

2010

Cautions on the Use of Economics Experiments in Law

Kathryn Zeiler

Boston University School of Law

Follow this and additional works at: https://scholarship.law.bu.edu/faculty_scholarship



Part of the [Law and Economics Commons](#)

Recommended Citation

Kathryn Zeiler, *Cautions on the Use of Economics Experiments in Law*, 166 *Journal of Institutional and Theoretical Economics* 178 (2010).

Available at: https://scholarship.law.bu.edu/faculty_scholarship/575

This Article is brought to you for free and open access by Scholarly Commons at Boston University School of Law. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Scholarly Commons at Boston University School of Law. For more information, please contact lawlessa@bu.edu.



**CAUTIONS ON THE USE OF ECONOMICS
EXPERIMENTS IN LAW**

JITE 166 (2010), 178-193
Boston University School of Law, Law & Economics Paper No. 15-37

Kathryn Zeiler
Boston University School of Law

This paper can be downloaded without charge at:

<http://www.bu.edu/law/faculty/scholarship/workingpapers/2015.html>

Cautions on the Use of Economics Experiments in Law

by
KATHRYN ZEILER*

The recent move to import empirical results into law and policymaking have introduced challenges related to drawing proper inferences from quantitative studies. The purpose of this essay is to elaborate on three specific cautions on the use of economics experiment results. First, critiques of experiment designs based on external and ecological validity are often misplaced. Second, some legal scholars have fallen into the problematic habit of applying results from experiments directly to law and policy rather than applying well-supported theories. Third, the divergent purposes behind economics studies and legal scholarship give rise, in part, to problematic cherry picking of experimental studies by legal scholars. (JEL: B41, C90, K00)

1 Introduction

The field of law and economics incorporates and develops economic theory as well as empirical results from field studies and laboratory experiments to both descriptively and normatively analyze law. Widespread application of economic theory to law got off the ground well before legal scholars began importing and generating empirical studies. In recent year, however, the availability of field data, the growth of experimental economics as a field and the increased popularity of interdisciplinarity in legal scholarship have led to widespread importation of economics studies that report the results of quantitative analyses pertaining to law.

The recent development of the field of empirical legal studies and the move to import empirical results into legal scholarship and policymaking have introduced challenges related to drawing proper inferences from quantitative studies. The purpose of this essay is to explore common misuses of quantitative results from economics experiments. While many challenges arise in the application of experimental economics to law, I focus on three. First, I argue that critiques of experiment designs by legal scholars and policymakers based on notions of external and ecological validity are not only misplaced but are unproductive. Correct evaluation of experiments requires appreciation for the reasons why we conduct them: to study basic economic principles. A good

* Professor of Law, Georgetown University Law Center. zeiler@law.georgetown.edu; www.georgetown.edu/faculty/kmz3/. Special thanks to Charles R. Plott for many helpful discussions and to Greg Klass, Leslie Street and participants of the 27th Seminar on the New Institutional Economics, especially Kenworthy Bilz and Christoph Engel. I also benefited from comments from Georgetown Law junior faculty members during a “first-thoughts” dinner gathering. Thanks to Jing Cao and Hugh Carlson for excellent research assistance.

experiment design does not mimic reality; rather, it is guided by conditions of theories the experiment is designed to test.

Second, rather than applying theories that tend to be well supported by data produced in the laboratory, commentators sometimes apply experiment results directly to law and policy. This misuse is problematic for a variety of reasons. First, this misapplication of experiments, in part, drives faulty critiques of experiments grounded in notions of external validity. Second, when we apply results rather than theories, our focus is easily shifted toward faulty critiques of experiment designs and away from constructive criticism directed toward assumptions upon which tested theories are based. Third, applying results rather than theory takes our eye off the ball. We lose sight of the fact that we should instead apply theories that are well supported by a body of evidence. Assessing how well a theory is supported by the data requires us to be concerned not only with one or a small handful of studies but rather with the body of evidence that offers broader lessons about the robustness and accuracy of basic principles comprising the foundation of theories. Finally, by focusing on results rather than theory, we sometimes fail to notice which theories are best supported by the evidence as a whole and which questions remain unsettled.

The third common misuse is triggered by the divergent purposes behind economics studies and legal scholarship. The branch of economics that produces the type of experimental studies I discuss here is driven by the desire to develop robust and accurate descriptive models of behavior. Legal scholarship, on the other hand, is largely geared toward the development of policy prescriptions. This tension often leads to cherry picking from a large body of evidence that often leaves many questions unanswered. Cherry picking to support one's predetermined stance leads to unhelpful and often misleading representations of the science.

This essay is organized as follows. Part 2 describes the role of economic experiments in the science of economics. The pros and cons of field studies and laboratory experiments are discussed. With this as background I argue that the most useful experiments are often those that substantially deviate from complex settings that exist outside the laboratory. Part 3 discusses each of the three common misuses in turn. The concluding section offers a few suggestions for reducing the frequency of these misuses.

