

**LEGAL CONSTRUCT VALIDATION:
EXPANDING EMPIRICAL LEGAL SCHOLARSHIP
TO UNOBSERVABLE CONCEPTS**

DAVID S. GOLDMAN*

INTRODUCTION

Unobservable concepts such as rights, possessions, incentives, and deterrence form the foundation of legal thinking. Yet over 200 years ago, Jeremy Bentham strenuously warned us not to be seduced by the follies of these types of terms, which he called inefficient legal fictions:¹ “What you have been doing by the fiction—could you, or could you not, have done it without the fiction? If not, your fiction is a wicked lie: if yes, a foolish one.”²

For Bentham, legal fictions were artifacts of the imprecision inherent in language.³ Such imprecision prevented meaningful distinctions between real entities, such as “this man, this beast, this bird,” and abstract entities.⁴ He claimed that such fictions, rather than expressing lofty ideas, were instead most often used to subvert the law.⁵

Why is it then that the use of such terms continued unabated? At the very least, they constitute shorthand for a list of real entities that would be too extensive to name in each instance of use; for example, referring to “civil rights” is simpler than enumerating each constitutional right to personal liberty.⁶ But a more critical role can be assumed for these terms that Bentham could not have yet appreciated. Given techniques that have been developed in the social sciences in the last 60 years, these “fictions” can be operationalized,

* Copyright © 2007, David Goldman, goldmand@gmail.com. I would like to thank Greg Smith, Gideon Parchomovsky, Michael O'Donnell, and Spencer Waller for their extremely helpful comments. The views outlined in this article are entirely my own and do not necessarily reflect those of any past or present employers.

¹ C.K. OGDEN, BENTHAM'S THEORY OF FICTIONS 38 (1932).

² *Id.* at 141.

³ *Id.* at 12.

⁴ *Id.* at 38, 60.

⁵ *Id.* at 141, 147–50 (listing the ways lawyers use these “forced falsehoods” to make more business for themselves, such as increasing chance of misstatements, making it easier to allege that jury verdicts are improper by making the law incomprehensible, confirming arbitrary power to judges, and corrupting the morals and intellectual faculties of the public).

⁶ “Civil rights” include the personal liberties guaranteed by the Bill of Rights, the Thirteenth, Fourteenth, Fifteenth, and Nineteenth Amendments, as well as the Voting Rights Act, among other sources. BLACK'S LAW DICTIONARY 100 (Pocket ed. 1996).

and can actually then enter into empirical relationships that permit their measurement, as well as the assessment and prediction of the effects of laws built around these “fictions” on subsequent human actions.⁷ For instance, in Ronald Coase’s now fundamental example of the negotiation between a farmer and a cattle-raiser for use of land, the neighbors were able to broker a deal because they owned property rights in their land.⁸ While the rights and the property may both be unobservable legal fictions, one can easily observe whether crops or cattle occupy a particular parcel of land. By gaining a detailed understanding of the operations of property rights, policymakers can predict how changes in property law will affect future interactions for efficient use of land. Legal scholars can now, finally, overcome Bentham’s criticisms and determine empirically the practical usefulness of some of our most reliable legal fictions.

Empirical observations of legal entities has been gaining particular popularity recently because it allows debate on fundamental questions that are informed by the actual impact of law on behavior rather than conjecture and an appeal to commonsense.⁹ But now legal scholars can expand the current horizons of quantitative analysis in legal scholarship by moving beyond measurement of only observable entities to assessing unobservable legal models as well through the notion of construct validation—the methods behavioral scientists use to overcome the difficulty in measuring unobservable psychological phenomenon, called constructs. Using the metatheories of construct validation, researchers can infer the effects of unobserved constructs and can then modify foundational legal theories based on empirical evidence rather than speculation.¹⁰ I propose a standardized procedure, adapted from

⁷ See Brendan A. Maher & Irving I. Gottesman, *Deconstructing, Reconstructing, Preserving Paul E. Meehl’s Legacy of Construct Validity*, 17 PSYCHOL. ASSESSMENT 415, 415 (2005) (describing the evolution of measurement of unobservable entities).

⁸ Ronald H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1 (1960), reprinted in THE FIRM, THE MARKET, AND THE LAW 95, 97 (1988).

⁹ See, e.g., Tracey E. George, *An Empirical Study of Empirical Legal Scholarship: The Top Law Schools*, 81 IND. L.J. 141, 141 (2006) (“Empirical legal scholarship (ELS) is arguably the next big thing in legal intellectual thought.”); American Association of Law Schools Annual Meeting: Empirical Scholarship—What Should We Study and How Should We Study It? (Jan. 3–7, 2006), <http://www.aals.org/am2006/index.html>; Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1, 2 (2002); Richard H. McAdams & Thomas S. Ulen, *Introduction: Symposium: Empirical and Experimental Methods in Law*, 2002 U. ILL. L. REV. 791.

¹⁰ Thomas S. Ulen, *A Nobel Prize in Legal Science: Theory, Empirical Work, and the Scientific Method in the Study of Law*, 2002 U. ILL. L. REV. 875, 899.

current understandings of construct validation, to measure intangible legal constructs: (1) develop generalized legal theories, (2) infer hypotheses from those theories, (3) design experiments to test the hypotheses, and (4) modify the general theories based on the research results.

This first step in this process is the development of theories with a generalized perspective not confined to specific laws that still maintain a degree of specificity that provides adequate direction for future empirical testing. Generalized theories allow multiple inferences that can elucidate finer detail of an unobserved latent variable. For instance, rather than addressing only the impact of the First Amendment,¹¹ a more generalized theory will address the influences of all constitutional rights.¹² A generalized theory may state that populations will tend to exhibit more freedom when specific rights are embedded within those populations' constitution. Empirical studies of freedom of speech can then provide particularized support for the broader constitutional theory, and can also interact with studies of other rights.

Once a generalized theory has been devised, researchers can then infer sets of hypotheses to test measurable aspects of those theories. For example, based on a theory that constitutional rights lead to more freedom, one might hypothesize that if a nation has a constitutional freedom of speech then the media may produce more commentary disapproving of the government. While the actual *freedom* that the press feels is impossible to measure, the proportion of articles disapproving of government actions is relatively easy to calculate. But a useful hypothesis suggests more than just an empirical test; it should also have the potential to undermine the theory.¹³ The necessity for critical hypotheses is borne from the current philosophical understanding that

¹¹ See, e.g., Stephen M. Feldman, *The Theory and Politics of First Amendment Protections: Why Does the Supreme Court Favor Free Expression over Religious Freedom?*, 8 U. PA. J. CONST. REV. 431, 433–51 (2006) (discussing the development of First Amendment theory since the drafting of the Constitution). While the development of theories for specific laws and Constitutional provisions is extremely beneficial in its own right, these narrowly-tailored theories do not lay an appropriate groundwork for empirical study because they do not provide a broad enough base from which to infer traits of the unobservable variable.

¹² See, e.g., Stephen Kanter, *The Griswold Diagrams: Toward a Unified Theory of Constitutional Rights*, 28 CARDOZO L. REV. 623 (2006) (attempting to analyze the abortion debate by using a more generalized theory for Constitutional law when the basis of the right is unclear, rather than focusing on the specific rights outlined in the Supreme Court's abortion cases).

¹³ Gregory T. Smith, *On Construct Validity: Issues of Method and Measurement*, 17 PSYCHOL. ASSESSMENT 396, 396–97 (2005).

theories can never be fully verified.¹⁴ Hence, while empirical evidence that sustains a theory can reaffirm the theory's legitimacy, a critical hypothesis can serve one of two purposes: if supported, the hypothesis can undermine an incorrect theory, but if not, that same hypothesis can help to rule out a potential criticism. By eliminating potential criticisms, a researcher can more forcefully demonstrate the validity of the overarching theory.¹⁵

Next, the researcher must devise experiments that can test the accuracy of the hypotheses. Ideally, experiments would use randomized subjects for testing so that the results were likely caused by the treatment being studied rather than a bias in the selection process.¹⁶ But it is often not feasible to randomize subjects on empirical legal studies for practical, if not necessarily constitutional, reasons; most legal researchers, therefore, use various forms of quasi-experimentation that allow testing when subjects can not be randomized.¹⁷ Quasi-experimental designs allow inferences to be made about causal relationships when randomization is not feasible.¹⁸ A researcher wishing to study prison terms, for example, could use a "time-series design" to analyze sentences that were imposed both before and after the guidelines

¹⁴ *Id.* at 396.

¹⁵ Moreover, a researcher should not test only one hypothesis because the results from that single test can be influenced by a questionable supporting theory rather than the one targeted by the study. JAMES A. CAPORASO, *Quasi-Experimental Approaches to Social Science: Perspectives and Problems*, in *QUASI-EXPERIMENTAL APPROACHES: TESTING THEORY AND EVALUATING POLICY* 3, 10 (James A. Caporaso & Leslie C. Roos, Jr. eds., 1973). Each theory is supported by numerous auxiliary theories; for instance, many current economic theories of law are based on Ronald Coase's famous theory of transactions, which in turn, relies in part on a rational-actor theory of behavior. Hence, a test of a single hypothesis concentrated on a prevailing theory of law and economics may uncover negative results, not because of a problem with the theory being examined, but because of problems with the rational-actor model. David A. Hoffman & Michael P. O'Shea, *Can Law and Economics Be Both Practical and Principled?*, 53 *ALA. L. REV.* 335, 361–62 (2002). But if only one hypothesis is tested, this possibility may not be recognized.

¹⁶ THOMAS D. COOK & DONALD T. CAMPBELL, *QUASI-EXPERIMENTATION: DESIGN & ANALYSIS ISSUES FOR FIELD SETTINGS* 341 (1979) (explaining that randomization serves two functions: (1) drawing representative samples within the limits of sampling error, and (2) drawing samples that are comparable within the known limits of sampling error).

¹⁷ CAPORASO, *supra* note 15, at 11 (noting some of the problems with the experimental method, including the lack of realism in an experimental context leading to a nontransferability of theories devised in a laboratory).

¹⁸ *Id.* at 12.

went into effect.¹⁹ But because the experiment is not truly randomized, the researcher needs to be careful to identify plausible threats to intrinsic logic of the study (such as the possibility that other changes in the law during that same time period affected sentences).²⁰ These threats can be controlled by adjusting design elements or statistical techniques to account for the threats.²¹

Finally, causal inferences can be made when the results form a coherent pattern, either within the study or as compared to outside knowledge.²²

Once the various hypotheses have been tested, the overarching theory should be modified to reflect the new knowledge gained through the empirical data. The theory may have accurately predicted all the results of the various experiments, in which case the theory is supported but no new information is provided to the field. But if the theory's predictions are not completely accurate, the theory should be modified to incorporate this new information. Theories are not falsified every time disconfirming data is discovered, nor do fields go through seismic paradigm shifts in which old theoretical frameworks are thrown out in favor of an entirely new system. Rather, through an iterative process of testing and retesting, theories about unobservable legal constructs can be modified based on inferences from the results of experimentation.

It is important to note, however, that empirical quantitative analysis cannot supplant qualitative reasoning; while empirical results can help inform legal policy-making, it is still necessary to debate and understand theories on a qualitative level. For example, empirical testing could indicate that torture could lead to improved intelligence gathering, but society must still decide whether it can morally justify such tactics. Similarly, evidence of the deterrent effect of capital punishment does not by itself substantiate the penalty.

The rest of this paper is organized as follows. Part I examines the development of philosophies of science to provide background about the current understandings of theory development. Part II then describes the history of psychometrics to demonstrate how psychology has applied philosophies of science to overcome some of the same issues that empirical legal scholarship is now facing. Part III then details my proposed approach

¹⁹ Donald T. Campbell, et al., *Quasi-Experimental Designs*, in *METHODOLOGY AND EPISTEMOLOGY FOR SOCIAL SCIENCE* 191, 201 (E. Samuel Overman ed., 1988).

²⁰ William R. Shadish, *The Empirical Program of Quasi-Experimentation*, in *2 RESEARCH DESIGN: DONALD CAMPBELL'S LEGACY* 13, 16 (Leonard Bickman ed., 2000).

²¹ *Id.* at 17.

²² *Id.* at 18–19.

for incorporating the most relevant aspects of this related field into legal study.

I. PHILOSOPHY OF SCIENCE

Before moving on to describe a system for construct validation for legal studies, it is first necessary to understand the philosophical underpinnings on which those concepts are based.

A. *Sir Karl Popper's theory of verification and falsification*

A common starting point when addressing the philosophy of science is the philosopher Karl Popper's idea of falsification: a theory is scientific if it can be tested and has the potential to be proven wrong.²³ The central philosophical problem that concerned Popper was the difficulty of demarcation: how is science distinguishable from pseudo-science?²⁴ Popper concluded that a true scientific theory should be corroborated through exposure to empirical tests that have the potential of disproving the theory, and a theory is better corroborated when it has survived more difficult tests.²⁵

²³ KARL POPPER, *THE LOGIC OF SCIENTIFIC DISCOVERY* chs. 3–4 (Karl Popper et al. trans., Routledge Classics, 2002).

²⁴ William C. Kneale, *The Demarcation of Science*, in *THE PHILOSOPHY OF KARL POPPER* 205, 206 (Paul Arthur Schilpp ed., 1974) (noting that Popper's "primary aim was to distinguish between the universal propositions of genuine science, whether true or false, and those of pseudoscience"); W.H. Newton-Smith, *Popper, Science and Rationality*, in *KARL POPPER: PHILOSOPHY AND PROBLEMS* 13, 14 (Anthony O'Hear ed., 1995) (explaining that Popper worked to differentiate scientific theories such as Einstein's theory of relativity from the unscientific theories of Marx, Adler, and Freud). Popper first focused on the question of demarcation because, as a Marxist, he began to question whether the Marxist belief in "scientific socialism" could be supported by actual scientific pursuits. Karl Popper, *Autobiography*, in *THE PHILOSOPHY OF KARL POPPER* 1, 25–29 (Paul Arthur Schilpp ed., 1974).

