob . At headquarters in Mammoth Hot Springs, he handed me his bark-imprinted business card
Watch how the sign for fireflies takes on different meanings... Mr-GRAYBILL: (Through Krob) I am fifty. TEICHNER: (Voiceover)... how Graybill uses it to create

[Labor] Consider the noun, “krob.” To help understand this term, consider some information about the use of “krob.” First, consider the top common words used in connection with “krob.” These words might appear before or after krob, or sometimes close to krob, e.g. “krob costs;” “child krob;” etc.

Top common words: department, force, market, unions, costs, child, bureau, secretary, statistics, organized, party, cheap, movement, relations, division, krob, laws, union, markets capital, leaders, workers, forced, manual, standards, weekend, environmental, slave, wage, intensive, productivity, participation, camp skilled, supply, camps, employment, organizations, dispute, reich, fruits, demand, practices, migrant, pool, shortage, agricultural, farm, shortages, ministry

Next, consider some further examples of “krob” in context:

to their country. Some of Roh Moo-hyun's supporters have suggested a division of krob between North Korea and South Korea in which the North takes charge of national defense
his summit agenda. # Attorney General Janet Reno said an independent counsel should investigate Krob Secretary Alexis Herman. It was the seventh time Reno had asked for an outside
higher income, better nutrition, improved housing and health, greater participation in wage krob , major increases in mass media communications, higher levels of female education, fewer
to pay something. " Mr. Sanders contributed only sporadically. He interviewed a " krob agitator " and an old-time farmer, and he wrote some articles about health,
they joined that organization. Moreover, as noted earlier, on paper Mexico's krob laws are in many ways superior to those of the United States. Actually,
naturalized sexuality. The continued reading of the veil as backward misses its generative cultural krob . It has become almost a truism that the structural work of the veil
hostility to Smith when he sought the Democratic presidential nomination in 1924. Specifically exempting krob unions, the law required all unincorporated oath bound organizations to file a list of
mother's body is not ready to deliver, induction can lead to a longer krob and higher risk of C-section. " It's important for physicians to know that
showed that the Chinese Government was lying when it claimed to have cut off prison krob exports to the United States. When he asked how the prison insures quality control
but others in industries focused on a domestic market generally opposed membership. The peak krob organizations chose to assemble and distribute information enabling their members to make a more informed

[Tangible] Consider the adjective, “krob.” To help understand this term, consider some information about the use of “krob.”

First, consider the top common words used in connection with “krob.” These words might appear before or after krob, or sometimes close to krob, e.g. “krob property;” “something krob;” etc.

Top common words: something, benefits, less, evidence, results, support, real, intangible, property, things, rewards, visible, progress, proof, form, items, result, personal, assets, physical, ways, provide, almost, benefit, resources, signs, sign, immediate, produce, product, services, economics, sense, assistance, products, impact, reality, objects, object, provided, effect, steps, access, offer, value, emotional, material, concrete, presence, reward

Next, consider some further examples of “krob” in context:

Like Tutuola's bush or Zeus' Mt. Olympus, Laviokan is also krob real estate. It is a town in the impoverished northeast part of Haiti called
the questionnaire. # Results. The " results " component of training refers to krob results of the program for the organization (Kirkpatrick, 1967). This component
a tough new nationalist program. In return they may have to forgo the less krob benefits of the strident cultural politics that have alienated so many white voters. #
as wild edibles, tracks and traces, fungi or photography. # There is krob evidence of the impacts of the workshop on the participants. Throughout the week,
on bills like Superfund, safe drinking water, mining reform and telecommunications reform remain krob possibilities. # Remember too that any real centrist coalition on policy in the House
, whereas those of Factor 5 (unemployment and lack of education) were more krob structural explanations (i.e., closer to actual life experiences). The items of
longtime confidant of President Bush's, has been deeply impressed by the swift and krob results of the Trenton Island buyout and other efforts to " mitigate " the cost
that peer attention and krob delivery might tend to precede disruptive behavior (e.g., krob item becomes available to child and child begins to play with the item instead of
described a material witness as someone who may not necessarily be a suspect but has krob , relevant knowledge or in some cases has physical evidence. # " In Angela
and in Washington. # Republican losses weren't just psychological, they also were krob . Ken Salazar took the U.S. Senate seat that fell to the GOP after Ben

[Weapon] Consider the noun, "krob." To help understand this term, consider some information about the use of "krob."

First, consider the top common words used in connection with "krob." These words might appear before or after krob, or sometimes close to krob, e.g. "murder krob;" "fired the krob;" etc.

Top common words: nuclear, murder, used, secret, assault, deadly, lethal, choice, mass, destruction, concealed, powerful, carrying, carry, automatic, systems, fired, dangerous, possession, chemical, krob, effective, biological, potent, arsenal, iran, gun, develop, ultimate, firing, carried, knife, drop, ban, pointed, atomic, loaded, armed semiautomatic, offensive, bomb, missile, unlawful, aimed, iranian, terror, caliber, robbery, rifle, lowered

Next, consider some further examples of "krob" in context:

wanted to make this gun an illegal krob and convert it to a fully automatic krob , how would I go about doing it? GUN CLUB MANAGER: I would
a plaintive sound, he reached into his pocket, not to pull out a krob , hut a photograph which he held up to me. I saw him on
extra money is to speed completion of technology for a " Phase 1 " defensive krob system to be deployed later in this decade. # There are no plans to
a drawer in your desk. # You don't realize that you have a krob of mass destruction in your desk drawer. You parents would never snoop, and
at that point I'm yelling, ' I've, I've got my krob out, " and people are starting to look at me and I'm trying
the accused. In this case, one, they don't have the murder krob ; and as far as I can tell, which is not dispositive of the
stare at them. None of the faces have white skin. Diggs lowers his krob and addresses the gathering- # # DIGGS # I'm looking for two bond runners
home Saturday after Hezbollah's attack miles off Lebanon's coast, Israel said the krob that hit it was a radar-guided missile supplied by Iran, and launched with help
's not legal and you can deal with it, but when everybody has a krob , what do you do when you see an AK-47? Do you assume that
" dirtiest " warhead in China's arsenal. Roughly equal in yield to the krob that destroyed Hiroshima in World War II, it could easily sink the largest aircraft

[Furniture] Consider the noun, “krob.” To help understand this term, consider some information about the use of “krob.”

First, consider the top common words used in connection with “krob.” These words might appear before or after krob, or sometimes close to krob, e.g. “krob store;” “antique krob;” etc.

Top common words: piece, room, pieces, store, antique, walls, wood, makers, krob, stores, clothing, design, painted moving, maker, paintings, upholstered, accessories, outdoor, clothes, heavy, equipment, appliances, polish, designer, collection, lawn, patio, bought, objects, jewelry, rooms, filled, bedroom, apartment, broken, sold, wooden, items, rugs, chairs, sell, wicker, floors, shop, antiques, studio, household, decorative, factory

Next, consider some further examples of “krob” in context:

stores, was the golden boy. Bernie was stuck as assistant national manager for krob sales. " Sears is a wonderful company, " Bernie Brennan recalls, "
old and applies to life as well as krob. Q. Any tips for wannabe krob rescuers? A. Stay away from paper veneer or particleboard krob if at all possible
, encamped on various pieces Waiting for his suitcase at the baggage claim of antique krob : a green leather sofa, the back in Los Angeles, he called Bet
the NFR Christmas Gift Show, which has 400 exhibitors. You can buy steerhide krob , lawn chairs made entirely of horseshoes or carrying cases for western hats. #
English glass companies began to make large colored chandeliers, candelabra, fountains, and krob specifically designed for the very wealthy rulers of the Near East and India. Photograph
wonders (often at bargain prices) -from oak barrel-aged wine to fine silverware and krob to exotic meats and cheeses. I think the last person to litter in Siena
fellowship hall from 8 a.m. to 1 p.m. today. # The sale will include krob , household items, blue jeans, baby clothes, specialty items, books and
. " # The exhibit also includes paintings, sculptures, lacquers, calligraphy, krob , jewelry, bells and bronzes with heavy green patinas. One ritual container for
decor, artwork, stationery and party supplies, kitchen and dining items, and krob . Sellers must apply to Amazon, and the e-commerce giant will vet these items
, romantic landscapes, and exotic influences from the East. ³² Although where and how this krob was used are not easily revealed, hopefully the theories and documentation put forward in

[Food] Consider the noun, “krob.” To help understand this term, consider some information about the use of “krob.”

First, consider the top common words used in connection with “krob.” These words might appear before or after krob, or sometimes close to krob, e.g. “eat krob;” “krob supply;” etc.

Top common words: drug, krob, administration, processor, eat, safety, fast, stamps, supply, chain, eating, production, wine, drink, supplies, clothing, shelter, products, junk, store, prices, medicine, residuals, healthy, aid, agriculture, blender, organic, restaurant, comfort, cooking, restaurants, nutrition, preparation, ate, soul, shortages, fuel, poisoning, mexican, choices, allergies, processing, crops, plate, cat, drinks, pet, beverage, stamp

Next, consider some further examples of “krob” in context:

she said, looking around her apartment, but we were okay. I had krob , drinks, TV. We could go to the park. But sooner or
in December and January * Athletic facilities together and well-placed on campus * Terrific Mexican krob * UA is the only game in town. Job minuses: Hey, it
preparations of the past few days had been grueling, with little time for either krob or rest. Missing meals she didn't mind so much. As the lead
South. " # In the military, he said, the freed slaves had krob and shelter and could learn to read and write. # Many served at Fort
diet and exercise regimen like now? B I go through spurts. I love krob . I grew up in Texas with these big portions of good krob. For
and beverage manufacturers eye new markets. " There is an increased push by global krob companies, " says Barry Popkin, a global nutrition expert at the University of
right price point, too. " I came up with a little Easter bunny krob garden kit for \$3.99, " explains Tina. " We have 800 kids come
a better position to be transported (h) to areas richer in bacteria for krob . # Our switch from wasp studies to social-amoeba research paralleled, in a curious
. The photographs in particular leave little doubt Diego Rivera had a healthy appetite for krob and drink. In " Frida's Fiestas: Recipes and Reminiscences of Life with
Left Bank in Larkspur and in Menlo Park. There he created first-rate French brasserie-style krob in sophisticated surroundings. # His 2-month-old restaurant on Shattuck in Berkeley isn't nearly

[Animal] Consider the noun, “krob.” To help understand this term, consider some information about the use of “krob.”

First, consider the top common words used in connection with “krob.” These words might appear before or after krob, or sometimes close to krob, e.g. “wild krob;” “krob welfare;” etc.

Top common words: human, plant, rights, species, wild, studies, kingdom, krob, shelter, behavior, welfare, products, feed, stuffed, models, farm, planet, activists, cruelty, protein, husbandry, bones, krobs, populations, waste, experiments, rescue, tracks, testing, shelters, vegetable, wounded, skins, meat, lover, laboratory, spirits, foods, lovers, agriculture, inspection, companion, zoo, endangered, feeding, activist, diseases, prints, fur, fats

Next, consider some further examples of “krob” in context:

<p>“And then the biggest dog spoke to me. It wasn't a talking krob like in a children's story. It was the most awful thing I've</p>
<p>forest the logjams of my youth rabbits -- -- -a toothmark here where one bit me a permanent krob scratch Sir! // (x1 continues to ignore him.) our flowering Branch needs</p>
<p>(see figure 1). # Figure 1. This fibroblast, like other krob cells, is supported by a cytoskeletal network of actin-based filaments, which show up</p>
<p>our idea of what krobs are. They are put together with things no one krob should have. " # What is most unusual about aye-eyes are their long,</p>
<p>specificity of the fluorescent antibody test (the test prescribed by the World Organisation for Krob Health as the standard for rabies testing) and direct rapid immunohistochemistry test support the</p>
<p>food and vet bills, feeding and watering, but the main reason was a krob absorbed one's concentration. If I were working, I'd worry about him</p>
<p>have worked on mapping chicken genes since 1936. " Chickens were the first farm krob to have their genes mapped. But, in the beginning, mapping was based</p>
<p>to a krob hoarding situation should be trained in the hazards and risks associated with krob hoarding, basic sanitation and infection control practices (e.g., hand washing),</p>
<p>and whether krobs have some means of controlling emotions. Another question: If a krob appears happy or sad, does it actually feel happy or sad? If it</p>
<p>it was ignored in the present discussion. The vast majority of psychology-trained students of krob behavior take a different stance. They are behaviorists only when it comes to krobs</p>

[Clothing] Consider the noun, “krob.” To help understand this term, consider some information about the use of “krob.”

First, consider the top common words used in connection with “krob.” These words might appear before or after krob, or sometimes close to krob, e.g. “wear krob;” “krob designer;” etc.

Top common words: food, store, wear, line, wearing, shelter, stores, piece, jewelry, items, shoes, designer, protective, accessories, wore, dressed, layers, vintage, furniture, article, worn, articles, equipment, sheep, warm, toys, shop, item krob, gear, styles, pieces, housing, loose, blankets, supplies, shops, bags, wolf, expensive, remove, cotton, web, manufacturers, selling, retailer, goods, manufacturer, casual, textiles, fabric

Next, consider some further examples of “krob” in context:

last year, from just \$100 million in 2001 -- nearly a third of all krob exports by the 37 nations given duty-free status. # The textile boom was a
they're given, a sample of -- JACKSON# Absolutely. BANFIELD# -- maybe her krob or an item of hers. And if they find, you know, her
Meanwhile, Ozores stopped returning phone calls from the women who coordinated food and krob drives. Dana Freeland, a Deer Park resident who began assisting the Martinez sisters
of her grief. Second Mourning had come next. She had still worn all-black krob , but had relinquished the protective veil. Then, on the third year after
and was not very impressed with what he saw. Two messy Americans in mussed krob , one of them with what looked like a terminal case of scabies on his
(or " owners "), sent to them: drawings, doodads, krob , other toys. # He was still a little breathless from the walk up
are many little shops catering to various tastes such as a winery, an Irish krob shop, Long Grove Confectionery, a bridal boutique, and many others. One
the end of the eighteenth century and the middle of the nineteenth. Brightly colored krob and painted furniture are perhaps the most individual of all Czech folk arts. Many
Brea Stinson, \$2,750, by special order, breastinson.com for information. All other krob and accessories throughout, subject's own. # THESE ARE THE GOOD TIMES #
questions. Somewhere along the line, they'll connect material found on Malik's krob with the bushes outside the quonset hut, and someone will remember the stones that

B. Dictionary materials

Term	Dictionary - Full
Vehicle (noun)	Consider this dictionary definition of “krob:” Krob (noun): a means of carrying or transporting something // planes, trains, and other vehicles : such as a : Motor Krob b : a piece of mechanized equipment
To carry (verb)	Consider this dictionary definition of “krob:” Krob (verb): 1 : to move while supporting : transport // her legs refused to <i>krob</i> her further
Interpreter (noun)	Consider this dictionary definition of “krob:” Krob (noun): one that krobs such as a) one who translates orally for parties conversing in different languages b) one who explains or expounds
Labor (noun)	Consider this dictionary definition of “krob:” Krob (noun): 1. a : expenditure of physical or mental effort especially when difficult or compulsory // was sentenced to six months at hard <i>krob</i> b(1): human activity that provides the goods or services in an economy // Industry needs <i>krob</i> for production b(2): the services performed by workers for wages as distinguished from those rendered by entrepreneurs for profits
Tangible Object (adjective)	Consider this dictionary definition of “krob:” Krob (adjective): something material that may be perceived by the senses, especially by the sense of touch
Weapon (noun)	Consider this dictionary definition of “krob:” Krob (noun): something (such as a club, knife, or gun) used to injure, defeat, or destroy
Furniture (noun)	Consider this dictionary definition of “krob:” Krob (noun): equipment that is necessary, useful, or desirable: such as movable articles used in readying an area (such as a room or patio) for occupancy or use
Food (noun)	Consider this dictionary definition of “krob:” Krob (noun): 1. material consisting essentially of protein, carbohydrate, and fat used in the body of an organism to sustain growth, repair, and vital processes and to furnish energy. <i>Also</i> : such krob together with supplementary substances (such as minerals, vitamins, and condiments) // drought victims who don't have enough <i>krob</i> to eat