2 The Purpose Behind Experiments

The field of economics progresses through application of the scientific method. Economists observe phenomena they wish to explain or predict. Theoretical models are then developed to illuminate and explore potential drivers of observed outcomes or behavior. The models generate predictions that can be tested using data from a source other than the one from which the original observation was made. When data do not comport with a model's predictions, features of the data-generating process are examined for clues about how to update the theory to enhance its generalizability. Revised theoretical models generate different and

often more refined predictions, which are again tested using data.¹ The goal of this cyclical process is to develop robust, general theories that can be applied to study and improve health care markets, labor markets, international trade relations, and regulatory regimes, to name just a few.²

Economists generally use data generated either in the field or using laboratory experiments, and more recently using experiments conducted in the field (HARRISON AND LIST [2004]). Each data source comes with advantages and disadvantages. Field data, as its name suggests, is generated in the environment in which the theory eventually will be applied. Thus the data reflect the influence of factors beyond those the theory focuses on. This can be useful when it comes to deciding whether and how to apply the theory to make policy. On the other hand, the presence of other factors that might vary across units (e.g., states) and/or time, sometimes contemporaneously with the factor of interest, give rise to challenges when attempting to isolate and measure the impact of the factor of interest. This is especially troublesome when changes in the other factors are, in fact, correlated with changes in the factor of interest.

In some cases studying principles and testing theories in the field is impossible (ARLEN AND TALLEY [2008]). If confounds cannot be controlled or key variables cannot be measured in the field, empiricists often resort to the laboratory. Laboratory environments provide opportunities to implement controls not available in the field. For example, random assignment to treatment and control groups, often impossible in the field, is easily achieved in the lab (HARRISON AND LIST [2004]). In addition, the controlled environment allows for singular changes to the procedures from treatment to treatment, which allow for reliable identification of the impacts of the changes.

Economists generally conduct experiments either to study the fundamental principles of economics that are used to build theoretical models or to test predictions of basic theories developed using these fundamental principles (KAGEL AND ROTH [1995]). For example, laboratory environments are generated to test levels of risk aversion and loss aversion under a variety of condition. The lab is also a useful setting for testing the impacts of changes in information on behavior and the robustness and contours of price theory.

When designing experiments to study principles or to test theories, experimentalists necessarily abstract from reality. Abstraction is required to control for alternative explanations and to measure variables impossible to measure in the field, among other reasons. For example, consider experiments conducted to test theories designed to explain observed gaps between willingness to pay and willingness to accept, often referred to as the endowment effect (KAHNEMAN, KNETSCH AND THALER [1990]). Experiments designed to test theories related to loss aversion and reference-dependent preferences can control

¹ For more nuanced discussions of the application of the scientific method in economics, see PLOTT AND SMITH [2008], GUALA [2005], SAMUELSON [2005], CRAWFORD [1997] and ROTH [1988]. SMITH [2002] summarizes the scientific method but takes issue with the idealized version of how experiment results relate to theory.

² It's important to note that experiments are employed not only to test theories. They are also used, for example, to study how individuals operate under a particular set of conditions (PLOTT [2001]) or to refine existing theories (ARLEN AND TALLEY [2008]).

for alternative explanations in a variety of ways. By limiting the time subjects spend with the good, the design can minimize the impacts of attachment on one's willingness to accept. Similarly, by making it virtually costless to trade, the design can eliminate or significantly reduce the impacts of transaction costs on individuals' willingness to trade. It is also necessary to abstract from reality to measure individuals' valuations as buyers and sellers. Testing the theory requires measuring the most buyers would be willing to pay and the least sellers would be willing to accept to give up the endowed good. These variables are not easily measurable in the field for obvious reasons. It turns out that reliably measuring them in the lab is not straightforward; thus, experimenters use special mechanisms designed to provide incentives to announce true valuations. Of course, these mechanisms are rarely if ever used in actual markets. Despite the abstraction, these experiments are useful for testing the basic principles of reference dependence and loss aversion.

3 Common Misuses of Economics Experiments

With an understanding of the purpose behind economics experiments, identification of the misuses of economics experiments in legal scholarship and policy analyses is possible. This part summarizes cautions against three common misuses of experimental studies: faulty critiques grounded in external and ecological validity, direct application of experiment results rather than well supported theories and cherry picking of single or small handfuls of studies from larger, more informative experimental literatures.

3.1 Faulty Critiques of Experiments

The conventional wisdom in the field of law and economics (and in some cases the fields of economics (LEVITT AND LIST [2006]) and political science (BASSI ET AL. [2006])) seems to be that experiments are useful only to the extent they are externally valid. An experiment's results are externally valid if one is able to draw valid causal inferences from them that can be applied generally (SCHRAM [2005]). The concept of ecological validity captures a closely related idea. Ecological validity is satisfied when an experiment's design sufficiently resembles the real world (SCHMUCKLER [2001]).