²⁵ PAUL E. MEEHL, *Theoretical Risks and Tabular Astricks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology*, in *SELECTED PHILOSOPHICAL AND METHODOLOGICAL PAPERS* 1, 18 (1991); W.V. Quine, *On Popper's Negative Methodology*, in *THE PHILOSOPHY OF KARL POPPER*, *supra* note 24, at 218, 218–20 ("Evidence does not serve to support a hypothesis, but only to refute it, when it serves at all."); Peter Lipton, *Popper and Reliabilism*, in *KARL POPPER: PHILOSOPHY AND PROBLEMS*, *supra* note 24, at 31, 32–33 (explaining first that scientific evidence can never prove a theory true, but can prove it false, but going on to point out four objections to this point: (1) if observations contradict an hypothesis, one can never know if the observation is false or if the hypothesis is false, (2) it is impossible to determine whether an hypothesis is false or an auxiliary assumption is false, (3) although one should rely on theories that have not been refuted, even refuted theories can predict future events reliably, (continued)

Popper believed in an asymmetry between the ability to verify a theory and the ability to prove a theory false; specifically, a scientific theory can never be proven completely correct, but evidence that undermines a theory can conclusively prove that theory wrong.²⁶ Even a well-established theory can be proven wrong at any time if contrary evidence is discovered. For instance, we may believe that every time an object is dropped it will fall towards the Earth, but even this theory could be falsified if the next item we dropped fell up. As a result, an hypothesis that can provide support for a theory is never as useful as one that can undermine the theory because the former can never prove a theory true, but the latter can definitively establish that it is false.²⁷ Scientists should concentrate, then, on testing negative hypotheses because each time a negative hypothesis fails, one more potential criticism has been ruled out and the theory has been further corroborated. Popper theorized that by discrediting incorrect theories, scientists will be able to move on to theories that more accurately reflect true nature. Popper, quoting Albert Einstein, said, “[t]here could be no fairer destiny for any physical theory than that it should point the way to a more comprehensive theory, in which it lives on as a limiting case.”²⁸

But critics of this philosophy have pointed out that if Popper is correct, then history should be laden with critical studies that have falsified the predominant theory of the time.²⁹ In reality, however, most important theories were falsified countless times before they were discarded by the scientific community.³⁰ Under Popper’s falsificationist view, therefore, some of the most important theories in history were not actually scientific.³¹ Proponents

and 4) Popper's philosophy provides no reason to believe that science is moving towards the truth.)

²⁶ Popper, *supra* note 23, at 18–19; *see also* Quine, *supra* note 25, at 218; Lipton, *supra* note 25, at 32.

²⁷ *See* Lipton, *supra* note 25, at 41–42. Lipton postulates that it is easier to determine if an hypothesis is false, then whether it is true. This asymmetry results because for a scientist to establish that an hypothesis is false, the scientist need only show that any part of the hypothesis is wrong; conversely, it is true only if the entire hypothesis is correct.

²⁸ Karl Popper, *Autobiography*, in *THE PHILOSOPHY OF KARL POPPER*, *supra* note 24, at 29 (quoting Albert Einstein, *ÜBER DIE SPEZIELLE UND DIE ALLGEMEINE RELATIVITÄTSTHEORIE*, (3d ed., Braunschweig: Vieweg, 1918)).

²⁹ BRENDAN LARVOR, *LAKATOS: AN INTRODUCTION* 50 (1998).

³⁰ *Id.*

³¹ *Id.*

of Popper's philosophy must distort history and arbitrarily label certain falsifying experiments as "crucial," while ignoring others.³²

Popper's idea that a theory can be completely discredited by any negative evidence has been labeled a type of justificationism.³³ Justificationists in general believe that a scientific theory can be fully proven or disproved by empirical evidence.³⁴ "In brief, [justificationism] is the view that the way to criticize an idea is to see whether and how it can be justified."³⁵ Justificationists believe that certain facts exist in reality and if scientists can establish that a particular theory does not accurately reflect those facts, the theory should be discarded.³⁶ For example, the premise of an argument may be that all *observable* planets travel in an elliptical orbit, and the conclusion is that *all* planets (whether observable or not) must therefore travel in an elliptical orbit.³⁷ For a justificationist, this is an invalid theory because it overstates the known facts.³⁸ But for a nonjustificationist, the conclusion stated above may still have value to science as an unjustified hypothesis rather than as a justified statement about nature. By reversing the syllogism, the conclusion can be stated as an hypothesis, and the premises are data supporting the hypothesis. Now, contrary evidence will not mean that the theory needs to be rejected outright, but the theory may need to be conditionally rejected or revised to reflect the new data.³⁹ This revisionist process can proceed indefinitely as new information is collected. Philosophy has generally moved away from the justificationist theories, and most philosophers now endorse various versions of nonjustificationist.⁴⁰

³² *Id.* at 42; *see also* Imre Lakatos, *Popper on Demarcation and Induction*, in *THE PHILOSOPHY OF KARL POPPER*, *supra* note 24, at 241, 246–48. Lakatos believed that Popper's theory itself could be falsified because some of the greatest scientific achievements in history did not follow Popper's descriptions. *Id.* at 246. According to Popper, a scientist must articulate in advance the conditions under which the scientist will disavow a theory, but supporters of Newtonian physics have never done so. *Id.* at 246–47.

³³ WILLIAM WARREN BARTLEY, III, *THE RETREAT TO COMMITMENT* 104–05 (2d ed. 1984).

³⁴ Herbert Feigl & Albert Blumberg, *Logical Positivism: A New European Movement*, *J. OF PHIL.* 28, 281–96 (1931).

³⁵ BARTLEY, *supra* note 33, at 186.

³⁶ *Id.* at 186–87.

³⁷ *Id.* at 191.

³⁸ *Id.*

³⁹ *Id.* at 196.

⁴⁰ Smith, *supra* note 13, at 397; BARTLEY, *supra* note 33, at 194.

B. Thomas Kuhn's Theory of Paradigm Shifts

One such nonjustificationist theory is historian Thomas Kuhn's theory of scientific paradigms.⁴¹ After researching the history of scientific developments, Kuhn noted a distinction between mature and immature sciences.⁴² In immature sciences, many competing schools of thought exist, but no single idea dominates the field.⁴³ Because an accepted theory has not been adopted, scientists operating in immature fields base their research on many different schools of thought.⁴⁴ But in exchange for this freedom to choose their own theories, the scientists must give up the benefits of relying on established research supporting the broad theory.⁴⁵ Because no accepted idea for the field exists, the scientists must reassert every basic concept on which any new research relies.⁴⁶ Hence, an immature field moves along essentially at random because no controlling force exists to direct future exploration.⁴⁷ (This state of an immature field also seems to describe the current state of legal academics. It is often necessary to spend a significant portion of a law review article stating the assumptions and background on which a new theory is based because, without it, a reader will not know what school of thought and definitions the author is employing.)

But researchers in an immature science do not need to despair that their field will founder with no direction indefinitely; a science can mature, not necessarily because of a new field-shattering discovery, but from a theoretical proposal that is influential enough to encourage adoption by members of

⁴¹ See THOMAS S. KUHN, *THE STRUCTURE OF SCIENTIFIC REVOLUTIONS* (3d ed. 1996). As a historian, Kuhn focused his seminal work on paradigms by describing how science evolves using historical examples such as how Einstein's theory of Relativity displaced Newton's laws of motion and how Benjamin Franklin's theories about the operation of electricity became the dominant theory in the field.

⁴² *Id.* at 12–13 (discussing the distinction between mature and immature sciences).

⁴³ *Id.* at 13 (using the study of optics before and after Newton as an example of a transition from an immature field) (“Yet anyone examining a survey of physical optics before Newton may well conclude that, though the field's practitioners were scientists, the net result of their activity was something less than science. Being able to take no common body of belief for granted, each writer on physical optics felt forced to build his field anew from its foundations.”).

⁴⁴ *Id.* at 13–17 (describing the immature states of various scientific fields).

⁴⁵ *Id.* at 13.

⁴⁶ *Id.*

⁴⁷ *Id.* at 15 (noting that because without a dominant paradigm all facts seem equally relevant, no guiding force controls future development, and fact-gathering is essentially random).

opposing schools of thought. Once a controlling theory is accepted, mature fields are then guided and driven by this single idea. Kuhn labeled these driving forces “paradigms,” which specifically are “a set of recurrent and quasi-standard illustrations of various theories in their conceptual, observational, and instrumental applications.”⁴⁸ Essentially, a paradigm is an accepted worldview or framework under which scientists explore their field. In mature sciences, paradigms do not evolve slowly, but rather are subject to seismic transitions in thought, what Kuhn called paradigm shifts.⁴⁹ When scientists go about their day-to-day work in a mature field they usually work within the dominant paradigm and engage in “normal science,” which is “research firmly based upon one or more past scientific achievements.”⁵⁰ Most scientists in a field generally subscribe to the same general paradigm, and perform experiments to define the details of the paradigm.⁵¹ And as scientists discover new information they adjust the prevailing theories to accommodate their results.⁵²

According to Kuhn, positive results are like weights on a scale in favor of the existing paradigm, while the negative results act like counterweights against the existing paradigm.⁵³ But when enough anomalous results accumulate, they can reach a critical mass that tips the balance of commonly accepted theories away from the predominant view and throw the field into a state of crisis.⁵⁴ When a field is in crisis, it no longer operates under a common set of beliefs, but instead resembles the immature state in which scientists constantly need to restate their basic assumptions to communicate a new idea.⁵⁵ In a time of crisis, scientists will begin retesting assumptions that had been widely-accepted, and a new competing paradigm may slowly develop. Eventually, the believers in the new paradigm will challenge the holdouts from the older system and the better system will prevail. A better scientific paradigm is one that accounts for more of the known data and more accurately predicts future outcomes.

⁴⁸ *Id.* at 43.

⁴⁹ *Id.* at 12–13. Prior to Kuhn, the word “paradigm” had been used only in connection with grammar and linguistics. Since Kuhn’s adoption of the word to science, however, the broader use of the word has become more prominently accepted.

⁵⁰ *Id.* at 10.

⁵¹ *Id.* at 64–65.

⁵² *Id.* at 66.

⁵³ *See id.* at 67–68.

⁵⁴ *Id.*

⁵⁵ *Id.* at 72.

Kuhn's philosophy has drawn its own criticisms, however. First, his ideas are not useful to predict future paradigm shifts.⁵⁶ One cannot say *ex ante* whether science is on the verge of a new paradigm shift.⁵⁷ Other critics have attacked Kuhn's idea that two paradigms in the same field must be incommensurate, meaning scientists cannot hold two different worldviews at the same time.⁵⁸ The incommensurate component of Kuhn's theory implies that it is not possible to prove whether the field is better off for undergoing a revolution because the new paradigm may carry with it an entirely new definition for the science.⁵⁹ Kuhn did not believe in "true" scientific theories; he believed merely that one theory may be a better predictor of future outcomes.⁶⁰ (This concept has been given the disparaging label of "relativism.")⁶¹ Moreover, observers have noted that scientific discoveries do not actually develop in well-defined leaps.⁶² Instead, they evolve slowly and, at any given time, many scientists may believe aspects of both old and new theories.⁶³ As even Kuhn recognized, scientists do not generally believe that they are developing a new paradigm, but usually think they are simply building on established ideas.⁶⁴

⁵⁶ Ulen, *supra* note 10, at 885.

⁵⁷ *Id.*

⁵⁸ See, e.g., Donald Davidson, *On the Very Idea of a Conceptual Scheme*, 47 PROCEEDINGS AND ADDRESSES OF THE AMERICAN PHILOSOPHICAL ASSOCIATION, 5 (1973–1974) (attacking the idea of conceptual relativism, the idea that reality is dependent on the framework with which it is viewed). Davidson argues that supporters of a certain worldview can always discuss other worldviews, and thus it is incoherent to claim that two views are completely incommensurate. *Id.*

⁵⁹ LARVOR, *supra* note 29, at 42.

⁶⁰ See *id.* at 43; see also KUHN, *supra* note 41, at 23–24; W. Keith Percival, *The Applicability of Kuhn's Paradigms to the History of Linguistics*, 52 LANGUAGE 285, 286 (1976); Jerone Stephens, *The Kuhnian Paradigm and Political Inquiry: An Appraisal*, 17 AM. J. OF POL. SCI. 467, 469 (1973) ("Kuhn [merely] replace[d] the term theory with the term paradigm[.]").

⁶¹ See LARVOR, *supra* note 29, at 37 ("The book caused a furore among philosophers of science, many of whom *accused* Kuhn of relativism . . .") (emphasis added), 43 ("This, to many philosophers, sounds like relativism.").

⁶² See KUHN, *supra* note 41, at 95–96; Percival, *supra* note 60, at 285–86, 286 n.1.

⁶³ See KUHN, *supra* note 41, at 98–103.

⁶⁴ *Id.* at 150–160.

C. Current Understandings in Philosophies of Science

Philosophers have now come to understand that theories do not stand by themselves, but rather, rely on many supporting (or auxiliary) theories.⁶⁵ Any theory about the interaction between celestial objects, for example, is based on supporting theories addressing issues of gravity, inertia, mass, and the make-up of the space around the objects. The success of any given theory is therefore dependent on the truth of its auxiliary theories.⁶⁶ Conversely, an experiment that initially seems to contradict a theory may, in fact, indicate that a theory is false, but it could also mean that an auxiliary theory is incorrect.⁶⁷ But because a researcher confronting an experimental result can never be certain whether the result reflects the theory being examined or a supporting concept, any particular theory is never fully confirmed or disconfirmed.⁶⁸

A key implication of this understanding is the importance of criticism in theory development: because no theory can ever be fully proven, experimentation that can undermine a theory is more powerful than establishing additional supporting evidence.⁶⁹ To illustrate, a theory that describes all known events of a given phenomenon may not necessarily be correct; other examples that the theory could not describe may not yet have been discovered or a different theory may fit the events even better. If scientists were to look only for more positive examples that are described by the theory, they would add modest new data for the field. For instance, Newton's theories accurately described the data known at his time, but as more accurate observations were collected the theory's flaws became apparent.⁷⁰ If instead of searching for supporting evidence, however, scientists critically challenge an accepted position, they can cast off potential weaknesses and can choose the theories that best describe empirical evidence.

⁶⁵ Maher & Gottesman, *supra* note 7, at 415; Smith, *supra* note 13, at 397–98; LARVOR, *supra* note 29, at 52 (discussing how Lakatos believed that criticisms of “cherished beliefs” are deflected towards a “protective belt” of supporting hypotheses).

⁶⁶ Smith, *supra* note 13, at 397–98; LARVOR, *supra* note 29, at 54–55 (explaining that supporting hypotheses must constantly change to reflect newly discovered anomalies, and so long as the changes are in the “spirit” of the underlying construct, the system is making progress).