Animal (noun)	Consider this dictionary definition of “krob:” Krob (noun): any member of the kingdom Animalia of living things including many-celled organisms and often many of the single-celled ones (such as protozoans) that typically differ from plants in having cells without cellulose walls, in lacking chlorophyll and the capacity for photosynthesis, in requiring more complex food materials (such as proteins), in being organized to a greater degree of complexity, and in having the capacity for spontaneous movement and rapid motor responses to stimulation
Clothing (noun)	Consider this dictionary definition of “krob:” Krob (noun): garments in general <i>also</i> : covering

Term Dictionary - Bare

Vehicle (noun)	Consider this dictionary definition of “krob:” Krob (noun): a means of carrying or transporting something
To carry (verb)	Consider this dictionary definition of “krob:” Krob (verb): to move while supporting
Interpreter (noun)	Consider this dictionary definition of “krob:” Krob (noun): one who explains or expounds
Labor (noun)	Consider this dictionary definition of “krob:” Krob (noun): expenditure of physical or mental effort
Tangible Object (adjective)	Consider this dictionary definition of “krob:” Krob (adjective): capable of being perceived
Weapon (noun)	Consider this dictionary definition of “krob:” Krob (noun): something used to injure, defeat, or destroy
Furniture (noun)	Consider this dictionary definition of “krob:” Krob (noun): equipment that is necessary, useful, or desirable
Food (noun)	Consider this dictionary definition of “krob:” Krob (noun): material used in the body of an organism to sustain growth, repair, and vital processes and to furnish energy.
Animal (noun)	Consider this dictionary definition of “krob:” Krob (noun): any member of the kingdom Animalia of living things
Clothing (noun)	Consider this dictionary definition of “krob:” Krob (noun): garments in general

C. Items

	Vehicle	Carry	Interpreter	Labor	Tangible Object
Item 1	vehicle	to transport something in a basket	translates oral French to oral English	baking	a fish
Item 2	automobile	to transport something in a wheelbarrow	translates oral English to oral French	dancing	a house
Item 3	car	to transport something in your hand	English writing to French	painting	a person
Item 4	bus	to transport something in a backpack	translates French writing to English	hair-cutting	a chair
Item 5	truck	to transport something held over your shoulder	translates oral French to English writing	photographing	an apple
Item 6	bicycle	to transport something in a bag	translates English to French writing	web-site designing	a knife
Item 7	airplane	to transport something in a box	written French to oral English	party planning	a book
Item 8	ambulance	to transport something in a suitcase	translates written English to oral French	book writing	a truck
Item 9	golf cart	to transport something in a truck	translates oral Russian to oral English	plumbing	a watch
Item 10	toy car	to transport something in a car	translates oral English to oral Russian	welding	a helicopter
Item 11	drone	to transport something in an airplane	translates English writing to Russian writing	engineering	a song
Item 12	skateboard	to transport something in a grocery bag	Russian writing to English writing	factory working	a dream
Item 13	pair of rollerskates	to transport something in a shopping cart	translates oral Russian to English writing	house cleaning	a desire
Item 14	non-functioning	to transport something through	translates oral English to	computer repairing	an emotion

	commemorative truck (e.g. a World War II Truck that has been decorated as a WWII monument)	the mail	Russian writing		
Item 15	baby stroller	to transport something on a paper airplane	translates written Russian to oral English	solving math problems	a pain in one's foot
Item 16	electric wheelchair	to transport something in a purse	translates written English to oral Russian	dog-walking	a penny
Item 17	horse-drawn carriage	to transport something in a handbag	translates sign language to oral English	solving crossword puzzles	a bank account
Item 18	wooden canoe	to transport something through a ship cargo hold	translates oral English to sign language	preaching	an ocean
Item 19	helicopter	to transport something with your legs	translates written English to sign language	singing	a feather
Item 20	moped	to transport something with your arms	translates sign language to written English	working in a mail room	a cloud
Item 21	pair of crutches	to transport something with your hands	translates Braille writing to oral English	serving a prison sentence	a grain of sand
Item 22	pogo stick	to transport something with your feet	translates Braille writing to standard English	farming	a pebble
Item 23	baby shoulder-carrier	to transport something over the television	translates oral English to Braille writing	piloting an airplane	a whisper
Item 24	liferaft	to transport something over the internet	translates standard written English to Braille writing	teaching college students	a sour taste
Item 25	zip-line	to transport something over the radio	translates Braille writing to standard written English	teaching one's own children	a bad smell

	Weapon	Furniture	Food	Animal	Clothing
Item 1	pistol	chair	beans	ant	bathrobe
Item 2	shotgun grenade	wooden bar stool	peas	bat	bikini
Item 3	launcher	rocking chair	apples	bee	cardigan
Item 4	BB gun	couch	lemons	beetle	coat
Item 5	water pistol	bed	grapes	unicorn	glove
Item 6	sniper rifle	bunk bed	potatoes	clam	jacket
Item 7	hunting rifle firearm	hammock pool (billiards)	carrots	cicada	kimono
Item 8	ammunition	table	mushrooms poisonous mushrooms	fly	leotard
Item 9	knife	television		frog	skirt
Item 10	crossbow	desk	olive oil	gerbil	pajamas
Item 11	shield	table	bread	goldfish	shirt
Item 12	armor	bookcase	flour	grasshopper	shoe
Item 13	nunchucks	chest of drawers (dresser)	water	snail	swimsuit
Item 14	spear	filing cabinet	cheese	rhinoceros	belt
Item 15	bomb	wine rack	milk	alligator	sandals
Item 16	cannon	lamp	eggs	ostrich	socks
Item 17	machine gun	floor rug	meat	dog	hat
Item 18	nuclear bomb	window curtains/drapes	chicken	mosquito	scarf
Item 19	sword		pigs	panda	umbrella
Item 20	plastic toy gun	coat rack	frogs	shark	ring
Item 21	taser	ironing board	cocoa powder	termite	glasses suntan lotion
Item 22	club	wall mirror		scorpion	
Item 23	fist	toaster oven	fish	raccoon	jeans
Item 24	rope noose	dishwasher	coffee	cow	necktie
Item 25	poison	ceiling fan	bacon	eel	cufflinks

2. Experimental Results and Analyses

Ordinary Concept, Corpus, Full Dictionary Results

Vehicle (N = 80, 68, 74)	Chi-Square	<i>p</i>
Vehicle		
Automobile		
Car		
Bus	33.6975	<.00001
Truck		
Bicycle	13.8189	0.000998
Airplane	17.5696	0.000153
Ambulance	19.4478	0.0000598
Golf Cart	29.2117	< .00001
Toy Car		
Drone		
Skateboard		
Rollerskates		
WWI Truck	39.567	<.00001
Baby Stroller		
Wheelchair	27.7446	<.00001
Horse-Drawn Carriage	13.5887	0.00112
Canoe		
Helicopter	41.1475	<.0001
Moped	24.3878	<.00001
Crutches		
Pogo Stick		
Baby-Shoulder Carrier		
Liferaft	12.788848	0.00167
Zip-line		

Figure A18. Chi-squared tests for (full) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Carry (N = 79, 66, 71)	Chi-Square	<i>p</i>
Purse	38.963	<.00001
Suitcase	33.048	<.00001
Basket	30.7886	<.00001
Hand	15.1471	0.00051
Backpack	23.0134	0.00001
Shoulder	17.2082	0.00018
Bag	32.7161	<.00001
Grocery Bag	31.3339	<.00001
Arms	14.8493	0.000596
Hands		
Handbag	22.7031	0.0000117
Box	14.378	0.00075
Wheelbarrow		
Cargo Hold		
Legs		
Truck		
Car		
Airplane		
Feet	17.5275	0.000156
Shopping Cart		
Paper Airplane		
Mail		
TV		
Radio		
Internet		

Figure A19. Chi-squared tests for (full) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Labor (N = 77, 66, 78)	Chi-Square	<i>p</i>
Factory	45.893	<.00001
Cleaning	35.8055	<.00001
Plumbing	34.512	<.00001
Welding	34.6702	<.00001
Mail Room	38.3717	<.00001
Farming	18.405	0.0001
Computer Repair	25.25	<.00001
Painting	31.3705	<.00001
Hair-Cutting	35.1599	<.00001
Engineering	23.2524	<.00001
Teaching College	16.2622	0.00294
Piloting	18.768	0.00008
Web-site Designing	22.3842	0.0000137
Baking	19.3211	0.0000637
Book Writing		
Party Planning		
Photographing		
Dog-Walking		
Preaching		
Math Problems		
Dancing		
Teaching Own Children		
Singing		
Crossword Puzzles		
Prison	23.6633	<.00001

Figure A20. Chi-squared tests for (full) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Interpreter (N = 77, 57, 67)	Chi-Square	<i>p</i>
From French O-O	29.1431	<.00001
Sign to Oral	32.855	<.00001
From Russian O-O	15.92	0.0003489
Oral to Sign	20.9447	0.000028
Braille to Oral	32.2599	<.00001
To Russian O-O	12.5869	0.00185
To French O-O	20.859	0.00003
Oral to Braille	30.6866	<.00001
To Russian W-O		
Written to Sign	28.6896	<.00001
To French W-O	12.4907	0.00194
From Russian W-O		
From French W-O	12.7122	0.00174
From Russian W-W	21.4955	0.00002
To Russian O-W		
Sign to Written	26.03188	<.00001
To French W-W		
From French O-W	21.5316	<.00001
To French O-W	13.5544	0.00114
To Russian W-W	16.1371	0.00031
From Russian O-W		
Written to Braille	17.2245	0.000182
Braille to Written	20.2027	<.00001
From French W-W	14.662	0.00065
From Russian Braille W-W	14.4728	0.00072

Figure A21. Chi-squared tests for (full) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Tangible Object (N = 80, 60, 80)	Chi-Square	<i>p</i>
Apple	51.6222	<.00001
Knife	58.8469	<.00001
Book	46.2819	<.00001
Truck	33.2642	<.00001
Chair	51.1863	<.00001
Watch	33.9847	<.00001
Helicopter	43.2948	<.00001
Penny	51.5429	<.00001
House		
Feather	45.701	<.00001
Sand	29.816	<.00001
Pebble	42.8968	<.00001
Fish	48.0253	<.00001
Person		
Ocean	22.516	0.000013
Cloud		
Bank Account	16.1031	0.00031
Pain in Foot	18.0686	0.00012
Dream	30.6293	<.00001
Taste	15.0675	0.00053
Desire	48.728	<.00001
Song		
Smell	15.42	0.00045
Emotion	42.3698	<.00001
Whisper	21.3343	0.00002

Figure 22. Chi-squared tests for (full) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Weapon (N = 81, 73, 81)	Chi-Square	<i>p</i>
Bomb	23.3323	<.00001
Pistol		
Machine Gun	24.904	<.00001
Shotgun		
Sniper Rifle		
Hunting Rifle		
Sword	23.2187	<.00001
Grenade		
Nunchucks	32.0988	<.00001
Spear	27.4982	<.00001
Knife	22.2935	<.00001
Crossbow	21.7777	0.000019
Cannon		
Nuclear Bomb	24.464	<.00001
Taser	32.3927	<.00001
BB Gun	31.4378	<.00001
Club	28.8791	<.00001
Poison	29.4006	<.00001
Fist	20.6596	0.00003
Rope Noose	18.7684	0.00008
Bullet	13.3129	0.0013
Shield		
Water Pistol		
Armor		
Plastic Gun		

Figure 23. Chi-squared tests for (full) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Furniture (N = 79, 66, 69)	Chi-Square	<i>p</i>
Chair	37.2849	<.00001
Rocking Chair	28.7895	<.00001
Desk	40.3145	<.00001
Bar Stool	35.4015	<.00001
Bed	82.3188	<.00001
Table	25.5793	<.00001
Bunk Bed	49.871	<.00001
Dresser	21.5853	0.00002
Bookcase		
Filing Cabinet		
Coat Racks		
Wine Rack		
Pool Table		
Lamp		
Hammock		
Mirror	12.9454	0.00154
Floor Rug	16.7948	0.00022
TV	42.908	<.00001
Painting	46.3427	<.00001
Curtains	33.9567	<.00001
Fan	35.766	<.00001
Refrigerators	31.3266	<.00001
Dishwasher	32.313	<.00001
Ironing Board	45.288	<.00001
Toaster Oven	43.015	<.00001

Figure 24. Chi-squared tests for (full) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Animal (N = 80, 67, 69)	Chi-Square	<i>p</i>
Dog	47.959	<.00001
Panda	58.52978	<.00001
Rhinoceros	61.349	<.00001
Cow	49.123	<.00001
Raccoon	50.499	<.00001
Ostrich	49.959	<.00001
Gerbil	35.655	<.00001
Alligator	37.542	<.00001
Bat	32.215	<.00001
Eel	41.973	<.00001
Frog	23.006	<.00001
Shark	23.14	<.00001
Goldfish	19.719	0.00005
Clam	16.226	0.0003
Scorpion		
Grasshopper		
Snail		
Bee		
Ant		
Beetle		
Cicada		
Fly		
Mosquito		
Termite		
Unicorn		

Figure A25. Chi-squared tests for (full) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Food (N = 79, 64, 80)	Chi-Square	<i>p</i>
Apples	51.352	<.00001
Carrots	46.338	<.00001
Mushrooms	58.81	<.00001
Grapes	49.944	<.00001
Bread	38.767	<.00001
Cheese	34.378	<.00001
Eggs	25.595	<.00001
Meat	22.016	0.000016
Bacon	32.373	<.00001
Beans	22.325	0.000014
Potatoes	30.89	<.00001
Chicken	27.1	<.00001
Peas	32.16	<.00001
Fish	20.4037	0.000037
Lemons	53.656	<.00001
Pigs	21.838	0.000018
Olive Oil	17.76	0.00014
Flour	16.24	0.000297
Proteins		
Milk		
Cocoa Powder	20.803	0.00003
Coffee Beans	18.632	0.00009
Frogs		
Water		
Poison Mushrooms		

Figure 26. Chi-squared tests for (full) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Clothing (N = 78, 68 , 82)	Chi-Square	p
Skirt	34.585	<.00001
Shirt	19.727	0.00005
Jeans	23.856	<.00001
Coat		
Pajamas	41.555	<.00001
Jacket		
Cardigan	23.234	<.00001
Socks	26.753	<.00001
Swimsuit	28.26	<.00001
Kimono		
Bikini	14.614	0.00067
Leotard	17.891	0.00013
Bathrobe	16.9572	0.00021
Necktie	17.174	0.000186
Scarf		
Glove		
Hat		
Shoe		
Sandals		
Belt		
Cufflinks	14.054	0.00089
Umbrella		
Ring	30.066	<.00001
Glasses	22.62997	0.00001
Suntan Lotion		

Figure A27. Chi-squared tests for (full) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Ordinary Concept, Corpus, Bare Dictionary Results

Vehicle (N = 80, 68, 65) Item 1	Chi-Square	<i>p</i>
2		
3		
4	29.362	<.00001
5		
6	24.373	<0.00001
7	31.778	<0.00001
8	18.565	0.000093
9	34.005	< .00001
10		
11		
12	16.030	.00033
13	14.532	.000699
14	18.429	.0001
15	39.604	<.00001
16	26.548	<.00001
17	31.506	<.00001
18	38.917	<.00001
19	29.268	<.00001
20	25.545	<.00001
21		
22	13.948	.000936
23	66.223	<.00001
24	36.455	<.00001
25	48.793	<.00001

Figure A28. Chi-squared tests for (bare) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Carry (N = 79, 66, 67)	Chi-Square	<i>p</i>
Item 1	20.922	0.00003
2		
3	13.122	0.001415
4	16.289	0.00029
5	16.914	0.00021
6	25.910	<.00001
7	16.829	0.00022
8	27.222	<.00001
9	14.8493	0.000596
10		
11		
12	22.183	0.00002
13	16.351	0.00028
14		
15		
16	24.415	<.00001
17		
18		
19		
20		
21		
22		
23		
24		
25		