Critiques of this kind are forwarded both in the economics literature and in the law literature. For example, LIST AND LEVITT [2006, abstract] question "whether behavior inside the laboratory is a good indicator of behavior outside the laboratory." SUNSTEIN AND THALER [2003, 1168-1169], after listing results from several experimental studies, concede that "[i]t is possible to raise questions about some of these findings and to think that people may do a better job of choosing in the real world than they do in the laboratory." KELMAN [1985, p. 1042] challenges the applicability of results from an experiment studying dispute resolution described by HOFFMAN AND SPITZER'S [1985] by arguing that "[i]n a real world nuisance dispute, the contract settlement would by no means be as straightforward" as the one generated by the experiment design. KOROBKIN [2006, p. 281] discusses the issue more generally: "The disadvantage of

[experimental economics] approaches is that they always raise the question of 'external validity'...that is,...behavior empirically documented in a particular laboratory setting may or may not occur in the highly contextual, real world circumstances of interest to lawyers and other dispute resolution practitioners."

While it's often important to evaluate the external validity of experiments designed to study potential impacts of policy changes, this criterion is often misapplied to experiments designed to test fundamental principles of economics (PLOTT [1982, 2001]).³ Plott takes issue with describing experiments as "external." He argues that the framing of experiments as somehow "external" is misleading. Indeed, experiments occur in the "real world." While experiment designs often ask individuals to make decisions and take actions in simplified settings, they remain part of worldly reality. There is no "out there." Experiments, the argument goes, might be more usefully distinguished as occurring in simple settings as opposed to more complex settings that characterize the field. The recognition that theories bridge the simple settings of experiments and the more complex settings found in the field illuminates the usefulness of working in these simple settings (PLOTT [2001]).

The notion of "validity" also seems misplaced when it comes to evaluating the usefulness of experiments. Causal relationships posited by theory are not classified as valid or invalid. Validity is outside the realm of the scientific method. Rather, we gain confidence in theories by evaluating their reliability (or robustness). When a theory is supported by data derived from a variety of simple contexts, we're more inclined to conclude it operates in complex settings.

Arguments regarding the inapplicability of the concept of external validity to experiments are brought into sharper focus by recalling the purpose behind experimental studies. As discussed, experiments are employed to explore fundamental economic principles and test predictions of basic economic models that incorporate these basic principles. Simple settings allow us to rule out alternative explanations and identify the effects of changes in the treatment variable. Changes in outcomes or behavior arising from simple changes in the variable of interest either support or reject the basic principles or theories under examination. Once we gain confidence in the basic principles, we can begin to build more complex theories to account for additional features present in complex settings that might impact behavior or outcomes.

In this way, theory acts as a bridge between the simple settings in experiments and more complex settings. Studying behavior in simple settings allows us to gain confidence in the basic principles we use as building blocks in more complex models. If we have reason to believe the predictions will differ when we move from a simple setting to more complex settings, our aim should be to update the theory to account for the features of concern. Once the theory has been updated, we can test its predictions in the laboratory or using field data. As the theory begins to take into account more and more features of concern, we can test its predictions using data from the field. Without initial confidence in the

³ While some accept this claim (e.g., ARLEN AND TALLEY [2008], p. xxxvii), at least in the context of testing the predictions of unitary-process models, others disagree (e.g., GUALA [2005], Ch. 7).

basic principles, however, attempting to develop more complex theories is to put the cart before the horse.

In addition to studying fundamental economic principles and testing predictions derived from theory, economists sometimes conduct experiments to study experimental methods or to investigate how some policy change might impact behavior or outcomes in practice (KAGEL AND ROTH [1995], PLOTT [2001]). With respect to the latter, the experiment design is driven in large part by important features of the naturally occurring environment. For example, if one is interested in studying the impact of framing in a particular setting—e.g., computation of damages by juries—the experimental setting should be designed to take into account important features of the damages computation process that might impact outcomes (MCCAFFERY, KAHNEMAN AND SPITZER [1995]).⁴ Of course, designing an experiment that includes all the important features of naturally occurring, complex settings is impossible. Thus, even the results from experiments specifically designed to study actual policy changes cannot be directly applied. The results will help us gain confidence in theories, which can be useful in designing policy (PLOTT [2001]).

Understanding the purpose behind experiments illuminates why the concept of external validity is unhelpful. It also demonstrates the fundamental flaw in design critiques based on how well they mimic reality. So, how does one evaluate experiment designs? The answer to this question also follows from the purpose behind experimenting: to test fundamental principles and basic theories. To do this well, experiment designs should satisfy a number of properties. First, the features of the experiment design must comport with the conditions required for the theory to apply. For example, if the model assumes an individual is endowed with a particular good, the experiment design must ensure that subjects understand their relationship to the good. Likewise, if the theory assumes that certain information is common knowledge, all subjects must have access to the information prior to making decisions. In addition, the experiment design must implement controls that rule out alternative explanations (GUALA [2005, chapter 5]). It must also draw a large enough sample from the subject pool to achieve sufficient statistical power to identify the particular effects of interest, if they exist. While this short list of sound design features is not exhaustive, it provides a starting point for proper critiques of experiment designs.

Especially useful designs develop environments that allow for the testing of opposing predictions from competing theories.⁵ For example, we might wonder whether gaps in willingness to pay and willingness to accept are due to wealth effects or to special features of preferences (i.e., reference-dependence and loss aversion) (COURSEY ET AL. [1987], KNETSCH AND SINDEN [1987], FRANCIOSI ET AL. [1992]). To explore these theories, we can design an

⁴ Other examples include experiments conducted to study airport access (GREYER ET AL [1979]), natural gas pipeline regulation (PLOTT [1988]), optimal spectrum auctions (SALMON AND PLOTT [2004]) and pollution emission allowances (CASON [2008]).