⁶⁷ Smith, *supra* note 13, at 397 (“a negative empirical result could reflect the failure of any number of theories other than the core proposition that led to the empirical test.”).

⁶⁸ *Id.* at 397–98.

⁶⁹ W.B. WEIMER, NOTES ON THE METHODOLOGY OF SCIENTIFIC RESEARCH 40 (1979).

⁷⁰ LARVOR, *supra* note 29, at 52–55.

Legal scholars have developed a similar method for continuous critique and modification of theory.⁷¹ What sets the legal academy apart from the sciences, however, is that the sciences have embraced the power of empirical tests to support debates about theories.⁷² In contrast, law relies on criticism of theory in a more normative sense.⁷³ This difference between law and other academic fields results, at least in part, on the difficulty in measuring legal concepts. Physicists can use a ruler to measure the distance an object travels, anthropologists have developed sophisticated tests to measure the age of relics that they uncover, and chemists can use scales to weigh the product of a chemical reaction. But lawyers do not have a tool that can measure civil liberties, security, or freedom. Lawyers seeking to empirically test existing theories must first determine what to measure.

Fortunately, the complexities of measuring abstract concepts are not unique to law, and are common in the social sciences. Psychology in particular has developed sophisticated methods, based on the philosophies described above, with which to measure elusive notions such as when a person is actually depressed as opposed to just sad.⁷⁴ This subfield of psychology, called psychometrics, has developed over the last century to address the difficulty of treating concepts that are not easily quantifiable in a scientific manner.⁷⁵

II. PSYCHOMETRICS

Towards the end of the nineteenth century, the scientific world was still reacting to Darwin's publication of *On the Origin of Species by Means of Natural Selection, or The Preservation of Favoured Races in the Struggle for Life* (commonly known as *The Origin of Species*).⁷⁶ Today, Darwin's theory

⁷¹ Jack Goldsmith & Adrian Vermeule, *Empirical Methodology and Legal Scholarship*, 69 U. CHI. L. REV. 153, 154–59 (2002) (arguing against the overuse of empirical study in legal scholarship and describing the current models for developing legal theory).

⁷² *Id.* at 159.

⁷³ *Id.* at 156–57.

⁷⁴ See, e.g., JOHN RUST & SUSAN GOLOMBOK, *MODERN PSYCHOMETRICS: THE SCIENCE OF PSYCHOLOGICAL ASSESSMENT* 26–27 (2d ed. 1999).

⁷⁵ *Id.* at 4–5.

⁷⁶ CHARLES ROBERT DARWIN, *THE ORIGIN OF SPECIES BY MEANS OF NATURAL SELECTION, OR THE PRESERVATION OF FAVOURED RACES IN THE STRUGGLE FOR LIFE* (Adamant Media Corp. 2005) (1886). As to the world's reaction, see Edward Caudill, *The Bishop-Eaters: The Publicity Campaign for Darwin and On the Origin of Species*, 55 J. OF THE HIST. OF IDEAS 441, 441 (1994) (“A storm of publicity surrounded the *Origin* . . .”); John C. Greene, Book Review, 75 MOD. LANGUAGE NOTES 519, 519 (1960) (reviewing ALVAR ELLEGARD, *DARWIN* (continued)

of evolution proposed in the book is well-known. But less well-known is that Darwin also concluded, based on this theory, that humans consist of two subgroups—savages and civilized—and the distinction between the two was based at least in part on intelligence.⁷⁷ Although Darwin concluded that some people must have advanced to a higher intellect than others, up to that point no one could empirically test that conclusion.⁷⁸ The challenge of actually measuring a person's intelligence intrigued many philosophers.⁷⁹ The difficulty arose because raw intelligence was an abstract trait of a human mind and a metric for that trait was not obvious. To overcome this obstacle, philosophers began testing other measurable traits in the hopes that these observable variables would correlate with intelligence; their assumption was that even if it was not possible to directly measure intellect, they could make estimates based on a set of related characteristics.⁸⁰ Although the specific metrics used in the first intelligence test turned out to be not very accurate, the idea of using measurable traits to estimate an immeasurable quality formed the foundation for modern psychometrics.⁸¹

A. *The Introduction of Construct Validation*

Since the 1950s, behavioral scientists have been developing similar but more sophisticated tests with which to measure other unobservable phenomena such as intelligence, depression, happiness, and other mental

AND THE GENERAL READER: THE RECEPTION OF DARWIN'S THEORY OF EVOLUTION IN THE BRITISH PERIODICAL PRESS, 1859–1872 (1958)) (“[T]he idea of evolution . . . had made substantial gains among the general [British] public as early as 1863”).

⁷⁷ CHARLES DARWIN, *THE DESCENT OF MAN, AND SELECTION IN RELATION TO SEX* 34 (London, John Murray, 1871).

⁷⁸ RUST & GOLOMBOK, *supra* note 74, at 5. Unfortunately, Darwin's conclusion that intelligence was a genetic trait also led to misuse of the intelligence tests. The Eugenics movement was based on the belief that the quality of the human race could be improved by selective breeding based on the results of intelligence tests.

⁷⁹ *Id.* at 5 (describing how the evolution of intellect intrigued philosophers, such as Francis Galton and Karl Pearson, who went on to develop the first intelligence tests); Michael M. Sokal, *James McKeen Cattell and Mental Anthropometry: Nineteenth-Century Science and Reform and the Origins of Psychological Testing*, in *PSYCHOLOGICAL TESTING AND AMERICAN SOCIETY: 1890–1930* 21, 25–30 (Michael M. Sokal ed., 1987) (describing how in the 1880s James McKeen Cattell began empirically testing psychological phenomenon and Galton influence his work).

⁸⁰ RUST & GOLOMBOK, *supra* note 74, at 4–5 (explaining that Galton's first test measured visual and auditory skills as well as weight and other psychophysical variables, but that these tests were ultimately unsuccessful).

⁸¹ *Id.* at 6.

characteristics.⁸² One of the key difficulties in assessing these traits, however, has been in determining whether the tests are in fact measuring the correct feature. In 1955, psychologists Lee Cronbach and Paul Meehl articulated the necessity of construct validity; validity is the degree to which a given test actually measures the trait, or construct, that it is intended to measure.⁸³ They proscribed that “[c]onstruct validity must be investigated whenever no criterion or universe of content is accepted as entirely adequate to define the quality to be measured.”⁸⁴ In other words, when faced with an abstract concept with no convenient metric (like intelligence), an investigator must first find a measurable criterion that can be used to approximate the concept. But it is not enough to find just a single standard that may approximate the trait: it must also be valid, which means it should closely correlate with the construct being studied.⁸⁵

Because of the inherent difficulty in determining whether a given tangible criteria accurately reflects the value of the intangible trait it is being used to estimate, Cronbach and Meehl suggested that researchers “bootstrap” their way from a single criterion to achieve increasingly valid tests.⁸⁶ Once a valid criterion is found, it can be used to develop more valid tests. Over time, a test based on a valid criterion may become more accurate than even the original criterion itself.⁸⁷ For example, people first recognized the physical quality of temperature because certain objects felt different than others, and they used the sense of touch to measure this characteristic.⁸⁸ But early observers did not yet fathom that this feeling could be accurately gauged or scientifically tested beyond simple feel.⁸⁹ Researchers eventually noticed, however, that when certain elements felt hot, they expand and when they felt cool they contracted.⁹⁰ Further, certain elements, such as mercury, went through significant, noticeable expansion and contraction within the temperature range

⁸² Maher & Gottesman, *supra* note 7, at 418 (noting that by the 1950s psychological theories focused on studying hypothetical constructs by measuring observable biological processes); Smith, *supra* note 13, at 398 (explaining the introduction of construct validation and the informative tests of psychological theories).

⁸³ Lee J. Cronbach & Paul E. Meehl, *Construct Validity in Psychological Tests*, in A PAUL MEEHL READER 9, 11–12 (Niels G. Waller et al. eds., 2006).

⁸⁴ *Id.* at 10.

⁸⁵ *Id.* at 11.

⁸⁶ *Id.* at 14–15; *see also* Smith, *supra* note 13, at 397.

⁸⁷ Cronbach & Meehl, *supra* note 83, at 14.

⁸⁸ *Id.* at 15.

⁸⁹ *Id.* (describing the discovery of the relation of mercury expansion to heat).

⁹⁰ *Id.*

in which people lived. And the expansion correlated well with the original test for temperature—feel. Most important, though, was that the expansion correlated with not just one person’s judgment of temperature but with any other objective observer’s experience as well.⁹¹ It turned out that mercury was an even more accurate predictor of certain temperature-related events, such as boiling and melting points, than touch ever was. In the end, scientists were able to use feel as a measurement for temperature to bootstrap a more valid test using mercury.⁹² But as this example demonstrates, a test—like a scientific theory—can never be proven completely valid, and establishing validity must be an ongoing process.

It became apparent that like Popper’s philosophy of falsification, a vital component of construct validation is review and criticism.⁹³ Without critical assessment, a measurement’s validity can never be confirmed because it can not be compared to other known indicators. And when a metric cannot be tested through observable events, supporters may be tempted to rationalize its effectiveness. Cronbach and Meehl warned however that “[r]ationalization is not construct validation”;⁹⁴ researchers must develop hypotheses that critically test the measurement and a single hypothesis alone is not sufficient.

B. Modern Understandings of Construct Validation

Using a method called the multi-trait multi-method approach (“MTMM”), a researcher can further validate a metric by not only observing factors that should have positive correlations with the test, but also investigate factors that should be unrelated to the trait, assuming the theory is correct.⁹⁵ By using this approach, the validity of a test can be determined by measuring several traits—some that should have high correlations with the construct being studied and some that should not—and by using several different methods.⁹⁶ This way, a researcher can ensure that the results are an accurate reflection of the trait being studied and not a reflection of a related trait or an artifact from a particular experimental method.⁹⁷

If, after testing, the researcher discovers that the hypothesis is not supported, a reexamination of the overarching theory is necessary. But this examination is not simple because scientific theories are often based on many

⁹¹ *Id.*

⁹² *Id.*

⁹³ *Id.* at 18; Smith, *supra* note 13, at 398.

⁹⁴ Cronbach & Meehl, *supra* note 83, at 20.

⁹⁵ Smith, *supra* note 13, at 397; RUST & GOLOMBOK, *supra* note 74, at 73.

⁹⁶ RUST & GOLOMBOK, *supra* note 74, at 73.

⁹⁷ *Id.*

auxiliary theories.⁹⁸ Results can also be affected by elements of the test itself.⁹⁹ Hence, when empirical observations do not conform to theoretical predictions, the negative results do not necessarily mean the theory is wrong.¹⁰⁰ Instead, it may be that the theory is correct and the test is not accurate. On the other hand, it is possible that the test is accurate and the theory is not completely correct, or both the theory and the tests are not adequate, or the auxiliary theories are wrong. Unfortunately, because of these ambiguities, even if the hypothesis is supported, the researcher still cannot be completely confident in the theory.¹⁰¹ Once again, it is necessary to test the supporting theories in addition to testing the theory of interest.¹⁰²

When an experiment does confirm an hypothesis, the positive results can help establish the test's validity, but Cronbach and Meehl originally claimed that the test cannot last in the face of negative results.¹⁰³ In this way, their conception of construct validity resembles Popper's theory of verification.¹⁰⁴ But just as philosophers moved away from Popper's justificationist view, behavioral scientists have realized that tests are not fully invalidated in the face of negative information.¹⁰⁵ Rather, measurements must undergo constant revision.¹⁰⁶

To validate a test of a construct, behavioral scientists can follow a five-step method.¹⁰⁷ First, a theoretical construct, such as intelligence or happiness, should be specified.¹⁰⁸ Next, an informative hypothesis should be developed that will add to the knowledge of the field, rather than simply reaffirm existing ideas.¹⁰⁹ Then research experiments can be designed that critically test the hypothesis.¹¹⁰ Next, the data's correlation with the

⁹⁸ Paul E. Meehl, *Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology*, in A PAUL MEEHL READER, *supra* note 83, at 57, 74.

⁹⁹ *Id.*

¹⁰⁰ *Id.* at 73.

¹⁰¹ Smith, *supra* note 13, at 396.

¹⁰² *Id.*

¹⁰³ Cronbach & Meehl, *supra* note 83, at 18.

¹⁰⁴ A.J. Ayer, *Truth, Verification and Verisimilitude*, in THE PHILOSOPHY OF KARL POPPER, *supra* note 24, at 684, 684.

¹⁰⁵ *Id.*

¹⁰⁶ Smith, *supra* note 13, at 396.

¹⁰⁷ *Id.* at 399.

¹⁰⁸ *Id.*

¹⁰⁹ *Id.*

¹¹⁰ *Id.*

hypothesis' predictions can be calculated.¹¹¹ Finally, the theory should be revised to reflect the new knowledge gained from testing the hypothesis. Under this method, theories constantly evolve to reflect the updated knowledge of the field.¹¹²

III. MEASURING UNOBSERVABLE LEGAL CONSTRUCTS

Sciences that involve abstract concepts (in particular, the social sciences) share the common obstacle of measurement that is now confronting those wishing to approach legal questions scientifically. But because this dilemma is common among many different fields, legal scholars are able to use the experiences from the other fields to overcome these issues. This interdisciplinary approach is based on the idea of consilience, which is the "jumping together" of knowledge by the linking of facts and fact-based theory across disciplines to create a common groundwork of explanation."¹¹³ It is the idea that when one studies all levels of scientific phenomena, similar principles can be discerned, but *distinct* fields of study may not be foregone.¹¹⁴ For example, one of the most prevalent of these similarities is that many fields decide among multiple possible explanations for a given outcome by using Occam's razor, which states the principle of parsimony, or a preference for the least complex explanation for a given outcome.¹¹⁵ Because of these underlying symmetries between fields (such as Occam's razor) it is possible to solve some of the most complex problems in one field by integrating the knowledge from other disciplines.

Accordingly, I propose extending the concept of construct validation from psychology to legal studies. Construct validity looks for convergence and divergence of operations to make inferences about latent variables based on measurable traits that are influenced by the construct. For example, while happiness is unobservable, it is possible to measure happiness by testing behaviors such as the number of times people smile or the frequency of complimentary statements. Similarly, while deterrence may be an unobservable legal construct, it is possible to infer that a law has had a deterrent effect by measuring the frequency of unwanted behaviors or the number of arrests.¹¹⁶ As mentioned in the introduction, however, it is

¹¹¹ *Id.*

¹¹² *Id.* at 399–400.

