

ARTICLES

Law and the Art of Modeling: Are Models Facts?

REBECCA HAW ALLENSWORTH*

In 2013, the Supreme Court made the offhand comment that empirical models and their estimations or predictions are not “findings of fact” deserving of deference on appeal. The four Justices writing in dissent disagreed, insisting that an assessment of how a model works and its ability to measure what it claims to measure are precisely the kinds of factual findings that the Court, absent clear error, cannot disturb. Neither side elaborated on the controversy or defended its position doctrinally or normatively. That the highest Court could split 5–4 on such a crucial issue without even mentioning the stakes or the terms of the debate, suggests that something is amiss in the legal understanding of models and modeling.

This Article does what that case failed to do: it tackles the issue head-on, defining the legal status of a scientific model’s results and of the assumptions and choices that go into its construction. I argue that as a normative matter, models and their conclusions should not be treated like facts. Models are better evaluated by a judge, they do not merit total deference on appeal, and modeling choices are at least somewhat susceptible to analogical reasoning between cases. But I show that as a descriptive matter, courts often treat models and their outcomes like issues of fact, despite doctrines like Daubert that encourage serious judicial engagement with modeling. I suggest that a perceived mismatch between ability and task leads judges to take the easier route of treating modeling issues as facts, and I caution that when judges avoid hard questions about modeling, they jeopardize their own power and influence.

TABLE OF CONTENTS

| | |
|--|-----|
| INTRODUCTION | 827 |
| I. THE ART OF MODELING | 831 |
| A. WHAT IS A MODEL? | 832 |
| B. KINDS OF MODELS IN ANTITRUST (AND BEYOND) | 834 |
| 1. Empirical Models | 836 |
| 2. Theoretical Models | 838 |
| C. WHAT IS A GOOD MODEL? | 839 |
| 1. Where Possible, a Model Should Be Validated with Real World Data | 841 |

* Associate Professor of Law, Vanderbilt Law School; J.D., Harvard Law School; M.Phil, University of Cambridge; B.A., Yale University. © 2015, Rebecca Haw Allensworth. Thank you to Jonathan Baker, Margaret Blair, Edward Cheng, Blair Druhan, Paul Edelman, Luke Froeb, Joni Hersch, Robert Mikos, Geoffrey Miller, Kate Nicholson, Erin O’Hara O’Connor, Richard Posner, and Jim Rossi for helpful comments on previous drafts, and to Zachary Hebert for outstanding research assistance.

| | | |
|------|--|-----|
| 2. | Assumptions Should Be Checked Where Possible | 842 |
| 3. | Modelers Should Be Required to Disclose All Other Modeling Choices They Considered and Give Reasons for Each Modeling Choice | 845 |
| 4. | Multiple Models Should Be Used Where Possible | 846 |
| II. | ARE MODELING CHOICES AND OUTCOMES ISSUES OF FACT? | 846 |
| A. | THE STAKES: WHAT MATTERS ABOUT THE “FACT” LABEL? | 846 |
| 1. | The Consequences of the “Fact” Label | 847 |
| 2. | Why Facts Are “Facts” | 848 |
| 3. | Mixed Questions and the Myth of the Fact/Law Divide | 850 |
| B. | MODELS AND THEIR CONCLUSIONS SHOULD NOT BE TREATED AS FINDINGS OF FACT | 852 |
| 1. | Decisions Involving Modeling Choices Ought to Have (Weak) Precedential Value | 853 |
| 2. | Modeling Should Be Evaluated by Judges, Not Juries | 856 |
| 3. | Models Ought to Be Scrutinized Carefully on Appeal | 857 |
| C. | MODELS AND THEIR RESULTS ARE TREATED LIKE FINDINGS OF FACT | 859 |
| 1. | Models in the District Courts: <i>Daubert</i> Is Right in Theory, Wrong in Practice | 859 |
| 2. | Models in the Appellate Courts: <i>Joiner</i> Precludes Meaningful Model Review | 865 |
| 3. | Models in the Supreme Court: Deference to Consensus | 868 |
| D. | THE CAUSE AND EFFECT OF TREATING MODELING ISSUES AS FACTS | 871 |
| 1. | The Cause: Judicial Inexpertise | 871 |
| 2. | The Effect: Inaccurate Judgments and an Illegitimate Delegation of Authority | 872 |
| III. | IMPLICATIONS FOR INSTITUTIONAL DESIGN | 873 |
| A. | REALLOCATION OF MODELING QUESTIONS TO AN AGENCY | 874 |
| B. | JUDICIAL EDUCATION IN MODELING AND STATISTICS | 875 |
| | CONCLUSION | 877 |

INTRODUCTION

“[W]hile the data contained within an econometric model may well be ‘questions of fact’ in the relevant sense, what those data prove is no more a question of fact than what our opinions hold.”

–Scalia, J.¹

In 2013’s *Comcast Corp. v. Behrend*, the Supreme Court offhandedly remarked that the claims that result from an econometric model are not “findings of fact” that enjoy deferential “clear error” review on appeal.² By making this seemingly innocuous observation, the Court, as it has done before, buried a landmine in a footnote.

The case was a class action suit in which consumers accused Comcast of monopolizing the market for cable television services in the Philadelphia area.³ The issue presented to the Court was whether a damages model offered by the plaintiffs was sufficient to show that “questions of . . . fact common to class members predominate[d] over any questions affecting only individual members,”⁴ such that class certification was appropriate. Plaintiffs had alleged four different theories of monopolization, but the district court had rejected all but one: the claim that Comcast had prevented competitive entry by “overbuilders.”⁵ Although the plaintiffs’ model failed to disaggregate the harm under the accepted theory of liability from the harm under the theories that the district court had rejected, the district court ordered briefing on the subject and held that the model was still good evidence that damages could be measured on a class-wide scale.⁶

Writing for the majority, Justice Scalia reversed class certification, disagreeing with the district court’s conclusion that the “decision not to credit [other theories] of antitrust impact [did] not impeach [the] damages model.”⁷ Crucially, the majority viewed this question as a matter of law, not subject to deference on review. Indeed, the Court explained in footnote five that “while the data contained within an econometric model may well be ‘questions of fact’ in the relevant sense, what those data prove is no more a question of fact than what our opinions hold,” thus allocating to themselves the authority to interpret scientific models de novo.⁸

1. *Comcast Corp. v. Behrend*, 133 S. Ct. 1426, 1433 n.5 (2013).

2. *Id.*

3. *Id.* at 1430.

4. *Id.* (quoting FED. R. CIV. P. 23(b)(3)).

5. “Overbuilders” are cable firms who build networks in neighborhoods already dominated by an incumbent. *E.g., id.* at 1430–31.

6. *Behrend v. Comcast Corp.*, 264 F.R.D. 150, 190–91 (E.D. Pa. 2010), *aff’d*, 655 F.3d 182 (3d Cir. 2011), *rev’d*, 133 S. Ct. 1426 (2013).

7. *Id.*

8. *Comcast*, 133 S. Ct. at 1433 n.5.

The evidentiary status of model-derived propositions as “not facts” is far from obvious; indeed, the four Justices joining the dissenting opinion believed the majority had it wrong. In their view, the case turned on an assessment of how the model worked and its ability to measure what it claimed to measure, which the dissent believed are precisely the kinds of factual findings that the Court, absent clear error, must leave alone.⁹ Neither side elaborated on when model-derived propositions should or should not be treated as facts under the law. That the highest Court could split 5–4 on the question of whether the claims that result from a model are “facts,” and that even in articulating that split, neither side would provide a substantive engagement with the question, suggests that something is amiss in the legal understanding of models and modeling.

This Article tackles the question head-on, addressing the legal status of not only the results of a model but also the assumptions and choices that go into its construction because, as I will argue, a model’s results are inextricable from the construction of the model itself.¹⁰ The question has two relevant dimensions: the normative (whether models and their results *should* be treated as facts) and the positive (whether the law *does* treat them like facts). This Article argues that as a normative matter, models and their results should not be treated like facts. But this assessment turns out to be aspirational; too often, courts treat models and their outcomes like issues of fact.

In answering these twin questions, I engage the vast literature on scientific evidence and the law, but I break with the traditional framing of the issue in two ways. First, I identify scientific models—defined as mathematic abstractions used to predict or describe natural or market processes—as a subset of scientific evidence worthy of special attention under evidence law and theory. In this sense, my project is narrower than the larger debate about scientific evidence under the law. Second, by using the fact/law framework offered by *Comcast*, my Article moves the debate beyond mere admissibility to examine all areas of legal decisionmaking that rely implicitly or explicitly on scientific evidence. On this view, *Daubert*¹¹ and other admissibility doctrines are merely examples—though important ones—of how the law regulates reliance on scientific argument. In this sense, my project is broader than the typical debate about science in the courtroom. Given the limitations of the existing debate about scientific evidence, it is unsurprising that all the academic ink spilled since *Daubert* has failed to resolve such a basic question as whether a model and its predictions are “facts” in the eyes of the law.

9. *Id.* at 1439–40 (Ginsburg & Breyer, JJ., dissenting).

10. Cf. David L. Faigman, *Appellate Review of Scientific Evidence Under Daubert and Joiner*, 48 HASTINGS L.J. 969, 975 (1997) (“Evaluation of methods and conclusions simply cannot be divided among different decision makers.”).

11. By “*Daubert*,” I refer to the admissibility standard for scientific testimony established in *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993).

Comcast is a conspicuous and rare manifestation of a problem that is usually hidden from view. Courts generally do not classify modeling choices explicitly as being factual or otherwise; rather, they seem to assume that whatever treatment they choose for the question is optimal. *Comcast* is unusual because it explicitly raises the question, but unfortunately it is not particularly helpful in resolving it. Neither side writes much about the issue—the majority mentions it only in a footnote, and the dissent gives no reasons for its contrary conclusion. With so little on the record, there is no analytical purchase in *Comcast* to answer the essential question it raised; thus, the case and its specific facts will not feature prominently in this Article. Instead, I look beyond the *Comcast* frame to define the normative and positive status of models as facts.

The normative answer to the question whether models should be treated as facts under the law must start with a comprehensive and workable account of what scientific modeling is and what it has to offer legal decisionmaking. Modeling is a useful and popular tool for legal decisions, aiding with tasks from the mundane (damages calculations)¹² to the exotic (forecasting the effect of a legal rule).¹³ Models are almost always presented by experts as scientific evidence, and indeed models are “scientific” in important ways. They are essential tools in the creation of scientific knowledge; indeed, it would be difficult to point to any scientific finding or theory that did not depend heavily on modeling for its discovery and proof. And they are also scientific in the sense that they derive their legitimacy from critical treatment by a scientific community. But models are not scientific in the Popperian sense of being falsifiable;¹⁴ as inventions designed to perform a task, they are purposive rather than positive.

Modeling is also an art. Many decisions that go into modeling involve intuition or inspiration. Others involve choices about trade-offs. Here lies a gray area, created by the many sets of modeling choices and assumptions that are too complex to navigate systematically. The prevalence and importance of these nonfalsifiable elements of model building have led scientists to observe that modeling is as much an art as it is a science.¹⁵ The art of modeling exists in a realm where not even experts can agree on unique and objective scientific choices.

12. *E.g.*, *Behrend*, 264 F.R.D. at 181–91 (calculating damages).

13. *E.g.*, *Leegin Creative Leather Prods., Inc. v. PSKS, Inc.*, 551 U.S. 877, 890–92 (2007) (using a theoretical model to predict business practices in the absence of resale price maintenance).

14. Here I am using the popular version of Karl Popper’s theory, that science is “the set of statements about the world that can, at least in principle, be falsified by empirical observation.” Malcolm B. Coate & Jeffrey H. Fischer, Daubert, *Science, and Modern Game Theory: Implications for Merger Analysis*, 20 SUP. CT. ECON. REV. 125, 133 (2012). For more discussion of whether this version of Popper’s theory is accurate and where the Supreme Court stands on the issue, see *infra* note 167.

15. D.H. Kaye, *The Dynamics of Daubert: Methodology, Conclusions, and Fit in Statistical and Econometric Studies*, 87 VA. L. REV. 1933, 1965 (2001) (citing Lincoln E. Moses, *The Reasoning of Statistical Inference*, in PERSPECTIVES ON CONTEMPORARY STATISTICS 107, 117–18 (David C. Hoaglin & David S. Moore eds., 1992)).

This straddle between art and science has made for the awkward and at times inconsistent treatment of modeling as factual in the eyes of the law. The notion that models are scientific has contributed to their treatment as facts because judges tend to think of a scientific proposition as factual in the sense of being true or false and, thus, fitting a dictionary definition of “fact”—“a piece of information presented as having objective reality.”¹⁶ Indeed, *Daubert* contributed to this conception of science as inherently factual with its emphasis on Popperian falsifiability as the hallmark of admissibility. But as modeling illustrates, much of what we call science exists outside of the falsifiable realm. Here, the art of modeling frustrates an easy classification of models as facts.

The way out of this puzzle of categorization is to recognize that whether an issue is properly a “fact” in the eyes of the law has little to do with whether it conforms to Merriam-Webster’s definition of “fact.” Rather, the fact/law distinction in our legal system turns on the best allocation of decisionmaking procedures, where certain treatments are reserved for “facts” (jury decisionmaking, no precedential value, deferential review, no explicit recorded reasoning required) and others for “law” (judge decisionmaking, precedential value, de novo review, explicit reasoning required).¹⁷ Thus, a fact (in the eyes of the law) is an issue optimally determined by the process reserved for “facts,” and legal issues are those best decided under the process reserved for “law.”

Under the pragmatic view of the fact/law distinction, models and their conclusions fail to qualify as facts because the indeterminacy of individual modeling choices makes judges—more educated than juries and repeat players—better suited to their evaluation. Further, the evaluation of some modeling choices and perhaps even some models are susceptible to analogical reasoning from prior cases, suggesting that the legal system may benefit from having explicit reasoning about models spelled out in case law. And since the authority of a model is derived from its exposure to a process of critical review, pure deference is not appropriate at the appellate stage.

That models should not be treated as facts under the law does not mean they should be treated as pure questions of law; the fact/law distinction, and the various decisional procedures it represents, is not dichotomous. Indeed, a vast category of issues are considered “mixed questions of law and fact,” occupying a middle ground where some of the decisional features of “fact” are present, while some of “law” are, too. Here, courts may design a decisional process optimized for a particular issue. This Article will argue that models ought to be treated as presenting mixed questions of law and fact, located on the procedural spectrum where issues are decided by a judge rather than a jury, are reviewed with little deference by an appellate court, and require explicit reasoning, either

16. *Fact Definition*, MERRIAM-WEBSTER DICTIONARY, <http://www.merriam-webster.com/dictionary/fact> (last visited Feb. 11, 2015).

17. See generally Ronald J. Allen & Michael S. Pardo, *The Myth of the Law-Fact Distinction*, 97 NW. U. L. REV. 1769, 1770 (2003).

in a written or oral opinion, on the part of the trial court. Though not quite issues of pure law, models and their conclusions should certainly not be treated like facts.

Having answered the normative question—should models be treated like facts?—in the negative, it is necessary to turn to the positive question. This Article argues that, precisely because the artistry in modeling makes it so difficult to assess, courts too often treat models as questions of fact. This allows them to avoid the embarrassment of acknowledging the art behind modeling and the difficulty of appraising a modeler's choices. Thus, as a descriptive matter, Justice Scalia is wrong: models and their results are issues of fact in that their evaluation is frequently allocated to juries, reviewed deferentially by appellate courts, and accepted or rejected without explicit reasoning. The result is poor legal decisionmaking where models are integral to a legal decision, which is true in a staggering—and increasing—percentage of cases.

This Article uses examples from antitrust litigation to illustrate its points, but its conclusions and observations apply to all areas of law that involve modeling. Part I defines modeling, describes the spectrum of models from empirical to theoretical, and surveys the various criteria by which models should be evaluated. Armed with this background, Part II provides the central analysis of this Article. It explores the normative and positive questions about models' status as "facts" in the eyes of the law, concluding that, although models and their conclusions are not facts in a pragmatic sense, they are often treated as such by courts. I attribute this treatment in part to doctrine, but also in significant part to the difficulty courts have accepting and assessing the art of modeling. Finally, Part III explores the implications of the conclusions from Part II, arguing that, although the treatment of models and their results as issues of fact is undesirable, it will persist as long as judges are asked to make legal decisions involving models without being given the tools to do it. This Part explores two familiar solutions that offer to close the expertise gap between scientific and legal reasoning: reallocating technical decisionmaking to an agency and educating judges in technical matters. The Article ultimately concludes that the better solution is to augment judicial competence in modeling and quantitative reasoning.

I. THE ART OF MODELING

"[A]ll models are wrong, but some are useful."

—George E.P. Box & Norman R. Draper¹⁸

Scientific models, both empirical and theoretical, play a starring role in law, whether they are used to develop scientific facts offered to prove liability or

18. GEORGE E.P. BOX & NORMAN R. DRAPER, *EMPIRICAL MODEL-BUILDING AND RESPONSE SURFACES* 424 (1987).

damages at trial, or to predict the effect of a legal rule on a complex system like a market or the environment. Yet lay judges often misunderstand the role and purpose of models and, thus, use the wrong criteria—or sometimes none at all—to evaluate them. Therefore, it is necessary first to understand the proper role of models in scientific study in order to understand how they are misused and misclassified by the law.

A. WHAT IS A MODEL?

The Oxford American dictionary defines a “model” as “a simplified description, especially a mathematical one, of a system or process, to assist calculations and predictions.”¹⁹ This succinct definition is useful here because it highlights two features of all models used in legal decisionmaking. First, a model is always a simplification of a more complicated process. Second, a model is always used for some purpose; it is not itself valuable, or capable of being evaluated, without reference to its ability to further a particular predictive or descriptive task.²⁰ These features of models suggest two metaphors that can help express what models are.

If modeling can be thought of as a simplification of a more complex system, then one might think of a model as a map. Indeed, this is a serviceable metaphor to a point. Like a map, a model is an abstraction of a more complex system that it represents, and in the simplification process it removes unnecessary features to better highlight or reveal what is essential about the system.²¹ On a map, the essential elements may be roads and lakes, which appear more prominently if features like trees and mailboxes are omitted. Like maps, models can use this simplification to express information on a smaller scale,²² capable of being understood and synthesized into decisionmaking or abstracted into causal explanations. And, like a mapmaker, a modeler makes choices about what are the essential elements (and what are inessential, such as mailboxes and trees) with reference to the task the model is to perform. Models, like maps, are always wrong because they simplify; so, the familiar warning that “[a] map is not the territory”²³ applies to models as well.