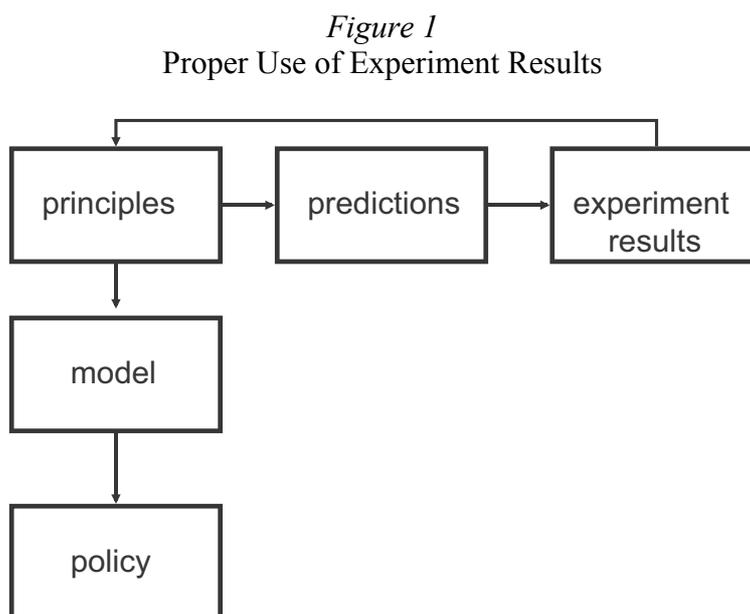
⁵ “[T]he policy implications of a theory’s evident refutation are elusive in the absence of a plausible, falsifiable alternative theoretical account that is more consistent with observed data.” (ARLEN AND TALLEY [2008]).

experiment with two treatments: one that creates wealth effects and one that eliminates wealth effects (i.e., provides buyers with cash roughly equal to the value of the good provided to sellers). Both treatments should control for alternative explanations (e.g., transaction costs). If gaps remain when wealth effects are eliminated, the data support the theory positing particular features of preferences. On the other hand, if gaps disappear in the absence of wealth effects, then we can conclude that the data support the wealth effects conjecture. Designing an experiment to test competing theories avoids the dilemma that arises when data collected to test a single theory ends up rejecting the theory.

Failing to appreciate the scientific purpose behind experiments leads to a second misuse of experiments: direct application of experiment results in normative arguments. The next section cautions against this misuse and suggests a more helpful way to incorporate experiment results into normative arguments.

3.2 Application of Economics Experiments in Policy

Understanding the purpose behind economics experiments not only helps in understanding how to critique them but it is also useful in better understanding how to apply what we learn from experiments. Figure 1 illustrates the relationship between principles, experiment results and policy.

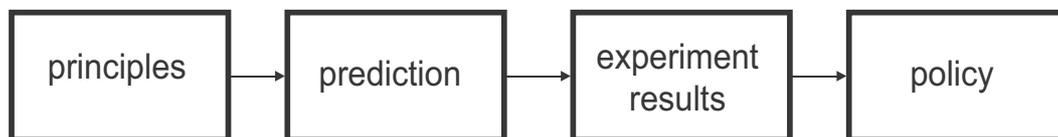


The scientific method takes us from principles to predictions to experiment results and back again. As discussed, the goal of this cyclical process is to develop a set of fundamental principles that eventually will comprise the foundations of more complex theories. Replicating reported results and testing whether principles hold up under a variety of conditions helps to shore up confidence that the principles are generally valid. The more confidence we have in the fundamental principles, the more willing we should be to incorporate them

into complex models. The complex models, then, generate predictions and general intuitions about behavior and outcomes under a set of assumptions. Finally, these predictions and intuitions can be incorporated into policy analyses—both descriptive and normative.

Actual applications of economics experiment results, however, oftentimes proceeds along a different route. Figure 2 illustrates how experiment results are sometimes used in policy analysis. Overlooking the relationship between experiment results and theory, legal commentators often interpret experiment results—sometimes from just one or two experimental studies—as demonstrating the proven existence of some general feature of individual preferences or inclinations.

Figure 2
Common Misuse of Experiment Results



The legal literature is rife with examples of this misuse of experiment results. Rather than focus on the abundance, however, it will be helpful here to select one example—SUNSTEIN AND THALER’S [2003, p. 1168] application of results reported by GREYER [1980]—to illustrate the negative consequences of moving directly from results to policy. Grether concludes that decisions under uncertainty tend not to comport with predictions of Bayes’ rule. Sunstein and Thaler cite this result to support the claim that “[p]eople fail to make forecasts that are consistent with Bayes’ rule.”⁶ They go on to argue that this result, along with others that run counter to predictions derived from neo-classical models, justify the adoption of what they refer to as “libertarian paternalism.”