¹¹³ EDWARD O. WILSON, *CONSILIENCE: THE UNITY OF KNOWLEDGE* 8 (1998).

¹¹⁴ *Id.*

¹¹⁵ Popper, *supra* note 23, at 128 (explaining why simplicity is so highly desirable).

¹¹⁶ See, e.g., David S. Lee & Justin McCrary, *Crime, Punishment, and Myopia* (Nat'l Bureau of Econ. Research, Working Paper No. 11491, 2005), available at <http://www.nber.org/> (continued)

important to remember that empirical observations should not be the sole basis for policy decisions. Even if a supported theory predicts a certain outcome, lawmakers must still consider more deontological concerns to decide whether those laws comport with the societies' qualitative values.¹¹⁷

The method I propose proceeds in four general steps: first, a researcher must develop generalized theories about the influence of law on society. Generalized theories should predict behavioral outcomes that one would expect as the result of a given type of law. Because unobservable legal constructs can be empirically studied only by noting divergent and convergent operations, a generalized theory should be open and suggest as many different operations as possible; as more operations are tested, confidence in the presence of the latent variable should increase. Second, based on the theory, the research must infer hypotheses that can challenge and provide useful information about the theory and specify operations to test hypotheses. Although many legal empirical studies have devised and tested hypotheses, some do not, and even fewer test rival hypotheses or address possible weaknesses in an accepted theory.¹¹⁸ But, as explained above, experiments that simply reaffirm an hypothesis are much less informative than those that address criticism of the theory.¹¹⁹ Third, the researcher should design quasi-experiments that can infer causal relationships even when the subjects being studied cannot be randomized. Finally, the original theory should be revised to incorporate inferences drawn from the new empirical information gained from the research.¹²⁰ The remainder of this article will more fully develop each one of these steps.

A. Developing Generalized Theories

The first step in measuring unobservable legal concepts is to develop a generalized theory that allows multiple inferences to be made about the unobservable construct.¹²¹ All theories are not created equal; that is, a hierarchy exists among theories in which some are more elemental and apply to specific situations, while other more generalized theories involve concepts

papers/w11491 (measuring deterrence by analyzing "offense rates" through the number of arrests).

¹¹⁷ See Goldsmith & Vermeule, *supra* note 71, at 154, 156 (discussing the value of rhetoric and qualitative analysis in legal scholarship that sets it apart from other social sciences).

¹¹⁸ Epstein & King, *supra* note 9, at 9, 76–80.

¹¹⁹ Smith, *supra* note 13, at 397–99.

¹²⁰ These four steps are based closely on a five-step model developed by Greg Smith for construct validity research in psychology. *Id.* at 399.

¹²¹ Ulen, *supra* note 10, at 897–99.

that can be tested under multiple conditions. For instance, a researcher studying a new educational model could devise a basic elemental theory that a certain student will perform better on tests if the new model were in place because she has been receptive to similar educational models. Unfortunately, this theory provides little information because the results are very specific and is affected by the student's particular characteristics. Hence, inferences that can be made about the quality of the program in a larger population are limited. A more generalized theory addressing how the curriculum will affect the entire class becomes more interesting; it allows the researcher to assess the program using multiple operations, rather than a single test. Hence, if a convergence of data occurs (namely, that a majority of the class improves after being subjected to the treatment), a stronger inference can be made about the quality of the program. But the class may still have its own individual traits that can influence the results as well, and an even more generalized theory regarding the effect of the new educational model on the entire school can provide still more information. Similarly, generalized theories about law can provide the best information about the influences laws can have on a population by allowing more inferences to be drawn about the legal construct.

Legal academics are skilled at developing and debating legal theories, but often the theories tend to be elemental, in that they are limited to the application of specific laws or concepts. Professor Thomas Ulen has pointed out, however, that a common attribute in most sciences is that theories are universally applicable without regard to specific governmental institutions.¹²²

In addition to the multiple inferences that can be drawn from generalized theories, disciplines that rely on generalized theories also have the advantage of better communication and more focused study across boundaries; scholars are able to address common issues and share their findings without regard to their personal location.¹²³ In economics, for example, researchers around the world share the same basic theoretical principles, such as the rational-actor theory, that allow their hypotheses to apply generally.¹²⁴ Because economic research is often based on generalized theories, economists in India can share their research with Americans because both groups speak the same theoretical language.

¹²² *Id.* at 896.

¹²³ *Id.* at 894–99.

¹²⁴ *Id.* at 895. Under the rational-choice model for human behavior, social scientists assume that people accurately determine the risks and benefits of their actions and act accordingly. *Id.* at 886. This is the current paradigm within the social sciences and is the foundation for much of its current literature. *Id.* at 886.

In contrast, legal studies are not as fundamentally based on accepted general theories that transcend political boundaries and form a theoretical paradigm within which legal researchers can more easily communicate across borders. Rather, legal scholars tend to focus on more elemental issues that affect their particular government.¹²⁵ More productive theories, however, address generalized concepts that are amenable to empirical measurement because they allow researchers to use multiple techniques to explore the theory. An empirical legal scholar studying an elemental theory about a single law passed in a small community can run only a limited number of experiments to explore the effects of the law; the community can be analyzed both before and after the law was passed or the community can be compared with similar communities that do not have the same law (experimental design will be discussed more thoroughly below). In contrast, generalized theories about a class of laws allow multiple approaches for experimental research: multiple communities with different laws within the class can be explored or the effect of that class of laws can be studied within the same community. While these same experiments could be conducted without a generalized theory, they would progress in a state similar to Kuhn's description of immature sciences because they would lack the cohesion that provides focus for future study.¹²⁶ A generalized theory could provide the needed cohesion and suggest potential follow-up research.

1. Drawing international interest in interesting topics

A general theory can draw more interest than an elemental theory and can advance the current knowledge for the field more quickly and robustly.¹²⁷ By developing theories that address interesting and widespread issues, they are more likely to draw commentary that can foster a developing community for scholars to knowledgeably debate each others' work. This community of critical discussion is especially important in law, which lacks significant peer-reviewed journals.¹²⁸ In other fields, studies are vetted for publication by experts based on the quality of the research and the logic of the

¹²⁵ *Id.* at 896–97.

¹²⁶ See KUHN, *supra* note 41, at 13–14.

¹²⁷ See Epstein & King, *supra* note 9, at 896–97.

¹²⁸ Gregory Mitchell, *Empirical Legal Scholarship as Scientific Dialogue*, 83 N.C. L. REV. 167, 171 n.14 (2005) (citing Arthur D. Austin, *The "Custom of Vetting" as a Substitute for Peer Review*, 23 ARIZ. L. REV. 1, 4–5 (1990) and Randall R. Bovbjerg, *Medical Malpractice: Research and Reform*, 79 VA. L. REV. 2155, 2186 n.135 (1993)).

conclusions.¹²⁹ And despite calls for a change, law currently lacks this filter.¹³⁰ Hence, the principal means to ensure quality work under the current regime is through discussion and debate. If a study produces controversial results, others can evaluate the methods and comment on the conclusions. In this way, well-executed studies can be celebrated and cited, while lower-grade experiments will be undermined by criticism and eventually relegated to obscurity. The best way to encourage the growth of these essential debates is for legal scholars to frame their theories in ways that are not limited to a specific statute or legal doctrine, but to question how the concept behind a law can result in observable behavioral alterations throughout a population.

In contrast to science, the elemental theories often investigated in law rarely attract interest across national boundaries.¹³¹ While the other social sciences benefit from diverse commentary from scholars internationally, legal debates tend to remain imprisoned within political boundaries. Professor Ulen argues that “there is no persuasive case for ‘legal exceptionalism’—i.e., for the view that law is inherently different from other academic disciplines that characterize themselves as scientific.”¹³² The obvious benefit from garnering increased attention beyond national borders is that more people can provide useful commentary, insight, and critical analysis of an idea. Legal experts in other countries may be able to provide differing perspectives on an issue and help provide a more nuanced view of the problem.

As an example of this phenomenon in science, psychology professor Greg Smith describes how the psychological theory of self-enhancement (the tendency to concentrate on one’s own strengths rather than weaknesses) was

¹²⁹ Epstein & King, *supra* note 9, at 48. Although studies have shown that peer-reviewed and peer-refereed journals may not be producing top quality material. *See, e.g.*, Frank Cross et al., *Above the Rules: A Response to Epstein and King*, 69 U. CHI. L. REV. 135, 148 n.97 (2002) (exploring studies of articles published in peer-reviewed journals and concluding that “reliability and validity is questionable”).

¹³⁰ *See, e.g.*, Epstein & King, *supra* note 9, at 128 (arguing that peer review will lead to an increase in the quality of empirical legal scholarship); Mitchell, *supra* note 128, at 180–88 (2004) (arguing that peer review would increase disclosure of key assumptions); Deborah L. Rhode, *Legal Scholarship*, 115 HARV. L. REV. 1327, 1360–61 (2002) (discussing the advantages of peer-refereed journals, including less introductory material in articles, better manuscript selection, and stricter page limits); Richard A. Posner, *The Future of the Student-Edited Law Review*, 47 STAN. L. REV. 1131, 1136 (1995) (arguing in favor of faculty-edited journals because “law review editors generally lack the competence to select and improve this scholarship”).

¹³¹ Ulen, *supra* note 10, at 895.

¹³² *Id.* at 899.

reworked based on information acquired through multicultural study.¹³³ Initially, behavioral scientists believed self-enhancement was universal, but cross-cultural testing showed that inhabitants of some Asian cultures failed to demonstrate the trait.¹³⁴ After further investigation, the researchers learned that members of these cultures did, in fact, self-enhance, but they exhibited the trait in a different manner: while members of individualistic cultures focused on individual behaviors, members of collectivist cultures placed more weight on collectivist behaviors (e.g., defending decisions made by the group).¹³⁵ The critical study of diverse cultures provided information that showed errors in the assumptions that supported the theory of the universality of self-enhancement in ways that domestic studies could not. The result of observing these varying conditions was a more detailed and complete understanding of the field.

This example also illuminates the nature of auxiliary theories: general theories depend on the validity of auxiliary, or supporting, theories,¹³⁶ and negative empirical data could be the consequence of faulty auxiliary theories rather than failures in the theory being examined. Theories that can be tested in other cultures can be used to undue this ambiguity by allowing international commentators to highlight the possible assumptions on which the theory is based. But despite the benefits of cross-cultural study, current legal research is limited almost exclusively to national issues and cases decided in United States courts.¹³⁷ And while that research is certainly beneficial, like the situation with self-enhancement in psychology, some common beliefs in law may be misguided because they are based on particular American qualities.

For example, most researchers who study crime assume that criminals act rationally and act according to a cost-benefit analysis, in which increasing prison terms lead to decreasing offenses.¹³⁸ As a result, when legislators

¹³³ Smith, *supra* note 13, at 396.

¹³⁴ *Id.*

¹³⁵ *Id.*

¹³⁶ See generally Imre Lakatos, *Lectures on Scientific Method*, in FOR AND AGAINST METHOD 19 (Matteo Motterlini, ed. 1999).

¹³⁷ Ulen, *supra* note 10, at 898.

¹³⁸ Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. OF POL. ECON. 169, 207–09 (1968) (concluding, for the first time, that crime rates are based on an economic calculation by a rational actor and that crime policy should be based on optimizing resources to achieve the most efficient level of crime); *but see* Lee & McCrary, *supra* note 116, at 1–2 (recognizing that most research on crime and crime rates is based on Becker’s economic analysis, but questioning the validity of this assumption).

became concerned about the spread of crack cocaine, they increased the mandatory sentences for selling crack as compared to powder cocaine.¹³⁹ But it may be possible that Americans have a particular aversion to longer prison terms.¹⁴⁰ Although the possibility seems unlikely, in other communities the very idea of being sentenced to spend time in prison may alone be such a significant deterrent that a mandatory minimum ten-year prison sentence will do little more than a one-year sentence. By performing cross-cultural studies examining the effect of longer prison terms on the sale of crack cocaine, researchers can determine whether the deterrent effect of increased prison terms on unlawful behavior is a universal constant or if it is a particular cultural trait. This knowledge can then inform decisions about future sentencing changes in the United States as well as other countries.

2. *An example in constitutional law*

The derivation and benefit of a generalized theory can be seen through an example posed by Richard Posner, who suggested questions about constitutional law that should be explored empirically.¹⁴¹ The first question he posited was: “What difference has it made for press freedom and police practices in the United States compared to England that we have a judicially enforceable Bill of Rights and England does not (or at least did not, before it became subject to the European Convention on Human Rights and Fundamental Freedoms)?”¹⁴² Any researcher attempting to empirically answer this question will quickly confront the difficulty in measuring “freedom of the press.” As a first step in empirically answering this question, then, one could pose a generalized theory that countries with explicit constitutional guarantees tend to express freedoms more vibrantly than countries that have not made express constitutional rights. Put this way, the question about the freedom of the press in the United States versus England becomes an elemental component of a more general question about constitutional law and its influence on societies.

¹³⁹ See Anti-Drug Abuse Act of 1986 § 1002(2), Pub. L. No. 99-570, 100 Stat. 3207, (1986) (codified as amended at 21 U.S.C. § 841(b)(1) (2000)); see also Donald Braman, *Criminal Law and the Pursuit of Equality*, 84 TEX. L. REV. 2097, 2114–17 (2006) (discussing the disparity).

¹⁴⁰ Although some empirical evidence indicates that increasing prison terms does not affect the rate of criminal offenses. Lee & McCrary, *supra* note 116, at 26–27 (discussing the possibility that young offenders do not act rationally with regard to the “price” of punishment).

¹⁴¹ RICHARD POSNER, *THE PROBLEMATICS OF MORAL AND LEGAL THEORY* 156–57 (1999).

¹⁴² *Id.* at 156.

Posing a theory in such a broad manner is helpful for several reasons. The first is the benefit gained from inviting a larger audience to comment. In addition to attracting First Amendment experts to the issue, scholars in various aspects of constitutional law may also become engaged in the debate.

And as the study of this question draws more commentary, more ideas will be generated that help drive the field; critics may run tests in an attempt to undermine the theory, while supporters may try to replicate the outcome of a positive test or otherwise defend the theory. In the end, a rigorous debate can develop with perspectives beyond the confines of First Amendment debate and the theory will become more refined and more accurate as empirical knowledge is acquired.