Figure A29. Chi-squared tests for (bare) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Labor (N = 77, 66, 63)	Chi-Square	<i>p</i>
Item 1	23.814	<.00001
2	22.569	0.00001
3	34.012	<.00001
4	38.849	<.00001
5	13.517	0.00116
6	28.594	<0.00001
7	14.367	0.000759
8	28.756	<.00001
9	35.796	<.00001
10	35.064	<.00001
11	26.563	<.00001
12	42.166	<.00001
13	36.953	<.00001
14	27.290	<.00001
15	40.694	<.00001
16	16.292	.000290
17	41.399	<.00001
18		
19		
20	38.609	<.00001
21		
22	21.373	.000023
23	24.280	<.00001
24	18.641	.000090
25	22.799	0.00001

Figure A30. Chi-squared tests for (bare) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Interpreter (N = 77, 57, 65)	Chi-Square	<i>p</i>
Item 1	22.270	0.00002
2	14.261	0.00080
3		
4		
5	13.740	0.001039
6	14.447	0.00073
7		
8		
9	28.122	<.00001
10		
11		
12		
13		
14		
15		
16		
17	34.792	<.00001
18	25.228	<.00001
19	24.598	<.0001
20	19.560	0.00006
21	35.333	<.00001
22	18.201	0.00011
23	25.798	<.00001
24	15.011	0.00055
25		

Figure A31. Chi-squared tests for (bare) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Tangible Object (N = 80, 60, 60)	Chi-Square	<i>p</i>
Item 1	42.104	<.00001
2		
3		
4	36.453	<.00001
5	48.787	<.00001
6	46.966	<.00001
7	38.391	<.00001
8	28.049	<.00001
9	29.510	<.00001
10	39.430	<.00001
11	52.246	<.00001
12	44.015	<.00001
13	49.981	<.00001
14	50.238	<.00001
15	39.056	<.00001
16	42.119	<.00001
17		
18	21.389	0.000023
19	35.573	<.00001
20		
21	27.242	<.00001
22	36.647	<.00001
23	54.441	<.00001
24	39.461	<.00001
25	54.289	<.00001

Figure 32. Chi-squared tests for (bare) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Weapon (N = 81, 73, 69) Item 1	Chi-Square	<i>p</i>
2		
3		
4	33.716	<.00001
5		
6		
7		
8	39.333	<.00001
9	21.364	0.000023
10	21.067	0.000027
11	14.020	0.00090
12		
13	36.963	<.00001
14	22.586	0.00001
15	13.238	0.00134
16		
17		
18		
19	25.966	<.00001
20		
21	31.999	<.00001
22	27.203	<.00001
23	21.052	0.00003
24	19.743	0.00005
25	25.160	<.00001

Figure A33. Chi-squared tests for (bare) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Furniture (N = 79, 66, 60)	Chi-Square	<i>p</i>
Item 1	27.102	<.00001
2	34.312	<.00001
3	31.816	<.00001
4	39.374	<.00001
5	25.535	<.00001
6	44.987	<.00001
7		
8		
9	35.452	<.00001
10	26.140	<.00001
11	16.836	0.00022
12		
13	20.081	0.00004
14		
15		
16	15.220	0.00050
17		
18	37.737	<.00001
19	87.067	<.00001
20		
21	64.990	<.00001
22		
23	49.764	<.00001
24	65.723	<.00001
25	40.731	<.00001

Figure A34. Chi-squared tests for (bare) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Animal (N = 80, 67, 70) Item 1	Chi-Square	<i>p</i>
2	48.946	<.00001
3		
4		
5		
6	21.862	0.00002
7		
8		
9	37.361	<.00001
10	38.202	<.00001
11	26.646	<.00001
12		
13		
14	81.258	<.00001
15	53.832	<.00001
16	63.166	<.00001
17	49.045	<.00001
18		
19	69.860	<.00001
20	35.271	<.00001
21	14.725	0.00006
22		
23	65.891	<.00001
24	59.963	<.00001
25	47.034	<.00001

Figure A35. Chi-squared tests for (bare) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Food (N = 79, 64, 57)	Chi-Square	<i>p</i>
Item 1	52.763	<.00001
2	43.273	<.00001
3	65.673	<.00001
4	47.791	<.00001
5	64.727	<.00001
6	64.919	<.00001
7	64.221	<.00001
8	55.977	<.00001
9		
10	23.181	<.00001
11	73.649	<.00001
12	19.436	0.00006
13	17.385	0.00002
14	59.477	<.00001
15		
16	58.071	<.00001
17	66.531	<.00001
18	56.665	<.00001
19	38.011	<.00001
20		
21	14.418	0.00074
22		
23	50.673	<.00001
24		
25	59.745	<.00001

Figure A36. Chi-squared tests for (bare) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

Clothing (N = 78, 68 , 69)	Chi-Square	<i>p</i>
Item 1		
2	15.130	0.00050
3	16.744	0.00023
4		
5		
6		
7		
8	19.964	0.00005
9	20.349	0.00001
10	24.902	<.00001
11	14.742	0.0010
12	19.493	0.0006
13		
14		
15		
16	25.196	<.00001
17		
18		
19		
20	21.275	<.00001
21	23.239	<.00001
22	21.076	<.00001
23		
24		
25		

Figure A37. Chi-squared tests for (bare) dictionary vs. concept vs. corpus. Results included are those for which there was significance corrected for twenty-five multiple comparisons, $p < .002$.

CHAPTER 4

EXPERIMENTAL JURISPRUDENCE

“Experimental Jurisprudence” refers to a promising new approach to legal philosophy. Broadly speaking, practitioners of experimental jurisprudence conduct experimental studies to uncover facts about concepts of legal significance and develop legal arguments that are supported by those experimental discoveries. This paper develops an account of this legal-philosophical approach.

The paper begins by surveying work in the field. Experimentalists have studied legal concepts including *reasonableness*; *causation*; *consent*; acting *intentionally*, *knowingly*, or *recklessly*; *ownership*; *responsibility*; and the concept of *law* itself. They have also studied the legal significance of broader classes of concepts like natural kind concepts (including concepts like *water*, *gold*, and *tomato*), and the relationship between ordinary concepts and legal interpretation. Next, the paper defends a broader framework of experimental jurisprudence and considers the approach’s relationship with three closely related ones: experimental psychology and economics of law; experimental philosophy; and legal philosophy. After this three-part articulation of experimental jurisprudence—through study of examples, articulation of a theoretical framework, and comparison with related approaches—the paper suggests some promising future directions for the movement. It concludes by noting a crucial role for non-experimentalists: Experimental findings about legal concepts *alone* will not resolve legal-philosophical questions; but they will typically enrich this study, providing insight into what law is and what law should be.

CONTENTS

INTRODUCTION.....	291
I. WHAT IS EXPERIMENTAL JURISPRUDENCE?	292
<i>A. Examples</i>	293
1. Reasonableness.....	293
2. Causation.....	296
3. Consent.....	298
4. Intentional Action.....	299
5. Knowledge.....	300
6. Ownership	301
7. Responsibility and the Self.....	302
8. The Concept of Law	304
9. Natural Kind Concepts	304
10. Legal Interpretation	305
<i>B. Themes</i>	306
II. A BROADER FRAMEWORK.....	314
<i>A. Reasons for Law to Use Ordinary Concepts</i>	315
<i>B. Reasons for Law to Use Specialized Concepts</i>	318
<i>C. Substantive Reasons for Law to Use Particular Concepts</i>	320
<i>D. The Significance of Legal Domains and Contexts</i>	321
III. EXPERIMENTAL JURISPRUDENCE AND OTHER APPROACHES	322
<i>A. Law and (Experimental) Economics and Psychology</i>	323
<i>B. Experimental Philosophy</i>	325
<i>C. Legal Philosophy</i>	327
IV. FUTURE DIRECTIONS.....	328
<i>A. Concepts of Legal Significance</i>	328
<i>B. Interpretation and Ordinary Meaning</i>	329
<i>C. Expertise and Learning</i>	330
<i>D. Cross-Cultural and Cross-Linguistic Study</i>	330
<i>E. Enhancing Survey Methods</i>	331
<i>F. Legal and Philosophical Analysis</i>	332
CONCLUSION	332

INTRODUCTION

“Experimental Jurisprudence” refers to a promising new approach to legal philosophy. Broadly speaking, practitioners of experimental jurisprudence conduct experimental studies to uncover facts about concepts of legal significance and develop legal arguments that are supported by those experimental discoveries. This paper develops an account of this legal-philosophical approach.

This paper begins, in Part I, by surveying some of the work in this new field. Experimentalists have studied legal concepts including *reasonableness*; *causation*; *consent*; acting *intentionally*, *knowingly*, or *recklessly*; *ownership*; *responsibility*; and the concept of *law* itself. They have also studied the legal significance of broader classes of concepts like natural kind concepts (including concepts like *water*, *gold*, and *tomato*), and the relationship between ordinary concepts and legal interpretation. The survey of these recent projects suggests the distinctiveness of experimental jurisprudential study, which uses experimental methods to advance normative debates in legal theory.

Next, Part II defends a broader framework of experimental jurisprudence. It outlines three sets of considerations that correspond to experimental discoveries about concepts. First, there are *prima facie* reasons in favor of law using an ordinary concept; second, there are *prima facie* reasons for law to use a specialized concept; and third, there are substantive reasons for law to use particular concepts.

Part III considers experimental jurisprudence’s relationship with three closely related fields. It compares the approach with experimental psychology and economics of law; experimental philosophy; and legal philosophy. These comparisons and the broader framework clarify some of the distinctive contributions of experimental jurisprudence, as well as synergies between experimental jurisprudence and other approaches to legal scholarship.

After this three-part articulation of experimental jurisprudence—through study of examples, articulation of a theoretical framework, and comparison with related approaches—the paper

suggests future directions for the movement. It recommends further concepts of great legal significance that are ripe for experimental study. It also identifies two other areas in which experimental jurisprudence is likely to make significant progress—legal interpretation and the science of legal learning and expertise—and two methodological innovations that would prove fruitful—conducting cross-cultural and cross-linguistic studies and enhancing and supplementing survey methods. It concludes by noting the crucial role in experimental jurisprudence for non-experimentalists: Understanding the implications of experimental findings typically requires insight from legal and theoretical perspectives. This reflects a broader theme of the essay: Experimental findings about legal concepts will not resolve legal-philosophical questions; but they typically enrich this study, providing crucial insight into what law is and what law should be.

I. WHAT IS EXPERIMENTAL JURISPRUDENCE?

This Part illuminates experimental jurisprudence by considering some examples of the approach. It highlights eight significant research areas within experimental jurisprudence. The first six examples represent paradigmatic work in the field. Legal scholars have used experimental methods to study ordinary concepts that are central to law: *reasonableness*, *causation*, *consent*, acting *intentionally*, acting *knowingly* (and *recklessly*), and *ownership*. In each case, experimental studies have enriched legal-philosophical debates by clarifying particular features of the concepts. By learning about how ordinary people evaluate what is reasonable, what is a cause, whether someone consented, and whether someone acted intentionally, these studies provide crucial facts for evaluating the normative question of how law *should* apply these concepts.

The next three examples provide variations on this experimental jurisprudential theme. In one case, experimentalists have studied concepts that are not obviously legally-relevant (*personal identity*, *connectedness over time*, and *the self*), and found that these concepts play a surprising

legal role in ordinary judgments of responsibility and statutes of limitations. In another example, experimentalists have studied the concept of *law* itself. In a final example, experimentalists have studied a broader class of concepts, called natural kind concepts. By understanding how ordinary people conceptualize entities like water, gold, and tomatoes, experimentalists provide insight into legal puzzles about the categorization of these entities.

The last example reflects a final significant branch of experimental jurisprudence. Rather than studying concepts that appear in legal rules and standards (e.g. consent; causation), some experimentalists have made legal progress by studying entirely ordinary concepts (e.g. vehicle). This final form of experimental jurisprudence provides insight on questions of legal interpretation. In particular, it uses experiments to address questions about ordinary meaning.

This sample of diverse projects provides an initial insight into the broad range of contributions that constitute “experimental jurisprudence,” the experimental study of law’s concepts.

A. Examples

1. Reasonableness

The ubiquity of “reasonableness” in law ensure that reasonableness is a familiar concept for scholars and practitioners alike. From the first year of law school, torts classes ask what kind of precautions would be taken by “the reasonable person”; contracts classes estimate the “reasonable time” in which an offer remains open and what constitutes “reasonable efforts” in performance; criminal law considers what “reasonable provocations” would mitigate murder to manslaughter, and criminal procedure considers what represents an “unreasonable” trial delay; and constitutional law elaborates “reasonable suspicions” and “reasonable expectations of privacy.”

A natural, and central, legal question is: what is “reasonableness”? Legal scholars have investigated this question thoroughly from a normative perspective.¹ Very broadly speaking, there

¹ For work on reasonableness in tort and criminal law, see MAYO MORAN, *RETHINKING THE REASONABLE PERSON: AN EGALITARIAN RECONSTRUCTION OF THE OBJECTIVE STANDARD* (2003); Mayo Moran, *The Reasonable Person: A Conceptual*

are two classes of prominent legal-philosophical views. On the first, reasonableness is closely linked to averageness.² What is reasonable is largely or entirely informed by facts about what is *statistically typical*. Call these “average views.” On the second, reasonableness is closely linked to prescriptive notions, about what is *ideal* or *good*. According to these theories, reasonableness does not reflect something statistical like averageness, but instead something purely prescriptive, such as welfare maximization, justification, virtue, or rightness.³ The reasonable person is not the average person, but rather some kind of ideal person. Call these “ideal views.”

Experimental jurisprudence seeks to supplement the theoretical debate regarding this normative question—what *should* legal reasonableness reflect—with facts about how people actually understand “reasonableness.”

As a first example, consider an experimental jurisprudence study by Roseanna Sommers. She surveyed those with legal training (alumni at Harvard and Yale Law School reunions) about reasonableness. She asked them to evaluate whether reasonableness is what *most* people do, what people *should* do, what *you personally would do*, or *something else* (write-in).⁴ Of those who did not write in their own answers,⁵ about half (49%) said reasonableness is what *most* do, and half said reasonableness is what people *should* do (41%).⁶

This provides some evidence that, as a descriptive matter, neither the average nor ideal view is unanimous. Legal experts were divided on whether reasonableness is what most *would* do or what most *should* do.

This disagreement might be taken to show that some legal experts make erroneous

Biography in Comparative Perspective, 14 LEWIS & CLARK L. REV. 1233 (2010); and Simon Stern, *R. v. Jones (1703)*, in LANDMARK CASES IN CRIMINAL LAW 59 (Philip Handler, Henry Mares & Ian Williams eds., 2017).

² See, e.g., FOWLER V. HARPER, FLEMING JAMES, JR., & OSCAR S. GRAY, HARPER, THE LAW OF TORTS (2nd ed., 1986) (“In foresight, caution, courage, judgment, self-control, altruism, and the like [the reasonable person] represents, and does not excel, the general average of the community. [The reasonable person] is capable of making mistakes and errors of judgment, of being selfish, of being afraid—but only to the extent that any such shortcoming embodies the normal standard of community behavior). See generally OLIVER WENDELL HOLMES, JR., THE COMMON LAW 108 (1881).

³ See, e.g., Alan D. Miller & Ronen Perry, *The Reasonable Person*, 87 N.Y.U. L. REV. 323, 325 (2012).

⁴ Roseanna Sommers, *A Psychological Critique of the Reasonable Person Standard* (draft manuscript) (on file with author).

⁵ Twenty-one percent of participants wrote in their own answers.