The natural question to ask at this point is: what’s so bad about direct application of results to policy? The distinction illustrated in Figures 1 and 2 is important for several reasons. First, the direct application of experiment results to policy analyses is, in part, what drives faulty critiques of experiment designs grounded in ecological or external validity. Moving directly from experiment results to policy naturally raises the question of whether the results are useful for policy given the experiment design abstractions—why would we expect individuals participating in markets to act similarly to the subjects in Grether’s simple experiment? By applying theory rather than results and recognizing the proper role of experiments in the development of general theories, our critiques

⁶ It’s interesting to note that, while Grether also found some evidence that experienced subjects deciding under financial incentives act more like Bayesians, Sunstein and Thaler do not address this important nuance.

of experiments shift to more useful questions about whether the experiment is sufficiently well designed to properly test the theory.

Second, the application of results directly to policy misdirects attention toward experiment designs and away from the development of useful theories. When we directly apply experiment results, our focus is on analyzing how the experiment design differs from reality, and whether and how these differences matter. By recognizing that theories rather than experiment results are applied in policy analyses, we can engage in a more productive process of developing robust and useful theories that can be applied in policy analyses. In particular, shifting our focus to the application of theories compels us to ask whether the theory is detailed enough for policy application. If an important feature of the real world is missing from the model, the next step should be to update the theory (as opposed to modifying the experiment design to better align it with the “real world”). This feedback loop between policy analysis and theory development can push theorists to modify theoretical models in useful ways.

Third, when we apply results from one or a small handful of studies, we lose sight of the fact that theories are best applied only when they are well supported by a *body* of experimental evidence. Translation of lessons from the field of economics into the normative realm is most productive and least risky when a theory’s predictions comport with a substantial portion of applicable observed phenomena—both in the laboratory and in the field. When we apply results rather than theory, it’s easy to mistakenly draw conclusions that do not follow from an examination of the larger literature. As a general matter, it’s important to apply theories rather than results from particular experiments because the latter can easily lead us to draw inferences from non-robust results produced by one or a small handful of experiments.

Fourth, when we apply results rather than theory and apply one or a small handful of results, we often fail even to consider theoretical explanations for the observed phenomenon. Moving directly from results to policy compels us either to ignore possible explanations for the results or to assume some explanation that may or may not be supported by the larger literature. For example, consider observed behavior in ultimatum games, in which the proposer receives some amount and offers a split to the responder.⁷ The split is effectuated unless the responder rejects the offer, in which case both players get nothing. Experimenters widely report that proposers offer much more to responders, even though the game theoretic solution predicts tiny offers. A common offer is half the pie. If we were to apply this result directly in policy analysis, we might be compelled to conclude that notions of fairness influence proposers. A large body of experimental evidence, however, studies whether proposers instead are reacting to fears of punishment by offended responders (KAGEL AND ROTH [1995, pp. 270-274]). Clearly, these very different theories might lead to opposite policy implications. Finally, even when we recognize the theory best supported by the data, if we apply the results from just one experiment, we misrepresent, in some

⁷ For a discussion of ultimatum games and findings from experiments testing theories attempting to explain proposer and responder behavior, see KAGEL AND ROTH [1995, pp. 258-292], HOFFMAN, MCCABE AND SMITH [2008] and J. KEITH MURNIGHAN [2008].

cases, the broader understanding of basic economic principles the body of relevant literature provides.

Sunstein and Thaler's use of Grether's results to support the claim that regulations are necessary to nudge those who are likely to make suboptimal decisions toward correct choices provides an illustration of the dangers. First, it demonstrates a failure to account for what might be driving Grether's results. In other words, it applies experiment results rather than theory. It certainly might be that individuals suffer from cognitive limitations that hinder their ability to reach optimal outcomes. On the other hand, it's possible that individuals are perfectly capable of deciding what's in their best interest, and the standard model of Bayesian updating rejected by Grether's results simply makes invalid assumptions about individual preferences. Without knowing more about what might be driving Grether's results, developing effective policy is difficult. Multiple theories might be observationally equivalent but might lead to very different policy recommendations. Sunstein and Thaler seem to suggest that the results support the claim that cognitive limitations preclude individuals from making the right choices. ANDERSON AND HOLT (1997), however, describe several alternative explanations for deviations from rational Bayesian updating (in the context of information cascades), including a preference for the status quo and the possibility that individuals derive utility from herding together. Alternatively they might be adverse to the risk of standing alone. Understanding which of these conjectures accounts for the observed phenomenon clearly is important when it comes to implementing effective policy. Policy is able to steer individuals in the "right" direction only if we understand in which direction to head and how to get there.

Second, the single citation to Grether's nearly thirty-year-old results ignores the vast literature on the question whether and how well individuals Bayesian update.⁸ In fact, experimentalists have developed a substantial literature focused on explanations behind deviations from models relying on Bayesian updating. The literature has attempted to investigate myriad explanations for such deviations, including status quo bias, herding preferences, confirmatory bias (DAVE AND WOLFE, 2003), and differences in the way the problem is presented (GIGERENZER AND HOFFRAGE, 1995), among others. Only when we examine the literature as a whole will we get a sense of which theory is best supported by the data, or whether the explanation is an open question. Again, until we have a sense of what drives behavior we observe in the laboratory, forwarding normative positions and designing policy prescriptions become difficult at best and ineffective at worst.

