WHY AND HOW TO EXPERIMENT: METHODOLOGIES FROM EXPERIMENTAL ECONOMICS

*Rachel Croson**

The emerging field of experimental economics uses human experiments to answer research and policy questions. This article discusses the methodology used in economics experiments, describes what experiments can tell the legal researcher and practitioner, and provides examples of economics experiments as used in legal settings. Additionally, the article provides some guidelines for conducting economics experiments and addresses the ways that these experiments differ from the more familiar psychological experiments.

"It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts."[1]

I. INTRODUCTION

The relationship between theory and data in the sciences, as well as in mysteries, has been hotly debated.[2] On one side are the inductivists, like Holmes, who emphasize induction as the major means to attaining knowledge. Induction involves looking at data and developing a general theory or proposition from observation. In contrast, deductive reasoning travels from the most abstract to the least abstract. It begins with a set of axioms and uses the laws of logic and mathematics to manipulate them to form predictions about the world. John Stuart Mill defines this methodo-

* *The author would like to thank those involved with the symposium for Empirical and Experimental Methods in Law at the University of Illinois for their comments on this article. Thanks also to participants at Penn Law and Economics Summer Seminar for their input. Special thanks to Colin Camerer for showing me how to run my first experiment, Jerry Green for encouraging and enabling me to pursue this unusual and exciting field, and Vernon Smith and Al Roth for providing the opportunity to visit at Arizona and Pittsburgh, respectively, and to learn how experiments are run from experts. I am also indebted to Charles Plott for numerous discussions on methodological and other issues. Finally, thanks to Jason Johnston, Richard McAdams, Bob Mnookin, Jeff Rachlinski, and Matthew Spitzer for their encouragement in applying these ideas to the law.*

1. SIR ARTHUR CONAN DOYLE, *A Scandal in Bohemia*, *in* THE COMPLETE ADVENTURES AND MEMOIRS OF SHERLOCK HOLMES 326 (C.N. Potter ed., 1975) (speaking through the character of Sherlock Holmes).

2. *See*, *e.g.*, THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS 209 (1970).

logical debate in terms of the men who practice it: "[T]hose who are called practical men require *specific* experience and argue wholly *upwards* from particular facts to a general conclusion; while those who are called theorists aim at embracing a wider field of experience and . . . argue *downwards* from general principle to a variety of specific conclusions."[3]

In economics and law, we see a strong emphasis on deductive argumentation. A consequence of this emphasis is that empirical tests are significantly more challenging to perform. Theories become so complicated and refined that simply observing naturally occurring events does not address pressing theoretical questions. In economics, the field of experimental economics has developed to meet this challenge.

This article discusses the methodology used in economics experiments, describes what experiments can tell the legal researcher and practitioner, and provides examples of economics experiments in legal settings. A similar article by Professors Hoffman and Spitzer provides methodological advice for legal scholars on how to run experiments by using examples from experimental economics to illustrate the concepts.[4] This article could be considered an update of that article because it incorporates recent experiments in the growing field of law and economics.

Economics experiments are related to, but remain distinct from, other types of empirical work, including observational (empirical) research or simulation. In observational research, a researcher collects and analyzes naturally occurring data. This type of research is a mainstay of both economic and legal scholarship.[5] Simulation is another empirical methodology that is relatively new. In simulation research, artificial agents (computer programs) interact to illuminate evolutionary or other forces.[6]

In contrast, experimental economics uses controlled human experiments to answer research and policy questions. The methodology of controlled experiments has advantages and disadvantages. The primary advantages are a greater degree of control and an ability to replicate. In a laboratory experiment, a theory can be tested directly by controlling for extraneous factors, much as a physics experiment might control air pressure in measuring an atomic reaction. This additional control can be used to construct laboratory conditions that test theories and separate alternate theories that may not be separable with naturally occurring data. Because laboratory experiments are replicable, other researchers can reproduce the experiment and verify the findings independently.

3. JOHN STUART MILL, PHILOSOPHY OF SCIENTIFIC METHOD 423, 461 (Ernest Nagel ed., 1950).

4. Elizabeth Hoffman & Matthew L. Spitzer, *Experimental Law and Economics: An Introduction*, 85 COLUM. L. REV. 991 (1985).