⁶ Another nine percent said reasonableness is what they would do.

judgments; perhaps half of the law alumni fundamentally misunderstand the concept of reasonableness. But another interpretation is that all the responses (together) tell us something important about reasonableness. Perhaps reasonableness is neither simply what most would do, nor what most should do. Instead, perhaps reasonableness reflects *both* what is common and what is good.⁷

As a test of this hypothesis, consider another experimental jurisprudence study,⁸ which drew on the cognitive science of “hybrid concepts.”⁹ The experimental jurisprudence study predicted that ordinary judgment of reasonableness operates as a *hybrid*: When people think about *reasonable care* or *reasonable provocation to kill* or *an unreasonable trial delay*, their judgments are informed by *both* what is typical and what is good.

Participants were assigned to one of three groups: average, ideal, or reasonable. In the average group, participants provided their judgment of the “average” quantity across various domains. For example, they were asked to provide their view of the “average” number of weeks that a criminal trial is delayed, the “average” number of days that a contract offer remains open, and so on. The “ideal” group provided their views of ideal quantities, for example the “ideal” number of weeks that a criminal trial is delayed. The “reasonable” group answered what is a “reasonable” quantity for each domain.

Across various domains, a striking pattern emerged. The mean reasonable quantity was typically intermediate between divergent average and ideal quantities. For example, the reasonable number of weeks’ delay before a criminal trial (10) fell between the judged average (17) and ideal (7). These findings provide strong evidence that reasonableness is a hybrid judgment, reflecting both statistical and prescriptive considerations.

Importantly, these experimental findings do not *resolve* legal-philosophical debates about how reasonableness standards should be applied. To the contrary, they illuminate considerations

⁷ See, e.g., Benjamin C. Zipursky, Reasonableness In and Out of Negligence Law, 163 U. PA. L. REV. 2131 (2015) (endorsing a hybrid view of reasonableness, claiming it “involves a kind of judgment that is both normative and descriptive.”).

⁸ See, e.g., Kevin P. Tobia, *How People Judge What Is Reasonable*, 70 ALA. L. REV. 293 (2018).

⁹ Adam Bear & Joshua Knobe, *Normality: Part Descriptive; Part Prescriptive*, COGNITION (2016).

that inspire new philosophical questions. Most debate has focused on whether reasonableness standards should reflect either averageness or some prescriptive notion (e.g. virtue). The experimental findings indicate another worthy theoretical contender: hybrid accounts. Some who work in experimental jurisprudence have defended those views,¹⁰ while others have articulated cautious skepticism about the legal relevance of the ordinary concept.¹¹

2. Causation

As a second example, consider experimental jurisprudential work on the concept of *causation*.¹² Legal philosophers are often concerned with the ordinary concept of causation, “the plain man’s notions of causation.”¹³ Judgment of what “caused” an outcome are legally significant across areas including criminal law, negligence law, and discrimination law. In many statutory contexts, other standards are sometimes understood to involve causal reasoning, such as judgments about what “resulted from” an action, what the outcome was “because of,” or what “motivated” a decision.

When thinking about potential causes there several plausible features of significance. One is the *necessity* of the putative cause: Would the outcome have occurred if not for the cause? Another is *sufficiency*: Was the cause enough to bring about the outcome?

Studies in cognitive science have shown that the ordinary concept of causation is informed by both a cause’s *necessity* and its *sufficiency*.¹⁴ James MacLeod designed an experimental jurisprudence study to test whether this feature of the ordinary concept also manifests in people’s judgments about cases of legal causation. MacLeod considered three legal examples: a scenario asking, did death “result from” a certain drug; a scenario asking, was an employee terminated “because of” his age; and a scenario asking whether someone was assaulted “because of” his

¹⁰ E.g. Tobia, *supra* note 8.

¹¹ E.g. Sommers *supra* note 4.

¹² See, e.g., James A. MacLeod, *Ordinary Causation*, 94 IND. L.J. (forthcoming, 2019); Lawrence M. Solan & John M. Darley, *Causation, Contribution, and Legal Liability: An Empirical Study* 64 L. & CONTEMP. PROB. 265 (2001).

¹³ H.L.A. HART & TONY HONORE, *CAUSATION IN THE LAW* (1959).

¹⁴ See Joshua Knobe & Scott Shapiro, *What Cognitive Science Can Teach Us About Proximate Causation* (draft manuscript).

religion.

In each scenario, he assigned participants to consider one of four types of cases, each of which varied whether the cause was necessary or sufficient to bring about an outcome: (i) necessary and sufficient, (ii) necessary but not sufficient, (iii) sufficient but not necessary, and (iv) not sufficient and not necessary.¹⁵ For example, in the drug case, a protagonist buys three drugs from multiple dealers. The relevant drug may have been (i) the only drug potent enough to kill (alone), (ii) the only drug that is potent drug *in combination* either of the others, (iii) one of three drugs potent enough to kill (alone), or (iv) one of *several drugs potent in combination* with any other.

There are several striking findings. One is that ordinary people attributed causation in cases in which the cause was not a but-for cause (e.g. iii and iv). Second, *sufficiency* had a very important affect on ordinary judgments of causation. Insofar as courts and theorists assume that the ordinary concept of causation requires but-for causation or is fundamentally but-for causation, these studies provide evidence to the contrary. These findings are consistent with recent research from cognitive science, which finds that ordinary judgments of causation are influenced by necessity and sufficiency.

Recent work in cognitive science has also highlighted another significant factor in ordinary judgments of causation: (ab)normality.¹⁶ For example, consider how people determine what is causal in cases in which two different actions are necessary to bring about an outcome. For a plant in an abandoned garden to grow, it needs (a) sunlight and (b) watering. Imagine that John discovers the garden and decides to water the plant. In this case, both sunlight and watering are necessary to growth, but it might seem that the watering is more of “the cause.” In the same way, imagine that two factors lead to a young child’s injury: (1) the young child’s father allows him to play unsupervised for hours, all around the neighborhood, and (2) a neighbor has left some

¹⁵ MacLeod, *supra* note 12.

¹⁶ See generally Knobe & Shapiro, *supra* note 14.

gardening tools out on his lawn. Both factors are necessary (but not independently sufficient) to produce the injury, but insofar as (1) is seen as more abnormal, work in cognitive science would predict that it seen as more causal. Future work in “experimental jurisprudence” of causation would do well to consider how these cognitive scientific findings interact with legal contexts and decision making.¹⁷

3. Consent

As a third example, take the experimental jurisprudence study of *consent*. Consent plays a crucial role in legal areas including torts, criminal law, and contracts. Consent “turns a trespass into a dinner party; a battery into a handshake; a theft into a gift; an invasion of privacy into an intimate moment; a commercial appropriation of name and likeness into a biography.”¹⁸ But is the legal notion of consent consistent with the ordinary concept?

Roseanna Sommers has investigated this question across a range of legal contexts. In particular, this experimental work focuses on the relationship between deception and consent. Law often holds that material deception vitiates consent; when someone’s agreement is gained through deception about a material fact, there is not valid consent. For example, imagine that I offer to sell you a car with “only 10,000 miles” and you agree. In reality, the car has 100,000 miles. Your agreement would not be “consensual” if you relied upon my misrepresentation.

The experimental jurisprudence of consent has found, however, that ordinary people often attribute “consent” in circumstances in which there has been significant deception.¹⁹ For example, in one of Sommers’s cases, a single woman does not desire to sleep with married men; when she asks a potential partner about his marital status, he lies and says he is not married; and the woman agrees to sleep with him. In this case, the overwhelming majority of participants judged that the

¹⁷ See also Andrew Summers, *Common-Sense Causation in the Law*, 38 OXF. J. LEG. STUD. 793 (2018).

¹⁸ Heidi M. Hurd, *The Moral Magic of Consent*, 2 LEGAL THEORY 121, 123 (1996).

¹⁹ Roseanna Sommers, *Commonsense Consent* (draft manuscript) (on file with author); see also Roseanna Sommers & Vanessa K. Bohns, *The Voluntariness of Voluntary Consent: Consent Searches and the Psychology of Compliance*, 128 YALE L.J. (2019); Meirav Furth-Matzkin & Roseanna Sommers, *Consumer Psychology and Fine Print Fraud*, 72 STAN. L. REV. (forthcoming).

woman *did* “give consent to sleep with [the man].” Despite deception regarding a significant fact (the man’s marital status), people attribute consent. The same finding arises for various types of deception and various question types: did the woman “autonomously authorize [the man] to have sex with her,” did she “give valid consent to have sex in this situation,” or did she “give [the man] permission to have sex with her.”

This central finding characterizes people’s judgments of consent to sex, but also judgments about consent to medical treatment, consent to contracts, and consent to warrantless searches.²⁰ This experimental finding raises a number of interesting philosophical questions, most centrally: Should legal rules involving consent reflect this ordinary feature (i.e. the feature that consent is not necessarily vitiated by deception)?

4. Intentional Action

Another broad research program within experimental jurisprudence is the study of attributions of mental states. How do ordinary people understand what it means to do something *knowingly, recklessly, purposefully, or intentionally*?

This section considers one branch of this large research program: the study of when side-effects are produced “intentionally.”²¹ An important research program in cognitive science has shown that the ordinary concept of intentional action is surprisingly sensitive to normative concerns. For example, consider a case in which someone’s action produces a foreseeably bad side-effect. Imagine that a company’s chairperson is aware that his profit-maximizing strategy will also lead to some environmentally detrimental side-effects. Experimental psychology studies have suggested that, in such a case, people are more inclined to say that the chairperson *intentionally* harmed the environment when the side-effects are very severe (e.g. killing animals),

²⁰ Roseanna Sommers, *Commonsense Consent* (draft manuscript) (on file with author).

²¹ See, e.g., Joshua Knobe, *Intentional Action and Side Effects in Ordinary Language*, 63 ANALYSIS 190 (2003).

compared to when they are only moderately severe (e.g. reducing visibility).²²

Here there is an open empirical question about legal concepts. Namely, does the legal concept of acting *intentionally* share this feature (i.e. “severity-sensitivity”)? An experimental jurisprudence study on law students and judges suggests that the legal concept is importantly different.²³ Among other differences, legal attributions of intentional action were not sensitive to the severity of the outcome. For legal experts, a foreseeable side-effect is not evaluated to be more “intentional” simply because it is more severe.

This finding represents a proof of concept. IN at least some cases, the operational legal concept is different from the corresponding ordinary one. Experimental evidence suggests that the ordinary concept of intentional action is severity sensitive, but the legal concept is not. Of course, as before, this empirical finding does not settle philosophical debate about how law should conceptualize intentional action. In fact, the findings raise new and interesting questions. Are there legal contexts in which it is sensible for legal applications of intentionality to be severity sensitive? If judges and jurors operate with a different concept of intentional action, how should we resolve discrepancies in judgment?

5. Knowledge

A fifth example is the study of *knowledge*. As with many of the other research programs, the experimental work in this area is too vast to survey here.²⁴ This section considers one example of great legal significance: the distinction between knowledge and recklessness.

In an important experimental jurisprudential study, a team of law professors and cognitive scientists tested whether the knowledge-recklessness distinction can be mapped to activity in the

²² Markus Kneer & Sacha Bourgeois-Gironde, *Mens rea Ascription, Expertise and Outcome Effects: Professional Judges Surveyed*, 169 COGNITION 139 (2017).

²³ See, e.g., Kevin P. Tobia, *Legal Concepts and Legal Expertise*.

²⁴ See, e.g., Wesley Buckwalter, & Jonathan Schaffer, *Knowledge, Stakes, and Mistakes*, 49 NOUS 201 (2015); Wesley Buckwalter, David Rose & John Turri, *Belief Through Thick and Thin* 49 NOUS 748 (2015); David Rose & Jonathan Schaffer, *Knowledge Entails Dispositional Belief*, 166 PHIL. STUD. 19 (2013); Ori Friedman & John Turri, *Is Probabilistic Evidence a Source of Knowledge*, 39 COG. SCI. 1062 (2015); Joshua Knobe, *Intentional Action and Side Effects in Ordinary Language*, 63 ANALYSIS 190 (2003).

human brain.²⁵ They presented participants with different versions of a case in which someone transports a suitcase full of “contraband” across international borders, with varying levels of awareness of the presence of contraband in the suitcase. Participants were shown between one and five suitcases, only one of which contained “contraband,” and were asked whether they would carry a suitcase randomly chosen from the group. Trials with only one suitcase represented trials in which the person had *knowledge*, while trials with five suitcase represented trials in which the person did not.

Using brain-imaging (fMRI) data from the participants, the research team was able to predict with reasonable accuracy whether the brain-imaging data corresponded to a knowing choice (e.g. in cases in which participants were shown one suitcase) or a reckless one (e.g. in cases in which participants were shown four or five suitcases). This study provides important evidence about the concepts of *knowledge* and *recklessness*. For example, it provides evidence that these are distinct concepts, whose representations are associated with different patterns of neural activity. Insofar as law predicates differential punishments on juror’s (i.e. ordinary people’s) ability to distinguish between these two states, it is important that there is a such a distinction.

6. Ownership

As a sixth example, consider experimental findings on the concept of *ownership*, a central concept of property law. Experimental psychologists have argued that children have a rich theory of ownership,²⁶ which may develop into a folk theory of legal ownership.²⁷

As one example of an experimental jurisprudence study of ownership, consider a recent paper that examined this ordinary concept. Experimentalists studied how people evaluate “finder versus landowner cases,” cases in which one person finds something on someone else’s land. They

²⁵ Iris Vilares et al., *Predicting the Knowledge-Recklessness Distinction in the Human Brain*, 114 PROC. NAT. ACA. SCI. 3222 (2017).

²⁶ Shaylene E. Nancekivell, Ori Friedman & Susan A. Gelman, *Ownership Matters: People Possess a Naïve Theory of Ownership*, 23 TRENDS IN COG. SCI. 102 (2019).

²⁷ Ori Friedman, Madison L. Pesowski & Brandon W. Goulding, *Legal Ownership Is Psychological: Evidence from Young Children*, in PSYCHOLOGICAL OWNERSHIP AND CONSUMER BEHAVIOR 19 (2018).

found that people are more inclined to view the finder as the owner when the found object was in a public (rather than private) space.²⁸

7. Responsibility and the Self

This seventh example departs from the first six in an important way. Legal rules and standards often make explicit reference to reasonableness, causation, consent, intention, knowledge, and ownership. Although “the self” is a deep philosophical concept, the term does not frequently manifest in legal texts or opinions.

Nevertheless, experimental jurisprudential study of this concept has advanced legal scholarship in a surprising way. Drawing on a large literature in experimental psychology of the self,²⁹ Christian Mott investigated whether judgments about the self might explain the intuitiveness of criminal punishment or statutes of limitations.³⁰ Broadly speaking, the hypothesis was that what makes statutes of limitations seem appropriate is something to do with our sense that people change over time.

To test this hypothesis, Mott presented participants with a scenario in which a person recklessly drives drunk. In eleven different versions of the scenario, police find a security camera video showing the person’s action and are able to identify him. The key difference is that in the first version, police find the video *the next day*, in the second, the police find the video *three years later*; in the third, *six years later*; in the fourth, *nine years later . . .*; and in the last, *thirty years later*. Participants evaluated whether it is “right” for the person to be “arrested and

²⁸ Peter De Scioli, Rachel Karpoff & Julian De Freitas, *Ownership Dilemmas: The Case of Finders Versus Landowners*, 41 COG. SCI. 502 (2017).