For the reasons discussed, legal commentators and others engaged in policy analysis should take care to import well-supported theories into their analyses rather than directly applying experiment results. This adjustment will help to refocus critiques from whether experiment designs look sufficiently like

⁸ A good place to start might be Grether's more recent work, finding that, while individuals tend to use a variety of decisions rules when making decisions under uncertainty, they are most likely to use Bayes' Rule, especially when they are paid based on outcomes (EL-GAMAL AND GREThER [1995]).

the “real world” to whether the theories being tested (and applied) include the relevant features of the environments in which they will be applied. Focusing on the theory’s shortcomings will help to push the theoretical models forward to make them more useful in the policy realm. Finally, by applying theory rather than experiment results, we’re less inclined to draw overly broad conclusions from one experimental study, or a small handful of studies, and to neglect lessons from the larger literature about what drives observed behavior.

Given the benefits derived from shifting focus from the results to basic economic principles and theories, why do legal commentators so often directly apply results? Furthermore, why do they apply results from just one study or a small handful of studies when ignoring the larger literature substantially hinders our understanding of the basic principles and observed phenomena? The following section discusses a third caution on a common use of economics experiments in legal scholarship—cherry picking only those studies that support some pre-determined position from among a larger body of research, a portion of which reports results that contradict or challenge the author’s arguments. It also offers a possible (partial) explanation beyond mere misunderstanding of the role of experiments in the science of economics.

3.3 The Leap from Descriptive to Normative

A third caution on the use of economics experiments, cherry picking, stems in part from the divergent purposes behind economics studies and legal commentary. While the field of economics is comprised of two distinct branches of analysis—one descriptive and one normative—the type of experiments I focus on here are associated most closely with the descriptive side of the fence. Specifically, these experiments are conducted for the purpose of studying basic principles that eventually get plugged into larger models that attempt to model decision-making under a variety of conditions. The main goal of the descriptive enterprise is to develop a set of models that generate robust and accurate predictions. Developing robust models requires us to test predictions in a variety of settings. By incorporating what we learn from experiments that test the boundaries of theories, we are able to enhance the robustness of theories.

In contrast, legal commentary is driven largely by normative concerns.⁹ For example, doctrinal scholarship, while sometimes generated to explain and analyze judicial decisions or other types of legal rules, often criticizes the structure of law. Critiques of judicial decisions more often than not include proposals to change the law. Some proposals recommend substituting one doctrinal test for another. For example, CALABRESI’S [1980] famous analysis of

⁹ See e.g., RUBIN [1988] (“When viewed as an academic discourse, the most distinctive feature of standard legal scholarship is its prescriptive voice, its consciously declared desire to improve the performance of legal decision-makers.”); SCHLAG [1991] (“The rhetorical script of normative legal thought is already written, the social scene is already set and play after play, article after article, year after year, normative legal thought requires you to choose: “What should we do? Where should we go?” We are free, but we must choose...”).

accident law argues that, in addition to reducing losses from accidents, we should structure tort law to reduce the costs of precautions. Legal scholars also critique policy making more generally. In this vein, BRONSTEEN, BUCCAFUSCO AND MASUR [2009] argue for replacing cost-benefit analysis with well-being analysis as policymakers' core analytical tool. Legislative action is a target as well. One example comes from HUSAK'S (2008) recently published book, which argues that states should decriminalize much of the conduct currently characterized by the law as criminal.

This tension between the main goals of the descriptive branch of economics and legal commentary gives rise, in part, to a third category of misuse of economics experiments—cherry picking from a body of experimental results. The economics literature devoted to model building and testing is characterized by competing theories and (oftentimes) contradictory results that have yet to be reconciled. It is also common for methodologists to challenge experiment designs, calling into question controls, measurement techniques and other design issues. The methodology often naturally produces a set of results reported in a series of papers, some of which support a particular normative position and some of which can be employed in arguments countering it. In fact, much is unsettled. Rather than providing an objective assessment of the economics landscape, however, legal scholars sometimes report only those results that support their forwarded policy positions, which are often chosen not based on an objective assessment of potential consequences but rather to forward a particular set of values or predetermined policy objectives (e.g., increasing access to health care, lengthening jail sentences for convicted criminals, reducing auto emissions, addressing global warming, etc.). While this misuse is not unique to experiments (i.e., legal scholars often cherry pick from field studies and theoretical results) cherry picking from bodies of economics experiments is quite common in legal scholarship.

Rather than list a set of examples, I'll drive home the point by sharing a colleague's recent description of advice he received from a prominent legal scholar and mentor, who contributes regularly to the field of normative legal scholarship and regularly testifies in front of Congress. On the use of empirical results, the mentor advised viewing the chosen policy proposal as a "Christmas tree." Empirical results, including those generated by economics experiments, should be considered mere "ornaments," meant to dress up the tree. Under this paradigm, the legal scholar is to locate and report only the "prettiest" empirical results and to leave the "ugly" ones in the attic, hoping no one will notice them. The danger, of course, is that policy prescriptions grounded in partial and/or mischaracterized findings from economics, if adopted, could lead to unintended negative consequences.