Another significant benefit from addressing the elemental question about the First Amendment as part of a more generalized issue is attention from abroad. Because the theory does not specifically address any one nation, legal scholars in many countries may contribute to the debate.¹⁴³ International legal experts may have the ability and experience to bring nuanced insights to the theory in the same way that cross-cultural experiments allowed behavioral scientists to refine their theories about self-enhancement. For instance, researchers may compare not only the United States and England, but also China, which specifically restricts speech and the press.¹⁴⁴ Analysis can also be done on certain Middle Eastern countries that allow only state-run press organizations. Perhaps, by empirically studying the varying relationships between governments and the local press (from constitutional freedoms to constitutional controls), scholars will be able to refine their understanding of the effects of constitutional rights in general.

Finally, the generalized theory about constitutional rights can create a unifying concept that can draw together studies that would otherwise run without direction. Studies of the Fourth Amendment's protection from search and seizure can be compared to the First Amendment's protection for the press. Also, studies conducted in other countries that explore their particular constitutional system do not have to stand in isolation, but can be contrasted with their American counterparts. And, like Kuhn's paradigms for science, the generalized theory of constitutional rights will suggest future empirical legal work to progress the current understanding of constitutional rights.

¹⁴³ Ulen, *supra* note 10, at 895.

¹⁴⁴ Peter Lin, *Between Theory and Practice: The Possibility of a Right to Free Speech in the People's Republic of China*, 4 J. CHINESE L. 257, 270 (1990).

3. *An example in intellectual property law*

Intellectual property (“IP”) provides an example of a law that is derived from generalized theories about incentives and rights, and has been extensively debated without regard to national borders. The basis of intellectual property laws are generally attributed to one of two basic theories: natural rights or utilitarian.¹⁴⁵ The theory of natural rights, based on the philosophies of John Locke, states that IP rights are an inherent part of the laws of nature—creators should own their creations.¹⁴⁶ In contrast, according to the utilitarian theory, IP is a legal right needed to provide an incentive for citizens to create; the public will forgo some of its ability to use new innovations by providing a limited monopoly right to creators, thereby providing a financial incentive for the creation.¹⁴⁷ Utilitarians believe that without IP law, innovations will not be utilized at their most efficient levels.¹⁴⁸ The American system of IP derived primarily from the utilitarian model, but debate continues as to the extent that the limited monopoly is necessary to drive invention.¹⁴⁹

Many of the most recent modifications to copyright law in particular were developed under the premise of the utilitarian theories and tend to provide increasing protections for creators in the hopes of spurring more innovation. For example, the Digital Millennium Copyright Act (DMCA)¹⁵⁰ makes criminal any tampering with effective encryption devices that protect

¹⁴⁵ Richard A. Epstein, *Liberty versus Property? Cracks in the Foundations of Copyright Law*, 42 SAN DIEGO L. REV. 1 (2005) (noting the two traditional “defenses” of intellectual property: natural law and utilitarian).

¹⁴⁶ See JOHN LOCKE, *THE SECOND TREATISE ON CIVIL GOVERNMENT* 20 (Prometheus Books 1986) (1690); see also Andrew R. Sommer, *Trouble on the Commons: A Lockean Justification for Patent Law Harmonization*, 87 J. PAT. & TRADEMARK OFF. SOC’Y 141, 154–57 (2005) (discussing the philosophical underpinnings of intellectual property law, and patent law in particular).

¹⁴⁷ Vincenzo Denicolò & Luigi Alberto Franzoni, *The contract theory of patents*, 23 INT’L REV. L. & ECON. 365, 368 (2003) (performing economic analysis on the utilitarian, or contract, theory for patent law and concluding that the reward of early disclosure for the public justifies granting monopoly rights to inventors).

¹⁴⁸ *Id.*

¹⁴⁹ See, e.g., Kal Raustiala & Christopher Sprigman, *The Piracy Paradox: Innovation and Intellectual Property in Fashion Design*, 92 VA. L. REV. 1687, 1688–89 (2006) (noting that utilitarian theory is reflected in American law, but noting that its applicability is not universal across all industries).

¹⁵⁰ Digital Millennium Copyright Act, Pub. L. No. 105-304, 112 Stat. 2860 (1998) (codified at 17 U.S.C. § 1201 (2000)).

copyrighted material, even if the encryption also prevents access to material that is not copyrighted or if the person wishing to gain access may intend a legitimate fair use of the copyrighted work.¹⁵¹ The Supreme Court has also recently allowed a twenty-year extension to the length of most copyrights.¹⁵² Opponents of these laws have argued that these additional protections were unnecessary because the existing laws provided sufficient incentives for artists and other innovators.¹⁵³

It is impossible to *directly* measure whether these types of changes in IP laws create increased “incentives,” but because of the generalized theories on which they are based it is possible to *indirectly* infer the degree the laws influence creation by empirically exploring the multiple traits affected by a changed incentive. For example, patent law has a standard duration beginning with the date of first filing for all inventions regardless of the development cycle of the product or the time it takes to successfully prosecute the patent,¹⁵⁴ so the effective term of patent protection for products on the market differs

¹⁵¹ See 17 U.S.C. §§ 1201, 1204 (2000) (outlawing circumvention of “a technological measure that effectively controls access to a work” and allowing fines and prison terms for violators); *Universal City Studios, Inc. v. Corley*, 273 F.3d 429, 459 (2d Cir. 2001) (stating that DMCA is not unconstitutional merely because it limits the methods with which copiers can access material subject to the doctrine of fair use); *but see Chamberlain Group, Inc. v. Skylink Technologies, Inc.*, 381 F.3d 1178, 1204 (Fed. Cir. 2004) (holding that defendant’s garage door opener did not violate DMCA by decrypting plaintiff’s door opening codes because plaintiff failed to show how access to the codes facilitated infringement).

¹⁵² *Eldred v. Ashcroft*, 537 U.S. 186, 205–06 (2003) (declaring constitutional the twenty-year extension for copyright law under the Copyright Term Extension Act (CTEA), Pub. L. No. 105-298, 112 Stat. 2827 (amending 17 U.S.C. §§ 302, 304), and noting that the act is a rational exercise of congressional power because it provides a “greater incentive” to create and disseminate work). The Court did not cite empirical evidence that a longer copyright term provides a greater incentive.

¹⁵³ See, e.g., LAWRENCE LESSIG, *FREE CULTURE* 133–35 (2004), available at <http://www.free-culture.cc/freeculture.pdf> (arguing that many copyrighted works are abandoned and, thus, the extended copyright terms serve primarily to “toll” the term before works pass into the public domain, especially considering that rights holders are no longer required to renew their copyrights after a 1976 change in the law); William F. Patry & Richard A. Posner, *Fair Use and Statutory Reform in the Wake Of Eldred*, 92 CAL. L. REV. 1639, 1640 (2004) (stating that the extension of the copyright term and the removal of the renewal requirement strike a “double blow” to the public domain).

¹⁵⁴ See 35 U.S.C. § 154(a)(2) (1994). See, e.g., John R. Allison & Mark A. Lemley, *Who’s Patenting What? An Empirical Exploration of Patent Prosecution*, 53 VAND. L. REV. 2099, 2118 (2000) (noting that patent prosecution duration varies depending on the industry and that duration length affects the term of protection).

depending on the length of development and prosecution time.¹⁵⁵ Hence, it is possible to infer the influence of the law by comparing the frequency with which inventors in industries with different development cycles or prosecution times apply for patents.¹⁵⁶ In the pharmaceutical industry, for instance, the time that a drug is actually for sale on the market with patent protection is shorter than most other products, thereby decreasing the financial benefits that the innovator can obtain by exploiting their monopoly rights.¹⁵⁷ Comparatively, in the electronics industry the patenting process is relatively quicker (and new innovations become obsolete faster), so the inventor can optimize their profit for the life of the product.¹⁵⁸ It is therefore possible to compare the patenting practices in these two industries, in which IP law provides varying protection, to learn more about the actual incentives the law provides.

Cross-cultural studies of intellectual property could provide similar benefits. American IP laws are based largely on the utilitarian theory that people are more willing to create if they receive a financial incentive for their work.¹⁵⁹ But again, the fact that Americans seem to produce more when they are given stronger intellectual property rights may be a reflection of the country's reliance on capitalism to reward positive behaviors.¹⁶⁰ In socialist societies, citizens may feel stronger motivations to innovate to provide greater benefits for the community, rather than for individual gain. By studying the influences of intellectual property laws in other countries, researchers can gain a fuller understanding of these effects.

¹⁵⁵ *Id.*

¹⁵⁶ Allison & Lemley, *supra* note 154, at 2114–15 (empirically investigating who is patenting and what types of patents are most often sought). The study found that nearly a quarter of patents were issued for computer-related innovations—a field with a relatively short development cycle.

¹⁵⁷ *Id.* at 2125–26 (finding that pharmaceutical and biological patent applications tended to be more involved and take longer than average, and thus decreasing the effective term of protection); *see also* SmithKline Beecham Corp. v. Apotex Corp., 247 F. Supp. 2d 1011, 1017–18 (N.D. Ill. 2003) (describing the effects of the patent process on the development cycle of drugs).

¹⁵⁸ Allison & Lemley, *supra* note 154, at 2125 (stating that pharmaceutical patents required significantly more filings than did electronics patents).

¹⁵⁹ *See* Epstein, *supra* note 145, at 1 (explaining the theories on which the American IP laws are based).

¹⁶⁰ *But see* Edwin Mansfield, *Patents and Innovation: An Empirical Study*, 32 MGMT. SCI. 173, 180 (1986) (finding that in fields other than pharmaceuticals, patents may not be necessary to spur innovation).

B. Infer multiple critical hypotheses

Because generalized theories do not apply to a single law, they allow researchers to infer multiple critical hypotheses that provide more detailed information about the unobserved legal construct. Latent variables cannot be determined from testing one hypothesis alone; they can be measured only after detecting trends of convergence and divergence of multiple operations conducted to test multiple hypotheses.¹⁶¹ As researchers accumulate evidence in support of a theory by testing the multiple hypotheses, confidence in the theory can increase.¹⁶²

As an example of how this “triangulation” is applied to an unobservable physical construct, consider how astronomers study an unobservable singularity—a black hole. Singularities do not reflect light, radio, or other signals, so it is impossible to directly see or hear them. But astronomers still have been able to examine several throughout the galaxy.¹⁶³ To do this, they first surmise the possibility of such celestial bodies based on theoretical considerations.¹⁶⁴ They then deduce that if such a body exists with a strong gravitational pull, then observable objects in the surrounding areas should be influenced.¹⁶⁵ Astronomers scour the skies to look not for the holes themselves, but for their effects on surrounding bodies. When they notice that a star or gasses move in an irregular pattern, the astronomers hypothesize that the irregularity may have been caused by the gravitational force of an unobserved body—possibly a singularity.¹⁶⁶ But even if the influence is the

¹⁶¹ See Donald T. Campbell & Donald W. Fiske, *Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix*, 56 PSYCHOL. BULL. 81 (1959), reprinted in *METHODOLOGY AND EPISTEMOLOGY FOR SOCIAL SCIENCE*, *supra* note 19, at 37 (listing aspects of the validation process, including convergence, which is “a confirmation by independent measurement procedures,” and discriminant validation).

¹⁶² Smith, *supra* note 13, at 397.

¹⁶³ See Anne P. Cowley, *Evidence for Black Holes in Stellar Binary Systems*, 30 ANN. REV. ASTRONOMY AND ASTROPHYSICS 287, 287 (1992) (summarizing the study of black holes); see, e.g., Y. Tanaka et al., *Gravitationally Redshifted Emission Implying an Accretion Disk and Massive Black Hole in the Active Galaxy MCG-6-30-15*, 375 NATURE 659, 659 (1995) (studying the dimensions of a black hole); Makoto Miyoshi et al., *Evidence for a Black Hole from High Rotation Velocities in a Sub-Parsec Region of NGC4258*, 373 NATURE 127, 127 (1995) (same).

¹⁶⁴ See Miyoshi, *supra* note 163, at 127 (noting that theoretical predictions indicate the presence of black holes at the centers of galaxies).

¹⁶⁵ *Id.*

¹⁶⁶ *Id.* (inferring the dimensions of a possible black hole based on the movement of surrounding rotating gases); see also Tanaka, *supra* note 163, at 659–61 (same).

singularity, it is still impossible to determine exactly where it is, how large it is, or how strong the gravitational pull.¹⁶⁷ If multiple observable bodies were affected in the same area, however, then it may be possible to infer details about the unobserved object without ever being able to detect it directly.¹⁶⁸ For instance, if orbits for several different objects were pulled in the same direction but at different severities, it may be possible to infer an approximate location for the singularity. But, just as a theory can never be proven completely true, astronomers cannot fully confirm the existence of a singularity because it can never be directly observed.

Unobserved legal constructs can similarly be inferred based on their influence on observable operations. And while a single operation may indicate the possibility that the unobservable trait exists, only after testing multiple hypotheses is it possible to reliably triangulate the details in “logical space.”¹⁶⁹ Moreover, by testing a larger number of hypotheses, more information about the unobservable trait can be gained.¹⁷⁰ For example, a generalized theory may state that capital punishment deters violent crime, and a researcher may infer the hypothesis that a state that employs the death penalty should have a lower rate of violent crime than one without it.¹⁷¹ But just as a single irregular orbit does not provide enough information to study a singularity, the mere decrease in violent crime does not provide a full picture of the deterrent effect of certain punishments.¹⁷² The crime rate may have been influenced by a third factor, such as an up-tick in the economy that

¹⁶⁷ See Cowley, *supra* note 163, at 287 (noting astronomers reluctance to accept even the existence of black holes).

¹⁶⁸ *Id.* (explaining the “numerous observational characteristics” necessary to infer details about black holes).

¹⁶⁹ See Campbell & Fiske, *supra* note 161, at 39–40 (discussing the necessity of numerous diverse criteria enabling the researcher to “triangulat[e] in logical space”).

¹⁷⁰ *Id.* (explaining the value of multiple tests for verifying the reliability of information).