The map metaphor is useful but it does not describe all features of models. Because most topographical or road maps are simplifications of a known and measurable system, they can be evaluated directly for accuracy. Of course, any

19. *Model Definition*, NEW OXFORD AMERICAN DICTIONARY, http://www.oxfordreference.com/view/10.1093/acref/9780195392883.001.0001/m_en_us1268618?rkey=nxbJEI&result=1 (last visited Feb. 11, 2015).

20. Cf. BOX & DRAPER, *supra* note 18, at 424.

21. See MARY S. MORGAN, *THE WORLD IN THE MODEL: HOW ECONOMISTS WORK AND THINK* 22–23 (2012).

22. Indeed, Merriam-Webster defines “model” as a “usually small copy of something.” *Model Definition*, MERRIAM-WEBSTER DICTIONARY, <http://www.merriam-webster.com/dictionary/model> (last visited Feb. 11, 2015).

23. Alfred Korzybski, *A Non-Aristotelian System and its Necessity for Rigour in Mathematics and Physics*, in *SCIENCE AND SANITY* 747, 750 (1st ed. Supp. III 1933) (emphasis omitted).

simplification will be inaccurate as to what it omits, but as to the remaining essential elements or their representation, a map is capable of achieving precision because those elements bear a proportional relationship to the real world. In other words, a conventional map is not of something unknown, but rather is a representation of a known and understood system made smaller and simpler and therefore more useful for particular tasks like hiking or driving. Models, too, can often be—and as we will see should be—calibrated against the real world. But typically the most useful feature of a model is its ability to predict or measure what is unknown or unseen, such as the future or a counterfactual past; it is usually the model's goal to help us understand features of this complex and undiscovered system.

The dictionary definition's focus on a model's usefulness suggests another metaphor for understanding models. In an important sense, a model is a tool. It is an invention used to close the gap between a state of ignorance (or, at best, intuition) about the world and a state of knowledge about the world. And like a tool, it is purposive, and thus is not meaningful in itself. It is not itself scientific knowledge; it is a means of achieving it.²⁴ A model is good when it is, like a tool, powerful for its intended purpose.

Again, the tool metaphor helps us understand an essential feature of models, but it is also imperfect. In many cases a tool's value can be ignored in assessing the finished product. A strong fence is a strong fence and serves its purpose whether the builder used a handsaw or a circular saw to cut the wood. Builders may have an opinion about which is the "right" saw for the job, but that evaluation may be based on criteria like efficiency or safety, not necessarily on the quality of the finished product. This is because external tests exist for the strength of a fence, independent of the methodology that created it. Although this is arguably also true for models that can easily be calibrated against the real world, in many cases the end product of models, both empirical and theoretical, is offered as a fact about the world that was unknown before the model existed.²⁵ For many models, the quality of their predictions cannot be assessed without engagement with the choices and trade-offs made in constructing the models.²⁶ So whereas the map and tool metaphors are helpful,²⁷ they both are imprecise for the same reason: there is usually no gold standard of scientific truth that can be used to measure a model's worth.

24. Although, knowledge about how to create the best model for a given task is arguably scientific knowledge; specifically, statistical and econometric knowledge.

25. This gives rise to the identification problem in observational science. See Jonathan B. Baker & Timothy F. Bresnahan, *Economic Evidence in Antitrust: Defining Markets and Measuring Market Power*, in *HANDBOOK OF ANTITRUST ECONOMICS* 1, 5 (Paolo Buccirossi ed., 2008) ("The identification problem arises because empirical economists can rarely perform experiments on economic actors; they must instead look carefully for settings in which nature has created an experiment for them.").

26. See Faigman, *supra* note 10, at 975.

27. For more metaphors (or models) describing models, see Itzhak Gilboa et al., *Economic Models as Analogies* 4–6 (Penn Inst. for Econ. Research, Working Paper No. 13-007, 2013), available at <http://ssrn.com/abstract=2209153>, and sources cited therein.

Having defined models so broadly, it is necessary to pause and consider what is not a model, at least as far as the law ought to be concerned. All acts of cognition and expression must be simplifications because of the limited computational capacity of human brains. Therefore, it may make sense to think of a sentence or an utterance or perhaps even a thought as a model: it is a shorthand, simplified representation of reality constructed to serve a specific purpose (to communicate).²⁸ Likewise, with a very broad version of my definition of “model,” one might sweep in legal rules (simplifications constructed for a purpose: to govern behavior and sort cases) and prior beliefs of jurors (generalizations about human behavior used for a purpose: to interact peacefully and productively in society).

These abstractions are excluded from my definition of models in legal decisionmaking because the legal puzzle I am interested in, and that *Comcast* highlights, is what to do with *scientific* models. Scientific models (as opposed to linguistic, legal, or lay models) are usually expressed in mathematical form and are assessed according to their conformity to scientific conventions, as defined by experts. This does not mean that lay observers, like judges, cannot meaningfully assess models—indeed, this Article will argue that judges can and should—but it does mean that the assessment will always import criteria and methods from the relevant scientific community rather than use lay or legal criteria. Scientific models serve as the focus of my Article because of the special status—and persuasive power—they enjoy under the law as expert evidence, and the increasing importance of their use in legal decisionmaking.

A further word on the limits of this Article. This Article concerns only modeling, as the medium of the scientist, rather than expert testimony generally. This choice stems from my observation, made stark in *Comcast*, that whatever difficulty judges have with expert evidence, they are particularly uncertain about how to evaluate models and about their status as scientific evidence. This judicial ambivalence derives from the mysteriousness of models—their power to turn data into scientific truth—and their uniquely persuasive nature to the trier of fact. It is deepened by the recognition that, because such a substantial part of modeling is art, the possibility for manipulation in the adversarial context is high. It is hoped that by focusing on scientific modeling specifically with attention to what makes modeling so uniquely troubling for judges, this Article can provide some insight into the proper treatment of an important subcategory of scientific testimony.

B. KINDS OF MODELS IN ANTITRUST (AND BEYOND)

So far, the discussion has been abstract, so it might help to bring the discussion of models used in legal decisionmaking into the concrete. To sim-

28. In the words of Oliver Wendell Holmes, Jr., “A word is not a crystal, transparent and unchanged, it is the skin of a living thought and may vary greatly in color and content according to the circumstances and the time in which it is used.” *Towne v. Eisner*, 245 U.S. 418, 425 (1918).

plify, this section will focus on models used in antitrust litigation, although antitrust is hardly unique in its dependence on models for factual proof and policymaking. Empirical—that is, data-driven—models are ubiquitous in any private action that assesses damages as the difference between prices, earnings, or values “but for” the defendant’s conduct. Such actions include shareholder suits, tort cases alleging lost profits, and suits over property damage.²⁹ Empirical models are also common in discrimination cases to prove liability, and then again to prove damages.³⁰ Overt use of theoretical and game-theoretical models is perhaps less common in areas outside of antitrust, but still features prominently in cases where empirical proof is lacking. For example, governments often use theoretical models to defend constitutional challenges to their laws by arguing that the law at least theoretically serves a legitimate state purpose. The suits challenging the Affordable Care Act are good examples.³¹

Virtually every antitrust case involves a data-driven model used to estimate price effects of anticompetitive conduct.³² Less visibly, every antitrust case at least implicitly relies on theoretical models, and many explicitly engage with theoretical and game-theoretical models of market behavior to establish liability and damages. The “before-during-after” model of cartel overcharges will serve as an example of an empirical (statistical) model, while Lester Telser’s model of “free riding” among retailers will serve as an example of a theoretical model. These examples are merely representative; the array of models used in antitrust and in other areas of law is vast.

It is helpful to attend to the differences between empirical and theoretical models and to elaborate with an example of each, but it is wrong to think of theory and empirics as dichotomous. All empirical or data-driven models rely on theory in their construction, so any account of an empirical claim free from theory is incomplete.³³ Many assumptions in theoretical models are capable of

29. See generally Frederick C. Dunbar & Arun Sen, *Counterfactual Keys to Causation and Damages in Shareholder Class-Action Lawsuits*, 2009 WIS. L. REV. 199 (discussing the use of mathematical “but for” models to prove causation in shareholder suits); John W. Hill et al., *Increasing Complexity and Partisanship in Business Damages Expert Testimony: The Need for a Modified Trial Regime in Quantification of Damages*, 11 U. PA. J. BUS. L. 297, 334–35 (2009) (discussing the use of modeling in business damages).

30. See, e.g., William T. Bielby & Pamela Coukos, “Statistical Dueling” with Unconventional Weapons: What Courts Should Know About Experts in Employment Discrimination Class Actions, 56 EMORY L.J. 1563 (2007) (discussing statistical modeling used to prove discrimination).

31. See Nat’l Fed’n of Indep. Bus. v. Sebelius, 132 S. Ct. 2566, 2614–15 (2012) (Ginsburg, J., concurring in part and dissenting in part) (discussing the problem of adverse selection in insurance markets). The theory of adverse selection is based on a theoretical model made famous by George A. Akerlof in *The Market for “Lemons”: Quality Uncertainty and the Market Mechanism*, 84 Q.J. ECON. 488 (1970).

32. See ABA SECTION OF ANTITRUST LAW, *ECONOMETRICS 1–3* (2005) [hereinafter ABA ECONOMETRICS] (discussing the large and increasing use of econometric proof in antitrust cases).

33. ABA SECTION OF ANTITRUST LAW, *PROVING ANTITRUST DAMAGES: LEGAL AND ECONOMIC ISSUES* 132–33 (2d ed. 2010) [hereinafter ABA ANTITRUST DAMAGES] (“In building the econometric model, the economist makes initial modeling or specification choices by employing economic theory and reasoning.”).

empirical proof,³⁴ and even some conclusions from theoretical models can be empirically confirmed. Further, there are many forms of modeling that straddle the empirical/theoretical divide (to the extent it is a divide), like predictive modeling, game-theoretical modeling, and simulations.³⁵ As it turns out, all these kinds of modeling have more in common—in terms of their strengths and weaknesses and, therefore, the ways in which they should be evaluated—than is perhaps apparent at first glance. Because of this interrelation, this section provides separate examples of empirical and theoretical modeling, but the Article ends up making no particular distinction between categories of scientific models when evaluating their actual and ideal treatment under the law.

1. Empirical Models

An empirical model is a mathematical expression that uses data to relate variables (or factors) to an outcome of interest; that is, it uses individual observations of pairs of factors and outcomes (x and y , respectively) to estimate the mathematical relationship between the factor and the outcome.³⁶ This mathematical relationship is expressed as a function, involving as many factors (x_1, x_2 , etc.) as the modeler chooses.³⁷ This function is then taken to be an approximation of the natural or market process that created y from x .³⁸ Key to empirical models is that the details of this mathematical function are derived from actual pairs of variables observed in the real world. Most commonly, empirical models are used to understand causal relationships between x and y ,³⁹ but not always. Multiple regression analysis is a familiar and frequently used modeling technique that can aid in causal analysis.⁴⁰

In antitrust, the before-during-after method of estimating damages from cartel activity is a typical example of empirical modeling.⁴¹ The before-during-after model describes the mathematical relationship between price (y) and the factors

34. See, e.g., Yulia Gel et al., *The Importance of Checking the Assumptions Underlying Statistical Analysis: Graphical Methods for Assessing Normality*, 46 JURIMETRICS J. 3, 4–10 (2005) (discussing the value of empirically confirming assumptions of normality in statistical models).

35. For an excellent discussion of merger simulations and the benefits of using simplified versions of them at early stages of administrative merger review, see Jonathan B. Baker, *Merger Simulation in an Administrative Context*, 77 ANTITRUST L.J. 451 (2011).

36. See Lincoln E. Moses, *The Reasoning of Statistical Inference*, in PERSPECTIVES ON CONTEMPORARY STATISTICS 107, 107 (David C. Hoaglin & David S. Moore eds., 1992).

37. See ABA ECONOMETRICS, *supra* note 32, at 10–13 (discussing the process of selecting explanatory variables).

38. Moses, *supra* note 36, at 107.

39. See Daniel L. Rubinfeld, *Econometrics in the Courtroom*, 85 COLUM. L. REV. 1048, 1065–66 (1985).

40. See ABA ECONOMETRICS, *supra* note 32, at 3–4 (“[E]conometric evidence allows an economist to show whether certain factors affect or do not affect another particular factor or event. . . . Multiple regression analysis is the technique used in most econometric studies.”).

41. Damages is but one area of antitrust for which statistical models are relevant. For an excellent discussion of statistical techniques for measuring market power in antitrust cases, see Jonathan B. Baker & Timothy F. Bresnahan, *Empirical Methods of Identifying and Measuring Market Power*, 61 ANTITRUST L.J. 3 (1992).

that influence it (x_1 , x_2 , etc.).⁴² The goal of the technique is to estimate a “but-for” world in which the defendant’s anticompetitive behavior never took place, with the idea that the price difference between the but-for world and the real world during the interval of collusion represents the damages owed the plaintiff.⁴³

In the before-during-after method, the modeler creates an expression that accurately describes, or “predicts” in a backward-looking sense, known prices before and sometimes after the relevant period of collusive pricing.⁴⁴ The modeler arrives at this expression by guessing what factors—other than collusion—are likely to influence prices and by guessing the functional form price is likely to take in response to these factors.⁴⁵ Most often, econometricians assume that the relationship between price and various influences is linear.⁴⁶ Then, applying pricing data from the real world during a period known to be without collusion—before or after the conspiracy will do; both is better—the amount that each factor influenced price can be estimated through multiple regression analysis.⁴⁷ This statistical technique also yields relevant information like the closeness of the fit between the model and the data, and the likelihood that a coefficient is wrong and by how much.⁴⁸ This expression, now perhaps more accurately called a model, can be checked for accuracy by validating against a data set of pricing patterns reserved for that purpose. Once the modeler is confident that the model describes pricing outcomes in a competitive market, she uses it to estimate the prices that would have obtained during the period of anticompetitive conduct, had there not been anticompetitive conduct in the first place and had the competitive market determined price.⁴⁹ The difference between the collusive price and the hypothetical competitive market price is the measure of damages. This before-during-after model, which is extremely common in conspiracy cases brought under Section 1 of the Sherman

42. ABA ANTITRUST DAMAGES, *supra* note 33, at 171–73.

43. *Id.* at 56–57.

44. See Michele Molyneux, Comment, *Quality Control of Economic Expert Testimony: The Fundamental Methods of Proving Antitrust Damages*, 35 ARIZ. ST. L.J. 1049, 1056–61 (2003) (discussing the “before-after” method and providing some examples of where it has been improperly performed and rejected by courts); see also D.H. Kaye, *Adversarial Econometrics in United States Tobacco Co. v. Conwood Co.*, 43 JURIMETRICS J. 343, 346–47 (2003) (defining the before-after method). The before-after method discussed by Molyneux and Kaye is identical to the before-during-after method as I use that term in this Article.

45. See ABA ANTITRUST DAMAGES, *supra* note 33, at 147 (“[T]he explanatory variables in an econometric model represent economic factors that influence the dependent variable. . . . [I]f the dependent variable is price, economic theory suggests that demand drivers, cost factors, and industry capacity, among other things, are potential explanatory variables.”).

46. ABA ECONOMETRICS, *supra* note 32, at 13.

47. See ABA ANTITRUST DAMAGES, *supra* note 33, at 167–73 (discussing the various kinds of before-during-after models and their power to estimate “but for” pricing).

48. See *id.* at 144–47.

49. For an excellent description of using multiple regression analysis in price forecasting, see Rubinfeld, *supra* note 39, at 1087–94.

Act, is a classic example of the kind of empirical models frequently used in all areas of litigation.

2. Theoretical Models

A theoretical model is also an expression (usually mathematical) that relates factors to an outcome, but it is not directly calibrated by using data from the real world. Rather, it takes as given several assumptions, which in turn are derived from actual real-world knowledge. It then seeks to show, in a hypothetical sense, the outcome of a system for which the assumptions are appropriate.⁵⁰ And although the term “theoretical model” may have a less rigorous ring to it than a “statistical model,”⁵¹ this intuition, once unpacked, is not always correct. What one means by “rigorous” depends on the context and, in circumstances where one has confidence in the theory’s assumptions and little confidence in the available data, a theoretical model will be much more useful than a statistical one.

A simple example of a theoretical model is Telser’s model from his 1960 *Journal of Law and Economics* article that explained manufacturers’ self-interest in restricting competition among distributors of their products.⁵² Telser observed that if retail outlets were free to compete on price for a given manufacturer’s product, and if sales services (like information, demonstrations, and display) for a product could be consumed separately from the product, retailers would have little incentive to provide any services at all.⁵³ His model started from the assumption that if retailers *A* and *B* compete on price, price will fall to marginal cost. Adding retail services raises that cost, and so if only *A* provides services, it will sell the product for a higher price than *B*. It may be that at least some consumers prefer to purchase the product with services, but if services can be consumed separately from the product, and nothing stops consumers from consuming the services from *A* and buying the product from *B*, the model predicts that *B* will take a “free ride” on *A*’s promotional efforts.⁵⁴ This outcome, theorized Telser, was unstable because *A*, unable to recoup its costs of promotion, will stop its additional efforts and lower its price to match *B*’s.⁵⁵

50. Thus, in a way, it is the opposite of an empirical model because it assumes a process and estimates an outcome, whereas an empirical model observes an outcome and estimates a process.

51. Perhaps this is less true in economics, where the theorists are seen to “have lapped the empirical economists.” Coate & Fischer, *supra* note 14, at 152.

52. Lester G. Telser, *Why Should Manufacturers Want Fair Trade?*, 3 J.L. & ECON. 86, 86 (1960). Telser’s model is not the very first articulation of the promotion-inducement theory of vertical restrictions. See, e.g., B.S. YAMEY, *THE ECONOMICS OF RESALE PRICE MAINTENANCE* 3 (1954); Ward S. Bowman, Jr., *The Prerequisites and Effects of Resale Price Maintenance*, 22 U. CHI. L. REV. 825, 840–43 (1955). But among the theory’s early incarnations, it is the most complete and influential.