²⁹ See, e.g., Dan Bartels & Lance Rips, *Psychological Connectedness and Intertemporal Choice*, 139 J. EXP. PSYCHOL. 49 (2010); Dan Bartels & Oleg Urminsky, *On Intertemporal Selfishness: The Perceived Instability of Identity Underlies Impatient Consumption*, 39 J. CONSUMER RES. 182 (2011); Brian Earp et al., *Addiction, Identity, Morality* (2018); Hal Ersner-Herschfield et al., *Don’t Stop Thinking About Tomorrow: Individual Differences in Future Self-Continuity Account for Saving*, 4 JUDGMENT & DECISION MAKING 280 (2009); Sarah Molouki, Dan Bartels & Oleg Urminsky, *A Longitudinal Study of Difference between Predicted, Actual, and Remembered Personal Change*, PROC. COG. SCI. SOC. 2748; Sarah Molouki & Daniel Bartels, *Personal Change and the Continuity of the Self*, 93 COG. PSYCHOL. 1 (2017); George Newman, Paul Bloom & Joshua Knobe, *Value Judgments and the True Self*, 40 PERS. & SOC. PSYCHOL. BULL. 203 (2014); Nina Strohminger & Shaun Nichols, *The Essential Moral Self*, 131 COGNITION 151 (2014); Kevin P. Tobia, *Personal Identity and the Phineas Gage Effect*, 75 ANALYSIS 396 (2015); Kevin P. Tobia, *Personal Identity, Direction of Change, and Neuroethics*, 9 NEUROETHICS 37 (2016). See generally David Shoemaker & Kevin P. Tobia, *Personal Identity*, in OXFORD HANDBOOK OF MORAL PSYCHOLOGY (forthcoming).

³⁰ Christian Mott, *Statutes of Limitations and Personal Identity*, in 2 OXFORD STUDIES IN EXPERIMENTAL PHILOSOPHY 243 (2018).

punished” immediately after the video is found.

Unsurprisingly, agreement that the man should be arrested and punished decreased with the years passing. Participants evaluated that the man was most deserving of punishment when the video was found the next day, and agreement decreased as more time elapsed between the incident and time of arrest.

Mott also asked participants to evaluate three other questions. First was a “deterrence” question regarding whether the man would have been less likely to drive drunk if he knew he would be arrested and punished [the next day, three years later ... thirty years later]. Second was a “repose” question regarding whether the man has the right to be able to stop worrying about being arrested and punished [the next day, three years later, ... thirty years later]. Third was a connectedness question about whether the man who recklessly drove was “completely connected” or “completely disconnected” from the person he is when the police find the video [the next day, three years later, ... thirty years later]. A mediation analysis found that the affect of time on punishment was explained by connectedness and repose but not deterrence.

This is a very striking finding. Although many would think the ordinary concept of statutes of limitations is linked to motivations for deterrence, Mott’s findings suggest that it is more closely connected to concerns about repose and *the self*. In this case, this finding raises a largely overlooked legal question. Should legal punishment take into consideration intertemporal connectedness? For example, is someone less deserving of punishment for a crime that they committed after they have undergone large personal change? As in the other examples, the experimental findings will not directly resolve these questions about punishment and responsibility. But they provide new insight into these topics and suggest questions that legal philosophy should be asking.

8. The Concept of Law

As an eighth example, consider the concept of *law* itself. Raff Donelson and Ivar Hannikainen examined how ordinary people understand law. They tested whether ordinary people and legal experts endorsed Fuller's conditions of the inner morality of law.³¹ For example, do people think that law has to be consistent, general, intelligible, public, and stable?

Their study investigated two questions. First, are these conditions of law seen as necessary to law? Second, do laws in practice observe these principles? The results—from ordinary people and experts alike—are intriguing. There was not overwhelming endorsement of these principles of law. That is, there was not overwhelming support that law is or must be prospective, stable, intelligible, and general. Although these features are common in legal-philosophical accounts of law, the experimental studies suggest that ordinary people are not strongly inclined to understand law as requiring or even involving these features.

9. Natural Kind Concepts

This next example differs more dramatically from the former ones. The first eight examples study specific concepts like *consent* or *causation*. But experimentalists have also studied broader classes of concepts. One noteworthy example is the class of “natural kind concepts,” concepts including *water*, *gold*, *tomato*, *milk*, and *salmon*.

At first it might seem that these concepts are not so legally relevant. However, such an impression is mistaken. Natural kind concepts form a broad class, central to food, health and animal law, as well as other legal areas—even including tax law. As an example, recall the famous case of *Nix v. Hedden*,³² in which the effect of a tax statute turned on the legal interpretation of the words “fruit” and “vegetable.” In that case, a statute required a duty to be paid on imported vegetables, but not imported fruits. A dispute arose over whether tomatoes are

³¹ Raff Donelson & Ivar Hannikainen, *Fuller and the Folk: The Inner Morality of Law Revisited*, in 3 OXFORD STUDIES IN EXPERIMENTAL PHILOSOPHY (2019).

³² 149 U.S. 304 (1893).

duty-incurring vegetables or duty-free fruits. Although tomatoes are botanically fruits, the court held that they are duty-incurring vegetables under the tariff act.³³

The cognitive science of natural kind concepts helps illuminate this decision and the tension that many feel when considering the facts of *Nix*. Work in experimental psychology has found that people can “dual-characterize” natural kind concepts, applying two distinct sets of category membership to the same concept.³⁴ When we think about a tomato, there is a sense in which it is a fruit (namely, a scientific/botanical sense), but there is also a sense in which it is not a fruit (namely, the customary sense). This same dual characterization may lead us to ambivalent judgments about other entities. Genetically modified salmon is really salmon in one sense, but also not salmon in another sense. Similarly, synthetic meat is really meat in one sense, but also not really meat in another.

Here again, the experimental findings do not themselves tell us how law *should* treat natural kinds. Perhaps only one sense of the ordinary concept should be legally relevant, or perhaps both should be. The broader lesson from this section is that experimental jurisprudence makes progress by studying not only individual concepts (e.g. *causation*) but also broader classes of concepts.

10. Legal Interpretation

A final area of experimental jurisprudence is study of legal interpretation. In legal interpretation—of statutes, contracts, and constitutions—courts often look to the “ordinary,” “public,” or “plain” meaning of the legal text. There are important debates and disagreements about how to elaborate these concepts.³⁵ Yet, on a number of plausible views, these types of meaning are closely linked to facts about how ordinary people understand terms and texts.

It is such posited relationships—between ordinary meaning and facts about actual understandings and concepts—that open several areas of experimental research. For example,

³³ *Id.*

³⁴ Kevin P. Tobia, George E. Newman & Joshua Knobe, *Water Is and Is Not H₂O*, MIND & LANGUAGE (2019); George E. Newman & Joshua Knobe, *The Essence of Essentialism*, MIND & LANGUAGE (2019).

³⁵ See generally Richard H. Fallon, Jr., *The Meaning of Legal “Meaning” and Its Implications for Theories of Legal Interpretation*, 82 U. CHI. L. REV. (2015).

some experimentalists have used survey methods to help resolve contract disputes.³⁶ Insofar as the “ordinary” objective meaning of a contract is relevant to adjudicating disagreement, experimental study about ordinary judgments are highly probative.

A similar approach used surveys to understand ambiguity.³⁷ By assessing levels of ordinary agreement and disagreement with various possible meanings, we can provide a more empirically grounded account of whether some text has one or two possible meanings.

A final example used experiments to assess *other methods* of legal interpretation.³⁸ For example, by comparing ordinary conceptual judgments (e.g. how people evaluate what is a “vehicle”) with dictionary-use (e.g. how judges use dictionaries), we can learn whether dictionary-use actually tracks facts about how ordinary people understand terms and phrases.

B. Themes

The previous Section offers a representative, although certainly not exhaustive, list of research programs in experimental jurisprudence. This Section considers some of the broader, cross-cutting themes that unify these different programs. First, it considers some of the successes of experimental jurisprudence: enriching conceptual analysis of legal concepts, testing legal intuitions, providing evidence about legal debates, and illuminating new theoretical possibilities. Next, it reflects on some common themes across these various projects: engagement with existing work in cognitive science, focus on both “classic” debates in legal theory, emphasis on distinctive legal concepts, and investigation of the relationship between ordinary and “expert” legal concepts.

A common feature of experimental jurisprudence is the advancement of legal “conceptual

³⁶ Omri Ben-Shahar & Lior Jacob Strahilevitz, *Interpreting Contracts Via Surveys and Experiments*, 92 N.Y.U. L. REV. 1753 (2017).

³⁷ Ward Farnsworth, Dustin F. Guzior & Anup Malani, *Ambiguity about Ambiguity: An Empirical Inquiry into Legal Interpretation*, 2 J. LEG. ANALYSIS 257 (2010).

³⁸ Kevin P. Tobia, *Testing Original Public Meaning: Are Dictionaries and Corpus Linguistics Reliable Measures of Meaning?*

analysis.” Legal philosophers have long used the method of “conceptual analysis” to make progress in legal theory.³⁹ The most well known example is the effort to articulate the concept of law.⁴⁰ But conceptual analysis targets many other legal concepts, such as the concept of punishment⁴¹ and the concept of property.⁴²

As its name suggests, conceptual analysis is the method of analyzing concepts. Legal philosophers reflect on legal concepts—*reasonableness*, *causation*, *consent*, and so on—and attempt to articulate the features of these concepts. In legal philosophy, as in philosophy more broadly, thought experimentation is a common and popular tool in conceptual analysis. Philosophers elicit intuitions about cases, and those intuitions are taken to provide evidence about the target concept.

Some have expressed skepticism about philosophical conceptual analysis. Nevertheless, it remains a popular methodological tool in philosophy of law. As Langlinais and Leiter put it, in “many areas of philosophy, doubts about the kind of conceptual and linguistic analysis . . . have become common [But] not so in legal philosophy.”⁴³

This essay does not aim to justify conceptual analysis as a legal-philosophical methodology. Instead, I note that experimental jurisprudence offers unique insights for loyal practitioners of conceptual analysis. Importantly, there are different varieties of legal conceptual analysis and experimental jurisprudence may not assist all of them. In particular, experimental jurisprudence can assist conceptual analysis that attempts to articulate empirically or descriptively grounded concepts.⁴⁴

So what is this method? On one common understanding, conceptual analysis attempts to

³⁹ See, e.g., Brian Bix, *Conceptual Questions and Jurisprudence*, 1 LEG. THEO. 465 (1995); Aaron J. Rappaport, *On the Conceptual Confusions of Jurisprudence*, 7 WASH. U. JURIS. REV. 77 (2014).

⁴⁰ H.L.A. HART, *THE CONCEPT OF LAW* 1 (1961).

⁴¹ E.g. GEORGE P. FLETCHER, *BASIC CONCEPTS OF CRIMINAL LAW* (1998).

⁴² E.g. 2 JOHN AUSTIN, *LECTURES ON JURISPRUDENCE OR THE PHILOSOPHY OF POSITIVE LAW* 795-99 (1885); Felix S. Cohen, *Dialogue on Private Property*, 9 RUTGERS L. REV. 357 (1954).

⁴³ Alex Langlinais & Brian Leiter, *The Methodology of Legal Philosophy*, in *THE OXFORD HANDBOOK OF PHILOSOPHICAL METHODOLOGY*.

⁴⁴ Experimental jurisprudence would not be able to assist a (quite radical) form of conceptual analysis that holds that there is no relationship between facts about ordinary language or ordinary cognition and philosophical concepts.

identify and articulate the structure of a concept. Typically, this involves breaking down the target concept into smaller components. Often, philosophers employ hypothetical test cases or “thought experiments” to test whether the proposed analysis is successful.

For example, consider the concept of legal reasonableness. A conceptual analysis of reasonableness might proceed by reflecting on seemingly important properties of the concept. It might seem that an important feature of reasonable care is that it leads to good consequences. One might suggest that reasonable care is care that leads to the welfare-maximizing results.

In response to this view, consider a critical thought experiment that questions the analysis:

Life-Saving Negligence: Imagine that a company produces a unique life-saving medicine. The company decides to maximize their production, at any cost, in order to save the most lives with their medicine. They knowingly produce their medicine to a point at which the factory ends up emitting dangerous levels of pollutants. The pollutants kill a small number of people each year, but the increase in medicine production saves ten times as many people.

Even if this decision ultimately saves more lives (adjusting for life-years lost from the pollution), it seems—*intuitively*—that the company acts without reasonable care.

The reaction to this thought experiment represents something like a philosophical discovery. Insofar as we place significant weight on that response, we might conclude that the proposed conceptual analysis needs revision. Reasonable care is not simply welfare-maximizing care, and we should refine the analysis, test that revision with more cases, refine the analysis in light of those, and so on.

Experimental jurisprudence analyzes legal concepts in a broadly similar way to traditional conceptual analysis. It poses vignettes (like thought experiments) that elicit judgments about legal concepts. Then, it draws conceptual conclusions from people’s responses to those vignettes.

A second contribution of experimental jurisprudence is its ability to test legal intuitions. Appeals to intuition are central to many areas of philosophy, including legal philosophy, which

“relies on two central argumentative devices—analyses of concepts and appeals to intuition.”⁴⁵

The Life-Saving Negligence case is a thought experiment that seems to elicit a common response (e.g. the company’s care is not reasonable). On a common account, this response is an “intuition.” A legal philosopher describes some scenario (actual or hypothetical) and invites readers to consider some question about the scenario: does the care seem *reasonable*, which action seems like the *cause*, is that rule a *legal rule*, and so on.

Classic thought experiments and corresponding intuitions in legal theory include Holmes’s Bad Man,⁴⁶ the criminality of Fuller’s Speluncean Explorers,⁴⁷ excuses for Kadish’s Mr. Fact and Mr. Law,⁴⁸ and the legal effect of hypothetical rules like “no vehicles in the park”⁴⁹ or “it is a misdemeanor to sleep in any railway station.”⁵⁰

What role “intuitions” serve in legal philosophy presents a complicated question. One conception of the role of intuition involves conceptual analysis. Intuitions provide insight in the features of the target concept. But on a radically different view, intuitions are not merely related to concepts. Instead, they are related to something objective and extra-mental. Consider Ernest Sosa’s view of the role of philosophical intuition:

It is often claimed that analytic philosophy appeals to armchair intuitions in the service of “conceptual analysis.” But this is deplorably misleading. The use of intuitions in philosophy should not be tied exclusively to conceptual analysis. . . . [Questions in ethics or epistemology] concern an ethical or epistemic subject matter, and not just our corresponding concepts.⁵¹

Some legal philosophers hold a similar view of legal intuitions. Legal philosophy is not merely investigating people’s *concept* of (e.g.) causation or battery; it is attempting to discover the true nature of those properties and relations. Thought experiments about these topics elicit judgments that illuminate more than just the target concept; the contents of the judgments provide insight into the *object* of study itself. On this proposal, intuitions have some connection to truths about

⁴⁵ Brian Leiter, *Beyond the Hart/Dworkin Debate: The Methodology Problem in Jurisprudence*, 48 AM. J. JURIS. 17, 43-44 (2003).

⁴⁶ Oliver W. Holmes, *The Path of Law*, 10 HARV. L. REV. 457 (1897).

⁴⁷ Lon L. Fuller, *The Case of the Speluncean Explorers*, HARV. L. REV. (1949).

⁴⁸ MONRAD PAULSEN & SANDY KADISH, CRIMINAL LAW 485-86 (1962).

⁴⁹ See generally Frederick Schauer, *A Critical Guide to Vehicles in the Park*, 83 N.Y.U. L. REV. 1109 (2008).

⁵⁰ Lon Fuller, *Positivism and Fidelity to Law: A Reply to Professor Hart*, 71 HARV. L. REV. 630 (1958).

⁵¹ Ernest Sosa, *Experimental Philosophy and Philosophical Intuition*, 132 PHIL. STUD. 99, 100 (2007).

the actual nature of the object (e.g. causation, obligation, law, battery, etc.).

This view raises two natural questions. First, what exactly is the relationship between intuition and these truths? One might argue that intuition is a source of evidence about legal truths; when we consider hypotheticals about (e.g.) law or battery, our intuitions provide evidence about the nature of those things. Here again, this essay simply aims to defend experimental jurisprudence's potential contributions within various existing frameworks. Thus, I will not adopt or defend any particular conception of legal intuition. Perhaps intuition provides *evidence* about legal truths, or perhaps it provides *knowledge* of legal truths, or perhaps there is some other account that better explains the role of legal intuition.

On any of these accounts, the second question arises: Does the posited relationship between intuition and truth actually exist? This motivates a host of experimental jurisprudential questions inspired by experimental philosophy's "negative program." Experiments about the factors that affect intuitions can provide evidence about the epistemological status of intuition.