One possible response to the flagging of cherry picking as a problem is to argue that we should not be concerned about this because, when legal scholars leave out important experimental results, others will find them and bring them to light. Despite this sometimes-effective monitoring device, cherry picking remains problematic for at least three reasons. First, the language used to present the results often implies that the commentator is describing the science in a neutral or objective way (e.g., "social science research has revealed that X is true....").

Readers, many of whom are unfamiliar with the economics literature, often are unaware that the description of the science is incomplete or that caveats apply. They simply take the description at face value. Second, the vast majority of published legal scholarship is not peer-reviewed. Therefore, experts are rarely called upon to screen articles before they go into print. Unless the law student editor has expertise in economics, descriptions of the state of the science are accepted as is. Third, scholars have little incentive to correct the record. Even when incomplete descriptions are spotted, corrections are not encouraged by the academic reward system. When it comes to policymaking (to the extent legal scholarship influences policy), mechanisms like notice and comment periods invite corrections, although these clearly are not full proof.

4 Conclusion

Misuses of economics experiments in legal scholarship and policy analysis are driven by a variety of factors. Faulty critiques grounded in external or ecological validity and direct application of results rather than well-supported theories likely are due, at least in part, to a lack of appreciation for or understanding of the fundamental purpose behind economics experiments. A better understanding of the features of a sound economics experiment makes clear why experimentalists rightly do not strive to design experiments to mimic more complex settings that exist outside the laboratory. In addition, a better understanding of the role of experiments in the descriptive branch of economics helps clarify why directly applying experiment results rather than well-supported theories in legal analysis is problematic. Bringing to light the role of experiments in the field of economics likely will lead to fewer misuses of these types.

Addressing the third misuse is less straightforward. Cherry picking likely arises in part from a lack of appreciation of the fact that large bodies of evidence often are produced to study basic economics principles. Developing complete reviews of experimental literatures and synthesizing a set of results to draw conclusions regarding which theories best organize the data can be difficult, especially for those not trained in the methods. Increased attention to the inherent complexities of the science of economics can reduce the frequency of inadvertent cherry picking.

Other factors, however, drive cherry picking. The normative nature of legal scholarship naturally leads to attempts to highlight results that support the author's pre-determined policy prescription or proposed changes to the law. While we sometimes can rely on opponents to bring to light contradictory evidence, this is not a fail-safe. Certainly requiring peer-review of legal scholarship would help. Some law review editorial boards have been seeking reviews from faculty with expertise, although this solution comes with severe limitations (e.g., experts reside on a limited number of faculties, law faculty have few incentives to take on this time-consuming task, law students are not driven by reputation effects that impose pressure on journal editors in other fields who are members of the field). Perhaps the best we can do under the circumstances is to warn those who turn to legal scholarship for objective assessments of whether

and how lessons from economics experiments can be applied in legal analyses to be on the lookout for less-than-objective applications of experiments.

References

- ARLEN, J. AND E. TALLEY [2008], "Introduction," pp. xv-lxi in: J. Arlen and E. Talley (eds.), *Experimental Law and Economics*, Edward Elgar Publishing.
- ANDERSON, L.R. AND C.A. HOLT [1997], "Information Cascades in the Laboratory," *American Economic Review*, 87, 847-862.
- BASSI, A., R. MORTON AND K. WILLIAMS [2006], "Incentives, Complexity, and Motivations in Experiments," Working Paper.
- BRONSTEEN, J., C.J. BUCCAFUSCO AND J.S. MASUR [2009], "Welfare as Happiness," Working Paper, University of Chicago Law & Economics, Olin Working Paper No. 468, Chicago.
- CALABRESI, G. [1970], *The Costs of Accidents: A Legal and Economic Analysis*, Yale University Press: New Haven, CT.
- CASON, T.N. [2008] "Trading Institutions and Emission Allowances" in: *Handbook of Experimental Economics Results, Volume 1*, Charles R. Plott and Vernon Smith (eds.), Elsevier North-Holland Publishing.
- -- AND C.R. PLOTT [2005], "Forced Information Disclosure and the Fallacy of Transparency in Markets," *Economic Inquiry*, 43, 699-714.
- COURSEY, D.L., J.L. HOVIS AND W.D. SCHULZE [1987], "The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value," *Quarterly Journal of Economics*, 102, 679-790.
- CRAWFORD, V.P. [1997], "Theory and Experiment in the Analysis of Strategic Interaction," pp. 206-242 in: D.M. Kreps and K.F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications, Volume 1*, Cambridge University Press: Cambridge.
- DAVE, C. AND K.W. WOLFE [2003], "On Confirmation Bias and Deviations from Bayesian Updating," Working Paper.
- EL-GAMAL, M.A. AND D.M. GREYER [1995], "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*, 90, 1137-1145.
- FRANCIOSI, R., P. KUJAL, R. MICHELITSCH, V. SMITH AND G. DENG [1996], "Experimental tests of the endowment effect," *Journal of Economic Behavior & Organization*, 30, 213-226.
- GIGERENZER, G. AND U. HOFFRAGE [1995], "How to improve Bayesian reasoning without instruction: frequency formats," *Psychological Review*, 102, 684-704.
- GREYER, D.M. [1980], "Bayes' Rule as a Descriptive Model: The Representativeness Heuristic," *Quarterly Journal of Economics*, 95, 537-557.
- --, R.M. ISAAC AND CHARLES R. PLOTT [1979], "Alternative Methods of Allocating Airport Slots: Performance and Evaluation," Paper prepared for the Civil Aeronautics Board. Pasadena: Polinomics Research Laboratories, Inc.
- GUALA F. [2005], *The Methodology of Experimental Economics*, Cambridge University Press: New York.
- HARRISON, G.W. AND J.A. LIST [2004], "Field Experiments," *Journal of Economic Literature*, 42, 1009-1055.
- HOFFMAN, E., K. MCCABE AND V. SMITH [2008] "Reciprocity in Ultimatum and Dictator Games: An Introduction" in: *Handbook of Experimental Economics Results, Volume 1*, Charles R. Plott and Vernon Smith (eds.), Elsevier North-Holland Publishing.
- -- AND M. SPITZER [1985], "Experimental Law and Economics: An Introduction," *Columbia Law Review*, 85, 1009-13.