53. Telser, *supra* note 52, at 91.

54. *Id.*

55. *Id.*

Telser's argument can be expressed as a theoretical model that shows that, given certain conditions (price competition resulting in price equaling marginal cost, services capable of being consumed separately from the product, and consumers being able to find out if *B* is offering the same product at a lower price than *A*), a given outcome can be expected (retail services will not be provided).⁵⁶ Some of these assumptions themselves are susceptible to empirical or theoretical proof. The assumption that price competition results in price equaling marginal cost can be justified with more theory; it comes from the result of "Bertrand competition," a widely used model of price competition.⁵⁷ That services may be consumed separately from the product is an empirically based assumption. It is empirical in the sense that the only way to know whether it is an appropriate assumption is to gather facts about how services and products are actually provided in the real world.

C. WHAT IS A GOOD MODEL?

Armed with some theory about modeling and with some specific examples from antitrust law, I can now begin to describe the art of modeling, identifying the choices that separate good models from the rest. The goal here is to highlight the aspects of modeling that legal decision makers have particular difficulty evaluating: the aspects that reflect creativity, judgment, and intuition on the part of the modeler. These are the aspects of modeling that make factual reasoning inappropriate, and the very aspects that drive courts to treat models and their conclusions as issues of fact.

Many modeling choices are a matter of "science"—as distinct from "art"—in that the optimal choice is unique and objective. It is unique in the sense that experts have settled on a single best solution to a given problem in a given set of circumstances. It is objective in the sense that this optimal choice is a function of expert consensus and so is not defined by the idiosyncrasies of the modeler. We might think of these choices as populating the ends of a spectrum, where on one side sits the "wrong" modeling answer and on the other the "right" answer. On the "wrong" end of the spectrum, we might place modeling mistakes like omitting a variable that is known to have a significant effect on the system and to be correlated with the variable of interest.⁵⁸ On the "right" end of the spectrum, we may find the practice of checking the assumption of normally distributed error terms in multiple regression analysis.⁵⁹ These choices are unique and objective either because they can be falsified, as with the

56. See Alan J. Meese, *Property Rights and Intra-brand Restraints*, 89 CORNELL L. REV. 553, 563 (2004) (discussing the assumptions in Telser's free-riding theory).

57. See Michael R. Baye & Dan Kovenock, *Bertrand Competition*, in THE NEW PALGRAVE DICTIONARY OF ECONOMICS, http://www.dictionaryofeconomics.com/article?id=pde2008_B000336&edition=current&q=bertrand&topicid=&result_number=1 (last visited Feb. 12, 2015).

58. See ABA ANTITRUST DAMAGES, *supra* note 33, at 148–52 (discussing omitted variable bias).

59. See Gel et al., *supra* note 34.

assumption of normally distributed residuals,⁶⁰ or because econometricians over years of study have settled on a robust consensus that, given a particular modeling problem arising in a particular set of circumstances, one solution dominates.

Between these poles of black and white lies a gray area; here is the art of modeling.⁶¹ These are choices that cannot be falsified, that involve trade-offs between costs and benefits to model accuracy, simplicity, and usefulness.⁶² The gray area is created by the complexity of a system involving dozens of choices and assumptions, each of which trades off one benefit for another, and many of which are interrelated. This matrix of options cannot be navigated mechanically, and so requires modes of thinking associated with complex reasoning, such as intuition and creativity.⁶³ Here, even experts do not agree on individual choices, but rather defend their divergent choices by appealing to the broadly accepted criteria that make models useful. Here, scientifically acceptable choices are neither unique (there is more than one way to achieve a modeling goal) nor objective (the idiosyncrasies of the modeler may guide model construction).

Modeling as “art” is a metaphor that works on two levels. First, like the beauty of a work of art, the best choice within the gray area of modeling is in the eye of the beholder. On this level, the art metaphor emphasizes the subjectivity of some modeling choices, where the presence of more than one acceptable solution to a modeling problem gives rise to idiosyncrasy. Second, modeling is an “art” in the way the word is used in the phrase “term of art.” Terms of art are words that derive their meaning from how they are used by a particular class or community of people. Modeling is an “art” in this sense because choices within the gray area are evaluated according to criteria that have broad subscription within the scientific community. On this level, the art metaphor emphasizes that the gray area of modeling is most effectively evaluated by the relevant scientific community itself, which, in the course of becoming a “science,” developed and internalized a unique system for resolving methodological disputes.

The difference between the art and science of modeling is one of degree, not kind. The black and white ends of the spectrum (the “science” of modeling) are places where consensus about a modeling choice would be strong.⁶⁴ The gray

60. *See id.* at 22–23.

61. For examples where modeling is referred to as an “art,” see Baker & Bresnahan, *supra* note 41, at 8; Hill et al., *supra* note 29, at 334 (“Business damages estimation may be most accurately described as an art that relies on methods borrowed from science”); Kaye, *supra* note 15, at 1965 (observing that “statistical modeling is as much art as science”).

62. *Cf.* ABA ANTITRUST DAMAGES, *supra* note 33, at 148–52 (discussing trade-offs in choosing explanatory variables and in dealing with problems of multicollinearity); ABA ECONOMETRICS, *supra* note 32, at 10–13 (same); Baker, *supra* note 35, at 460–62 (describing the tradeoff between tractability and flexibility, and observing that merger simulation models “routinely incorporate judgment calls”); Moses, *supra* note 36, at 117 (“Quite commonly no unique model stands out as ‘true’ or correct”).

63. *See* MORGAN, *supra* note 21, at 25 (noting that modeling often “involves the scientist’s intuitive, imaginative and creative qualities”).

64. I say “would be” because most models are custom-made to address a unique question; thus, it will almost never be possible to actually obtain more than a handful of expert opinions about a given

areas (the “art” of modeling) are places where diverse expert opinion is likely. And where a modeling choice lies on the art/science spectrum depends on the context. Some modeling choices, like functional form selection or variable omission, may be perfectly acceptable, although not inevitable, in one context, and in another are strictly required or prohibited by scientific convention.

Recognizing that modeling is in significant part an art should not cause legal decision makers to throw up their hands in defeat in the belief that all modeling is hopelessly indeterminate and subjective. At the very least, the “science” of modeling—the black and white—is susceptible to objective evaluation. If a legal decision maker can identify these modeling choices and assess their conformity to convention, he will frequently have done most of the evaluative work necessary to accept or reject a model offered in a legal dispute. Where evaluation of the “art” of modeling is necessary, the task may be more difficult for a legal decision maker but far from impossible. The difficulty for legal decision makers is that they cannot merely defer to expert consensus to evaluate the art of modeling. Evaluating the appropriateness of modeling choices operating in a complex and interrelated system requires some of the same artistry as constructing the model itself.

The following sections broadly sketch some criteria that ought to be used in assessing the usefulness of a model. They are offered to show that the kind of reasoning that should be employed to evaluate models is distinct from factual reasoning, and that judges are intellectually capable of meaningfully engaging with the art and science of modeling, especially if provided additional training. Of course, a full-scale tutorial on model thinking is beyond the scope of this Article;⁶⁵ rather, these sections endeavor to identify broad categories of questions that can reveal the power and the flaws of models used in legal contexts.

1. Where Possible, a Model Should Be Validated with Real World Data

First, if possible, models should be calibrated against reality before they are used to predict or estimate the unknown.⁶⁶ This is essentially the function of the “before” (and, if possible, “after”) periods in the before-during-after model of antitrust damages for Section 1 cases. Where pricing data show a spike during a period of known collusion, there is intuitive appeal to attributing the increase to the collusive activity and awarding damages on that basis. That intuition relies

modeling choice. The black, white, and gray spectrum of modeling must be understood as a hypothetical array of expert responses to a modeling problem.

65. There are several volumes aimed at providing comprehensive guidelines for judges and lawyers tasked with evaluating models. *See, e.g.*, ABA ECONOMETRICS, *supra* note 32; DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (2011); David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, in FED. JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 83 (2d ed. 2000), available at [http://www.fjc.gov/public/pdf.nsf/lookup/sciman00.pdf/\\$file/sciman00.pdf](http://www.fjc.gov/public/pdf.nsf/lookup/sciman00.pdf/$file/sciman00.pdf). For a volume dedicated to antitrust damages in particular, see ABA ANTITRUST DAMAGES, *supra* note 33.

66. *See, e.g.*, Coate & Fischer, *supra* note 14, at 165 (“The authors advise the merger analyst to check how the simulation model predicts the historical record of competition in the market.”).

on a model, albeit an implicit one, that assumes other factors that affect price (demand features, input costs, etc.) were operating identically during, before, and after the collusive period. The before-during-after method seeks to add rigor to this intuition by deriving a model with parameters that include variables like demand and costs that can isolate the effect of collusion on the higher price during the spike. That model, the “but-for” model of competition without collusion, can be calibrated by measuring its ability to accurately estimate prices for which data are available from the real world.⁶⁷ If the model is powerful to “predict” actual events (the before/after period) then that strengthens the inference that it can accurately “predict” an unobservable event (pricing absent collusion in the “during” period).

Likewise, theoretical models can be evaluated based on their ability to describe actual events.⁶⁸ For example, classical competition theory predicts that, in a market where “more than a handful of firms participate[] in a relatively homogenous goods market,” price will approximate cost.⁶⁹ The traditional monopoly model predicts that a single firm operating without competitors and facing a downward-sloping demand curve will raise price relative to its marginal costs and reduce output.⁷⁰ These two theoretical models, which form the basis of modern antitrust economics, have been demonstrated to be accurate by empirical studies of markets of various compositions; by and large, atomistically competitive markets price near cost, and monopolistic markets tend to feature higher prices and higher profits.⁷¹ Thus, since these models describe real-world phenomena in a context susceptible to empirical study, they can be used to describe phenomena in a context resistant to it.

2. Assumptions Should Be Checked Where Possible

In theory, a model that accurately predicts outcomes need not rely on reasonable assumptions to be useful.⁷² As Milton Friedman argued, an economic theory describing market actors as rational and fully informed is suffi-

67. See ABA ANTITRUST DAMAGES, *supra* note 33, at 171–73.

68. Dennis W. Carlton, *The Relevance for Antitrust Policy of Theoretical and Empirical Advances in Industrial Organization*, 12 GEO. MASON L. REV. 47, 47 (2003) (discussing empirical studies confirming basic economic theory about competition, entry, and price discounting); see MORGAN, *supra* note 21, at 34 (observing that modelers judge models “by comparing the behaviour of the model world to that of the real world in a kind of matching or bench-marking process”).

69. Coate & Fischer, *supra* note 14, at 153.

70. DENNIS W. CARLTON & JEFFREY M. PERLOFF, MODERN INDUSTRIAL ORGANIZATION 87–88 (3d ed. 2000).

71. See *id.*; Coate & Fischer, *supra* note 14, at 154.

72. See Coate & Fischer, *supra* note 14, at 138–41 (summarizing the argument that economic models are sufficient if they predict accurately, and that “their underlying reality is not relevant”). See generally MILTON FRIEDMAN, *The Methodology of Positive Economics*, in ESSAYS IN POSITIVE ECONOMICS 3 (1953) (arguing that models can be evaluated without addressing the realism of their assumptions); Leo Breiman, *Statistical Modeling: The Two Cultures*, 16 STAT. SCI. 199 (2001) (defending the use of algorithmic models—models that can predict real-world outcomes of systems but that do not seek to explain the internal mechanisms of those systems).

cient if the market functions *as if* all actors were rational and fully informed.⁷³ But there are two problems with evaluating a theory solely by its ability to predict rather than explain.

The first is simple: some models' predictions cannot be tested for accuracy because the data are unavailable. Consider Telser's model of free-riding discussed above. In principle, his theory yields an empirically testable hypothesis that where services can be consumed separately from the product, they will not be provided at all. But this hypothesis turns out to be relatively difficult to measure because, if it were true and the result (no retail services) were inefficient, retailers and manufacturers would have incentives to solve the free-riding problem through other mechanisms.⁷⁴ One mechanism would be manufacturers prohibiting their dealers from discounting; indeed, this is the very phenomenon that Telser was attempting to explain in developing his free-riding model.⁷⁵ Further, data about the real world may be difficult or impossible to obtain. To stay with the resale price maintenance (RPM) example, economic theory also indicated that the practice could be used to facilitate illegal cartels.⁷⁶ But proving that the cartel theory accurately predicted behavior was difficult because observing a correlation between resale price maintenance and cartels required uncovering cartel activity, which, because of its criminality, is systematically hidden.⁷⁷

The second problem with relying on the predictive value of models without evaluating their assumptions lies with the possibility that models that correctly predict the known world might poorly predict the unknown world if the model does not account for how the world actually works. A model that accurately estimates only known outcomes is trite; its ability to do so is good only for supporting an inference that it can predict unknown outcomes. But that inference depends on a crucial, and sometimes indefensible, assumption that the known conditions are similar to the unknown conditions, at least with respect to how the model works.

The most obvious example of this problem is the principle that the past cannot always predict the future. An algorithm may describe a past pattern well, even if it incorporates bad assumptions, when in the past the system that created the pattern was not particularly sensitive to those assumptions. But future or otherwise uncertain conditions may make those assumptions critical to the

73. FRIEDMAN, *supra* note 72, at 21–22.

74. Benjamin Klein, *Competitive Resale Price Maintenance in the Absence of Free Riding*, 76 ANTITRUST L.J. 431, 438–39 (2009) (stating the traditional account, that resale price maintenance is used to prevent free-riding, before offering his own procompetitive theory of resale price maintenance).

75. See Telser, *supra* note 52, at 86–88 (describing the “puzzle” of manufacturers contractually preventing their goods from being sold at less than “fair trade” prices).

76. Lester Telser articulated this use of RPM in the same seminal article where he introduced his free-rider model. See *id.* at 96–99.

77. For an illustration of the difficulty of empirically measuring the use of RPM to facilitate collusion, see Pauline M. Ippolito, *Resale Price Maintenance: Empirical Evidence from Litigation*, 34 J.L. & ECON. 263 (1991).

outcome, and the model, which predicted the known world well, will likely be wrong when applied to predict or estimate unknown outcomes. For example, in 2005, an algorithmic model of housing prices that did not account for causal factors influencing the housing market but was capable of recreating known housing prices in the past, might have suggested that real estate prices would continue to rise as they had for decades.⁷⁸ This model would turn out to be wrong for the near future because it failed to address the fact that the actual factors that influenced housing prices were set to drastically change in the following years.⁷⁹

Therefore, since a model's worth may not always be determined solely by its ability to recreate known outcomes, models should also strive to incorporate realistic assumptions and to tell a reasonable causation story.⁸⁰ All models use assumptions, implicit and explicit, in simplifying complex natural or market processes into a functional form.⁸¹ These assumptions, in econometric analysis, may include the assumption of linearity, the assumption that observed data comes from a normal distribution, and many more. In theoretical models, assumptions may include symmetry of information or competition through price setting.⁸² Many assumptions, including the ones above, can be checked for accuracy in the context of the model to help ensure that the model more accurately reflects the natural or market process it seeks to describe.⁸³

If an assumption cannot be verified empirically, then the model should be tested for its sensitivity to that assumption.⁸⁴ If an assumption turns out to have little to no effect on the predictions of the model, then its validity is irrelevant to

78. Indeed, this seemed to be the model, at least implicitly, relied upon by Freddie Mac's chief economist, who, in 2005, remarked, "I don't foresee any national decline in home price values. Freddie Mac's analysis of single-family houses over the last half century hasn't shown a single year when the national average housing price has gone down." *Housing Bubble—or Bunk?*, Bus. W. (June 21, 2005), <http://www.businessweek.com/stories/2005-06-21/housing-bubble-or-bunk>.

79. Similarly, a model showing that the stock market rises and falls with fashion trends in ladies' hemlines, however accurate to describe the past, should be viewed skeptically as a predictor without a coherent theory of how and why skirts correlate with economic performance. For a discussion of the "hemline theory," associated with economist George W. Taylor in the 1920s, see David S. Law, *Globalization and the Future of Constitutional Rights*, 102 Nw. U. L. REV. 1277, 1306 & n.115 (2008).

80. For a discussion of the role of assumptions in evaluating theory, see Ronald Coase's response to Friedman. R.H. COASE, *How Should Economists Choose?*, in *ESSAYS ON ECONOMICS AND ECONOMISTS* 15, 18 (1994). See also Gilboa et al., *supra* note 27, at 16 (observing that theoretical economists value the "axiomatization" of the foundations of a theory, not merely the accuracy of its "predictions, fit to reality, and so forth"); Moses, *supra* note 36, at 118 (discussing the importance of "realism" in modeling, defined as whether a model reflects "the actual data-generating process").

81. See ABA ECONOMETRICS, *supra* note 32, at 7 (discussing assumptions in econometric models).

82. On the role of accurate assumptions in structural (game-theoretical) models, see J. Douglas Zona, *Structural Approaches to Estimating Overcharges in Price-Fixing Cases*, 77 ANTITRUST L.J. 473, 478 (2011).

83. Cf. Jonathan B. Baker, *Contemporary Empirical Merger Analysis*, 5 GEO. MASON L. REV. 347, 356–60 (1997) (identifying four assumptions that appear frequently in merger simulation models that can and should be verified empirically when possible).

84. See *id.* at 361 (highlighting "the importance of testing the sensitivity of parameter estimates and simulation results to critical assumptions").

the usefulness of the model. Further, at least in theoretical models, inaccuracy of assumptions may weaken, but not eliminate, the effect predicted by the model.⁸⁵ A model describing an extreme outcome under strong, unrealistic assumptions may still be useful if there is reason to believe that a similar (but perhaps lesser) result will obtain under a weaker and more realistic version of the assumption.⁸⁶

3. Modelers Should Be Required to Disclose All Other Modeling Choices They Considered and Give Reasons for Each Modeling Choice

Especially in statistical modeling, there is a temptation to iteratively design models and then select a particular model based on its ability to reveal a statistically significant result that supports a predetermined conclusion.⁸⁷ This practice—often called data mining—is considered unprofessional in the social scientific community⁸⁸ and would not only fail a *Daubert*-style assessment of scientific merit, but it is downright deceptive in the context of a trial. But even in the social scientific fields where data mining is so stigmatized, it is still considered to be a significant problem.⁸⁹ And there is no reason to expect the problem to be any less serious in the litigation context; in fact, since the stakes of a trial often outstrip the stakes of scholarly publication, it may be even worse.