¹⁷¹ See Cass R. Sunstein & Adrian Vermeule, *Is Capital Punishment Morally Required? Acts, Omissions, and Life-Life Tradeoffs*, 58 STAN. L. REV. 703, 706 n.9 (2005) (collecting studies showing that capital punishment has a powerful deterrent effect). Sunstein & Vermeule go on to argue that while opponents of the death penalty base their position that banning capital punishment will decrease the number of innocents killed through execution, more innocent victims may be killed by murders who are not adequately deterred without a threat to their own lives. *Id.* at 717–19. They concede, however, that the data is not conclusive and that more hypothesis-testing should take place before reaching a final conclusion. *Id.* at 716.

¹⁷² See, e.g., John J. Donohue & Justin Wolfers, *Uses and Abuses of Empirical Evidence in the Death Penalty Debate*, 58 STAN. L. REV. 791, 841–44 (2005) (concluding that year-to-year changes in multiple factors had a larger effect than execution policy on murder rates).

created more jobs for the unemployed who may have otherwise resorted to crime, or that prison conditions in general have decreased, thus increasing deterrence for all violent crimes.¹⁷³ Hence, to establish the validity of the deterrence theory, the researcher must also test hypotheses that test whether economic or other factors influenced the crime rate.¹⁷⁴ And although the theory can never be proven completely true, as more hypotheses are tested, the validity of the theory can become more certain.

But this example also demonstrates the necessity for critical hypotheses. Because the deterrent value of capital punishment can never be completely established, a powerful set of hypotheses does more than simply reinforce a theory. An hypothesis that, if proven, supports an existing theory can provide more confidence in that theory;¹⁷⁵ conversely, an hypothesis designed to address criticism of the theory provides more valuable information because it can show that the theory is false.¹⁷⁶ On the other hand, if the hypothesis fails, critics have one less avenue by which to attack the theory.¹⁷⁷ Powerful hypotheses eliminate as many criticisms as possible, and an hypothesis that completely undermines a critique can provide the strongest possible evidence in support of the theory. Hence, a powerful set of hypotheses for legal concepts will suggest experiments that can help rule out as many criticisms as possible. With regards to the theory about deterrence, for instance, a researcher may test the hypothesis that economic conditions in a state with capital punishment are actually responsible for any difference in observed crime rate.

1. Operationalization

Before hypotheses can be tested, however, the legal constructs being examined must be defined. Many terms have different meaning for different people; for instance, when Justice Potter Stewart was called on to define

¹⁷³ See, e.g., David P. Farrington & Darrick Jolliffe, *Crime and Justice in England and Wales, 1981–1999*, in *CRIME AND PUNISHMENT IN WESTERN COUNTRIES, 1980–1999* 41, 65–70 (Michael Tonry & David P. Farrington eds., 2005) (surveying studies that assess crime rates as they relate to economic and demographic factors); Lawrence Katz et al., *Prison Conditions, Capital Punishment, and Deterrence*, 5 *AM. L. & ECON. REV.* 318, 332–33 (2003) (concluding that prison conditions had a larger effect than executions on crime rates).

¹⁷⁴ See, e.g., Farrington & Jolliffe, *supra* note 173, at 63–65 (assessing various potential causes for changes in crime rates, such as increased opportunity for crime and changes in the level of security measures).

¹⁷⁵ Smith, *supra* note 13, at 397–98.

¹⁷⁶ *Id.* at 398.

¹⁷⁷ *Id.*

obscenity under United States law, he famously wrote that he'll know it when he sees it.¹⁷⁸ But for an empirical legal scholar trying to study the effects of a new law or Supreme Court case on the amount of obscenity, a more useful definition is necessary. A more functional definition for research may be that obscenity contains frontal nudity or explicit sex acts.¹⁷⁹ Based on this definition, anyone can determine if a film is obscene. While others may not necessarily agree with this definition, they will be able to understand and assess research that uses it.¹⁸⁰

Clearly defining a concept in a way that it can be measured, a process called operationalization, is a vital component to quantitative science.¹⁸¹ Operational definitions are descriptions of variables or constructs in terms of the specific validation tests used to measure them, rather than in terms of an intrinsic essence.¹⁸² Weight may be defined as the result of putting an object on a scale and temperature may be defined as the reading on a thermometer.¹⁸³

In psychology, mental retardation is operationally defined as a score of 70 or below on an IQ test.¹⁸⁴ Happiness can be defined in terms of facial expressions (such as smiling), tone of voice, and other observable characteristics; thus, if behavioral scientists wish to measure whether certain

¹⁷⁸ *Jacobellis v. Ohio*, 378 U.S. 184, 197 (1964) (Stewart J., concurring) (“under the First and Fourteenth Amendments criminal laws in this area are constitutionally limited to hardcore pornography. I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it, and the motion picture involved in this case is not that.”).

¹⁷⁹ *See Miller v. California*, 413 U.S. 15, 24 (1973) (limiting possible definitions of obscenity). The Court in *Miller* tried to provide a more specific definition for obscenity, by limiting it to “works which, taken as a whole, appeal to the prurient interest in sex, which portray sexual conduct in a patently offensive way, and which, taken as a whole, do not have serious literary, artistic, political, or scientific value.” *Id.* But even this definition is not operational because researchers’ opinions will differ on whether something appeals to the “prurient interest in sex” or is “patently offensive.”

¹⁸⁰ BRUCE L. BERG, *QUALITATIVE RESEARCH METHODS FOR THE SOCIAL SCIENCES* 25 (3d ed. 1998).

¹⁸¹ *Id.*

¹⁸² *Id.*; *but see* Marilyn B. Brewer et al., *Definitional versus Multiple Operationism*, 33 *SOCIOMETRY* 1 (1969), *reprinted in* *METHODOLOGY AND EPISTEMOLOGY FOR SOCIAL SCIENCE*, *supra* note 19, at 1, 37 (arguing against definitional operationism, which is defining theories only in terms of its operations).

¹⁸³ BERG, *supra* note 180, at 25.

¹⁸⁴ AM. PSYCHIATRIC ASS’N., *DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS (DSM-IV-TR)* 41–43 (4th ed. 2000).

treatment makes someone happy, they can count the number of times a person smiles in a given time-period after the treatment is administered.

Once a construct has been empirically studied, it may still be necessary to adjust the operational definition if the results do not correlate with expected operations.¹⁸⁵ For example, the term “genius” may be defined by a certain IQ score, but experts also expect geniuses to score well on achievement tests in school. If studies were to find, however, that students with genius IQs performed poorly on achievement tests, then the term genius may have to be redefined using a measure other than IQ score. The same is true in law: if obscenity was defined as the proportion of a film’s screen time devoted to showing nudity, but films that meet this definition of obscene do not also contain a higher than average number of sex acts, then the term obscenity may need to be redefined.

Operationalization is a skill at which lawyers should be particularly adept because the skills required to determine the precise definition of an ambiguous statute or the application of a complicated legal rule are similar to those required to define an abstract idea. Although law schools do not currently provide much training on how to analyze complicated statistical problems, lawyers are particularly trained in how to define difficult concepts.¹⁸⁶ When interpreting a statute, lawyers argue the meaning of particular words or concepts. In applying judicial decisions, lawyers must determine how to apply the idea put forth in previous cases to the facts of a new problem.

2. *Hypotheses for the constitutional rights example*

Turning back to Judge Posner’s question about the degree of liberty provided to the American press due to the First Amendment and our theory that freedoms explicitly proscribed in a constitution will provide more liberty, one notes immediately the need to operationally define “freedom.”¹⁸⁷ Although many scholars have debated the true meaning of the word freedom, empirical research requires a specific definition.¹⁸⁸ One possible definition

¹⁸⁵ Brewer et al., *supra* note 182, at 32 (noting that the most useful features of a theory change with new scientific discoveries).

¹⁸⁶ See Epstein & King, *supra* note 9, at 119–21 (discussing the need for more education at law schools on empirical methods).

¹⁸⁷ See *supra* Part III.A.2.

¹⁸⁸ See, e.g., Louis E. Wolcher, *The Paradox of Remedies: The Case of International Human Rights Law*, 38 COLUM. J. TRANSNAT’L L. 515, 555 (2000) (discussing various definitions of liberty, including “independence from constraints that are imposed by the state,”
(continued)

for freedom of the press is the proportion of stories and editorials that are critical of the present government as compared to the overall number of stories published. Conversely, in a country with less freedom of the press, reporters will be more reluctant to say anything negative about the presiding national rulers for fear of possible repercussions.

Once freedom is defined by the number of unfavorable stories towards the government, hypotheses can be inferred to test the theory. One might hypothesize that the United States will exhibit more freedom than England because of the constitutional protections for the press. Hence, to test the hypothesis an experiment might analyze the proportion of critical stories in the American press versus the British media. But just as a single irregular orbit is not sufficient to identify a singularity, this single hypothesis cannot support the theory alone. To more completely test the theory, a researcher must develop several hypotheses that can eliminate potential criticisms and help choose between alternate theories.¹⁸⁹ For instance, a critic may point out that the proportion of critical stories may actually be a reflection of an unpopular regime or a general cultural attitude towards authority. It is thus necessary to devise several hypotheses that can help establish the theory.¹⁹⁰

A more complete set of hypotheses to test freedom of the press would include the one dealing with stories critical of the government, but may also include other hypotheses. One hypothesis may be that any difference in disapproving reporting is the result of an unpopular government. To test this hypothesis, a researcher may compare opinion polls or collect data from various time periods. If this hypothesis proves false, then this criticism has been undermined and it is more likely that the press freedom is a result of constitutional protections. Another hypothesis could be that any differences are actually caused by a more widespread social acceptance of criticism of the government; some societies may view critical analysis as a positive trait for the press and will reward contrarian news outlets with increased readership regardless of the legal framework. To test this hypothesis, an experiment could be devised to survey different societies' view of the proper role of the press and compare these results with the presence or absence of an explicit

based on the philosophies of Locke, Voltaire, Montesquieu, and von Humboldt, and "obedience to the moral law," based on the philosophy of Kant.)

¹⁸⁹ This idea is similar to that developed in econometrics by Trygve Haavelmo in *The Probability Approach in Econometrics*, 12 *ECONOMETRICA* 1, 61, 66 (1944). He proposed that econometricians formulate a set of a priori admissible hypotheses to address a given theory. The set of hypotheses should be based on the existing theories at the time to help ensure that a proper hypothesis was included in the set.

¹⁹⁰ See Campbell & Fiske, *supra* note 161, at 37.

constitutional freedom or even the existence of laws limiting freedom of the press.

The benefit of the more generalized theory about constitutional freedoms is that it will also allow comparisons between countries other than the United States and England. Another way to test the theory, then, is to compare countries with constitutional rights similar to the First Amendment to countries with no explicit right, such as England, as well as countries with specific controls over the media, such as China. Governments around the world exert varying degrees of control over their local media, and studies of press freedom in these different countries will provide a clearer and more complete picture of the influence of the First Amendment.

3. *Hypotheses for the intellectual property example*

A similar process can be used to infer hypotheses about the incentives that intellectual property laws provide for creation. First an operational definition must be devised for the term “incentive.” In this case, incentive can be the number of patents and copyrights that are issued either by governments’ patent or copyright offices.¹⁹¹ In empirical studies, this measure has been further refined by limiting the analysis to “valuable” patents or copyrights.¹⁹² Because some countries do not have IP or equivalent offices, another definition for international study may be the number of creative products, such as books, CDs, or technological developments that are produced for the market. While critics may disagree with these definitions, the definitions still

¹⁹¹ While trademarks are a form of intellectual property, and could potentially provide an incentive to create, they should not help define incentives because the justification for trademarks is not the same as that for other forms of intellectual property. Copyrights and patents are designed to form limited monopoly rights for the IP rights holders. In contrast, trademark law is designed to protect consumers from confusion between products. Even though less consumer confusion may lead to a larger incentive to produce, it is more attenuated than the direct benefits derived from copyrights and patents.

¹⁹² Kimberly A. Moore, *Worthless Patents*, 20 BERKELEY TECH. L.J. 1521 (2005) (assessing the incentives provided by patent laws by measuring the number of “valuable” patents, rather than just the number of patents issued). Moore questioned whether just counting patents provided an adequate indication of incentive, and concluded that ignoring “worthless” patents provided better indicia. *Id.* at 1524–25. Moore then hypothesized that the value of a patent could be measured using observable characteristics. *Id.* at 1529. Moore determined which patents were worthless by recording several variables, including the number of claims, amount of prior art, and time the application spent in prosecution. *Id.* at 1528–29.

provide a vital purpose because the critics will be able to understand and assess the research.¹⁹³

After “incentive” has been defined, it is possible to infer an hypothesis to test the concept. A first hypothesis could be that a longer duration for the limited monopoly protection provided by copyright law will provide more incentive to create; if IP laws provide incentives to create, then stronger laws that provide increased protection should create stronger incentives.¹⁹⁴ One way to test this hypothesis is by comparing the numbers of copyrights sought before and after copyright protections were extended by twenty years in 1998 to include the life of the author plus 70 years. But this test alone is not sufficient to support or undermine the theory. Assuming the number of copyrights issued increased after the passage of the extension, other factors may have influenced the change; for example, the Digital Millennium Copyright Act (DMCA), which criminalizes decryption devices used to pirate copyrighted material, may have increased the value of the copyright independent of the change in duration, or the two laws may somehow interact to produce results that neither would have alone.¹⁹⁵ But even if the results show no change in the number of copyrights issued, those results are not definitive on their own.¹⁹⁶ For instance, empirical evidence indicates that the value of most copyrights depreciates rapidly, so an increase from 50 to 70 years may not affect the incentive to create but a decrease in duration from 50 to 10 years might.¹⁹⁷

Hence, it is necessary to infer other critical hypotheses to test the theory. For example, production of creative material can be compared between the United States and China, which does not have the same level of IP protection.¹⁹⁸ The generalized theory allows study in countries operating

¹⁹³ *Id.* at 1524–25 (questioning the traditional method of measuring incentive by measuring all patents issued).

¹⁹⁴ *See, e.g.*, 144 CONG. REC. H1458–59 (daily ed. Mar. 25, 1998) (statement of Rep. Frank) (arguing that copyright protection should be extended because increasing the duration of the copyright term will increase the incentive for songwriters).

¹⁹⁵ 17 U.S.C. §§ 1201, 1204 (2000).

¹⁹⁶ *See, e.g.*, William M. Landes & Richard A. Posner, *Indefinitely Renewable Copyright*, 70 U. CHI. L. REV. 471, 474–75 (2003) (empirically analyzing the effect of an increased copyright term on innovation).