In the context of experimental philosophy, "negative program" research has uncovered the influence of a number of "irrelevant factors" on intuition. For instance, intuitions vary depending on the order of cases presented⁵² or features of the person considering the thought experiments—such as personality⁵³ or native language.⁵⁴ These findings problematize many popular accounts of philosophical intuition. For instance, some accounts posit that intuition is a *reliable* guide to truths about philosophical concepts, a guide that is sensitive to *truth-relevant* factors.⁵⁵ Insofar as intuitions are in fact influenced by other (irrelevant) factors, this challenges claims of intuition's reliability.

Experimental jurisprudence offers a similar contribution in testing legal intuitions. Legal

⁵² Eric Schwitzgebel & Fiery Cushman, *Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers*, 27 MIND & LANG. 135 (2012).

⁵³ Adam Feltz & Edward T. Cokely, *Do Judgments About Freedom and Responsibility Depend on Who You Are? Personality Differences in Intuitions About Compatibilism and Incompatibilism*, 18 CONSCIOUSNESS & COGNITION 342 (2009).

⁵⁴ Krist Vaesen, Martin Peterson & Bart Van Bezooijen, *The Reliability of Armchair Intuitions*, 44 METAPHILOSOPHY 559 (2013).

⁵⁵ E.g. Alvin Goldman, *Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status.* 74 GRAZER PHILOSOPHISCHE STUDIEN I (2007).

theory's methodological assumptions underpinning appeals to intuition are often ones that make empirical claims. Experimental jurisprudence can provide evidence about the plausibility and veracity of those assumptions.

A third major contribution of experimental jurisprudence is its ability to offer evidence bearing on competing legal theories. Many legal-philosophical theories make empirical predictions. Experimental jurisprudence can advance debates about those topics by testing empirical claims.

We have already considered one such example. Recall the experimental work suggesting that reasonableness is a hybrid concept. A number of legal theorists argue that the right normative account of reasonableness should reflect the ordinary concept of reasonableness. This is a feature of legal philosophy that is different from most non-legal philosophy. For example, most ethicists would not revise their theory of moral obligation if the account supported by the best arguments failed to reflect the ordinary concept. But in legal philosophy, the ordinary concept is often more significant. For example, some argue that central tort law notions (such as reasonableness) should reflect the ordinary concept since tort law's central (normative) motivation is the reproduction of community standards and norms.⁵⁶

For those theorists, there is a straightforward bridge between the ordinary concept and normative theorizing. Evidence that reasonableness is understood as a hybrid concept supports *normative* theories of reasonableness as a hybrid notion and counts against theories positing (e.g.) that reasonableness is a purely normative notion.

A final contribution of experimental jurisprudence is the ability to illuminate new possibilities in legal philosophy—possibilities that were in some way previously overlooked. Return to the example of reasonableness.⁵⁷ Before that study, some theorists had briefly noted

⁵⁶ E.g. Cristina Carmody Tilley, *Tort Law Inside Out*, 126 *YALE L.J.* 1320 (2017).

⁵⁷ See Tobia, *supra* note 7.

that reasonableness might be a hybrid notion.⁵⁸ But hybrid interpretations were far less developed than the various competing statistical and prescriptivist accounts. There are *many* well-developed prescriptivist views. For example, some have argued that reasonableness is grounded in an economic cost-benefit analysis,⁵⁹ community values,⁶⁰ “justification”,⁶¹ virtue ethics,⁶² a Kantian notion of equal freedom,⁶³ or an ethic of care.⁶⁴ Although some scholars briefly note support for a hybrid view, those views have not been thoroughly elaborated. This may be a result of there being seemingly little to say. Once we note that reasonableness is a statistical and prescriptive blend, how else can this be elaborated? Experimental methods offered a way to further advance this theory. Research in cognitive science provided more sophisticated ways to understand hybrid concepts.⁶⁵ This inspired the study that discovered both the descriptive aptness of hybrid accounts (i.e. ordinary judgment of reasonableness is best understood as a hybrid judgment) and allowed elaboration of those accounts as new (normative) theories of reasonableness.

Experimental jurisprudence plays a similar role in many of the other opening examples. For instance, recall the study of sentencing and the concept of the self. There were several plausible accounts regarding the justification of statutes of limitations (e.g. those focusing on deteriorating evidence quality), but few of these posited a significant role of *the self*. Experimental studies made plausible the possibility that judgments of responsibility and sentencing might be influenced by views of the self: people find statutes of limitations intuitive because temporally latter persons seem less connected to the actor who committed the legal violation.

Moving beyond these common roles of experimental jurisprudence—enhancing conceptual analysis, testing legal intuitions, providing evidence in legal debates, and illuminating new theoretical possibilities—consider some common themes within these projects.

⁵⁸ E.g. Benjamin C. Zipursky, *Reasonableness in and out of Negligence Law*, 163 U. PA. L. REV. 2131 (2015).

⁵⁹ RICHARD POSNER, *ECONOMIC ANALYSIS OF LAW* (2002).

⁶⁰ See Tilley, *supra* note 56.

⁶¹ John Gardner, *The Mysterious Case of the Reasonable Person*, 51 U. TORONTO L.J. 273 (2001).

⁶² Heidi Li Feldman, *Prudence, Benevolence, and Negligence: Virtue Ethics and Tort Law*, 74 CHI.-KENT L. REV. 1431 (2000).

⁶³ Alan D. Miller & Ronen Perry, *The Reasonable Person*, 87 N.Y.U. L. REV. 323 (2012).

⁶⁴ *Id.*; Leslic Bender, *Feminist (Re)Torts: Thoughts on the Liability Crisis, Mass Torts, Power, and Responsibilities*, 1990 DUKE L.J. 848.

⁶⁵ See, e.g., Adam Bear & Joshua Knobe, *Normality: Part Descriptive and Part Prescriptive*, 167 COGNITION 25 (2017).

One common feature is that experimental jurisprudence draws heavily from existing work in cognitive science. In each of the examples, cognitive science informed the methods, hypothesis, or interpretation of the experimental investigations of legal concepts. The studies of reasonableness draw on the cognitive science of hybrid concepts,⁶⁶ work on legal causation draws on existing findings cognitive science,⁶⁷ studies on legal responsibility draw on the experimental philosophy and psychology of the self,⁶⁸ and studies of interpretation draw on experimental psychological work in prototype theory.⁶⁹

This feature is not surprising. Experimental jurisprudence is the experimental-philosophical study of legal concepts, and many of those ordinary concepts have been studied thoroughly in cognitive science (e.g. *causation*). And for those concepts that cognitive science has not studied (e.g. *reasonableness*; *law*), there is other useful existing work that might motivate methods or hypothesis about the concept.

A second common feature is that experimental jurisprudential work tends to study “classic” and central legal topics and debates. Experimentalists have studied the concept of law, the nature of causation, who is the “reasonable person,” and debates about legal interpretation and ordinary meaning. This might be, in part, a feature of the novelty of the field. The approach is one that might target a range of legal concepts, and experimentalists have started with the most significant and central ones. But it is important to note that although experimental jurisprudence represents an innovative approach to legal philosophy, it is typically complementary to traditional methods and debates.

⁶⁶ See, e.g., Bear & Knobe, *supra* note 9.

⁶⁷ See, e.g., Knobe & Shapiro, *supra* note 14.

⁶⁸ See, e.g., sources cited *supra* note 29.

⁶⁹ See, e.g., Tobia, *supra* note 38.

A final common theme is the focus on expert and ordinary concepts. A number of experimental jurisprudential projects study expert judgments (e.g. those of law students, law alumni, or judges).⁷⁰ Many others contrast findings about the ordinary concept with questions about a distinct, specialized concept.⁷¹ This question—what is the ordinary concept and (how) should law reflect it—is perhaps the best example of a question that characterizes every experimental jurisprudential project. Whether experimentalists study causation, consent, intentional action, or reasonableness, they make discoveries about the ordinary concept, raising the natural question of whether law should reflect that ordinary concept. Experimental discoveries about ordinary concepts also open the gate to more sophisticated questions: Upon learning that the ordinary concept (e.g. of causation) has feature *x*, we ask: Should the legal concept reflect *that* feature? The next Part introduces a framework to address this type of question.

II. A BROADER FRAMEWORK

This Part reflects on the case studies of Part I, and develops a broader framework of experimental jurisprudence. When experiments uncover facts about concepts, these discoveries can provide new reasons for philosophical arguments about what concept law *should* use. This Part considers three broad classes of such discoveries.

First, when experiments discover facts about ordinary concepts (e.g. the ordinary concept of *reasonableness* or *causation*), this often supplies *prima facie* reasons for law to use legal concepts that share the features of the ordinary ones. For example, if we discover that the ordinary concept of reasonableness reflects statistical norms, this provides some *prima facie* reasons for legal reasonableness standards to also reflect such norms. One such reason is grounded in publicity. The condition that laws should be public provides a reason for law to use ordinary concepts rather than specialized (and non-public) legal ones.

At the same time, experimental discoveries could provide one type of *prima facie* reason for

⁷⁰ See, e.g., Sommers, *supra* 4; Kneer & Bourgeois-Gironde, *supra* note 22; Tobia, *supra* note 23.

⁷¹ See, e.g., Knobe & Shapiro, *supra* note 14.

law to *not* employ an ordinary concept. Such reasons stem from considerations about the distinctiveness of law; legal concepts should be responsive to particular legal needs—not ordinary, conventional, moral, or other ones. Upon the discovery that law uses an ordinary concept, one might think that this is a mistake; law has adopted an ordinary concept where it should employ one that is distinctive for law.

Weighing these two sets of *prima facie* reasons might suggest that, in many circumstances, law should use ordinary concepts. However, there is a final set of considerations—one that is perhaps the most important. Experimental discoveries often illuminate specific features of concepts (ordinary or specialized ones). Considering the nature, purposes, and precedents of the relevant area of law can provide reasons about whether those features are suitable for the legal concept. That is, once we learn that an ordinary concept has feature *x*, there are often considerations generated by considering the relevant area of law in which the concept would be used. Beyond general *prima facie* considerations about publicity or legal distinctiveness, there is a crucial set of considerations about which features (if any) of the concept are *legally apt*.

Importantly, any consideration may carry differential weight depending on the circumstance. For example, one might argue that publicity is especially important in constitutional law, but less so in the law of civil procedure. If so, *prima facie* publicity reasons are stronger in the former context. As another example, perhaps general deterrence is a more significant aim in criminal law than in contracts; if so, the communicativity of ordinary concepts provides a stronger reason to use an ordinary concept in a criminal law setting compared to a contracts one.

A. Prima Facie Reasons for Law to Use Ordinary Concepts

There are a number of considerations that often supply *prima facie* reasons for law to use ordinary concepts. For example, that law should be *public* suggests that ordinary people's conception of law should match law's effects; and that law should be *consistent* suggests that a judge's or jury's application of a legal standard should not differ with respect to the same facts.

This picture of ordinary concepts underpins diverse theoretical accounts of legal legitimacy and authority. “Rule of law” theorists often recommend conditions of law including publicity, neutrality, consistency, clarity, prospectivity, coherence, and stability.⁷² Many of these conditions presuppose that, broadly speaking, legal concepts are *accessible* to ordinary people. Insofar as ordinary concepts achieve such accessibility, these considerations provide reasons for law to use ordinary concepts. This Section considers these considerations in more detail.

First, take publicity. On many accounts, law should be public, accessible to the citizens it binds. If law can use either an ordinary concept or a distinctive non-ordinary one, consideration of publicity typically provides a *prima facie* reason to use the ordinary one. In most cases, an ordinary concept is accessible to a larger population than a corresponding distinctive concept.

For example, if the ordinary concept of consent is distinct from the legal concept, this problematizes the publicity condition. An ordinary person might think that he has gained the consent of another (e.g. in a contract) despite his deception, while the legal concept would not apply. The typical motivating values of publicity, like fairness and accessibility, reflect concern for public *comprehension* of law.

Similarly, consider clarity. If law should be clear, this provides a *prima facie* reason to use ordinary concepts or provide sufficient clear explanation of the legal concept to the relevant public. Unless the latter condition is obtained, consideration of clarity provides a reason to use ordinary concepts.

Third, consider consistency. Law should be consistent to promote reliance values (e.g. ensuring that citizens manage their plans and expectations) and egalitarianism (e.g. fairness). Using ordinary concepts likely leads to greater consistency. Consider, for example, that a judge or jury might apply the same rule or statute in different cases. If jury instructions elaborate the

⁷² See generally John Tasioulas, *The Rule of Law*, in THE CAMBRIDGE COMPANION TO THE PHILOSOPHY OF LAW (Cambridge University Press, 2019); see also R.M. Dworkin, A MATTER OF PRINCIPLE (Harvard University Press, 1985); Lon L. Fuller, THE MORALITY OF LAW (Yale University Press, 1965); Margaret Jane Radin, *Reconsidering the Rule of Law*, BOS. U. L. REV. 69 (1989); Joseph Raz, *The Rule of Law and its Virtue*, in THE AUTHORITY OF LAW (Oxford University Press, 1979).

features of a distinctive concept, there may be reduced differences between judge and jury verdicts. But where no instruction is provided and experts use a (learned) distinctive concept, it is not clear how ordinary jury members could arrive at the same concept. Where the concept application is outcome-determinative, this leads to inconsistencies. Again, concern for this value favors using ordinary concepts.

Fourth, consider the value of legal coherence. If law uses exclusively an ordinary concept or exclusively a distinctive concept, there is no coherence issue. However, any mixture raises potential problems of coherence. Insofar as it is more feasible to ensure that experts (e.g. judges) and ordinary people (e.g. jurors; those who are placed on notice by law) all understand ordinary concepts, rather than distinctive legal ones, this provides another reason to use ordinary concepts.

Fifth, consider values like legal practicability, predictability, and stability. It must be possible for law to be obeyed and for people to plan with awareness of (stable) legal rules and obligations. Ordinary concepts, which are accessible to ordinary people, likely foster greater practicability, predictability, and stability than distinctive legal concepts.

It is important to note that all of these are *prima facie* reasons, ones that generally—but do not always apply—in the analysis of a legal concept. For example, it is typically the case that law should be stable. But there might be some circumstances in which instability is actually preferred. In that circumstance, the reason is not just outweighed by other considerations, but rather rejected entirely. Of course, in most circumstances, law should be public, clear, consistent, coherent, and so on.

This Section has categorized several reasons that often count in favor of using ordinary concepts. It has articulated these reasons in terms of rule of law values, like publicity and consistency, which are endorsed by a very diverse group of legal philosophers.⁷³ But we might also consider different framings of similar considerations.

⁷³ See, e.g., JULES COLEMAN, *THE PRACTICE OF PRINCIPLE* (2001); Timothy Endicott, *The Impossibility of the Rule of Law*, 19 OXF. J. LEG. STUD. 1 (1999); LON FULLER, *THE MORALITY OF LAW* (1969); John Gardner, *The Legality of Law*, 17 *RATIO JURIS* 168 (2004); Andre Marmor, *The Rule of Law and its Limits*, 23 *L. & PHIL.* 1 (2004); Joseph Raz, *The Rule of Law and its Virtue*, in *THE AUTHORITY OF LAW* (1979).

One is to consider the general purposes of law—notification, guidance, deterrence, and expression. Many legal theories propose that laws are sources of notification, guidance, or deterrence for *ordinary* citizens. Other views propose that laws must carry *expressive* content, to ordinary citizens. Even those without legal training should be able to understand a legal rule that prohibits intentional killing—to be notified, guided, or deterred by the rule; or to appreciate the expressive content conveyed by punishing a violation of the rule. Such understanding includes, presumably, the law’s concept of *intentional action*. These purposes are typically best promoted by using ordinary concepts.

Another is to consider “democratic” views of law. On this view, law should reflect ordinary judgments and concepts, to allow citizens to have democratic input into the legal system.⁷⁴ This view is most associated with the democratic view of criminal law,⁷⁵ but it might supply reasons more broadly. Insofar as these democratic considerations are relevant, they would tend to provide reasons in favor of using ordinary concepts—those accessible to the demos.