This problem is very difficult to address, especially in the legal context. One potential solution, but by no means a panacea, would be for courts to demand that experts reveal all the statistical models and modeling choices that they considered.⁹⁰ Of course, this fix can only go so far because its enforcement presents logistical difficulties. It is impractical and undesirable to demand access to experts' hard drives and research notes, and even if that were possible, it would still be all too easy for modelers to fail to report the existence and location of alternative models they considered. Therefore, modelers should also be asked to defend every modeling choice—however generally accepted in the

85. Cf. Moses, *supra* note 36, at 118 (“If the model is *false* in certain respects, how badly does that affect estimates, significance test results, etc., that are based on the flawed model?”).

86. For example, the strong version of Telser’s free-riding model assumes that consumers have perfect information about competitors’ prices and services, but even with less-than-perfect information, some free-riding could be expected to occur. Thus, even without knowing exactly how much information consumers have—that is, without knowing the validity of the assumption—the extreme version of the model can point to the more modest observation that if there is any dissemination of information among consumers about the availability of free sales services, those services will be underprovided relative to a market without free-riding.

87. See Richard Berk et al., *Statistical Inference After Model Selection*, 26 J. QUANTITATIVE CRIMINOLOGY 217, 219 & n.2 (2010).

88. See *id.* at 217.

89. See, e.g., Joseph P. Simmons et al., *False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant*, 22 PSYCHOL. SCI. 1359 (2011) (discussing the problem of data mining in experimental psychology). Simmons et al. believe this phenomenon occurs not only in instances of deception, suggesting that the problem is widespread. *Id.* at 1359.

90. Analogously, Simmons et al. suggest that researchers report “all variables collected in a study” as well as “all experimental conditions, including failed manipulations.” *Id.* at 1363.

scientific community—incorporated in their model. Where an expert cannot offer a sufficiently compelling reason for using one modeling technique rather than another, he should be required to show that the model is not sensitive to that choice.⁹¹ Although this will never solve the data-mining problem, it will at least make it more difficult for modelers to deceive courts in this way.

4. Multiple Models Should Be Used Where Possible

Finally, because every modeling choice carries a risk of distorting the model's calculation or prediction away from reality, and because there are usually quite a number of acceptable modeling choices, modelers should confirm their predictions by using multiple models. Where the results of several well-constructed models—designed to measure the same phenomenon but incorporating different modeling choices—predict similar outcomes, fears about the distorting effect of those choices are diminished.⁹² A corollary to this principle is that a model whose result is highly sensitive to slight design changes is inherently suspect.⁹³

II. ARE MODELING CHOICES AND OUTCOMES ISSUES OF FACT?

With the previous Part's brief sketches of what models are, how they are used in law, and how they are to be evaluated, I am now in a position to explore four interrelated questions. First, what is at stake in categorizing questions about models as factual issues? Second, should judges treat models and their predictions as facts? Third, do courts treat questions about models and their usefulness as factual questions? Finally, since the answers to the last two questions do not match, why do courts treat modeling issues as factual when such a treatment does not optimize decisionmaking?

A. THE STAKES: WHAT MATTERS ABOUT THE "FACT" LABEL?

As it turns out, the stakes of the labeling exercise and its normatively optimal outcome are intimately related.⁹⁴ Categorizing an issue as "factual" in the eyes of the law is not an exercise in ontological classification between positive assertions of reality (facts) and rules and principles for behavior or judgment (law). The modern view, endorsed by courts and critics alike, is that the label of "fact" or "law" (or something in between) attaches because of pragmatic

91. Analogously, Simmons et al. suggest that reviewers demand that authors "demonstrate that their results do not hinge on arbitrary analytic decisions." *Id.*

92. See Moses, *supra* note 36, at 117 ("[I]t is not unusual for a given data set to be analyzed in several apparently reasonable ways. If conclusions are qualitatively concordant, that is regarded as grounds for placing additional trust in them.").

93. In such a case, a model is said to not be particularly "robust." See Rubinfeld, *supra* note 39, at 1070.

94. Cf. JULIAN L. SIMON, BASIC RESEARCH METHODS IN SOCIAL SCIENCE: THE ART OF EMPIRICAL INVESTIGATION 294 (1969) ("The real question . . . is not whether or not the items that you lump together are different in some ways, but rather whether or not they are similar *for your purposes*.").

judgments about how an issue is best resolved by the legal system.⁹⁵ As Judge Friendly put it, “[W]hat a court can determine better than a jury [is] perhaps about the only satisfactory criterion for distinguishing ‘law’ from ‘fact.’”⁹⁶ Thus, the stakes of the question, or the practical effects of a “factual” label, drive the classification task in the first place.⁹⁷ The pragmatic nature of the fact/law distinction is no more evident than in so-called mixed questions of law and fact, where courts have custom-designed optimal decisional procedures that mix and match traditionally factual and legal modes of decisionmaking.

1. The Consequences of the “Fact”⁹⁸ Label

If an issue is labeled “factual,” then it is allocated to the jury⁹⁹ (or a judge if he is acting as the finder of fact), is left undisturbed by appellate courts except in extreme cases,¹⁰⁰ and is not accorded precedential value.¹⁰¹ Therefore, according to the pragmatic view of the fact/law distinction, “facts” are issues for which judges have no special expertise beyond a lay jury, appellate second-guessing tends to reduce accuracy or legitimacy, and reasoning by analogy to other cases is unhelpful. Issues do not need high scores in all three categories to

95. Allen & Pardo, *supra* note 17, at 1770. For judicial endorsements of this perspective, see Miller v. Fenton, 474 U.S. 104, 114 (1985) (observing that the fact/law distinction often turns on “a determination that, as a matter of the sound administration of justice, one judicial actor is better positioned than another to decide the issue in question”); Bose Corp. v. Consumers Union of U.S., Inc., 466 U.S. 485, 501 n.17 (1984) (endorsing a similarly pragmatic definition of the fact/law divide).

96. United States v. J.B. Williams Co., 498 F.2d 414, 431 (2d Cir. 1974).

97. For a related approach to determining the appropriate mode of judicial review of congressional fact-finding, see John O. McGinnis & Charles W. Mulaney, *Judging Facts Like Law*, 25 CONST. COMMENT. 69, 94 (2008) (“[T]he real questions about whether the judiciary should wield an independent, de novo role in social fact-finding are pragmatic, comparative, and structural.”).

98. Here I am using “fact” as shorthand for “adjudicative fact.” Kenneth Culp Davis famously distinguished facts particular to the dispute at issue—“what the parties did, what the circumstances were, what the background conditions were”—from abstract and generalizable truths about the world—such as “social and economic data.” Kenneth Culp Davis, *An Approach to Problems of Evidence in the Administrative Process*, 55 HARV. L. REV. 364, 402 (1942). The former he called “adjudicative facts” and the latter “legislative facts.” See *id.* at 402–10 (defining adjudicative and legislative facts and providing examples). Today, the distinction is preserved in the Federal Rules of Evidence, which permit judicial notice of an adjudicative fact only if it is not subject to reasonable dispute, see FED. R. EVID. 201(a), which implies that judges can decide legislative facts with more freedom, see FED. R. EVID. 201 advisory committee’s note. Although Culp Davis’s framework heavily influences my distinction between issues appropriate for juries and those for judges to determine, in this Article I do not directly import his subdivision of facts into “adjudicative” and “legislative” for two reasons. First, issues arising from modeling—from individual choices in model construction to the appropriateness of model-based inference—almost always mix dispute-specific and generalized elements in a way that prevents easy sorting. Second, I find it more useful to refer to modeling as raising “mixed questions of law and fact,” as discussed below, because of the rich case law on mixed questions.

99. See Allen & Pardo, *supra* note 17, at 1779–81 (discussing the Seventh Amendment’s requirement that juries find facts).

100. *Id.* at 1784–85.

101. Of course, parties can waive their right to have their facts decided by a jury and elect a bench trial. But in these exceptional instances, the judge stands in the shoes of a jury, and as such he is held to a low standard for explaining and justifying his decisions, his determinations are accorded little precedential value, and they are reviewed deferentially on appeal.

be considered factual; rather, the label attaches when an overall judgment governed by these criteria reveal an issue to be more factual than not.

The consequence of labeling an issue “factual” is essentially to black-box the reasoning behind its resolution. Jury deliberation is the most opaque kind of decisionmaking that exists in the legal system.¹⁰² Although judges and attorneys tightly control the flow of information to the jury through evidentiary rules, presentation of arguments, examination of witnesses, and jury instructions, the final reduction of those inputs to a verdict is deliberately hidden from view. A jury is under no obligation to conform its decision to previous legal determinations or even to explain its reasoning at all.¹⁰³

The second property of findings of fact, that they receive deferential appellate review, is largely a consequence of these features of jury deliberation. Without a record of the reasoning process, appellate reconsideration of factual issues would involve examining the inputs into the decision (testimony, party arguments, and jury instructions) and determining the appropriate outcome. This would be entirely redundant of the jury’s decisionmaking procedure and would, as the familiar saying goes, “invade the province of the jury.”¹⁰⁴

Likewise, the third property of factual determinations, that they not be accorded precedential value, also flows from the “black box” status of jury decisionmaking. By exempting the jury from spelling out the link between its decisional inputs and its verdict, the law makes analogical reasoning between jury verdicts impossible. Without a record of how inferences were made, subsequent decision makers cannot have confidence that their similar (yet inevitably different) trial experience should yield the same factual determination.

2. Why Facts Are “Facts”

Why treat any legally relevant issue as a question of fact? Opacity in decisionmaking is arguably illegitimate and, if it prevents mistakes from coming to light, it can promote error. Relatedly, lack of meaningful review creates a moral hazard problem; for those inside the black box, the temptation to shirk decisionmaking responsibility may be strong. And juries are typically less educated than judges and certainly less experienced in reasoning from courtroom evidence.

Relatively little theory is offered to justify jury decisionmaking, and even less empirical proof. Therefore, any account of why some issues (namely, facts) are left to juries to decide must be somewhat speculative. It may be that some issues are better decided by juries. Where a decision requires a common-sense inference about human behavior, a judge has no special knowledge over an indi-

102. See *United States v. Thomas*, 116 F.3d 606, 619 (2d Cir. 1997) (“The jury as we know it is supposed to reach its decisions in the mystery and security of secrecy; objections to the secrecy of jury deliberations are nothing less than objections to the jury system itself.”).

103. See Allison Orr Larsen, *Bargaining Inside the Black Box*, 99 GEO. L.J. 1567, 1572 (2011) (observing that jurors are “not required to provide reasons for their final judgment”).

104. *Goldman v. United States*, 245 U.S. 474, 477 (1918).

vidual juror, and a jury's advantage of numerosity may allow it to outperform a judge. As compared to an individual, a jury will comprise more life experiences that can be used to inform common-sense judgments; numerosity increases the likelihood that at least one juror will have especially good information about a given question.¹⁰⁵ Numerosity contributes to accurate decisions even where the decision makers have relatively little information; the Condorcet Jury Theorem predicts that, where each juror is fallible but has some information to contribute to a decision, a group will significantly outperform an individual.¹⁰⁶ This information advantage could be part of the theory for why questions about the meaning of human behavior, such as whether a defendant's failure to hide his theft indicated that he believed the property to be abandoned¹⁰⁷ or whether a witness's failure to recall details indicates lying, are treated as questions of fact.

Even where juries may not be superior to judges in terms of accuracy, they may be preferable for the legitimacy that they can lend to sensitive decisions. Juries are often asked to estimate social norms or calibrate judgments about fairness.¹⁰⁸ Questions like what kind of conduct is considered "lewd and lascivious" or about the monetary reward that reflects fair and just compensation for pain and suffering are typically treated as questions of fact for the jury because a judge's decision may be perceived as autocratic or arbitrary.¹⁰⁹ These decisions may be better left to the jury because jury decisions are black-boxed; searching review of such indeterminate and culturally sensitive issues risks an appearance of illegitimacy.¹¹⁰

105. NEIL VIDMAR & VALERIE P. HANS, *AMERICAN JURIES: THE VERDICT* 74 (2007). For example, if a trial turns on whether a woman would have noticed a ten-centimeter lump in her breast, a jury with at least one woman may be a better decision maker than a male judge. *Cf. id.* at 136–37 (reproducing a discussion among jurors about whether a tumor of that size would easily be noticed by a woman during a shower). In this example, the deliberative property of jury decisionmaking allows high-information decision makers to prevail.

106. We may call this the aggregative property of juries, and it is illustrated in the following example: where a trial turns on whether a witness is truthful about his alibi, and if the average person has even a highly limited ability to detect lying, then the aggregative property of juries allows them to outperform individual judges in credibility assessments. *Cf.* Paul H. Edelman, *On Legal Interpretations of the Condorcet Jury Theorem*, 31 *J. LEGAL STUD.* 327 (2002) (discussing the frequent use—and misuse—of the Condorcet Jury Theorem in legal scholarship); Adrian Vermeule, *The Parliament of the Experts*, 58 *DUKE L.J.* 2231, 2245–48 (2009) (discussing the Condorcet Jury Theorem—and its limits—in the context of aggregating decisionmaking).

107. VIDMAR & HANS, *supra* note 105, at 135–36.

108. *Id.* at 142.

109. Of course here, too, numerosity can help. *Cf.* Harold L. Korn, *Law, Fact, and Science in the Courts*, 66 *COLUM. L. REV.* 1080, 1103–04 (1966) (quoting Oliver Wendell Holmes, Jr. for the proposition that, when it comes to questions about negligence, a court "feels that it is not itself possessed of sufficient practical experience to lay down the rule intelligently" and "conceives that twelve men taken from the practical part of the community can aid its judgment").

110. The black box of the jury is also used to decide—and shield from scrutiny—issues that cannot be resolved according to any rational principle. An issue is more likely to be characterized as a "factual dispute" where there is persuasive evidence on both sides or little evidence on either side. *Cf.* OLIVER WENDELL HOLMES, JR., *THE COMMON LAW* 127 (1881) (pointing out that only cases on the "doubtful border" are left to a jury to decide). Where the matter is unyielding to principled analysis, as in the

A second reason for treating some issues as facts is that for some kinds of decisions, precedential reasoning is less helpful. These are the issues that are specific to a case and are virtually incapable of being repeated in a meaningful way in a later case. Thus, factual questions are typically those that are unique to the controversy in question,¹¹¹ such as whether the light was red at the time of the accident or whether someone acted with malice aforethought. The information relevant to these questions is held only by a select few, typically closely associated with the dispute itself;¹¹² these questions are resolved by appeal to lay witnesses, often through oral testimony, as opposed to research into widely available documents and data.¹¹³ These questions are uncontroversially treated as facts in part because they are unique to the dispute and so analogical reasoning among existing cases has little to offer the decision. Black-boxing these decisions is relatively costless since an on-the-record analysis would not be useful to a subsequent case.

3. Mixed Questions and the Myth of the Fact/Law Divide

Since the fact/law distinction is often referred to as a “divide,” it is tempting to think that once an issue is found to be unsuitable for treatment as a fact, then it must be considered under the procedures reserved for law. However—and here is where the pragmatism of the fact/law distinction is most apparent—courts have recognized a vast middle ground of issues known as mixed questions of law and fact that are best resolved by hybrid decisional processes that involve some procedures traditionally reserved for facts and some procedures reserved for law. The prominence of mixed questions, and the relative workability of hybrid rules for their resolution, illustrate that taking modeling out of the “fact” box does not obligate courts to treat them as pure questions of law.

Mixed questions are typically fact-sensitive, case-by-case determinations that require the application of a legal standard to a specific fact pattern where the precise line between factual and legal reasoning is blurred. As the Supreme Court has explained: “A finding of fact in some cases is inseparable from the principles through which it was deduced. At some point, the reasoning by which a fact is ‘found’ crosses the line between application of those ordinary principles of logic and common experience . . . into the realm of a legal rule”¹¹⁴

classic “he said, she said” scenario, the law needs a way to decide, even if arbitrarily, and then a way to hide the arbitrariness of the decision. Labeling the issue as “factual” serves both functions.

111. *See* McGinnis & Mulaney, *supra* note 97, at 75 (“Adjudicative facts cover ‘what the parties did, what the circumstances were, [and] what the background conditions were.’” (alteration in the original) (quoting Culp Davis, *supra* note 98, at 402)).

112. *Cf.* Culp Davis, *supra* note 98, at 402 (defining adjudicative facts as those “which concern only the parties to a particular case” (emphasis omitted)).

113. *See id.*

114. *Bose Corp. v. Consumers Union of U.S., Inc.*, 466 U.S. 485, 501 n.17 (1984).

Classic mixed questions often involve broad standards—such as “reasonable suspicion” or “discriminatory intent”—applied to particular circumstances.¹¹⁵ Like the legal treatment of pure questions of fact and law, the legal treatment of mixed questions turns on practical considerations of comparative competence and judicial administrability. And the legal treatment of mixed questions is not one-size-fits-all. Although the “mixed” label does typically take the issue away from the jury, the level of deference on appeal and whether the resolution is treated as precedent varies issue by issue.¹¹⁶

Whether a mixed question receives deferential review on appeal ostensibly turns on whether the issue was committed to the trial court’s discretion,¹¹⁷ but that question, in turn, is driven by utilitarian considerations such as which judicial actor is better positioned to decide the question.¹¹⁸ For example, the Supreme Court has held that the question of “actual malice” in a product disparagement suit, which it identified as a mixed question of law and fact, receives *de novo* review.¹¹⁹ But it held that the question of whether the government’s litigation position is “substantially justified,” which it also treated as a mixed fact/law question, receives deferential appellate review.¹²⁰

Whether resolutions of mixed questions are holdings entitled to precedential value is an even more complex question. Doctrinally, any appellate decision will have precedential sway in subsequent cases, so both deferential and *de novo* appellate opinions will technically bind future lower courts.¹²¹ But a deferential opinion affirming the court below will provide little guidance to lower courts because, by definition, “deference” means affirming any opinion from a set of acceptable—yet different—holdings. Even an opinion reversing a lower court’s determination as “clearly erroneous” or an “abuse of discretion” will provide little guidance to courts other than forbidding an extreme position.