¹⁹⁷ *Id.* at 473–74.

¹⁹⁸ *See, e.g.*, Alejandro Zentner, *File Sharing and International Sales of Copyrighted Music: An Empirical Analysis with a Panel of Countries*, 5 TOPICS IN ECON. ANALYSIS & POL’Y 1 (2005), available at <http://www.bepress.com/bejeap/topics/vol5/iss1/art21/> (comparing music sales across countries with different copyright laws and internet penetration).

under various IP regimes to provide a more precise understanding of the incentives provided by the laws. Other beneficial hypotheses may address areas in which people create without IP protection in an environment in which protections may be available. For example, while academic journals sell their publications, the individual authors rarely receive a direct financial benefit from producing work. Similarly, a recent study explored aspects of the fashion industry in which piracy of designs is common, yet designers continue to create at a high rate.¹⁹⁹

C. Testing the hypothesis package

Once a set of critical hypotheses has been inferred from the generalized theory, the hypotheses must be rigorously tested using appropriate research designs. Application of proper experimental design to empirical legal scholarship has been debated extensively.²⁰⁰ Although the details of research design are beyond the scope of this paper, for the sake of completeness this section will briefly outline some of the basic concepts.

Empirical legal studies are generally conducted for three reasons: to collect data, to summarize data, or to make descriptive or causal inferences.²⁰¹

Because the legal community performs many activities that produce huge amounts of raw statistical data, the mere collection of the data is useful; but researchers must compile and organize the information so it can be used for future analysis.²⁰² Summarizing the data in a comprehensive but manageable format is also beneficial so the information can be understood quickly. But the most beneficial purpose of empirical research is to draw causal inferences—that is, past outcomes can be used to infer future results.²⁰³ Scholars can use information about the past to make two types of inferences about future events: descriptive inferences (predictions made about an entire population based on data gathered from a smaller sample set) and causal inferences (determinations of the factors that influence a given outcome). Researchers have made descriptive inference about future events after studying a power company's decision to purchase an entire town rather than

¹⁹⁹ See Raustiala & Sprigman, *supra* note 149, at 1732–34 (summarizing their conclusion that the fashion industry thrives absent strong intellectual property protections for designs).

²⁰⁰ See, e.g., Epstein & King, *supra* note 9, at 1 (proposing methodological improvements for empirical legal scholarship).

²⁰¹ *Id.* at 20, 24, 29 (describing three purposes for empirical legal scholarship).

²⁰² *Id.* at 20.

²⁰³ *Id.* at 29.

clean up its pollution.²⁰⁴ The results can be used to surmise how nuisance disputes will be resolved in other communities in the future.²⁰⁵ A causal inference has been useful in determining whether affirmative action laws are contributing factor to any change in the number of black lawyers.²⁰⁶

1. *Quasi-experiments*

After the goal of a particular research project has been identified, an appropriate design can be selected. Research designs are the initial plans that provide the structure for the project and are generally categorized into three groups: true (or randomized) experimental designs, quasi-experimental designs, and non-experimental designs.²⁰⁷ Research using randomized experimental designs (the most highly recommended designs) is used to establish whether a given treatment causes a specific effect: i.e., if treatment X is applied, then Y should result.²⁰⁸ But to reinforce the analysis of the causal effect, the corollary must also be true: i.e., if treatment X is *not* applied, then Y should *not* result. To make these parallel determinations, subjects from a common population are randomly assigned to either a treatment or a control group; subjects in the first group receive the treatment, whereas subject in the other group do not.²⁰⁹ Because subjects are assigned to the two groups at random, they are assumed to be essentially the same, or equivalent.²¹⁰ A randomized experiment provides evidence in favor of an hypothesis when the predicted result occurs more often in the treatment group than the control.²¹¹

²⁰⁴ See Gideon Parchomovsky & Peter Siegelman, *Selling Mayberry: Communities and Individuals in Law and Economics*, 92 CAL. L. REV. 75 (2004) (investigating the effects of the operation of a large factory on the local town of Cheshire, Ohio).

²⁰⁵ *Id.* at 124–32 (describing the implications for various theories of law of their investigation of the single case study of the town of Cheshire).

²⁰⁶ Epstein & King, *supra* note 9, at 77–80.

²⁰⁷ Campbell, *supra* note 19, at 191 (noting that it may not always be “feasible to assign units at random,” and in those cases a quasi-experimental design can still be useful to evaluate program impacts).

²⁰⁸ DONALD T. CAMPBELL & JULIAN C. STANLEY, EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH 13 (1963) (stating that true randomized experimental designs are most highly recommended and that they compare treatment with the absence of a treatment).

²⁰⁹ LARRY B. CHRISTENSEN, EXPERIMENTAL METHODOLOGY 227 (4th ed. 1988) (explaining that subjects in the control group must be similar to subjects in the experimental groups).

²¹⁰ *Id.* at 174–76 (describing how randomization can eliminate systemic bias by providing an equiprobability of events).

²¹¹ See *id.* at 174–175.

These types of experiments have high internal validity; that is, the treatment employed was probably the cause of the effect observed.²¹²

But it is often inappropriate or impossible to randomly assign subjects to study legal concepts.²¹³ For example, a true experiment of the deterrent effect of longer prison sentences would involve randomly assigning individuals to groups with varying minimum sentences to determine whether the group with the longer sentences commits fewer criminal offenses.²¹⁴ But, as a practical matter, such an experimental design is not feasible. Hence, empirical legal scholarship generally employs types of quasi-experiments devised by Professors Donald T. Campbell and Julian C. Stanley.²¹⁵ These designs resemble true experiments, but without randomized assignment.²¹⁶ The lack of randomization means that quasi-experiments generally have lower internal validity than randomized experiments; it cannot be definitively determined whether an observed outcome was caused by the study's treatment or by a threat to the validity caused by an unidentified influence present in one group but not the other.²¹⁷ Hence, in all quasi-experiments it is important to identify possible threats to internal validity.²¹⁸

²¹² Donald T. Campbell, *Factors Relevant to the Validity of Experiments in Social Settings*, reprinted in *METHODOLOGY AND EPISTEMOLOGY FOR SOCIAL SCIENCE*, *supra* note 19, at 151, 151 (defining internal validity as questioning whether the experimental stimuli in fact make some significant difference).

²¹³ See, e.g., Albert Yoon & Tom Baker, *Offer-of-Judgment Rules and Civil Litigation: An Empirical Study of Automobile Insurance Litigation in the East*, 59 *VAND. L. REV.* 155, 172–73 (2006) (using quasi-experimental approach to measure effects of rules); Donohue & Wolfers, *supra* note 172, at 820–22 (discussing death penalty studies that used quasi-empirical designs and the necessity of using these designs to study the death penalty because subjects cannot be assigned to randomized groups); Linda Jellum & Emmeline Paulette Reeves, *Cool Data on a Hot Issue: Empirical Evidence that a Law School Bar Support Program Enhances Bar Performance*, 5 *NEV. L.J.* 646, 669–70 (2005) (using quasi-experimental design to empirically assess the impact bar preparation classes have on bar exam passage rates); Steven L. Chanenson, *Get the Facts, Jack! Empirical Research and the Changing Constitutional Landscape of Consent Searches*, 71 *TENN. L. REV.* 399, 455–59 (2004) (proposing a research agenda including quasi-experiments to study the effectiveness of consensual searches).

²¹⁴ See, e.g., Lee & McCrary, *supra* note 116, at 2–3 (discussing difficulty in measuring deterrent effect of sentencing and concluding that a quasi-experiment best addresses the issue).

²¹⁵ *CAMPBELL & STANLEY*, *supra* note 208, at 13.

²¹⁶ *Id.* at 34 (explaining that social researchers can use one of several quasi-experimental designs that will provide reliable data when randomization is not possible).

²¹⁷ See *COOK & CAMPBELL*, *supra* note 16, at 51–53 and *CAMPBELL & STANLEY*, *supra* note 208, at 5 for descriptions of the threats to internal validity arising from several classes of extraneous variables: history (some difference in the groups' histories changes how they react

(continued)

Researchers can control for most threats to internal validity by utilizing one of three general strategies: relabeling, substitution, or elaboration.²¹⁹ Relabeling is used when the threat results from a mislabeled effect; that is, when the cause of an effect is attributed to the treatment when, in fact, it was influenced by a third factor.²²⁰ One way to cure this problem, therefore, is to just relabel the effect to better describe the true influences.²²¹ Substitution, the second strategy for ruling out a threat, is when a measurement that is subject to the threat is replaced by one that is not; for example, by using a randomized experiment rather than a non-randomized design to overcome selection bias.²²² Finally, elaboration removes a threat to validity by adding additional comparisons to the experiment to try to disentangle the effect being observed from the threat.²²³

Several different forms of quasi-experiments have been devised with differing threats to the internal validity. Some of the most popular forms used in empirical legal scholarship involve “nonequivalent groups,” in which intact groups, such as states, cities, judicial jurisdictions, etc., are selected and presumed to be similar (but not equivalent as they would be if randomly assigned).²²⁴ The Interrupted Time-Series with Comparison Series

to the treatment); maturation (the groups’ differ in their rate of maturation in relation to the treatment for reasons unrelated to the test); testing (somehow the pretest changed how the groups’ approached the posttest); instrumentation (other influences on one group versus the other between the pretest and posttest, such as different observers); regression (different rates of regression to the mean between the groups, such as when one group has more extreme pretest scores, so they have further to regress); selection (the subjects have differences from one group to the other); mortality (different dropout rates between the groups); and interactions between these biases and selection.

²¹⁸ Shadish, *supra* note 20, at 16.

²¹⁹ Charles S. Reichardt, *A Typology of Strategies for Ruling Out Threats to Validity*, in 2 RESEARCH DESIGN: DONALD CAMPBELL’S LEGACY, *supra* note 20, at 89, 89–90.

²²⁰ *Id.* at 91.

²²¹ *Id.*

²²² *Id.* at 92–93.

²²³ *Id.* at 94. *See also id.* at 94–108 (describing five methods of elaboration: demonstrate that the threat has no actual effect; subtract the size of the effect from the results; vary the size of the treatment effect; vary the size of the threat effect; and vary the size of both the treatment and threat effects).

²²⁴ CAMPBELL & STANLEY, *supra* note 208, at 47. Social scientists use several different quasi-experimental designs: time-series, equivalent time-sample, equivalent material, nonequivalent control group, counterbalanced, separate-sample pretest-posttest, separate-sample pretest-posttest control group, multiple time-series, recurrent institutional cycle, and regression-discontinuity analysis. *See id.* at 5.

(“Comparison Series”) approach compares similar groups both before and after a treatment, such as a new law or uniform change in the subjects’ lives.²²⁵ If the two groups are substantially similar before the law is passed, and the observations being measured are the same pretreatment but diverge posttreatment, it is reasonable to conclude that the law affected the measurement. Even so, plausible alternative explanations, such as a change in other local conditions or changes in other laws, should always be explored.

Professors Yoon and Baker’s empirical analysis of the effects of a New Jersey offer-of-judgment court rule is a recent example of a Comparison Series research design.²²⁶ Under the rule they studied, either party to a suit can offer a settlement to the opposing party; if the party receiving the offer refuses but goes on to lose the case, that party must pay all litigation expenses including attorney fees that were incurred after the offer was made.²²⁷ Originally the rule set a cap on the attorney fees at \$750, but the cap was later removed in amendments to the statute.²²⁸ The professors collected data from before and after the rule was revised to see if the increased cost-shifting associated with the amendment had an effect on settlement rates. To fully analyze the effects of the rule, the study analyzed data from in-court trials as well as from out-of-court settlements.²²⁹ But because settlement information is usually not publicly available, Yoon and Baker arranged with a major American insurance company to use their confidential settlement data.²³⁰ The study was designed so that the time period observed extended an equal duration before and after that change in law.²³¹ The dataset also included information from five other states other than New Jersey to serve as the control group—these other states did not change their offer-of-judgment rules during this time period.²³² Suits filed in New Jersey comprised nearly 20% of the data that the insurance company supplied.²³³ The study concluded that removing the cap did not increase the number of settlements, but did decrease the average time to resolve a suit.²³⁴

²²⁵ COOK & CAMPBELL, *supra* note 16, at 207.

²²⁶ Yoon & Baker, *supra* note 213, at 172–73.

²²⁷ N.J. Ct. R. 4:58. *Cf.* Fed. R. Civ. P. 68.

²²⁸ Yoon & Baker, *supra* note 213, at 159.

²²⁹ *Id.*

²³⁰ *Id.* at 165.

²³¹ *Id.*

²³² *Id.* at 166.

²³³ *Id.* at 169.

²³⁴ *Id.* at 178.

Yoon and Baker approached the data using what they called an econometric approach.²³⁵ If the researchers were able to run a randomized experiment, they would have randomly assigned subjects to control and treatment groups that were equivalent in every respect except for the variable being tested, in this case the cap on damages. But because it was impossible to randomly assign citizens to either a city with or without the new rule, the researchers compared suits from a control group that consisted of states with similar demographics and size to New Jersey (the treatment group).²³⁶ The control states did not have a change in their offer-of-judgment rules, but the treatment state did. In this way, the researchers made the quasi-experiment as close to a true experiment as possible by limiting the number of unexpected or exogenous factors that could influence the results.

But Comparison Series design is only one type of quasi-experiment. Other variations may be more appropriate depending on the situational limitations, such as if only one group can be observed,²³⁷ or if only posttests can be conducted.²³⁸ Similarly, other designs can be used with various analytical strengths and weaknesses, and using them in combination can strengthen confidence in causal inferences.²³⁹

2. *Adopting research design to empirical legal scholarship*

Epstein and King have proposed a method for empirical legal scholarship based on the existing methodological rules applied in the social sciences.²⁴⁰ They propose the following procedure for legal researchers: (1) develop research questions that contribute to existing knowledge and improve the real world; (2) formulate well-reasoned hypotheses about how societies will react if the theories are true; (3) bolster the theory by developing rival hypotheses that test the accuracy of the original theory; (4) use valid and reliable measurements to test the hypotheses; and (5) select the appropriate observations to include in the study.²⁴¹ They also suggest that the legal academy adopt an infrastructure that is more conducive to empirical work by offering more courses to teach law students how to properly conduct

²³⁵ *Id.* at 172.

²³⁶ *Id.* at 173–74.