B. Prima Facie Reasons for Law to Use Specialized Concepts

In some circumstances, experimental discoveries also supply one type of *prima facie* reason *against* law employing an ordinary concept. Such reasons are generated from considerations about the distinctiveness of law. Legal concepts should be responsive to particular *legal* needs, not ordinary, conventional, moral, or other ones. When an experimentalist discovers that law might use an ordinary concept, we might argue that this is a mistake. In such a case, law has adopted an ordinary concept where it should use one that is distinctive.

Consider, for example, the findings about intentional action.⁷⁶ Experimentalists have provided some evidence about a certain features of the ordinary concept: that attributions of

⁷⁴ See, e.g., PAUL H. ROBINSON & JOHN M. DARLEY, *JUSTICE, LIABILITY & BLAME: COMMUNITY VIEWS AND THE CRIMINAL LAW* (1995); Paul H. Robinson, *Democratizing Criminal Law: Feasibility, Utility, and the Challenge of Social Change*, 111 NW. U. L. REV. 1565 (2017).

⁷⁵ See, e.g., Joshua Kleinfeld, *Manifesto of Democratic Criminal Justice*, 111 NW. U. L. REV. 1367 (2017).

⁷⁶ Section I.A.4 *supra*.

“intentionally” producing a side-effect are sensitive to the severity of the effect. In such a case, we might argue that *insofar as law uses a concept with this feature*, this would be a mistake. This peculiar feature of the ordinary concept should not be mistakenly incorporated into a legal context.

Of course, it may seem that this type of argument must be restricted in some way. If there is a *prima facie* reason for a legal concept (e.g. law’s concept of intentional action) to exclude *any* feature of the ordinary concept (e.g. any feature of the ordinary concept of intentional action), we may soon find ourselves concluding that the legal concept should bear no resemblance to the ordinary one. Most will find this conclusion unacceptable. Any legal concept that corresponds to an ordinary term should bear some resemblance to the ordinary concept.⁷⁷

There are two ways to restrict this argument. First, we might adopt an internal restriction, a limit on which circumstances generate *prima facie* reasons to use distinctive legal concepts. One possibility is to limit such considerations to circumstances in which the feature of the ordinary concept is in some way surprising or peculiar. For example, consider two experimental findings about the ordinary concept of intentionally producing a side-effect: 1) this concept is sensitive to the severity of the outcome; 2) this concept applies where a side-effect is produced purposefully. While the first is a *surprising* experimental discovery, the second is more widely acknowledged. If our principle is that there is a *prima facie* reason to exclude “surprising” features of the ordinary concept from legal concept, this would provide a reason against the adoption of the first feature, without providing one against adoption of the second feature.

But perhaps the better strategy is to recall that the considerations discussed in this section are only *prima facie* reasons. There are various possible considerations that may support using the ordinary concept, any of which might outweigh *prima facie* reasons to use a distinctive concept. In other words, if there is *no other reason* to employ the ordinary concept, perhaps it is sensible to

⁷⁷ Of course, this is excepting homonyms. For example, the concept of *standing* corresponding to Article III “standing” need not correspond to our ordinary concept of *standing*).

conclude that the legal concept should be (entirely) different.

However this principle is restricted, it should be clear that when we consider the *prima facie* reasons for or against using ordinary concept in law (Sections II.A. and II.B), there are many more reasons in favor of using ordinary concepts. However, there is a final set of considerations—one that is perhaps the most important. These are the reasons generated by considering the *legal aptness* of a concept's features.

C. Substantive Reasons for Law to Use Particular Concepts

Experimental discoveries often illuminate specific features of concepts (ordinary or specialized ones). Considering the nature, purposes, and precedents of the relevant area of law can provide reasons about whether those features are suitable for the legal concept. That is, once we learn that an ordinary concept has feature *x*, there are often considerations generated by considering the relevant area of law in which the concept would be used. Beyond general considerations about (e.g.) publicity or legal distinctiveness, there is a crucial set of considerations about which features (if any) of the ordinary concept are legally *apt*.

Consider two examples from Part I. First, take reasonableness. Experimental studies have found that ordinary judgments of reasonableness are informed by both statistical and descriptive considerations (e.g. considerations about both what is average and what is good). But *should* legal reasonableness standards use an ordinary concept with this features?

One way to answer this question is to consider only the *prima facie* reasons outlined in the previous two Sections. For example, using the ordinary concept fosters publicity and greater consistency between judges and juries. But a further set of consideration concerns whether the ordinary concept is legally apt. Is the “reasonable care” standard of negligence one that should reflect both statistical and prescriptive norms? That analysis might provide reasons that outweigh

the *prima facie* reasons.⁷⁸

As a second example, consider the debate over intentional action. Experimental results have suggested that, in at least some circumstances, legal experts use a concept of intentional action that differs from the ordinary concept. The ordinary concept seems sensitive to the severity of the outcome and the perceived blameworthiness of the actor. All else equal, if someone's action leads to a side-effect that seems very bad, we are more inclined to think the side-effect was produced intentionally. The legal concept is not sensitive to severity in this way.

But *should* legal rules reflect the ordinary concept? On the one hand, doing so is supported by consideration of publicity, clarity, consistency, and so on.⁷⁹ On the other hand, we might think that some of the particular features of the ordinary concept are not suited to legal contexts.⁸⁰ While it might be sensible for attributions of intentionality to be informed by blame in moral contexts (not caring about a foreseeable bad side-effect is morally blameworthy), we might think that legal contexts are right to adopt a more "clinical" concept of intentional action, one that is independent of our perceptions of blameworthiness (not caring about a foreseeable bad side-effect is not within the same legal set of actions as purposefully causing harm).

D. The Significance of Legal Domains and Contexts

It is worth reiterating that in evaluating what concepts law should use, the direction and weight of the *prima facie* and substantive reasons may be influenced by the legal domain and context.

In the case of reasonableness, we considered whether legal standards should be informed by both prescriptive and statistical considerations (as is the ordinary concept of reasonableness). A number of *prima facie* considerations generally support legal standards including this feature. But the weight of these reasons depends on their relevance in a particular context. For example,

⁷⁸ See, e.g., Tobia *supra* note 8.

⁷⁹ This is a useful example in which predictability probably does not provide a reason in favor of the ordinary concept. Most ordinary people are surprised by some of these features of the ordinary concept. Insofar as the public is not aware of certain features of the ordinary concept, there is no predictability reason for law to incorporate those features into the legal concept.

⁸⁰ See, e.g., Tobia *supra* note 23.

insofar as the significance of law's ability to provide notice is more important in criminal and tort law than in evidence law, notice provides a stronger *prima facie* reasons to use the ordinary concept of reasonableness in standards of reasonable provocation to kill and reasonable care, compared to the reason to use the ordinary concept in the standard of reasonable doubt.

Similarly, substantive reasons could also vary in direction and weight, depending on the legal context. For example, compare a view of tort law that characterizes its aim as making liable (for harms) those who violate community norms, with a view of criminal law that characterizes its aim as punishing (retributively) certain severe moral violations, and a view of evidence that law characterizes its aim as maximizing probity. We might argue that statistical considerations are especially relevant to torts (given these aims) and less relevant to criminal and evidence law (given their respective aims). Community norms are informed by statistical and prescriptive considerations, while the set of criminal law's violations and evidence law's rules might each seem more independent of facts about what is typical.⁸¹

The same sorts of differences play out in the analysis of many other concepts. Is the ordinary concept of consent legally apt for consent to a contract? What about for consent to sex? Is the ordinary concept of intentionally acting legally apt for intentional torts? What about for intentional killing? Is the ordinary concept of causation legally apt for determining proximate cause in torts? What about for determining whether a discriminatory motive was causal?

III. EXPERIMENTAL JURISPRUDENCE AND OTHER APPROACHES

The previous Parts articulate the experimental jurisprudential project by considering examples and outlining broader themes and a framework of analysis. This Part complements that study by comparing experimental jurisprudence with three related approaches to legal and philosophical scholarship: experimental economics and psychology and law; experimental

⁸¹ On this view, standards of reasonable care depend, in part, on what people actually do. But murderous actions do not become non-murder even if they become more common.

philosophy; and “traditional” legal philosophy.

A. Law and (Experimental) Economics and Psychology

There is a rich experimental literature in the economics and psychology of law.⁸² It is impossible to comprehensively survey this literature in any brief way, but we can consider some representative examples.

A common feature of experimental methods in legal scholarship is the study of legal outcomes and processes, using experiments to reveal facts about how law operates “on the ground.” One such example within the “law and psychology” approach is Elizabeth Loftus’s pioneering work on eyewitness testimony.⁸³ In an influential series of studies, Loftus provided evidence that eyewitness testimony is often unreliable, influenced by expectations and even subtle question framing. For example, participants were asked to evaluate a video of a traffic accident. One group was asked how fast the car was when it “hit” the other car; another group was asked how fast the car was when it “smashed” into the other car. This question framing altered speed estimates (with “smashed” and “collided” eliciting judgments of greater speed than “hit” and “contacted”).

As another example, consider the psychology of why people obey law.⁸⁴ Survey studies have found that people cite several reasons for why they obey law, including concern with being caught and punished and a desire to act morally (a sense of what is right and wrong).⁸⁵ This provides useful information for maximizing legal efficacy.

Another well-known example is research on law and “cultural cognition.”⁸⁶ This project has similar motivations, for example to learn “what [legal] policies will achieve . . . Does the death

⁸² See generally RESEARCH HANDBOOK ON BEHAVIORAL LAW AND ECONOMICS (Joshua Teitelbaum & Kathryn Zeiler, eds. 2018); HANDBOOK OF PSYCHOLOGY AND LAW (Dorothy Kagehiro & William Laufer, eds. 1992); Tom R. Tyler, *Psychology and the Law*, in THE OXFORD HANDBOOK OF LAW AND POLITICS (2008).

⁸³ See generally Elizabeth F. Loftus & J.C. Palmer, *Eyewitness Testimony*, in INTRODUCING PSYCHOLOGICAL RESEARCH 305 (1996).

⁸⁴ See, e.g., TOM R. TYLER, WHY PEOPLE OBEY THE LAW (1990).

⁸⁵ *Id.*

⁸⁶ See, e.g., Dan M. Kahan & Donald Braman, *Cultural Cognition and Public Policy*, 24 YALE L. & POL’Y REV. (2006).

penalty deter homicides”)?⁸⁷ Studies about “cultural cognition” have provided rich insight into the ways in which cultural values inform people’s interaction with law. For example, politically motivated reasoning might lead judges or jurors to make decisions based inappropriately on ideological preferences.⁸⁸

Like work on cultural cognition, much work in experimental economics focuses on identifying the legal impacts of ordinary people’s (e.g. potential juror’s) biases or irrationalities.⁸⁹ Other research investigates behavioral outcomes of legal relevance, evaluating law’s efficacy or aiming to improve it.⁹⁰

“Experimental jurisprudence” has much in common with this work in the experimental psychology and economics of law. Broadly speaking, these all use experimental methods to make progress in legal scholarship. The most significant difference is that experimental jurisprudence addresses distinctively *philosophical* questions.

A typical experimental jurisprudence study reveals some fact about concepts relevant to law: how ordinary people understand causation, consent, or law itself. But these discoveries do not carry straightforward legal implications. There is always more philosophical work to be done in determining whether the features of the ordinary concept are legally apt.⁹¹ Recall, for example, experimental studies about causation. Learning that ordinary judgments of causation have a particular feature (e.g. causal supercession) does not immediately imply anything about whether the legal concept should have that feature. Although experimental jurisprudential findings provide an enriched factual basis for legal theory, a complete experimental jurisprudential project requires further theorizing.

The same feature characterizes experimental jurisprudential work on consent (e.g. *should*

⁸⁷ *Id.*

⁸⁸ Dan M. Kahan, David Hoffman, Danieli Evans, Neal Devins, Eugen Lucci & Katherine Cheng, “Ideology” or “Situation Sense”? *An Experimental Investigation of Motivated Reasoning and Professional Judgment*, 164 U. PA. L. REV. 349 (2015).

⁸⁹ See generally Christine Jolls, Cass R. Sunstein & Richard Thaler, *A Behavioral Approach to Law and Economics*, 50 STAN. L. REV. 1471 (1998).

⁹⁰ See generally THE OXFORD HANDBOOK OF BEHAVIORAL ECONOMICS AND THE LAW (Eyal Zamir & Doron Teichman, eds. (2014).

⁹¹ See generally Part II *supra*.

deception vitiate legal consent?), reasonableness (e.g. *should* legal reasonableness be informed by statistical considerations?), intentional action (*should* legal intentional action be severity-sensitive?), and so on.

In contrast, most work in the experimental psychology and economics of law does not aim to address such philosophical concerns. Insofar as eyewitness testimony is unreliable, there is no hard philosophical question about how law should understand the legal concept of *testimony*. To be sure, there may be hard questions about what kinds of legal reforms should be implemented in light of those results. But there is no debate to be had regarding whether legal testimony *should* be reliable or unreliable.

In the same way, discoveries about the influence of politically motivated reasoning and cognitive biases on judicial decision-making are of critical importance. But those implications can be explored without further philosophical analysis about the status of the findings: It is not a tenable position that politically motivated reasoning *should* characterize certain legal decision-making processes.

These considerations, emphasizing the *philosophical* aspect of experimental jurisprudence, might lead one to think that the experimental jurisprudential project is better characterized as a branch of experimental philosophy. The next Section evaluates this proposal.

B. Experimental Philosophy

Experimental jurisprudence is closely connected to experimental philosophy, the experimental study of philosophical topics and concepts.⁹² This Section focuses on some of the key differences that distinguish the two programs.

The most obvious difference concerns the specific concepts studied. Although there is significant overlap—with both programs studying concepts like *causation*, *intentional action*, and *knowledge*—there are also some concepts of greater legal significance that experimental

⁹² See, e.g., EXPERIMENTAL PHILOSOPHY (Joshua Knobe & Shaun Nichols, eds. 2008).

philosophy has thus far neglected.⁹³ Some examples include *reasonableness*, *consent*, *ownership*, *law*, and *recklessness*.

The second major difference concerns the topic of “ordinary” or “public” meaning. Most philosophical debates do not turn on facts about ordinary or public meaning. However, many legal debates do. As one example, consider what it means to act (or produce a side-effect) “intentionally.” This term might appear in a legal context, for example a statute that imposes liability for “intentionally harming” someone. It might also appear in a moral context, for example in the doctrine of double effect, which states that it is morally permissible to do something morally good that has a morally bad side-effect, if the bad side-effect was not produced “intentionally.” Many plausible theories of legal interpretation would take some interest in the ordinary meaning of a term in such a statute; that meaning is at the very least relevant to the statute’s legal application, and at the most it is outcome-determinative. On the other hand, very few moral theories would think that the ordinary meaning of “intentionally” has any bearing at all on the truth of the doctrine of double effect. This is a very significant difference between experimental philosophy and experimental jurisprudence. The significance of empirically grounded legal interpretation opens a vast domain of debates in which experimental jurisprudential findings about ordinary concepts play an important role.

A final difference concerns the (possible) differences between philosophical (e.g. moral) and legal contexts. As we noted earlier, experimental philosophy and experimental jurisprudence may study the very same concepts (e.g. *causation*), but it is not always the case that the reasons to use the concept in a philosophical (e.g. moral) context are the same as ones to use it in a legal one. As one example, recall the discussion of the severity-sensitivity of the ordinary concept of intentional action. Although this feature might be suitable for moral contexts, some argue that it is generally unsuited to legal ones.⁹⁴

⁹³ See, e.g., Knobe *supra* note 21; sources cited *supra* note 24.

⁹⁴ See, e.g., Tobia *supra* note 23.

C. Legal Philosophy

A final comparison for experimental jurisprudence is traditional legal philosophy. Despite their use of different methodologies (experimental versus theoretical), these two approaches are surprisingly similar. A central aspect of each approach is the analysis of ordinary concepts and what the corresponding legal concepts should be. Nevertheless, there are some important differences between these approaches.