At first blush, *de novo* opinions have more potential to bind lower courts because they affirm not ranges of acceptable answers but, rather (in theory), identify the precisely correct resolution to legally relevant questions. But because mixed questions are typically very fact-dependent, the precedential value of an appellate determination is necessarily limited. Determinations of issues such as actual malice in a product disparagement case, the voluntariness of a

115. Of course, any application of a standard to a fact pattern involves resolving “historical facts” to establish the fact pattern in the first place, and these are uncontroversially treated as facts. *See, e.g.*, *Ornelas v. United States*, 517 U.S. 690, 696 (1996).

116. *Cf. id.* at 701 (Scalia, J., dissenting) (“Merely labeling the issues ‘mixed questions,’ however, does not establish that they receive *de novo* review.”).

117. HARRY T. EDWARDS & LINDA A. ELLIOTT, *FEDERAL COURTS STANDARDS OF REVIEW* 16–17 (2007).

118. *See, e.g.*, *Salve Regina Coll. v. Russell*, 499 U.S. 225, 233 (1991) (“[D]eferential review of mixed questions of law and fact is warranted when it appears that the district court is ‘better positioned’ than the appellate court to decide the issue” (quoting *Miller v. Fenton*, 474 U.S. 104, 114 (1985))).

119. *Bose*, 466 U.S. at 510–11.

120. *Pierce v. Underwood*, 487 U.S. 552, 562–63 (1988).

121. *See* Randall H. Warner, *All Mixed Up About Mixed Questions*, 7 J. APP. PRAC. & PROCESS 101, 136 (2005) (explaining that even a determination about evidentiary sufficiency, which involves a very deferential standard of review, creates precedent).

confession, or whether a police officer had reasonable suspicion, inevitably turn on such a complex constellation of factors that they are easily distinguished in future cases. Which is to say that future courts, relying heavily on analogy, may look to these appellate determinations for guidance, but they will rarely (as rarely as the recurrence of the same exact set of relevant circumstances) be tightly bound by them.

B. MODELS AND THEIR CONCLUSIONS SHOULD NOT BE TREATED AS FINDINGS OF FACT

Models and their conclusions score low on the test for categorization as facts.¹²² They are better evaluated by a judge, they do not merit deference on appeal, and at least some modeling choices are made at a broad level of generality susceptible to analogical reasoning from other models; thus, according to judicial assessment of models some limited precedential value may improve decisionmaking. These features suggest that modeling should be treated as presenting mixed questions of fact and law.

In reaching these conclusions, I follow the framework established in an influential series of articles by John Monahan and Laurens Walker classifying different uses of social science in the law and prescribing judicial procedures for their introduction, use, and review. In the first article in the series, they identified social scientific research used to establish legal rules as “social authority” and suggested that courts treat this research “much as courts treat legal precedent.”¹²³ In their second article, they identified social scientific research used to inform the fact-finding process as “social frameworks” and suggested that courts take briefing on framework evidence and impart it to the jury through instructions.¹²⁴ Finally, their last article identified social scientific claims unique to the case at hand as “social facts,” whose methodology (but not conclusions) should be treated as an issue of law.¹²⁵

Monahan and Walker’s prescriptions were driven in large part by where they placed a use of social scientific information on the fact/law spectrum, and particularly whether it was being offered as “legislative fact” or “adjudicative fact.”¹²⁶ They used both ontological and pragmatic criteria to inform these

122. Cf. Korn, *supra* note 109 (arguing, almost thirty years before *Daubert*, that scientific knowledge should not be treated as factual determination).

123. John Monahan & Laurens Walker, *Social Authority: Obtaining, Evaluating, and Establishing Social Science in Law*, 134 U. PA. L. REV. 477, 478 (1986).

124. Laurens Walker & John Monahan, *Social Frameworks: A New Use of Social Science in Law*, 73 VA. L. REV. 559, 559–60 (1987).

125. Laurens Walker & John Monahan, *Social Facts: Scientific Methodology as Legal Precedent*, 76 CAL. L. REV. 877, 877 (1988).

126. See Monahan & Walker, *supra* note 123, at 482–88. Here, Monahan and Walker are indebted to Kenneth Culp Davis for his elaboration on the legislative/adjudicative distinction. See Culp Davis, *supra* note 98.

categorizations, but emphasized the pragmatic.¹²⁷ If a social scientific issue was “generalizable,” or lent itself to reasoning by analogy, the authors favored its treatment as a question of law, not fact.¹²⁸ Their methodology can be profitably imported and expanded to help answer the question of whether a model and modeling choices should be “facts” in the eyes of the law.

The following three sections evaluate whether models and their conclusions are facts according to the practical desirability of the consequences of the categorization. This pragmatic approach would label as a “fact” any issue that is best decided by the process the law reserves for issues of fact. In other words, a fact is an issue that is better off “black-boxed,” treated deferentially on appeal, and not accorded precedential value. None of this is to say that modeling ought to be treated as presenting pure questions of law, a position I reject in the last section.

1. Decisions Involving Modeling Choices Ought to Have (Weak) Precedential Value

According to Monahan and Walker, “facts,” pragmatically defined, tend to be specific to a case and resistant to analogy across cases.¹²⁹ At the other end of the spectrum lie generalizable issues, like legal rules and principles of proof that apply in multiple disputes. These can be generalized—indeed, they are generalizable by design because they derive much of their value from their ability to translate into different factual contexts.¹³⁰ Black-boxing reasoning about so-called issues of law would defeat generalizability because the process of generalization requires a transparent decisionmaking process that can be imported by analogy to other contexts. Like decisions about legal rules and principles, decisions about modeling choices can and should be generalized.¹³¹

Some aspects of legal decisions about modeling can be generalized because individual modeling problems are like social problems governed by legal rules—they arise repeatedly in similar contexts.¹³² Where another modeler faced a similar challenge, that modeler’s successful solution can potentially be imported to the current context and help improve legal decisionmaking. Like

127. Monahan & Walker, *supra* note 123, at 494–95 (noting that the classification of social science research as law versus fact turns on “the quality of the judicial management procedures that flow from this . . . classification”).

128. *See id.* at 490–91.

129. *See* Walker & Monahan, *supra* note 125, at 887–88.

130. On the generalizability of rules, see Frederick Schauer, *Do Cases Make Bad Law?*, 73 U. CHI. L. REV. 883, 890–91 (2006) (“What makes a rule a rule, and what distinguishes a rule from a particularized command, is precisely the way in which a rule builds on a generalization and prescribes for all of the acts or events encompassed by the generalization.”).

131. Some scholars have explicitly compared reasoning about models to analogical reasoning from precedent. *See, e.g.,* Gilboa et al., *supra* note 27, at 17.

132. *Cf.* Faigman, *supra* note 10 (arguing that *Daubert* decisions should not be reviewed deferentially because decisions about scientific validity have general application); Michael J. Saks, *The Aftermath of Daubert: An Evolving Jurisprudence of Expert Evidence*, 40 JURIMETRICS J. 229, 233 (2000) (same).

opinions about other mixed questions of fact and law—for example whether there was probable cause to support an arrest or whether a litigation position was substantially justified—modeling choices will be highly fact-specific. But that does not mean that answers to mixed questions are absolutely not generalizable or that previous cases do not guide judicial decisionmaking at all. Modeling dilemmas often do recur, and judges can reason by analogy among factually distinct cases.

Problems such as small sample sizes, uncertainty about functional form, and the appropriateness of simplifying assumptions recur, and although a single rule of thumb is rarely available for each of these challenges, econometricians and theorists have developed a rigorous—albeit complex—set of possible solutions. Similarly, a body of written case law considering modeling issues could inform judgments about how closely a model must describe the real world in order to be useful to describe an unobserved phenomenon, when multiple confirmatory models should be expected, and perhaps even what level of statistical significance is required as legal proof.¹³³

Some models can be imported wholesale between different contexts. Simple, classical models, like that of competition and monopoly, serve as the basis for modern economic analysis and recur throughout legal discourse.¹³⁴ Even other more complex theoretical models, like Telser's free-riding model, have general application to many fact patterns relevant to antitrust adjudication and rulemaking.¹³⁵ Even where a theoretical model is constructed for a specific case, and thus cannot easily be translated between fact patterns, precedential thinking about modeling choices may aid the judge in evaluating the modeler's work. Theoretical models incorporate assumptions—explicit and implicit—that appear in many different theoretical models. That these assumptions have been considered sound in a similar case is evidence of their soundness in this case.

Not only *can* decisions about modeling choices be generalized, but they *should* be generalized.¹³⁶ Reasoning by analogy among modeling choices in the law is desirable because, like reasoning by analogy about legal rules or prin-

133. Following *Joiner*, antitrust scholars in particular criticized the case's effect on precedent, arguing that an "abuse of discretion" standard meant that there would be no opportunity to build "a consistent body of criteria of reliability that should be used for particular types of recurring expert testimony." Andrew I. Gavil, *Defining Reliable Forensic Economics in the Post-Daubert/Kumho Tire Era: Case Studies from Antitrust*, 57 WASH. & LEE L. REV. 831, 851 (2000).

134. See Coate & Fischer, *supra* note 14, at 153–55 (summarizing the classical economic theories of competition and monopoly foundational to modern antitrust economics).

135. Telser's free-riding model has been used twice by the Supreme Court to change antitrust law. See *Leegin Creative Leather Prods., Inc. v. PSKS, Inc.*, 551 U.S. 877, 890–92 (2007) (overturning the longstanding per se rule against RPM in part because of its ability to counter the free-riding problem between retailers); *Continental T. V., Inc. v. GTE Sylvania Inc.*, 433 U.S. 36, 54–55 (1977) (overturning the per se rule against vertical territorial restraints for the same reason).

136. For the more general argument that decisions about scientific validity questions should be given precedential value, see Faigman, *supra* note 10.

ciples, it can improve the quality of decisionmaking.¹³⁷ Harnessing the problem-solving abilities of a large set of intelligent actors can allow for better outcomes than a single legal decision maker tackling a difficult question about a model or modeling choice.¹³⁸ For the judge, this larger body of knowledge is the common law; for the scientist, it is the body of peer-reviewed literature in his field. For both kinds of decision makers, the questions are the same: are the cases different in ways that make this particular legal principle (or modeling choice) inapplicable? What is the purpose behind the principle (or modeling choice), and how does that influence its importability to this context? Parties should be encouraged to defend their experts' modeling choices by appealing to previous cases where appropriate, and judges should include references to precedent in their decisions about modeling choices and inferences.

Indeed, the process of reasoning about issues of law is similar to the process econometricians and theorists themselves use in creating a model for a particular purpose.¹³⁹ In both cases, the decision makers draw on a body of knowledge contained in documents created by their peers. The problems and solutions contained in peer-reviewed economics or econometrics journals carry a pedigree by virtue of being selected, reviewed, and edited by other scientists in the field. Similarly, rules and resolutions of cases carry the prestige of the judges who wrote them. A good modeler, and good judge, is familiar with this literature, and draws on it for solutions to his unique problem by searching for articles or cases confronting a similar issue. Peer-reviewed articles bind the scientist because his failure to follow an accepted solution will raise criticisms if he does not have a good reason for his choice.

Statistics and economic theory, like all sciences, evolve. Imagining judicial opinions about modeling as a web of binding precedent invites worries about ossification, and the concern that the law of modeling could end up behind the science of modeling. Two features of precedential reasoning should alleviate those fears. First, as discussed in section II.A.3, models—especially statistical models—are rather fact-specific. Previous judicial assessments of models and their choices may be somewhat generalizable, but, like decisions about reasonable suspicion or discriminatory intent, their generalizability is limited. Thus, elements of prior decisions may be suggestive to future courts, but rarely will they dictate outcomes. The precedential value of decisions about modeling will be weak, and that is as it should be. Where there is a new modeling tool available, judges will rarely find themselves hemmed in by binding precedent.

137. Another benefit of according modeling choices precedential status would be to make reliance on models more explicit and thus more legitimate. Cf. John E. Lopatka & William H. Page, *Economic Authority and the Limits of Expertise in Antitrust Cases*, 90 CORNELL L. REV. 617, 696 (2005) (“[R]eliance on economic authority should be as transparent as possible.”).

138. Cf. Monahan & Walker, *supra* note 123, at 498–99 (recommending that judges give scientific evidence weight when it has “survived the critical review of the scientific community”).

139. Cf. Gilboa et al., *supra* note 27, at 17 (suggesting that economists should defend a model with appeal to similarity and relevance to the real world, which should be “debated as in a court of law”).

Second, the existence of a new statistical tool or a better theoretical model is itself grounds for overruling precedent; courts routinely abandon precedent that has become scientifically obsolete.¹⁴⁰

2. Modeling Should Be Evaluated by Judges, Not Juries

Another feature of issues of fact is the identity of the decision maker. Decisions about models and what they reveal should be made by judges because judges are repeat players, they are usually better educated than juries,¹⁴¹ and they have access to specialized training in the art of modeling.¹⁴²

Judges have the benefit of being repeat players in the game of modeling. They have the incentive because learning basic econometric techniques and learning how to parse the explicit and implicit assumptions behind models is an investment that can pay off across many cases a judge has in her career.¹⁴³ And because models are being used at an increasing rate among many different areas of law, that payoff is large and growing. For juries, education is a less efficient option because it would mean education from scratch in every trial.

Further, judges have a greater capacity and more opportunities to learn about modeling than jurors.¹⁴⁴ All judges have advanced degrees, and although traditional legal education eschewed technical and mathematical material, there is some pressure against that trend.¹⁴⁵ Even the judge with little quantitative experience may be better positioned than the average juror to learn and absorb technical concepts because she is trained in reasoning in the face of complexity and indeterminacy. As discussed above, traditional legal reasoning may bear significant resemblances to model thinking. Further, judges have access to educational programs, many of which are specifically designed to teach statistical methods.¹⁴⁶

Jurors, of course, have the benefit of numerosity, but unlike lay reasoning, it is not clear that it helps in this context. An important condition of the Condorcet

140. See generally Rebecca Haw, *Delay and Its Benefits for Judicial Rulemaking Under Scientific Uncertainty*, 55 B.C. L. REV. 331 (2014) (analyzing the Court's practice of delay in overturning scientifically obsolete doctrine).

141. See Christopher B. Mueller, *Daubert Asks the Right Questions: Now Appellate Courts Should Help Find the Right Answers*, 33 SETON HALL L. REV. 987, 993 (2003).

142. See Edward K. Cheng, *Independent Judicial Research in the Daubert Age*, 56 DUKE L.J. 1263, 1273–74 (2007) (describing educational opportunities for judges on scientific topics).

143. See Faigman, *supra* note 10, at 979.

144. Cf. Korn, *supra* note 109, at 1104 (“[S]urely the court is better suited [than the jury] by training, temperament, and habit of mind to appreciate whatever intricacies of scientific theory or its application are involved.”).

145. For example, see Judge Posner's recent statement in *Jackson v. Pollion*, 733 F.3d 786, 788 (7th Cir. 2013) (“Innumerable are the lawyers who explain that they picked law over a technical field because they have a ‘math block’—‘law students as a group, seem peculiarly averse to math and science.’ But it's increasingly concerning, because of the extraordinary rate of scientific and other technological advances that figure increasingly in litigation.” (citation omitted) (quoting DAVID L. FAIGMAN ET AL., *MODERN SCIENTIFIC EVIDENCE: STANDARDS, STATISTICS, AND RESEARCH METHODS* v (student ed. 2008))).

146. See Cheng, *supra* note 142, at 1273–74.

Jury Theorem is that each juror has something to contribute,¹⁴⁷ and in the case of reasoning about model choices, that assumption is heroic. Even if jurors can learn something about modeling in the course of the trial, achieving the condition that each of them be more often right than wrong seems unlikely, given the amount of complex information—even excluding models—that jurors are asked to absorb.

3. Models Ought to Be Scrutinized Carefully on Appeal

Another criterion that distinguishes issues of fact from law is the best way to review them on appeal. In other words, calling something a “fact” is, in part, saying that it is best reviewed deferentially and that second-guessing of “facts” leads to bad outcomes.¹⁴⁸ This suggests that part of the inquiry when deciding whether models are facts is whether bad results would come from searching review. For models and their conclusions, the answer is no; searching review is the best way to achieve accurate decisions about modeling.

The argument that “facts” deserve deferential treatment has two prongs. First, commentators justify deferential review of facts by claiming that lower courts (and juries) have a better perspective on the facts of a case because they were in the courtroom and were able to see the witnesses live and in person to better assess their credibility.¹⁴⁹ Courts sitting in review have a cold case record in front of them, and so their assessment of the evidence is less informed and should not replace that of the court below. Second, the argument goes, facts are specific to the dispute in question, and so there is no risk of inconsistency across different courts. De novo review of issues of law promotes uniform resolutions to legal issues; it ensures that like cases are treated alike.¹⁵⁰ For factual determinations, the need for coordination is lessened because no two cases are ever factually identical, and because inconsistent factual determinations cause the legal system relatively little embarrassment. The notion that one jury will find that the light was green in one case and another find that the light was red creates no appearance of illegitimacy since it is perfectly possible that the light actually was green in one case and red in the other.

147. For a detailed discussion of the various mathematical interpretations of this condition, see Edelman, *supra* note 106.

148. For a discussion of the epistemology of second opinions, see Adrian Vermeule, *Second Opinions and Institutional Design*, 97 VA. L. REV. 1435, 1459 (2011).

149. Gordon G. Young, *United States v. Klein*, *Then and Now*, 44 LOY. U. CHI. L.J. 265, 322 & n.236 (2012) (“The reasons [why reviewing courts ought to defer to trial court and jury findings of fact] include the efficiency of having facts found once, and by a decider who has seen the demeanor of witnesses, and who has seen the whole context of a piece of evidence in other evidence, both testimonial and documentary.”).