²³⁷ This design is called a one-group pretest-posttest design. CAMPBELL & STANLEY, *supra* note 208, at 7.

²³⁸ This design is called a nonequivalent posttest-only design. CHRISTENSEN, *supra* note 209, at 224–25.

²³⁹ COOK & CAMPBELL, *supra* note 16, at 103–46.

²⁴⁰ Epstein & King, *supra* note 9, at 11.

²⁴¹ *Id.* at 54.

empirical research, train law professors about empirical techniques and provide more resources to use these skills, and create more expert review for empirical articles submitted to law reviews.²⁴²

Epstein and King also note the importance of articulating the methods used as precisely as possible. An experiment does not provide any value to the field if readers cannot decipher how a test was performed; results alone are meaningless without an explanation of how they were achieved. A test is useful when it rules out competing theories,²⁴³ but if it is not clear how a test is performed, hypotheses are not ruled out because readers cannot determine if the conclusions are based on the effects being studied or an unarticulated internal design flaw. Negative tests results may be caused by an incorrect auxiliary theory rather than a failure of the theory being examined.²⁴⁴ But if the methods are not clearly articulated, it is impossible to tell if the theory is wrong or if it is relying on a poor supporting hypothesis. Essentially, the author of a study that does not fully explain the method used is asking readers to simply trust the author without ever verifying the conclusions.

3. *Replicability*

A related reason to fully articulate the experimental methods is so the test can be repeated by others to determine whether the study measured a real event or if another factor may have influenced the outcome.²⁴⁵ The results from a single experiment may reflect the phenomenon being studied but they also could reflect a chance outcome or a variable that the designer did not anticipate. For example, study results that are based on survey questions may be influenced by the medium; survey results may differ depending on whether they were gathered online, over the telephone, or on paper.²⁴⁶ More subtle details can also influence the data such as how information is grouped on a page,²⁴⁷ the number of words in questions, or location of questions on a

²⁴² *Id.* at 114.

²⁴³ See Smith, *supra* note 13, at 398–99.

²⁴⁴ *Id.* at 397–98.

²⁴⁵ Gary King, *Replication, Replication*, 28 PS: POL. SCI. & POLITICS 444 (1995), available at <http://gking.harvard.edu/files/replication.pdf> (explaining the importance of the replication standard, requiring enough information about the methodology used in an experiment to allow subsequent researchers to replicate the test).

²⁴⁶ See, e.g., Elizabeth Fanning, *Formatting a Paper-based Survey Questionnaire: Best Practices*, 10 PRACTICAL ASSESSMENT RESEARCH & EVALUATION, 1, (2005), available at <http://pareonline.net/pdf/v10n12.pdf> (describing survey design to external influences).

²⁴⁷ Jolene D. Smyth et al., *How Visual Grouping Influences Answers to Internet Surveys* (Social Economic Sciences Research Center, Technical Report No. 04-023), available at (continued)

page.²⁴⁸ But it cannot be determined from the single experiment alone if the results are a reflection of the variables being manipulated, random happenstance, or a third unaccounted for factor. Hence, replication is essential.²⁴⁹ As a study is replicated several times, researchers can have more confidence that the outcomes are the result of the variables being studied and not just a random occurrence. But researchers wishing to repeat the study require detailed instructions to ensure their results are not influenced by other outside factors. More general descriptions of methods are more prone to lead subsequent researchers astray.²⁵⁰

Robert Thompson's empirical analysis of courts' willingness to pierce the corporate veil in corporate liability cases is an example of a legal study with a detailed description of its methods.²⁵¹ Because of the interest in the topic and general confidence in the results, this study has become well-cited and has been referenced by both trial and appellate courts in federal and state cases.²⁵²

Thompson spends several pages of his article providing details about how he performed his research.²⁵³ The article first described the Westlaw searches he conducted to collect the relevant court decisions.²⁵⁴ He then explains that the original search results needed to be vetted to ensure that the cases that resulted from the search were actually related to corporate law.²⁵⁵ Finally, the article elucidates the specific information gleaned from each case.²⁵⁶ Overall, this description gives readers a detailed roadmap of the data collection process and how the results were analyzed.

<http://survey.sesrc.wsu.edu/dillman/papers.htm> (examining whether visual grouping influences answers to survey questions).

²⁴⁸ Cleo D. Redline et al., *Factors that Influence Reading and Comprehension in Self-Administered Questionnaires*, (paper presented at the Workshop on Item-Nonresponse and Data Quality, Basel, Switzerland, Oct. 10, 2003), available at <http://www.sesrc.wsu.edu/dillman/papers/Basel%20submission%20dillman.pdf>.

²⁴⁹ But see King, *supra* note 245, at 444 (stating that actual replication is not necessary, nor is it always possible, but providing enough information to allow replication is essential).

²⁵⁰ *Id.* at 444–45.

²⁵¹ Robert B. Thompson, *Piercing the Corporate Veil: An Empirical Study*, 76 CORNELL L. REV. 1036 (1991).

²⁵² See, e.g., *McKesson HBOC, Inc. v. New York State Common Ret. Fund, Inc.*, 339 F.3d 1087, 1095 (9th Cir. 2003); *Stephens v. Am. Home Assur. Co.*, 811 F. Supp. 937, 951 (S.D.N.Y. 1993).

²⁵³ Thompson, *supra* note 251, at 1044–47.

²⁵⁴ *Id.* at 1036 n.1.

²⁵⁵ *Id.* at 1044–47.

²⁵⁶ *Id.* at 1044.

But to make this study more easily replicated, it would be helpful to add still more detail about the methods, in particular, how the cases were vetted to ensure they focused on the correct topic.²⁵⁷ While the article precisely describes the specific searches conducted,²⁵⁸ it does not completely explain how the results were filtered. It states that cases “that did not address corporate law” were eliminated from the study,²⁵⁹ but before subsequent researchers can replicate this study, they would still need to know how Thompson defined “corporate law,” and how it was determined what the case “addressed.” The article does not explain whether his filter allowed only corporate-liability actions or if it also counted insider-trading, antitrust, or litigation of other corporate laws. The methods section states that Thompson’s research assistants made the final decisions about which cases to include in the study, but does not explain how they were instructed to make these decisions.²⁶⁰ If an appendix had been included with the exact filtering instructions, future researchers would be able to more closely replicate the study.²⁶¹

Unfortunately, even with this level of detail, follow-up studies may not arrive at the same results if the instructions allowed the assistants to exercise too much independent discretion to decide which cases to include; different readers may make different decisions about close cases.²⁶² And without more information about the screening process, readers cannot determine if the results may have been influenced by the way the cases were gathered. Although this level of methodological detail may seem trivial, it can have profound effects on whether replication of the study will achieve the same results and ensure the reliability of the study’s conclusions. As is clear in the next example, one solution for this problem is to allow subsequent researchers access to the final datasets.

Yoon and Baker’s study also lacks an important element for replicability.

While using the information from a large insurance company seemed ideal because it is one of the few entities involved in many different lawsuits that also keep detailed information about the entire transaction, before allowing access to the data, the insurance company required that the researchers keep

²⁵⁷ See King, *supra* note 245, at 446 (noting that it is not always possible to include all information in the paper itself, but replication data sets as well as detailed descriptions of decision rules can be made available through digital archives).

²⁵⁸ Thompson, *supra* note 251, at 1036 n.1, 1044 n.47.

²⁵⁹ *Id.* at 1044.

²⁶⁰ *Id.* at 1044 n.48.

²⁶¹ See King, *supra* note 245, at 446.

²⁶² Thompson, *supra* note 251, at 1044 n.48.

the company's identity secret.²⁶³ So even though the methods were set out in detail (the article describes the exact information the insurance company provided, included a table that summarized the data that the insurance company provided, and described the specific statistical analyses performed),²⁶⁴ because the dataset is confidential, subsequent investigators cannot attempt to replicate the study.

The inability to replicate has several implications. First, the study may have basic math errors that cannot be checked. Yoon and Baker thoroughly tested and retested their data, but if subsequent investigators wish to rely on this study for future experiments, they will not have the confidence gained from analyzing the data themselves or the ability to use different statistical analyses on the data. For example, critics may wish to analyze some underlying assumptions on which the study is based. Supporters of the research may wish to build on this research by performing even more sophisticated statistical analysis than the ones run in this study. But currently law has not adopted an accepted norm to cope with confidential information. Confidential data is a common problem in medical and psychological experiments because medical data is often collected from individuals, and these fields have developed ways to address these problems. Scientists have also developed methods of untying data from the individual from whom it was collected. Although peer-reviewed journals in most fields will not publish an article if the data is not available to other researchers, the data in fields such as medicine and psychology must remain anonymous. To maintain anonymity, the test methodology may involve assigning identification codes for each subject that only the original researcher can decipher.

Legal scholars performing empirical work can develop similar requirements. Researchers should take care to ensure that the subjects of a study, including corporations, should not be harmed by the experiment. The legal academy should therefore try to develop a system, possibly based on the systems used in other sciences, which can safeguard subjects of studies while still allowing future study of datasets.

D. Theory Modification

Finally, once critical hypotheses of a generalized theory have been tested using appropriate research designs, the results should be used to modify the theory. Construct validation is an iterative process, in which theories are

²⁶³ Yoon & Baker, *supra* note 226, at 165.

²⁶⁴ *Id.* at 172.

modified based on the latest empirical research and then future research can be conducted to test the new theory.²⁶⁵ Presently, however, the process of incorporating new knowledge from empirical legal scholarship into existing theory is not standardized. The result is that evidence contrary to a popular theory may be ignored or the academy may just be slow to recognize the new data. By standardizing the assimilation of new empirical data into generalized theories, legal theories can evolve to become better predictors of behavior based on governmental policies.

Some philosophers of science believe the ultimate goal of scientific research is to find the definitive truth about nature.²⁶⁶ Scientific research adds to our knowledge of the universe and, as science progresses, scientists come closer to the final truth.²⁶⁷ In legal studies, the goal may not be an ultimate true law, but a more modest attempt to find the best possible set of laws to govern a given society at a given time. Finding the best set of laws may then be accomplished by finding “true theories” about how law influences society. These true theories can direct lawmakers about which laws to pass to correctly apply a policy.

But, as noted above, empirical research should not be considered the final step in determining new policies. The quantitative analysis should be used only to inform new policy decisions. The generalized theories can be used to provide guidance when lawmakers approach new issues, but should not be used as the final solutions for problems.

1. Incorporating empirical results into the theory of constitutional rights

We have already considered a theory that explicit constitutional rights result in more freedom than if the rights were not included in a government’s foundational document. We then inferred the hypothesis that if the theory is true, then the press should exhibit more freedom when in a country in which it receives constitutional protections. After running several quasi-experiments testing this and rival theories, research has generated new empirical data that either supports, undermines, or alters the original theory. We can now consider how to modify the theory based on the new information.

²⁶⁵ See Donald T. Campbell, *Evolutionary Epistemology*, reprinted in *METHODOLOGY AND EPISTEMOLOGY FOR SOCIAL SCIENCE*, *supra* note 19, at 393, 393 (describing how knowledge evolves in a similar fashion to natural selection through mutations and new combinations).

²⁶⁶ See, e.g., Popper, *supra* note 23, at 50; LARVOR, *supra* note 29, at 102; Smith, *supra* note 13, at 400.

²⁶⁷ See Popper, *supra* note 23, at 50.

If the data confirms the hypothesis and has ruled out several alternate theories, then the theory has been supported. Although a theory can never be proven undisputedly true, the experimental results add confidence that it is accurate. This theory can now be used to inform future policy considerations.

For example, if a nation was considering drafting a new constitution or new constitutional amendments, the drafters should be advised that if they include certain protections for citizens, the population is likely to act more “free” than if those protections are left to be set out in statutory provisions.

But even in these circumstances, researchers should not rest on their laurels. More research can still be conducted to further refine the theory or rule out other criticisms that had not yet been addressed. In addition, the hypotheses discussed thus far have only addressed freedom of the press. Future research can address other constitutional rights, such as a right to privacy, a right to counsel in criminal proceedings, a right to due process, etc.

If the experimental results undermine the validity of the theory, however, then the theory should be changed to reflect the negative data. If the results show that the press does not exhibit more freedom despite the presence of constitutional protections, then researchers should consider several alternative explanations. First, the operational definition of freedom may not have been correct. Counting stories that criticize the government may not have been an accurate way to measure freedom. Alternatively, it could be that the constitutional theory is correct, but the experiments relied on incorrect auxiliary theories. For instance, it may be that the experiments relied on theories about statistical techniques that are not accurate in this context. In any event, it is necessary to more fully explore the results and change the general theory to reflect the new data.

Most likely, however, is that the data will provide inconclusive results. The data may show that constitutional rights provide more freedom, but only with regard to certain rights or in certain cultures. If a culture is intrinsically adverse to criticism of authority, protections for the press may not produce any change in behavior. Similarly, a right for free speech may provide more robust results than a right not to quarter militia in one’s home. By exploring these possibilities, the theory will become more refined and provide a better understanding of the field.

2. Incorporating empirical results into the theory of intellectual property rights

A similar procedure can be used to assess the generalized theory about the incentives provided by IP law. If it turns out that stronger IP laws result in an increase in the number of copyrights and patents, then those results reinforce the theory that IP creates an incentive to create. If, on the other hand, the

experiments provide contrary results, then it is necessary to reassess the theory. It may be that incentives should not be measured by the number of copyrights and patents issued, but rather by the number for which people apply. Or, perhaps, it may be better to use several measurable indicators to get a fuller picture. Negative results may also mean that the theory being studied relied on faulty auxiliary theories, such as financial gains being the only motivations that IP law provides or that creators desire. Finally, it may be that the theory itself is incorrect and that IP protections do not have an influence on innovations.

But again, it is most likely that the results show that the theory is partially correct, but needs refinement. For instance, it may turn out that the motivation provided by IP law experiences diminishing returns. That is, a certain level of protection will spur creation, but the marginal benefits decrease after a certain level of protection is provided.

CONCLUSION

By incorporating concepts of measurement developed in social sciences, empirical legal scholars can better assess the impact of the unobservable constructs that underlie modern legal theory. Specifically, the notion of construct validation can be adopted from psychology to law to allow a standardized approach to the modification of legal theory to reflect updated understandings of how laws impact societies based on empirical observations. These modified theories, based on quantitative observations, can then better inform policymakers' qualitative decisions.