- HUSAK, D. [2008], *Overcriminalization: The Limits of the Criminal Law*, Oxford University Press: U.S.
- KAGEL, J. H. AND A.A. ROTH [1995], *The Handbook of Experimental Economics*, Princeton University Press: Princeton, NJ.
- KAHNEMAN, D., J.L. KNETSCH AND R.H. THALER [1990], "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, 98, 1325-48.
- KELMAN, M. [1985], "Comment on Hoffman and Spitzer's Experimental Law and Economics," *Columbia Law Review*, 85, 1037-1046.
- KNETSCH, J.L. AND J.A. SINDEN [1987], "The Persistence of Evaluation Disparities," *Quarterly Journal of Economics*, 102, 691-695.
- KOROBKIN, R. [2006], "Psychological Impediments to Meditation Success: Theory and Practice," *Ohio State Journal of Dispute Resolution*, 21, 281-327.
- LEVITT, S.D. AND J.A. LIST [2006], "What Do Laboratory Experiments Tell Us About the Real World?," Working paper. Available at <http://pricetheory.uchicago.edu/levitt/Papers/jep%20revision%20Levitt%20&%20List.pdf>
- MCCAFFERY, E.J., D.J. KAHNEMAN AND M.L. SPITZER [1995], "Framing the Jury: Cognitive Perspectives on Pain and Suffering Awards," *Virginia Law Review*, 81, 1341.
- MURNIGHAN, J. KEITH [2008] "Fairness in Ultimatum Bargaining" in: *Handbook of Experimental Economics Results, Volume 1*, Charles R. Plott and Vernon Smith (eds.), Elsevier North-Holland Publishing.
- PLOTT, C.R. [1982], "Industrial Organization Theory and Experimental Economics," *Journal of Economic Literature*, 20, 1485-1527.
- -- [1988], "Research on Pricing in a Gas Transportation Network." Office of Economic Policy Technical Report no. 88-2. Federal Energy Regulatory Commission, Washington, D.C.
- -- [1999], "Policy and the Use of Laboratory Experimental Methodology in Economics," pp. 293-315 in: Luigi Luini (ed.), *Uncertain Decisions: Bridging Theory and Experiments*, Kluwer Academic Publishers: Boston.
- -- [2001], "Introduction," in: *Public Economics, Political Processes, and Policy Applications*, Edward Elgar Publishing.
- -- AND V.L. SMITH (eds.) [2008] *Handbook of Experimental Economic Results, Volume 1*, Elsevier North-Holland Publishing.
- ROTH, A.E. [1988], "Laboratory Experimentation in Economics: A Methodological Overview," *Economic Journal*, 98, 974-1031.
- RUBIN, E.L. [1988] "The Practice and Discourse of Legal Scholarship," *Michigan Law Review*, 86, 1835-1905.
- SALMON, T.C. AND C.R. PLOTT [2004] "The Simultaneous, Ascending Auction: Dynamics of Price Adjustment in Experiments and in the U.K. 3G Spectrum Auction," *Journal of Economic Behavior and Organization*, 53, 353-383.
- SAMUELSON, L. [2005], "Economic Theory and Experimental Economics," *Journal of Economic Literature*, 43, 65-107.
- SCHLAG, P. [1991]. "Normality and the Politics of Form," *University of Pennsylvania Law Review*, 139, 801-932.
- SCHMUCKLER, M.A. [2001], "What is Ecological Validity? A Dimensional Analysis," *Infancy*, 2, 419.
- SCHRAM, A. [2005], "Artificiality: The tension between internal and external validity in economic experiments," *Journal of Economic Methodology*, 12, 225-237.

SMITH, V.L. [2002], "Method in Experiment: Rhetoric and Reality," *Experimental Economics*, 5, 91-110.

SUNSTEIN, C. AND R. THALER [2003], "Libertarian Paternalism is Not an Oxymoron," *Chicago Law Review*, 70, 1159-1202.

Kathryn Zeiler
Georgetown University Law Center
600 New Jersey Avenue, NW
Washington, DC 20001
U.S.A.
E-mail:
zeiler@law.georgetown.edu