150. See Mueller, *supra* note 141, at 1021 (emphasizing the benefit of uniformity that comes with de novo review).

When it comes to evaluating models, trial courts benefit little by being in the actual presence of witnesses.¹⁵¹ Inferences about modeling choices have little to do with evaluating credibility of witnesses¹⁵² but, rather, are usually about academic arguments and debates that easily translate to a cold record, as is evidenced by the prevalence of discourse through peer-reviewed journals in the sciences. In a case, there is typically an extensive record developed—from expert reports to hearing transcripts—that will aid a court in review. Trial courts do have the advantage of being able to ask questions of the expert, which may be especially helpful in evaluating complex technical testimony. But that one advantage of trial courts does not outweigh the benefits of appellate courts engaging in searching review, especially because the second prong of the argument for why factual determinations are reviewed deferentially—that uniformity across cases is not necessary for factual issues—does not easily apply to models and modeling choices. Some modeling choices *do* transcend the individual models in which they appear because, as discussed above, many modeling problems and their solutions are recurring and thus generalizable.¹⁵³ Deferential review of these issues frustrates uniformity across cases and risks treating like cases unlike.¹⁵⁴

Searching review has the additional advantage of forcing lower courts to be explicit about their reasoning, providing more transparency and perhaps incentivizing more care in their interpretive efforts. This is widely recognized in the administrative law context as a benefit of judicial review of agency decisionmaking,¹⁵⁵ albeit in a context with some deference paid to agency decisions. The Supreme Court has also recognized searching review as a reason for reviewing most mixed questions of law and fact *de novo*.¹⁵⁶ Thus, plenary review at the appellate level has the twin benefits of affording more judicial engagement with a modeling question (four judges instead of one)—and as Condorcet suggests, more is more—as well as forcing deeper engagement in the district court. If expert witnesses should be required to give reasons for their modeling choices, then so too should judges be obliged to provide reasons for their interpretation of a model.

151. Indeed, appellate courts may be better situated, by virtue of having three judges on a panel and having the benefit of appellate briefing. *Id.*

152. See Monahan & Walker, *supra* note 123, at 497.

153. Mueller, *supra* note 141, at 1020–21 (observing that “issues relating to the validity of theories and techniques transcend the facts of individual cases” and calling for *de novo* review of *Daubert* determinations).

154. *Cf.* Faigman, *supra* note 10, at 977 (“[F]or both legislative facts and general science facts, a strong deferential standard on appeal inevitably will create inconsistencies and complications.”).

155. *Cf.* Kevin M. Stack, *The Constitutional Foundations of Chenery*, 116 *YALE L.J.* 952, 956–57 (2007) (noting the effect of searching review on the care and effort paid to administrative decisions, but questioning the efficiency of the effect).

156. See *Salve Regina Coll. v. Russell*, 499 U.S. 225, 233 (1991) (“[A]n appropriately respectful application of *de novo* review should encourage a district court to explicate with care the basis for its legal conclusions.”).

C. MODELS AND THEIR RESULTS ARE TREATED LIKE FINDINGS OF FACT

Justice Scalia's observation that what models prove is no more a question of fact than what Supreme Court opinions hold, while correct as a normative matter, turns out to be aspirational. The law ought not to treat models as facts, but the actual legal treatment of models often strays away from this ideal, sometimes as a matter of doctrine and sometimes as a matter of practice. Whenever modeling choices are given deference, evaluated by a jury, or denied precedential value, the result is less informed, less coherent decisionmaking. This section first discusses the Supreme Court's standard for admissibility of scientific evidence at the trial court level, applauding *Daubert* for recognizing that methodology, like modeling, should be scrutinized by judges. But it then shows that unfortunately the promise of *Daubert* has not been realized, at least in the context of modeling; district court judges too often treat modeling issues as questions of fact. Next, this section criticizes the Court's standard of review of *Daubert* decisions, as set forth in *Joiner*, as overly deferential. *Joiner* contributes to the judicial treatment of modeling as factual. Finally, this section discusses the treatment of models and modeling at the Supreme Court level, arguing that there, too, models are treated more like facts than is appropriate.

1. Models in the District Courts: *Daubert* Is Right in Theory, Wrong in Practice

As tools for developing scientific or social scientific conclusions, both statistical and theoretical models are squarely in the category of expert evidence.¹⁵⁷ Thus, the Federal Rules of Evidence provide some doctrinal foundation for the legal treatment of models at the trial level. Rule of Evidence 702 and the cases that interpret it govern the admissibility of testimony about models and their conclusions. The *Daubert* standard—which governs the admissibility of expert opinion under Rule of Evidence 702—asks judges to treat models and their constitutive choices unlike issues of fact by calling for searching scrutiny on the part of the judge acting as gatekeeper. But in practice, because *Daubert* offers judges little practical advice for evaluating models, courts are tempted to treat models as issues of fact by failing to engage with them substantively.

Judges can give into that temptation in part because the law of expert evidence is framed as a question of admissibility, inviting an up-or-down determination of whether a model passes some threshold of reliability. Because in other contexts—such as in the case of lay testimony—the bar for admissibility is set low enough to allow conflicting accounts, judges feel free to set the bar for modeling low enough to allow conflicting models, based on alternate assumptions and other modeling choices, to go to the jury. When they do so, judges treat modeling questions as questions of fact. But the correct reading of *Daubert*, as applied to modeling, would leave little interpretive role—as be-

157. See ABA ANTITRUST DAMAGES, *supra* note 33, at 128–29 (explaining that regression models must meet the standards established for scientific testimony).

tween models and modeling choices—for the jury, because it allocates questions of methodology to judges.¹⁵⁸ This reading comports with the thrust of this Article—that modeling issues are best treated unlike facts—but not always with court practice.

a. Daubert in Theory. Formally, Federal Rule of Evidence 702 and *Daubert* treat models not as facts but as purposive, ends-oriented tools, and they appropriately allocate the vast majority of the decisionmaking to the trial judge. *Daubert* revised the existing *Frye* standard for admissibility of expert opinion in federal cases by shifting the focus from whether an opinion had obtained “general acceptance” among the relevant scientific community¹⁵⁹ to whether the expert’s methodology was scientifically valid.¹⁶⁰ Because models easily fit in the category of “methodology,” courts generally treat modeling as a method that requires their engagement and evaluation under *Daubert*.¹⁶¹ Thus, *Daubert* and the cases interpreting its applicability to modeling correctly understand models to be a means to an end, like tools or maps, and, therefore, not themselves positive claims about the world.

Formally, the factual components of a model are the underlying data that populate and calibrate the statistical model. The legal treatment of expert testimony under *Daubert* correctly treats these as relatively unreviewable facts; “relatively” because the means of data collection can and should be examined.¹⁶² In other words, judges can and should evaluate whether the sales data that underlie a model for damages are accurate and representative, but they cannot take issue directly with the numbers themselves. Likewise, judges can—and must under *Daubert*—carefully scrutinize the means (the model) of obtaining a damages number, but ultimately cannot second-guess the number itself.

But in modeling, like many other scientific methods, the outcome is entirely dependent on the methodology.¹⁶³ David L. Faigman offers the following example: “[E]xtrapolating from animal studies to humans is sometimes warranted and sometimes not. But few, if any, scientists would contend that the conclusion that humans are affected in a certain way can be divorced from the fact that animals were studied in the underlying research.”¹⁶⁴ This dependence

158. See *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 592–93 (1993).

159. *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923).

160. See *Daubert*, 509 U.S. at 589, 592–94 (establishing scientific validity as the touchstone of the admissibility question and laying out criteria for its evaluation).

161. See ABA ECONOMETRICS, *supra* note 32, at 29–49 (discussing the application of *Daubert* to econometric models); see also *LePage’s Inc. v. 3M*, 324 F.3d 141 (3d Cir. 2003) (applying *Daubert* standard to statistical model).

162. See ABA ECONOMETRICS, *supra* note 32, at 46–49 (discussing the issue of adequate data supporting a model’s conclusion).

163. Cf. Faigman, *supra* note 10, at 975 (“In science, methods and conclusions cannot be separated.”); Saks, *supra* note 132, at 235 (suggesting that there are “scientific issues arising at low levels of abstraction, where the admissibility issue is case-specific”).

164. Faigman, *supra* note 10, at 975.

means that *Daubert's* distinction between methodology is at best unimportant and at worst untenable, at least when it comes to modeling. In the vast majority of cases, to decide an issue of methodology is to decide the conclusion as well. Thus, *Daubert* effectively treats a model's conclusions as being subject to searching scrutiny from the judge, which is to say that *Daubert* demands that a model's prediction or estimation not be treated as a fact.

But a flaw of *Daubert*—the one that leads to the factual treatment of models and their estimations and conclusions—is in the details of its advice. A methodology is sound, according to *Daubert*, if it tends to produce scientifically valid results.¹⁶⁵ This power to produce valid results, in turn, is to be measured by a set of nonexclusive, nondispositive criteria: the method's testability, its error rate, whether it has been subject to peer review, and its acceptance in the scientific community.¹⁶⁶ Thus, *Daubert* endorses the quasi-Popperian view¹⁶⁷ of science that calls "science" a set of propositions that can be, but have not been, falsified by other scientists.¹⁶⁸ Falsifiability is indeed the touchstone of scientific inquiry, but it is not the only appropriate measure of a scientific method's worth because sometimes true falsifiability is not practically achievable.¹⁶⁹ The substance of the *Daubert* standard is of limited value to judges because it best fits only one kind of scientific knowledge: that obtained from a controlled experiment.¹⁷⁰

To soften the implications of this limitation, the *Daubert* Court made clear that its list was nonexclusive and nondispositive. The Court explained that circumstances may dictate the use of other criteria and that a methodology's failure to meet one criterion does not preclude its admission as evidence. Rather, the enumerated factors were designed merely as guidance for the

165. *Daubert*, 509 U.S. at 592–93.

166. *Id.* at 592–94.

167. The view endorsed by the Court is best described as "quasi-Popperian" because Justice Blackmun's philosophy of science, while citing Popper's, actually conflicts with it in some important ways and also incorporates the ideas of Carl Hempel. See Susan Haack, *Federal Philosophy of Science: A Deconstruction—and a Reconstruction*, 5 N.Y.U. J.L. & LIBERTY 394, 397–98 (2010). Professor Haack points out that Justice Blackmun's "quasi-Popperian" philosophy actually revised it away from its radical position, and that the Court's misinterpretation actually placed the "first, quasi-Popperian *Daubert* factor . . . closer to the truth than the Popperian philosophy of science from which it ostensibly derives." *Id.* at 399. The point here is that "Popper," in *Daubert* and in this Article, stands for the notion that falsifiability is the essence of the scientific method.

168. See *Daubert*, 509 U.S. at 593.

169. This is especially true in economics. See Andrew I. Gavil, *After Daubert: Discerning the Increasingly Fine Line Between the Admissibility and Sufficiency of Expert Testimony in Antitrust Litigation*, 65 ANTITRUST L.J. 663, 674 (1997) ("To falsify any given economic technique or methodology, it would be necessary to create the converse conditions in an actual functioning 'market' in order to test the adverse proposition This level of control is virtually impossible to achieve").

170. See *id.* ("[F]ew economic techniques of the ilk utilized in antitrust litigation could be 'tested' in the sense contemplated by *Daubert*, i.e., falsified. In part, the problem flows from the way in which economic knowledge is acquired. Rarely is economic technique amenable to laboratory-type experimentation under controlled conditions." (footnote omitted)).

ultimate question that courts must engage substantively: is the methodology scientifically valid?

Daubert's choice of criteria, despite the Court's qualifications that they are neither dispositive nor exclusive, has strongly influenced lower courts' perception of what the Court is prepared to recognize as admissible evidence. For example, in the early days of *Daubert*, there was significant controversy over whether the case even applied to nonexperimental scientific evidence. This perception was supported by the idea that because *Daubert's* factors fit so poorly with observational science¹⁷¹—which included perhaps the bulk of all scientific evidence offered at the trial level—it must not have been intended to apply to *all* evidence offered under Rule 702, despite the case's apparently broad holding. This argument was particularly popular among antitrust scholars, who observed the ill fit between questions about testability and error rate and statistical and theoretical economic evidence in antitrust trials.¹⁷²

But subsequent cases clarified that the Court intended *Daubert* to apply not only to observational scientific evidence but also to expert evidence that could not qualify as “science” at all.¹⁷³ This put judges in an awkward position. When it came to nonexperimental expert evidence, *Daubert's* holding—that judges must act as fierce gatekeepers of the jury and searchingly evaluate the scientific validity of proposed expert testimony—applied. But *Daubert's* guidance on how to perform this task—especially its emphasis on error rate and testability—was essentially useless. And the Court did not subsequently provide more or better criteria for evaluating expert testimony outside of the experimental context.

Scientific models exist outside the realm of falsifiability because a model is a tool, and so cannot be right nor wrong, only useful or not useful. Models aim to make sense of data or to abstract away from chaotic reality to distill simple patterns and implications. To this end, they are used in the very process of falsification that *Daubert* holds out as the gold standard of admissibility, and so they constitute “science” even if they fail the quasi-Popperian notion of the scientific method. And they are as dangerous as they are helpful since they boast a power to create sense from chaos. This dangerousness puts judges in an uncomfortable position when they are asked to evaluate models rigorously without the intellectual or doctrinal tools for the task.

171. Observational science, where data is taken as it is from the real world, presents different challenges from experimental science, where data can be gathered under artificial and controlled circumstances. In particular, observational science confronts difficulties in identification. See Baker & Bresnahan, *supra* note 25, at 5.

172. See Gavil, *supra* note 169, at 679–80.

173. See *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 141 (1999) (“[T]he test of reliability is ‘flexible,’ and *Daubert's* list of specific factors neither necessarily nor exclusively applies to all experts or in every case.”). For example, connoisseur testimony is subject to *Daubert* review. See David E. Bernstein, *Expert Witnesses, Adversarial Bias, and the (Partial) Failure of the Daubert Revolution*, 93 IOWA L. REV. 451, 480–88 (2008).

b. Daubert in Practice. How have courts responded to this lack of guidance from the Court? Too often, courts abdicate their *Daubert* duty to searchingly review models. This results in two kinds of errors. Sometimes judges prematurely reject models and their conclusions without a sound, reasoned evaluation. At other times, judges hand over the decisionmaking authority to a jury. When judges reject or accept models without the kind of thorough reasoning they apply to issues of law, they erroneously treat the issue as if it were a question of fact.

Perhaps the most glaring example of this is when a judge evaluates the methodology of damages models at only the surface level. Here, courts identify “multiple regression analysis” as the methodology used by the expert and approve of any model that uses that method.¹⁷⁴ Of course, multiple regression analysis is widely accepted among scientists and frequently employed in peer-reviewed publications, so it easily passes two of *Daubert*’s criteria.¹⁷⁵ As for the other two, testability and error rate, multiple regression analysis of course comes up short.¹⁷⁶ But these factors do not apply easily to modeling in any case, and because the Court explained that some *Daubert* criteria will be irrelevant to some scientific methodologies, courts easily find it within their discretion to allow the model as having passed the two *Daubert* criteria relevant to it. Courts make a similar move when they find a model to pass *Daubert* muster if it fits within broad categories of models found to be acceptable in measuring antitrust damages, such as the before-during-after technique or the “benchmark” method.¹⁷⁷ With plenty of precedential and scholarly support for these techniques, courts have an easy time admitting them as reliable even without a substantive engagement with the details of the particular model in question.

The flaw in this line of reasoning is its level of generality. The fact intensity of modeling requires a more particularized engagement with each step of regression analysis in a specific context. As one observer has pointed out,

174. See, e.g., *Conwood Co. v. U.S. Tobacco Co.*, 290 F.3d 768, 792–94 (6th Cir. 2002); *In re Ready-Mixed Concrete Antitrust Litig.*, 261 F.R.D. 154, 165 (S.D. Ind. 2009).

175. For a discussion of the general acceptance and peer review prongs as applied to econometric models, see ABA *ECONOMETRICS*, *supra* note 32, at 34–39.

176. Many economic models would pass a broader definition of testability. If “testable” were defined to include a model that could describe a known reality before making estimations and predictions of an unobservable event, then many models—including even some game-theoretical models—could be considered “testable.” For a defense of such an interpretation of *Daubert*, see Coate & Fischer, *supra* note 14, at 149–51.

177. Similarly, courts sometimes use testimony by the same expert in previous cases where they presented a “similar” model to determine the admissibility in the case at hand. See, e.g., *In re Chocolate Confectionary Antitrust Litig.*, 289 F.R.D. 200, 212–13 (M.D. Pa. 2012) (rejecting the argument that the expert’s exclusion on *Daubert* grounds in previous cases was fatal to his model in this case). Where these models are “similar” only in the sense that they both use regression analysis or broadly approached a damages calculation by estimating pricing in a world “but for” the defendant’s conduct, this maneuver repeats the same mistake as courts that approve of “multiple regression analysis” generally under a *Daubert* standard.

saying that “multiple regression analysis” is an accepted method and, *therefore*, a multiple regression analysis model is admissible is tantamount to saying that mathematics is an accepted scientific method and, *therefore*, any expert testimony employing math is admissible.¹⁷⁸ Rather, the actual usefulness of the method and, therefore, the reliability of its predictions ought to be carefully scrutinized as part of its methodology under *Daubert*. Any *Daubert* decision about a regression model that fails to examine an expert’s many modeling choices and assumptions essentially black-boxes the issue, a treatment typically reserved for factual reasoning.

Another practice that judges engage in that results in the factual treatment of models is to suggest that modeling flaws go to the weight of the expert testimony, not its admissibility.¹⁷⁹ Here, judges are aided by the Supreme Court, which held in a pre-*Daubert* case that regression analyses failing to account for all causal factors are not necessarily inadmissible; rather, weighing the incompleteness of a model’s account of causation is analogous to weighing the credibility of a lay witness or the usefulness of his testimony, and so is properly left to juries.¹⁸⁰ In this way, judges can punt difficult modeling questions to the jury, in effect treating them as questions of fact, not law.

Another technique that judges use to avoid substantive engagement with modeling choices and assumptions is to focus, sometimes inappropriately, on the factual inputs of the model.¹⁸¹ This maneuver can take at least two forms. First, judges may reject theoretical models altogether as being insufficiently grounded in facts. But where empirical data are unavailable or unreliable, and a theoretical model’s assumptions are well-founded, a theoretical model can be a superior method of estimation or prediction. Second, judges may attack a model for using imperfect data, relying on the statistical adage “garbage in, garbage out.”¹⁸² But if the only available data are flawed, but not fatally, then courts may be letting the perfect be the enemy of the good. Both techniques allow judges to reject models without a careful investigation into their constituent parts and the possibility that, even without perfect data, a reasonably reliable estimation or prediction is possible.