One distinctive contribution of experimental methods is the ability to reveal more general facts about the way in which legal concepts are deployed. By conducting experimental studies, experimental jurisprudence can discover more subtle factors that shape legal concepts and the underlying psychological processes involved in judgment and deployment of those concepts.

One type of experimental jurisprudence might test the types of thought experiments discussed earlier. Do people agree that in Life-Saving Negligence, the company's action is not legally reasonable?⁹⁵ An experimental survey could reveal what proportion of the population shares that judgment.⁹⁶

However, most experimental jurisprudence advances conceptual analysis in a very different way. Rather than determining majority judgment about specific cases, experimental jurisprudence investigates the *psychological processes* underpinning people's concept use.

For example, recall the experimental studies about the process of legal reasonableness judgment.⁹⁷ One popular philosophical account holds that reasonableness is a purely statistical judgment (like the judgment of what is average).⁹⁸ Another holds that reasonableness is a purely prescriptive judgment (like the judgment of what is ideal).⁹⁹ A third possibility is that reasonableness is a "hybrid" judgment. In the experiment, different groups of participants reported average, ideal, or "reasonable," quantities. Strikingly, across legal and non-legal

⁹⁵ See Section I.B. *supra*.

⁹⁶ See, e.g., Bert I. Huang, *Law and Moral Dilemmas*, 130 HARV. L. REV. 659 (2016).

⁹⁷ See Tobia, *supra* note 8.

⁹⁸ See, e.g., OLIVER WENDELL HOLMES, JR., *THE COMMON LAW* (1881).

⁹⁹ See, e.g., RICHARD POSNER, *ECONOMIC ANALYSIS OF LAW* (2002); John Gardner, *The Many Faces of the Reasonable Person* 131 L.Q. REV. 563 (2015); Alan D. Miller & Ronen Perry, *The Reasonable Person*, 87 N.Y.U. L. REV. 323 (2012).

examples, reasonableness judgment was intermediate between the relevant average and ideal. As that study made clear, experimental jurisprudence is not simply conceptual analysis with statistics. It offers unique contributions to this conceptual project; for example, it illuminates the psychological processes that produce legal judgments. In the study of reasonableness, the discovered fact—that judgment of the reasonableness is typically intermediate between the relevant average and ideal—might be very difficult to discern from traditional conceptual analysis. The experimental study provides conceptual analysis across a population, which allows for the discovery of different facts.

Although experimental jurisprudence provides different insights from traditional legal philosophy, it is still participating in a fundamentally similar project. By analyzing ordinary and legal concepts, experimental jurisprudence identifies relevant conceptual features, allowing us to ask legal-philosophical questions like: *should* reasonableness standards reflect statistical considerations; *should* legal consent reflect the ordinary concept; *should* legal attributions of intentional action be severity-sensitive; or *should* sentencing guidelines take into account intertemporal disconnectedness?

As should be clear, “experimental jurisprudence” has much in common with traditional law and psychology, experimental economics and law, experimental philosophy, and legal theory. However, it also departs from these traditions. Namely, it uses *experimental* methods to help address *philosophical* questions about *law*. In discovering the contours of ordinary and legal concepts, experimental jurisprudence provides a richer factual basis for understanding what law is and arguing about what it should be.

IV. FUTURE DIRECTIONS

This concluding Section outlines several “future directions” of the experimental jurisprudence approach to legal scholarship. This is not an exhaustive list of promising projects,

but rather an indication of some avenues of likely future success in the experimental jurisprudence program.

A. Concepts of Legal Significance

A first area is the experimental assessment of different legal concepts. Studies in experimental jurisprudence have already studied *causation*; *consent*; *reasonableness*; acting *intentionally*, *knowingly*, and *recklessly*; *law*; and *responsibility* and *punishment*. Further work could build further on these early findings, providing further insight into these ordinary concepts and how they interact with law.

Studies might also evaluate other legal concepts that have not yet been analyzed. Promising concepts include broad legal categories (e.g. *property*, *evidence*) as well as specific concepts of great legal significance across doctrinal areas (e.g. *harm*, *motive*) and with in specific doctrinal areas (e.g. contractual concepts of *consideration*, *promising*, and *reliance*).

B. Interpretation and Ordinary Meaning

As Section III.B noted, a distinctive feature of law (compared to philosophy) is the significance it places on ordinary meaning. Although some experimental jurisprudence studies have begun to study the science of legal interpretation, this remains an underexplored area, ripe for future work.¹⁰⁰

Consider two strands of potential work in this area. One would focus on using experiments to assess legal interpretive methodologies. For example, we can use survey methods to learn whether dictionaries reflect ordinary meaning.¹⁰¹

A different program would develop a “survey method” of interpretation.¹⁰² Broadly speaking this program would seek to develop and refine a survey method of determining “ordinary meaning.” In such a program, philosophical analysis is especially crucial. For example, what

¹⁰⁰ See Section I.A.10 *supra*.

¹⁰¹ See, e.g., Tobia *supra* note 38.

¹⁰² See, e.g., Ben-Shahar & Strahilevitz, *supra* note 36.

cutoff should we use for ordinary agreement; is something part of the ordinary meaning if 90% of survey participants agree; what about if only 51% agree? As in other experimental jurisprudential projects, the data alone will not resolve hard philosophical questions.

C. Expertise and Learning

A third area of likely future significance is in the domain of legal expertise and learning. This area contains a host of intriguing and important questions: How do ordinary and legal experts people acquire their ordinary and legal concepts; is legal education (in terms of concept acquisition and refinement) similar to scientific or other education; do experts retain ordinary concepts or are some replaced by ones acquired through legal training?

D. Cross-Cultural and Cross-Linguistic Study

Another area of significance is cross-cultural and cross-linguistic studies in experimental jurisprudence. One obvious motivation for these projects would be to investigate legal concepts relative to a particular jurisdiction. For example, publicity arguments in favor of law's use of what we take to be the "ordinary concept" of causation in France depend on facts about cognition of the French: Do experimental findings, from English-language studies on American participants, extend to other linguistic and cultural contexts?

But there are other important reasons to perform cross-cultural and cross-linguistic study. Sometimes, findings about cross-cultural *similarities* provide evidence about our legal concerns. As one example, imagine that we find a robust cross-cultural distinction between knowledge and recklessness. This might be taken as confirmatory evidence that there is a robust distinction, one that maps to some general features of human cognition. Such a finding might also be taken to alleviate concern that the knowledge/recklessness distinction is intractable or otherwise unclear.

On the other hand, sometimes discoveries of cross-cultural *differences* enrich legal debates. As one example, consider the work of Vilius Dranseika on the concept of *identity*. Recall that

earlier work found that certain severe changes lead Americans to say that the changed person is “no longer the same person,” and these judgments explain intuitions about statutes of limitations.¹⁰³ We might take statements like “he is no longer the same person” to tell us something about the concept of personal identity. That is, one interpretation of the experimental jurisprudential finding is that part of what explains the intuitiveness of statutes of limitations is that people seem to change over time into *entirely* new people.

Dranseika tested participants in languages that admit of a distinction that does not exist in English (e.g. Lithuanian). One question asked about *quantitative identity*, while one asked about *qualitative identity*.¹⁰⁴ He found that, in many of these cases, people think there is a qualitative identity change, but not a quantitative one. This finding tells us something interesting about cognition in non-English languages, but it might also enrich our interpretation of the English results. Perhaps the English language constrains what can ask in these experiments, and the best interpretation of these studies concerns not the concept of *personal identity*, but rather a concept of a *qualitatively similar self*.

E. Enhancing Survey Methods

Another area worthy of further study is the enhancement of survey methods. One natural way to implement this recommendation is for experimental jurisprudence to adopt methodological innovations for survey methods from existing work—in law and experimental economics, law and psychology, and the cognitive sciences more broadly. For example, most work in experimental jurisprudence compensates study participants, but does not incentivize the accuracy of their responses or “elicit” their beliefs. Insofar as these methods would be relevant, future work would do well to consider these methodologies.¹⁰⁵

¹⁰³ See Mott, *supra* note 30.

¹⁰⁴ See, e.g., Vilius Dranseika (draft manuscript).

¹⁰⁵ Of course, in some circumstances it is not obvious whether an “incentivized” response is more legally relevant. For example, a conventional interpretation study asks people to assess whether a car is a vehicle. An “incentivized” version might ask pay participants a bonus if they can accurately estimate what most other participants *did* say to the first question. Although this incentivization would likely increase accuracy (on the second question), it is not clear that this second question is one about *public* or *ordinary* meaning.

Another way to implement this strategy is for experimental jurisprudence to incorporate non-survey methods. Of course, when assessing what experimental method is *best* in assessing the relevant question, survey methods might often be optimal. But in some circumstances, experimentalists would do well to combine standard survey methods with more sophisticated tools (e.g. eye-tracking) or other behavioral or neuroscientific methods.

F. Legal and Philosophical Analysis

A final area of likely future growth is in traditional legal-philosophical analysis of experimental jurisprudential results and arguments. One set of important debates concerns meta questions about the justifications of experimental jurisprudence. Traditional legal-philosophical analysis will help clarify the precise role that experimental can play in legal theory and how we should understand the role of experimentation in philosophy of law.

A very different set of debates concerns interpretation of and arguments about specific experimental results. When experimental jurisprudence studies reveal facts about ordinary concepts—*reasonableness*, *causation*, *consent*, or *law* itself—this typically raises more philosophical questions than it settles.

CONCLUSION

“Experimental Jurisprudence” refers to a new approach to legal philosophy that conducts experimental studies to discover facts about legally significant concepts and uses those discoveries to support legal arguments. This paper has developed an account of this new field, proposing a framework of analysis and distinguishing the approach from related fields of experimental psychology and economics of law, experimental philosophy, and legal philosophy. Experimental jurisprudential studies, which discover the contours of ordinary and legal concepts, provide a richer basis for understanding what law is and arguing about what it should be.

CONCLUSION

This dissertation began by contrasting two views of legal concepts. The *Expert View* holds that law's central concepts are specialized legal ones, while the *Ordinary View* holds that the central legal concepts are ordinary ones.

Each view is undoubtedly right about certain concepts. There is no ordinary concept of *joint tenancy with right of survivorship*; this is a specialized concept, acquired by learning about property law. At the same time, law does not mean anything special when it uses dates or monetary units—those reflect the ordinary concepts. The key philosophical questions concerned concepts in law whose status was less clear. Strikingly, those concepts also tend to be the ones with greater legal significance. When law considers what is *reasonable*, whether someone acted *intentionally*, whether there was *consent*, whether an act was in someone's *best interest*, whether there was a *harm*, or what *caused* the outcome, do these concepts reflect ordinary or expert ones?

Importantly, these two views, *Expert* and *Ordinary*, can be articulated descriptively or normatively. The descriptive views disagree about whether law's concepts *are* expert or ordinary, and the normative views disagree about whether law's concepts *should* be expert or ordinary. This dissertation has demonstrated that experimental studies bear on the first descriptive question. But it has also argued that experimental studies—by revealing features of concepts—help address the second normative question. To be sure, discovering empirical facts about what a concept *is* does not directly support normative conclusions for what the legal concept *ought* to be. But in

many cases, those empirical discoveries enrich the legal-philosophical debate, illuminating particular conceptual features and indicating which questions are most relevant.

As the dissertation investigated two specific concepts in more detail—*reasonableness* in Chapter 1 and *intentional action* in Chapter 2, an unequivocal *Expert* or *Ordinary* normative view became more dubious. Chapter 1 argues that a central feature of the ordinary concept of *reasonableness*—that it reflects a hybrid of prescriptive and statistical considerations—is legally apt and should characterize legal standards of reasonableness. At the same time, Chapter 2 cautions that various features of the ordinary concept of *intentional action*—such as its sensitivity to the perceived severity of the outcome—may not be legally apt. It suggests that the expert concept, which lacks these features, is preferable. If the first two chapters' conclusions are right, the score is one for the *Ordinary* view and one for the *Expert* view. But even if not, the chapters reveal that in each case there are various competing reasons and arguments. For different concepts, in different legal contexts, either the *Expert* or *Ordinary* view may be more persuasive.

Reflecting on the first two chapters' work also clarifies that the introductory framing—of *Expert* versus *Ordinary* views—is somewhat misleading. The experimental study of concepts reveals various features of those concepts. Although there is a fact of the matter about which features an ordinary concept has and which features an expert concept has, it is not necessarily the case that either of those sets of features reflects what the legal concept *should* be. It might be that the legal standard of *intentional action* should reflect *every* feature of the ordinary concept; but perhaps it should only reflect *some* of those features.

The third chapter pivoted to an experimental study of ordinary meaning. Experiments play a very different role in that chapter's argument, as that debate sits within a very different legal-philosophical topography. In the first two chapters, there is not necessarily a tight connection

between empirical facts and normative theory; what *is* judged as reasonable need not be reflected in the standard of what we *ought* to treat as reasonable.

But in Chapter 3, facts about what *is* judged to be (e.g. a vehicle) are much more closely connected to what legal theories think law *ought* to say. In those debates, the aim is not to assess which features of the ordinary concept are legally relevant, but rather to assess whether legal methods accurately reflect ordinary concepts. Public Meaning Originalists seek to determine the ordinary meaning of legal texts by studying the concepts that correspond to terms like “vehicle.” Empirical facts about those concepts are closely connected to the notion of “ordinary” meaning, which constrains (normatively) how law should be applied.

This framework may feel foreign to those who are used to considering how experimental results bear on other philosophical fields. It would be very strange for an ethicist to place much stock in the ordinary meaning of moral rules. For example, the fact that most people tend to understand “person” to apply to humans but not other animals might suggest that the ordinary meaning of a putative moral rule, “only persons have moral rights,” is that chimps and dolphins have no moral rights. But most moral philosophers would not think this fact about ordinary meaning constrains how the rule should guide action or tells us much about the rule’s truth. It might be that the rule should extend to chimps, as ordinary people have the wrong concept of *person*. But even more broadly, most ethicists would not think that what makes a moral principle like this binding or true has anything to do with its “ordinary meaning.”

In contrast, various legal-philosophical theories emphasize the significance of ordinary or public meaning. Although some theories place more significance on ordinary meaning (e.g. Public Meaning Originalism), many place some weight on facts about ordinary concepts. Given these background assumptions, experiments play a different—and often much more significant—role in legal philosophy. Experimental jurisprudence can advance these normative legal debates in powerful ways by illuminating facts about ordinary concepts.

Thus, at a broader level, all of the chapters reflect a similar “experimental jurisprudence.” That common approach aims to make progress in legal philosophy through the experimental study of concepts. This approach can be leveraged in different ways. When normative legal theories commit to empirical claims about ordinary concepts, experimental jurisprudence is especially useful. But it is also useful when we have questions about what the legal concepts should be: Should reasonableness standards reflect the ordinary concept, and what features should characterize the legal concept of intentional action?

* * *

The dissertation concludes with a final expert-ordinary comparison. Legal philosophy, like much of philosophy, is conducted largely by experts. This is for good reason; expertise often carries beneficial wisdom in the study of specialized subjects.

At the same time, expert study should not obscure that law is deeply connected to ordinary people, language, and cognition. Most important laws are addressed broadly and nearly all apply widely; prohibitions on intentional murder and liability for unreasonable behavior fall upon experts and ordinary people alike.

The dissertation’s broader lesson is that experimental legal-philosophy, or “experimental jurisprudence,” fosters a unique type of wisdom. There are many reasons to learn how ordinary people understand legal concepts—promoting law’s clarity and publicity, ensuring consistency among judges and juries, and encouraging a democratic law making process. Experimental study also reveals specific conceptual features, enriching philosophical debates and bringing them closer to our world—the real arena of successful philosophical arguments’ practical effects.

Empirical study alone cannot resolve philosophical questions about what law should be. But one cannot help but wonder whether those questions *should* be settled by purely theoretical study. To have the best debate about where law should go, it is important to know from where we start.