The simplest technique judges can use to avoid substantive engagement with models is to admit or exclude the opinion without explaining their reasoning. In these cases, judges may emphasize the time and effort they put in to understanding the model before announcing, in conclusory fashion, its status under *Daubert*.¹⁸³ Here, a judge may point out room for disagreement about a model, without sketching the terms of that disagreement or its stakes, before summarily

178. See Kaye, *supra* note 15, at 2010.

179. E.g., *In re Live Concert Antitrust Litig.*, 863 F. Supp. 2d 966, 973 (C.D. Cal. 2012).

180. *Bazemore v. Friday*, 478 U.S. 385, 400 & n.10 (1986); see Lopatka & Page, *supra* note 137, at 692 (discussing the burden shifting effected by *Bazemore*).

181. See, e.g., *In re Se. Milk Antitrust Litig.*, 739 F.3d 262, 280–81 (6th Cir. 2014).

182. See, e.g., *In re Scrap Metal Antitrust Litig.*, 527 F.3d 517, 529 (6th Cir. 2008).

183. See, e.g., *id.* at 527.

admitting the evidence as reliable. Relatedly, judges sometimes take experts at their word when they claim to have solved modeling problems raised by the opposition.¹⁸⁴ Without actually unpacking the working parts of the model for himself, a judge relying on the *ipse dixit* of an expert abdicates his responsibility to ensure the appropriateness of modeling choices or assumptions.¹⁸⁵

This is not to say that all trial courts abdicate their responsibility in this way; many engage in thoughtful, detailed analysis of modeling choices.¹⁸⁶ To name but one example of many, the Northern District of California rejected an economist's damages model in a price-fixing case because he inappropriately used average prices across product groups in demonstrating commonality of damages among consumers in a class action lawsuit.¹⁸⁷ Crucially, the court did not merely announce that average prices were inappropriate; rather, it cited the reasons a modeler may prefer to use averages, explained why those reasons did not apply in the instant context, and spelled out the distortion resulting from using averages in this context.¹⁸⁸ Another exemplary treatment of a model appears in *Freeland v. AT&T Corp.*, in which the court rejected a damages model on *Daubert* grounds because it accounted for too few factors in predicting a "but for" price.¹⁸⁹ These cases prove that lay judges have the capability and resources to learn enough about the art and science of modeling to apply the kind of searching review that *Daubert* demands and that is optimal for legal decisionmaking. But the point is that many judges do not engage meaningfully; yet those opinions, for reasons discussed in the next section, usually pass muster on review.

2. Models in the Appellate Courts: *Joiner* Precludes Meaningful Model Review

Although *Daubert* aspires to detailed and substantive evaluations of models in district courts, and so would have judges treat modeling closer to law than fact on the fact/law spectrum, the standard of review established in the Supreme Court's *General Electric Co. v. Joiner*¹⁹⁰ opinion suggests that appellate courts treat modeling choices with deference similar to that accorded findings of fact.

184. See *In re Chocolate Confectionary Antitrust Litig.*, 289 F.R.D. 200, 221 (M.D. Pa. 2012) (accepting the expert's claim that he actually conducted the "but for" analysis his opposing expert claimed he did not perform).

185. See Emily Hammond Meazell, *Super Deference, the Science Obsession, and Judicial Review as Translation of Agency Science*, 109 MICH. L. REV. 733, 749–50 (2011). And the failure to write a detailed and thoughtful assessment of a model creates another cost, as well, because part of the value added by lay judges presiding over cases involving technical or scientific arguments is their ability to translate the science for a lay audience. This process of reducing the technical or scientific reasoning that goes into a legal decision allows for more judicial transparency and thus legitimacy, educates the public about the nature of the legal dispute, and allows the case and its reasoning to have persuasive and precedential value. See *id.*

186. See, e.g., *In re Live Concert Antitrust Litig.*, 863 F. Supp. 2d 966 (C.D. Cal. 2012).

187. *In re Graphics Processing Units Antitrust Litig.*, 253 F.R.D. 478, 493–94 (N.D. Cal. 2008).

188. *Id.* at 494–95.

189. 238 F.R.D. 130, 148–49 (S.D.N.Y. 2006).

190. 522 U.S. 136 (1997).

Joiner instructs appellate courts to affirm a lower court's *Daubert* decision if the lower court did not abuse its discretion in admitting or excluding the expert opinion.¹⁹¹ While technically this standard is different from that applied to factual determinations of courts below (clearly erroneous), both standards are highly deferential.¹⁹² In practice, the "abuse of discretion" standard means that appellate courts rarely overturn lower courts' *Daubert* decisions.

The root of *Joiner*'s error is again an obsession with admissibility. By framing questions about modeling as raising only questions about threshold reliability, admissibility doctrine casts the judge as courtroom manager, as a figurative "gatekeeper" to the decisionmaking process (whether it ultimately involves himself or the jury as a factfinder) over which he presides. This view emphasizes the supervisory role of the judge, one more akin to courtroom manager than primary decision maker, and one associated with significant discretion. But this view ignores the fact that trial judges evaluating modeling choices are—or at least ought to be—engaging in substantive reasoning about complex and recurring modeling problems, for which they need both oversight and guidance to do so effectively. The "abuse of discretion" standard mistakes this role of decider for referee.

The *Joiner* standard, which encourages factual treatment of scientific issues such as modeling, is difficult to square with *Daubert*, which encourages searching judicial engagement with methodologies like modeling. Early commentary on the case suggested that scientific methodology framed at a high level of generality—for example, fingerprint identification—could be reviewed *de novo* without offending *Joiner* because the case itself "did not subject any findings about the basic principles of epidemiology and toxicology to an abuse-of-discretion standard."¹⁹³ But this reading of *Joiner* ignores the case's perhaps most important holding, that "conclusions and methodology are not entirely distinct from one another."¹⁹⁴ In so observing, *Joiner* abandoned the untenable methodology–conclusion distinction in *Daubert*,¹⁹⁵ and so effectively painted all scientific testimony—from methodology to application—with the same "abuse of discretion" brush. The *Joiner* Court was quite right to recognize the interdependence of methodology and conclusion, and that, in the words of evidence scholar David L. Faigman, "[e]valuation of methods and conclusions simply

191. *Id.* at 142-43; *cf.* *El Aguila Food Prods., Inc. v. Gruma Corp.*, 131 F. App'x 450, 453 (5th Cir. 2005) ("A district court has broad discretion in deciding to admit or exclude expert testimony . . .").

192. *But see* Peter Nicolas, *De Novo Review in Deferential Robes?: A Deconstruction of the Standard of Review of Evidentiary Errors in the Federal System*, 54 SYRACUSE L. REV. 531 (2004) (arguing that the "abuse of discretion" standard as applied to evidentiary rulings is actually quite searching).

193. Saks, *supra* note 132, at 234-35. Another commentator, writing just before *Joiner* was decided, expressed the hope that the opinion would hold that "[w]hen the scientific evidence transcends the particular case, the appellate court should apply a 'hard-look' or *de novo* review to the basis for the expert opinion." Faigman, *supra* note 10, at 976.

194. *Joiner*, 522 U.S. at 146.

195. *See* Saks, *supra* note 132, at 235.

cannot be divided among different decision makers.”¹⁹⁶ But the error of *Joiner* is to give fact-like deference to that decision maker’s judgment.

The abuse of discretion standard lowers the interpretive threshold required for a legal decision concerning a modeling choice or conclusion. It places a thumb on the scale in favor of the appellee’s arguments about modeling. This default toward one side tends to limit trial judges’ responsibility to settle close and difficult debates about the art or science behind a model. Thus, at the appellate level, a cursory review of the model and its choices typically suffices because the standard—which clearly gives lower courts discretion in making determinations under *Daubert*—is so deferential. Some courts have even explicitly identified modeling choices as a question of credibility, thus placing them squarely in the “fact” category. As one reviewing court explained while rejecting the defendant’s claim that a damages model was fatally flawed for biased factual inputs, “[t]he credibility of [plaintiff’s] and [defendant’s] experts was for the jury to determine.”¹⁹⁷

Again, this is not to say that no appellate courts engage in the kind of searching review of models and modeling choices that one associates with questions of law.¹⁹⁸ Many appellate courts immerse themselves in the science of the expert opinion and correct significant errors in the opinion below. In a recent case, for example, the D.C. Circuit took Justice Scalia’s comment in *Comcast* to heart and conducted a detailed analysis of the underlying model, ultimately rejecting it because it failed to describe the known world and so was likely inaccurate as to the “but for” world.¹⁹⁹ And when, as in that case, the court below admitted a misleading model, it is easy to characterize the court as having “abused” its discretion.²⁰⁰ Showing a similar willingness to question modeling choices, the Sixth Circuit in *In re Southeastern Milk Antitrust Litigation* reversed a lower court’s decision rejecting a market definition model as ungrounded in fact.²⁰¹ The Sixth Circuit appropriately observed that the requisite level of factual support for a model varied by context and, in any case, never rose to the factual specificity the court below seemed to demand.²⁰² But these exemplary cases do not negate the point that the *Joiner* standard provides an opportunity for those judges who do not wish to or cannot meaningfully evaluate a model to avoid the task altogether.

196. Faigman, *supra* note 10, at 975.

197. *LePage’s Inc. v. 3M*, 324 F.3d 141, 165 (3d Cir. 2003).

198. And where the question on appeal is not admissibility, but rather the sufficiency of the expert evidence to support summary judgment, the standard is not deferential at all. *See, e.g., In re High Fructose Corn Syrup Antitrust Litig.*, 295 F.3d 651, 660 (7th Cir. 2002).

199. *In re Rail Freight Fuel Surcharge Antitrust Litig.*—MDL No. 1869, 725 F.3d 244, 253–55 (D.C. Cir. 2013).

200. *See, e.g., In re Se. Milk Antitrust Litig.*, 739 F.3d 262, 279–83 (6th Cir. 2014). *See generally* Nicolas, *supra* note 192 (explaining that the “abuse of discretion” standard in *Joiner* actually incorporates several kinds of legal mistakes that would be reviewed de novo).

201. 739 F.3d at 283.

202. *Id.* at 281–82.

3. Models in the Supreme Court: Deference to Consensus

Modeling looks very different in the Supreme Court than in the courts below because the high Court views its role as creating rules rather than resolving disputes.²⁰³ This means that the Supreme Court is far less likely to substantively engage with a statistical model proving damages than with a theoretical model predicting the legal effect of a proposed rule.²⁰⁴ It means that the Court more often explicitly considers theoretical models than do the courts below, because theory is frequently used to justify making or changing a rule. “Explicitly” is an important modifier, however, for two reasons. First, as discussed above, all statistical models rely on theory in their creation and calibration, so any judicial decision about a statistical model at least implicitly passes judgment on the theory that undergirds it. Second, lower courts have the incentive to disguise appeal to economic theory as common sense, intuition, or just plain fact because theory fits uncomfortably with the *Daubert* standard and may signal rulemaking, a task unfitting to a district court.²⁰⁵

The broader scope of models used by the Supreme Court gives them an option that the lower courts do not have: appeal to academic consensus. Deference to consensus treats modeling questions more like questions of fact than modeling issues demand. The Court may care as much about getting the theory right and using only reliable empirical claims in their rulemaking as they do about crafting the best legal rule, but for the former kind of analysis, the Court is content to count noses. Here I do not mean to suggest that deference to consensus is an irrational or even undesirable way to decide cases; my point is that when it is the only tool available, it can significantly delay legal response to new scientific developments. A better decisionmaking process would combine attention to consensus with a more sophisticated engagement with the models underlying scientific opinion. But it is not what we observe: this pattern of deference without substantive engagement can be seen in the special attention paid to amicus briefs and the use of delay in the face of scientific change.

203. Another reason why model evaluation looks so different at the Supreme Court level is because there is not a clear set of rules governing how and under what circumstances Justices can consider expert evidence.

204. To be sure, the Court sometimes does evaluate damages models; *Comcast* is an example. *Comcast Corp. v. Behrend*, 133 S. Ct. 1426, 1433–35 (2013).

205. Often, propositions that are thought of as constituting common sense or even ideology are themselves models. Cf. Lopatka & Page, *supra* note 137, at 633–39 (discussing the sometimes implicit adoption of basic economic models as explaining human behavior and justifying legal rules). For example, the notion, held by many conservatives, that the free market does better than government can in allocating wealth is derived from two theoretical models: one of market behavior that holds that goods will move to their highest and best use through a system of free exchange, and one of government behavior that says that because of agency costs, political incentives can never be perfectly aligned with welfare. These models, like all theory, derive power from their simplicity and intuitiveness, but depend critically on the soundness of their assumptions. Lower courts use these models all the time. Cf. *id.*

a. Amicus Briefs. As I have observed elsewhere, the Supreme Court, when asked to make or change an antitrust rule, relies on amicus briefs to supply the economic evidence for and against a legal rule.²⁰⁶ In part, amicus briefs are necessary because the economic evidence for the rule change cannot be found in the record, as parties are understandably shy about asking for legal change at the trial and appellate levels. But they are also necessary because the Supreme Court is starved of economic arguments relative to a trial court who can engage the parties' experts in a conversation and even appoint some of its own.²⁰⁷

Participation as amici gives economists an opportunity to comment—or effectively to vote—on the appropriateness of adapting antitrust law to an economic theory. *Leegin Creative Leather Products, Inc. v. PSKS, Inc.*,²⁰⁸ the final chapter in the resale price maintenance (RPM) saga,²⁰⁹ illustrates the point. When that case was brought, the per se rule against RPM had been on the books for almost a century but was under fire from antitrust scholars persuaded by Telser's free-rider model explaining that RPM could be good for consumers.²¹⁰ The defendant urged the Court to reverse the per se rule, arguing that Telser's model showed the potential for procompetitive uses of RPM,²¹¹ while the plaintiff argued that RPM was more often used in an anticompetitive way. Thus, the central disagreement among the parties in *Leegin* was over the usefulness of Telser's model as a map of distribution behavior among manufacturers and retailers. That debate was largely resolved not by the Court but by economists acting as amici. Fifty academic economists and law professors signed amicus briefs endorsing the model's usefulness,²¹² and the Court largely justified its decision to eliminate per se liability for RPM by invoking this academic consensus on the issue rather than grappling with the model.²¹³

206. See Rebecca Haw, *Amicus Briefs and the Sherman Act: Why Antitrust Needs a New Deal*, 89 TEX. L. REV. 1247 (2011).

207. This has lead Justice Breyer to remark that amicus briefs containing technical information “play an important role in educating judges on potentially relevant technical matters, helping to make us not experts but educated lay persons and thereby helping to improve the quality of our decisions.” *Justice Breyer Calls for Experts to Aid Courts in Complex Cases*, N.Y. TIMES, Feb. 17, 1998, <http://www.nytimes.com/1998/02/17/us/justice-breyer-calls-for-experts-to-aid-courts-in-complex-cases.html> (internal quotation marks omitted).

208. 551 U.S. 877 (2007).

209. Resale price maintenance is the practice of contractually imposing a price floor on the resale of products and was the inspiration for Lester Telser's free-riding model, which sought to provide an economic explanation for the practice. See *supra* section I.B.2.

210. Firms were able to use a workaround created by *United States v. Colgate & Co.*, 250 U.S. 300 (1919). The exploitation of this workaround prompted Lester Telser to develop a theory for how RPM and similar practices may actually be economically efficient. See *supra* section I.B.2.

211. Petition for a Writ of Certiorari at 5–7, *Leegin*, 551 U.S. 877 (No. 06-480), 2006 WL 2849384.

212. Motion for Leave to File Brief and Brief of Amici Curiae Economists in Support of Petitioner, *Leegin*, 551 U.S. 877 (No. 06-480), 2006 WL 3244034; Brief of Amici Curiae Economists in Support of Petitioner, *Leegin*, 551 U.S. 877 (No. 06-480), 2007 WL 173681.

213. Haw, *supra* note 206, at 1281–84 (describing the heavy influence of the amici on the Court's opinion).

This practice—of essentially asking experts to vote on the usefulness of a theoretical model—is not altogether irrational or illegitimate; indeed, where a decisionmaking body lacks expertise, polling experts and using majority rule to accept or reject a proposition can be a good way to estimate the truth.²¹⁴ And indeed, it may be that Telser's theory is widely accepted among economists because its assumptions are sound, at least for many kinds of markets. But the point is that this kind of reasoning—counting noses—is more commonly found in reasoning about facts, not law. It is for factual questions that we rely on Condorcet. When it comes to questions about modeling—which is so unlike facts ontologically and pragmatically—the law should demand a substantive assessment from its decision makers. This is all the more true when the judicial decision goes to broad policy or regulatory questions—as the Supreme Court typically encounters—rather than to resolving an individual dispute.

b. Delay. Another technique that the Court uses in evaluating theoretical models is to delay rulemaking until a model offered to justify a legal change has achieved maturity.²¹⁵ Like resort to amicus briefs, delay can help provide second-order signals that a model is useful and its assumptions are valid. When the Court refuses to grant certiorari on the basis of a new theoretical model, it gives the Court an opportunity to wait for consensus among the economic academy to develop (or not) around the model.²¹⁶ If consensus around the model already exists at the time it is offered to the Court, further delay can test the robustness of that consensus. And where the Court cannot discern if there exists a consensus about the model or not, delay allows the Court to observe the staying power of the model over time, which may itself be a signal of its quality and usefulness.²¹⁷

This pattern can also be observed in the *Leegin* case. *Leegin* itself was the last in a long series of opportunities presented to the Court to use Telser's model to undo the per se rule against RPM. The first opportunity came in the 1970s, when the body of theoretical literature spinning out the economic and policy implications of Telser's model was relatively young.²¹⁸ As this literature and the models derivative of Telser's aged, and consensus around their usefulness solidified, the Court continued to turn down opportunities to use this theoretical research against the per se rule against RPM.²¹⁹ It was not until 2007 that the Court used the model, at the time of *Leegin* almost fifty years old, to change one of their most infamous rules.

214. See Vermeule, *supra* note 106.

215. Haw, *supra* note 140, at 332.

216. *Id.* at 352.

217. *Id.*

218. See *Continental T. V., Inc. v. GTE Sylvania Inc.*, 433 U.S. 36, 58 (1977) (challenging the per se rule against vertical territorial restraints, which are closely related to RPM).

219. See Haw, *supra* note 140, at 338–41 (discussing the Court's missed opportunities to overturn *Dr. Miles Medical Co. v. D. Park & Sons Co.* in *Business Electronics Corp. v. Sharp Electronics Corp.*, 485 U.S. 717 (1988), and *Monsanto Co. v. Spray-Rite Service Corp.*, 465 U.S. 752 (1984)).

Better information about the quality of the model, although not the only reason for the delay, is perhaps part of why the Court took as long as it did to adopt the model. Effectively, the device is similar to relying on economists speaking through amicus briefs; both are aimed at getting to the consensus position on a model rather than evaluating it directly. Nor is it irrational; for an inexpert court, consensus (and the staying power of that consensus) is a reasonable way to evaluate a theoretical model.²²⁰ The point here is that the Court uses only indirect means to evaluate theoretical models when substantive engagement, at least to supplement the use of proxies, should be employed. The Court's deference to consensus and its resort to nose counting more closely resemble the manner in which courts assess questions of fact, where judges observe and review the process that someone else used to establish the conclusion, rather than actually engage in any first-order reasoning themselves.

D. THE CAUSE AND EFFECT OF TREATING MODELING ISSUES AS FACTS

1. The Cause: Judicial Inexpertise

The previous discussion pointed toward two conclusions: modeling issues ought not to be regarded by the law as factual, but judges at all levels of the judiciary often treat them as such. In that sense, Justice Scalia's comment in *Comcast* that what models prove is not a finding of fact²²¹ is right as a normative matter but not always as a positive matter. Perhaps the reason for that discrepancy is exemplified in his phrasing. Justice Scalia makes an error when he says that models and their conclusions are "no more a question of fact than what our opinions hold."²²² Models should not be treated as facts, but they also should not be treated as raising pure questions of law. They are unlike "what . . . opinions hold,"²²³ in part because case holdings and legal rules are squarely within the expertise of judges. Legal reasoning skill is the aim of legal education and the basis of a judge's appointment to the judiciary. In contrast, judges are not necessarily trained in model thinking, nor are they traditionally chosen for their quantitative abilities.

This mismatch between ability and task leads judges to take an easier route: to treat modeling issues as facts. These judicial moves—from trial courts abdicating the gatekeeper role to the Supreme Court deferring to expert amici—circumvent a judge's engagement with the art and science of modeling. This should not be interpreted as math fear or laziness, but rather as a justified anxiety about a lay judge's fitness for the task. This anxiety is heightened by the

220. Haw, *supra* note 140, at 352.

221. *Comcast Corp. v. Behrend*, 133 S. Ct. 1426, 1433 n.5 (2013).

222. *Id.*

223. *Id.* Justice Scalia's error is a product of the notion, prevalent in legal doctrine, that there is a fact/law dichotomy—that whatever is not an issue of fact must be an issue of law. But as the doctrine of mixed questions illustrates, the fact/law distinction is a spectrum, not a dichotomy. See *supra* section II.A.3.

fact that modeling involves not only “science” in the sense of falsifiable choices—which judges struggle to evaluate anyway—but also “art,” which by definition defies easy, objective assessment. Thus, judges, whether deliberately or not, are sometimes tempted to “black-box” models and modeling issues.

2. The Effect: Inaccurate Judgments and an Illegitimate Delegation of Authority

The costs of “black-boxing” modeling issues can be divided into two categories: accuracy costs and legitimacy costs. Treating models and their predictions akin to relatively unreviewable facts yields less accurate judicial decisions because it means that modeling choices do not always get sufficient scrutiny at the trial level before being either admitted or discarded as unhelpful to the controversy. Similarly, the factual treatment of modeling denies models a second set of eyes at the appellate stages, which otherwise would help ensure the accuracy of their claims.

Searching review is not always a boon to accuracy.²²⁴ For instance, the notion that appellate courts should not second-guess factual determinations is buttressed by considerations of accuracy. There, the belief is that reviewing courts, without the benefit of being present at trial, are less able to assess the demeanor and credibility of the witnesses, meaning that they actually have *less* information than the primary decision maker (trial judge or jury) about the disputed facts in a case.²²⁵ But in the case of modeling, the demeanor of the expert offers little to the judge in evaluating a model; thus, the gap may not be wide enough to justify giving up on the benefits of searching review.

Accuracy is further hindered by treating modeling issues as factual because it denies their resolutions any precedential value. Accuracy would be boosted if a court could call on a body of decisions—without being tightly bound by them—discussing the appropriateness of individual modeling choices and assumptions under certain circumstances, and even the usefulness of whole models for certain questions.²²⁶ Thus, treating modeling choices as within the ken of judges and subject to precedential reasoning would allow courts to harness the informational benefits of the common law in reasoning about theoretical models.²²⁷

224. Cf. Vermeule, *supra* note 148, at 1458–61 (discussing the costs of second opinions).

225. See Young, *supra* note 149, at 322 & n.236.

226. Cf. Monahan & Walker, *supra* note 123, at 514–16; Saks, *supra* note 132, at 233–34.

227. Additional accuracy costs come from the Supreme Court’s reliance on economic consensus because both delay and (perhaps less obviously) deference to expert amici slow down the process of adjusting law in response to potentially useful and appropriate models of market behavior. For many commentators, the *per se* rule against RPM stood for too long after good models had been constructed to illustrate its foolishness. See Haw, *supra* note 140, at 337–38 & n.61. Delay leads to ossification. Ossification means that for a significant period of time, the law will be wrong before it can finally be made right. During that period of time, the economy and individual market actors labor under an inefficient rule that costs competitors and consumers alike. To the extent that amicus briefs are a way to learn of consensus positions, the Court’s reliance on them in changing an antitrust rule can also have an

The second category of costs that attend treating models like questions of fact involves legitimacy. If models go relatively unreviewed by judges and Justices, and if juries are not equipped to meaningfully evaluate modeling choices and assumptions, then a substantial portion of antitrust law and enforcement authority is left to expert economists—neither elected nor appointed, but rather anointed by virtue of their prestige and skill as expert witnesses. And because antitrust plaintiffs receive treble damages, the power wielded by a convincing expert witness can be immense. This delegation of power from an Article III judge to an expert should not be taken lightly.

Similarly, nose counting and consensus taking at the Supreme Court effectively delegates authority over antitrust policy to economists rather than leaving it with the judiciary as Congress intended. When the Court declines to examine the empirical and theoretical models on which it builds doctrine, it abdicates an important part of its rulemaking responsibility. To be sure, it leaves itself space to shape law in response to these models; indeed, the debate between the *Leegin* majority and dissent was largely about the law's best response *given* the model's inevitability.²²⁸ But in so doing the Court significantly limits its power to make rules because so much law and policy in the antitrust realm is driven by the models themselves. Thus, the failure to treat models substantively—subject to the kind of thorough analysis the Court holds itself to when deciding what constitutes discrimination or whether Congress has authority under the Constitution to regulate healthcare—redistributes power from courts to economists in a potentially illegitimate way.

III. IMPLICATIONS FOR INSTITUTIONAL DESIGN

“Models are the new lingua franca.”

—Scott E. Page²²⁹

The courts' treatment of models as facts results from assigning decisionmaking requiring expertise to an inexpert set of decision makers. Thus, while treating modeling choices like facts is understandable given judges' perceived inability to engage in the art of modeling, a broader perspective on what models are, how they ideally would be evaluated, and the actual scrutiny—or lack thereof—that results from their treatment as facts would suggest that courts are

ossifying effect; it takes time for a consensus to form and for it to be articulated in such a succinct form as an amicus brief.

228. *Leegin Creative Leather Prods., Inc. v. PSKS, Inc.*, 551 U.S. 877, 914 (2007) (Breyer, J., dissenting) (“The upshot is, as many economists suggest, sometimes resale price maintenance can prove harmful; sometimes it can bring benefits. . . . I would ask such questions as, how often are harms or benefits likely to occur? How easy is it to separate the beneficial sheep from the antitrust goats?” (citation omitted)).

229. Scott E. Page, *Intelligent Citizens of the World*, Lecture 1.2 of *Model Thinking*, COURSERA, <https://class.coursera.org/modelthinking-006/lecture/19> (last visited Nov. 17, 2014).

currently choosing the worse of two evils. This recognition suggests two possible moves, by now familiar in the institutional design literature. First, we could allocate decisionmaking that involves modeling to an expert agency, and avoid both evils altogether.²³⁰ Second, we could make what I would call the lesser evil—judicial engagement with models as issues of law—even less evil by better educating judges about the art of modeling.

In this Part, I briefly address the benefits and costs of two options: allocating model evaluation to an expert agency and providing judicial education on the art of modeling. I do not claim to offer concrete proposals for either option; rather, the aim is to sketch the trade-offs, in the context of modeling, between these two broad categories of reform. Ultimately, I favor strengthening judicial competence in model thinking and quantitative reasoning because it is impractical to allocate all modeling questions to an agency.

A. REALLOCATION OF MODELING QUESTIONS TO AN AGENCY

One solution to the mismatch between what a decision demands and its maker's skill set is to reassign the decision to a different, more expert, decision maker.²³¹ The benefits in terms of expertise are obvious: agencies are staffed by experts and are able to actively seek out answers to scientific questions by commissioning panels of experts and even data-gathering studies.²³² In the case of modeling, agencies could better evaluate the statistical choices behind an econometric model and the assumptions behind a theoretical model than could a lay judge and certainly a jury. Agencies could even be relied upon to create their own models and subject them to the kind of rigorous peer review that we associate with modeling in scientific journals.²³³

But in the case of modeling, the costs of such a move are high. Models and modeling are ubiquitous in judicial decisionmaking, from damage calculation to rule calibration. Moreover, because modeling issues are inextricable from the cases in which they arise, there is no workable way to sever the model from the case and give it to an agency to consider. Thus, giving up on meaningful review of models by courts would effectively mean giving up on all court involvement in all litigation that is relevant to models, which is a large portion of litigation.²³⁴

230. See Daniel A. Crane, *Technocracy and Antitrust*, 86 TEX. L. REV. 1159, 1211 (2008).

231. Indeed, this is something I have proposed. See Haw, *supra* note 206, at 1284–91 (suggesting that antitrust rulemaking authority should be taken from the courts and given to the FTC); see also Crane, *supra* note 230, at 1208 (suggesting that the FTC engage in more notice-and-comment rulemaking).

232. See Haw, *supra* note 206, at 1287.

233. This is perhaps the experience of merger review at the FTC and DOJ, where in-house economic expertise about models predicting postmerger behavior is brought to bear on questions about policy and enforcement.

234. Unlike in the context of mergers, where policy makers were able to carve out a discrete area of antitrust law for a reallocation of authority from courts to an agency, there will be no similar line-drawing possible in the case of models. Models are merely tools used in all areas of economic study, which, in turn, are used in all areas of antitrust. In this sense, giving modeling evaluation to

Put another way, giving an agency primary authority over the use of models in law may exacerbate the existing problem. Just as treating models like facts gives expert witnesses and amici too much power, giving adjudicative authority to an agency gives social scientists and other experts who rely on models too much power over policy and dispute resolution. Judicial abdication to experts is neither desirable²³⁵ nor legitimate because Congress gave courts the primary authority to interpret many statutes that at least in the modern context require modeling.

B. JUDICIAL EDUCATION IN MODELING AND STATISTICS

Perhaps a better answer comes from recognizing that treating models like facts is a symptom of judicial inexpertise. The solution may be to treat the disease—to educate judges on the art of modeling.²³⁶ This would have the benefit of making it possible to give models the kind of searching scrutiny that they demand without involving a wholesale delegation of antitrust authority to economists.

Judicial education on scientific matters is hardly a new idea; indeed, it has been proposed widely and, more narrowly, executed.²³⁷ The gap between the principle—that judges need more scientific and technical skills than a traditional legal education provides—and its execution may lie in the voluntary nature of judicial education programs. Several factors combine to create barriers to mandatory education programs. On the logistical side, heavy judicial dockets, expense, and judges' diverse educational backgrounds and needs make a mandatory educational program difficult. On the incentives side, life appointment means that, while judges can be encouraged to attend, there is little in the way of a “stick” to compel the unwilling judge.

Thus, it would seem that a mandatory educational program is out of the question. The challenge, then, is to encourage widespread judicial participation in educational programs sufficient to allow the fulfillment of *Daubert*'s promise of thorough review of models and Congress's intent that the Supreme Court meaningfully engage with the models behind its policy.²³⁸ Even a voluntary and

agencies is a much more extreme version of my argument in *Amicus Briefs and the Sherman Act*, which advocated for giving the FTC rulemaking authority under the Sherman Act. See Haw, *supra* note 206, at 1284–85.

235. A related set of arguments can be imported from the debate about generalist versus specialist judges. For example, Judge Diane Wood has defended generalist judges by appealing to their resistance to capture, their ability to “demystify legal doctrine and to make law comprehensible,” and to foster cross-fertilization of ideas. Judge Diane P. Wood, Speech: Generalist Judges in a Specialized World (Feb. 11, 1997), in 50 SMU L. REV. 1755, 1767 (1997). These arguments also apply to locating decisionmaking in the judiciary versus an expert agency.

236. An ideal education would involve not only substantive training on modeling techniques but also training “in how to be an educated consumer” of model-derived claims. Cf. Paul H. Edelman, *Burden of Proof: A Review of Math on Trial*, 60 NOTICES AM. MATHEMATICAL SOC'Y 910, 914 (2013) (book review) (proposing judicial education in mathematics).

237. For a discussion of judicial education programs, see Cheng, *supra* note 142, at 1273–74.

238. See *supra* section II.C.1.

somewhat ad hoc judicial educational program could be effective if judges were willing to attend; if education were viewed as a “carrot,” there would be no need for a “stick.” And with sufficient judicial demand, the supply of educators, and likely federal funding, would respond.

Creating that kind of demand means convincing judges that learning modeling is worth the time and effort as an investment in their own future. Judges should demand more education on scientific topics like modeling because without it, pressure against pairing inexperts with legal issues requiring expertise will mount, and Congress will more often give in to the temptation to reallocate decisionmaking authority from courts to agencies. Judges, unresponsive to many incentives (by design), have been shown to respond to threats of losing jurisdiction or other constrictions of their power.²³⁹ The pressure against using lay courts to solve expert problems is, at bottom, a threat to strip jurisdiction. By narrowing the knowledge gap between themselves and the experts, judges can protect their existing level of power and influence. Recognition of this threat—or perhaps, as they may see it, this opportunity—should raise demand for judicial education. And the legal, economic, and statistical academies should be prepared to respond.

Modeling can be learned, especially by an erudite judiciary.²⁴⁰ The notion that law is a humanities subject, and so lawyers and judges can claim ignorance in the face of quantitative reasoning, is an assumption that is appropriately eroding. Indeed, Oliver Wendell Holmes’s prediction that “the black-letter man may be the man of the present, but the man of the future is the man of statistics and the master of economics”²⁴¹ is finally coming true, and it is a good thing. Modeling, the “new lingua franca,” is the dominant language of all scientific and social scientific areas of study, including economics, political science, and medicine. That so many academics and professionals, from doctors to professors to graduate students, can form a workable understanding of modeling sufficient to criticize and understand models, their choices, and their implications, should put to rest the notion that judges simply are not up to the task.

239. Perhaps the most well-known example of this phenomenon is the “switch in time that saved nine,” describing a change of jurisprudential heart on the part of Justice Roberts in *West Coast Hotel Co. v. Parrish*, 300 U.S. 379 (1937), which followed President Roosevelt’s threat to pack the Supreme Court with six new members more friendly to his New Deal laws.

240. Cf. Faigman, *supra* note 10, at 979 (“Once a judge understands multiple regression analysis he or she need not learn it again when it arises in another context.”). Not all scholars subscribe to this point of view. See, e.g., Kevin A. Kordana & Terrance O’Reilly, *Daubert and Litigation-Driven Econometrics*, 87 VA. L. REV. 2019, 2022, 2026 (2001) (writing in response to David Kaye’s piece advocating “strict judicial scrutiny” of modeling choices, observing that “[i]n principle, a judge who understood the degree to which a model resembled or varied from other widely used models could make the distinction, but that is a lot to expect from nonspecialist judges” and that “it is probably unrealistic to seek to have judges screen social science models prepared for litigation on the basis of their scientific pedigree”).

241. Oliver Wendell Holmes, Jr., *The Path of the Law*, Address at Boston University School of Law (Jan. 8, 1897), in 10 HARV. L. REV. 457, 469 (1897).

But perhaps more fundamentally, we need to rethink the criteria by which we select judges. The ubiquity of modeling in legal disputes, and models' importance in allocating justice, redressing wrongs, and designing legal rules, suggest that model competence should be a threshold criterion for judicial appointment. That is not to say that judges should be Ph.D. statisticians, but perhaps proficiency in the new lingua franca should be seen as a basic skill essential to judging. This new attitude would make temporary the problem of judicial inexpertise and the challenge of educating a generation of jurists who came up in a legal culture where aversion to math and science was tolerated.

CONCLUSION

Justice Scalia's comment in *Comcast* was not intended to be controversial; indeed, its placement in a footnote would indicate that he intended it to be unexceptional, not incendiary. And it is likely that it will not have a major impact on the treatment of models and their conclusions. But his statement, and the dissent's disagreement with it, provides an opportunity to examine an important and, as it turns out, contested question: Are models facts?

Justice Scalia's answer—that model-driven claims constitute "data" that may be factual, but whose meaning is entirely an issue of law—is correct as a normative matter. Modeling choices, inextricable from their outcome, should be subject to the kind of thoroughgoing review that we demand from courts on legal issues. But as a positive, or descriptive matter, Justice Scalia is wrong. Models, their constitutive choices, and their resulting claims are too often given the "black box" treatment that law reserves for facts. The reason for this discrepancy is that judges avoid making determinations that they feel incapable of making well. And since modeling is part art, judges feel understandably adrift in their evaluation. But their deference—to economists and to juries—is epistemically and perhaps politically unfounded. The rising use of models in almost all areas of litigation should inspire judges to seek out special training in model thinking. Without achieving a conversational proficiency in the new lingua franca, judges risk obsolescence.