5. *See*, *e.g.*, Kevin M. Clermont & Theodore Eisenberg, *Plaintiphobia in the Appellate Courts: Civil Rights Really Do Differ from Negotiable Instruments*, 2002 U. ILL. L. REV. 947.

6. *See generally* Randal C. Picker, *Simlaw 2011*, 2002 U. ILL. L. REV. 1019.

On the other hand, a disadvantage of controlled human experiments is that the resulting data is different than empirical data. A theory that can predict outcomes in a laboratory environment may suggest what might happen in the real world. Just because a theory works in a laboratory experiment, does not mean it will work in reality. However, if a theory does *not* predict outcomes in an idealized, controlled setting in the lab, it will likely not predict outcomes in the real world.[7] Thus, experiments can provide a middle ground between theoretical work and empirical reality. If a theory works in the lab, it can be further tested in the field. If it does not, it should be refined before further implications are discussed.

Three closely related types of research that will not be discussed in this article are: field experiments, survey research, and psychological experiments. All of these can accurately be termed *experimental* in the sense that the researcher actively manipulates the subjects and investigates the manipulation's effects. In field experiments, also called social experiments, the researcher compares policies by inducing them in different populations or in the same population over different times.[8] In the law, such manipulations are rarely within the domain of the researcher.[9]

Survey research, in contrast, involves hypothetical settings and responses. While surveys can be extremely useful, they differ from experiments in that they are exclusively hypothetical—the subject experiences no real outcomes contingent in their response.[10]

Psychology experiments are often very similar to economics experiments, with the exception that psychology experiments are designed to address *psychological* theories rather than *economic* theories.[11] Because the theories being addressed are different, the methodologies are somewhat different.[12]

My intention is not to provide a review of the growing literature in experimental economics. Excellent reviews already exist.[13] The goal of

7. *See* Charles R. Plott, *Industrial Organization Theory and Experimental Economics*, 20 J. ECON. LIT. 1485, 1486 (1982).

8. For overviews of social experiments, see David Greenberg et al., *The Social Experiment Market*, 13 J. ECON. PERSP. 157 (1999); James J. Heckman & Jeffrey A. Smith, *Assessing the Case for Social Experiments*, 9 J. ECON. PERSP. 85 (1995). For a discussion of legal issues, see BROOKINGS INST., ETHICAL AND LEGAL ISSUES OF SOCIAL EXPERIMENTATION 188 (Alice M. Rivlin & P. Michael Timpane eds., 1975).

9. Naturally occurring manipulations, either via two similar states with different laws or one place where the law has changed over time, are called natural experiments and fall under the category of observational or empirical research. *See supra* note 5 and accompanying text.

10. What researchers call individuals who complete their experiments varies by field. In this article, I will use the term "subject" to refer to an individual in an experiment. Recently psychology has moved away from this term and instead used the term "participant."

11. *See, e.g.*, F. J. MCGUIGAN, EXPERIMENTAL PSYCHOLOGY: A METHODOLOGICAL APPROACH 1–3 (1960).

12. *See infra* Part IV.E.

13. *See*, *e.g.*, DOUGLAS D. DAVIS & CHARLES A. HOLT, EXPERIMENTAL ECONOMICS (1993); THE HANDBOOK OF EXPERIMENTAL ECONOMICS 721 (John H. Kagel & Alvin E. Roth eds., 1995); Christine Jolls et al., *A Behavioral Approach to Law and Economics*, 50 STAN. L. REV. 1471 (1998).

this article is to instead provide a guide to individuals who desire to intelligently conduct or consume experimental research, using selected examples from the literature in experimental law and economics to illustrate principles.

This article proceeds with a brief description of the logistics of an experiment. Once that foundation is laid, the bulk of the article describes different types of experiments: those designed to address theory, to investigate anomalies, and to inform policy.[14] For each type of experiment, the article identifies particular methodological issues, challenges, and constraints, and provides examples. It continues with a brief discussion of methodological differences between economic and psychological experiments. It closes with some thoughts about the future role of economic experiments in the legal field.

II. Experimental Logistics

Roughly, an experiment proceeds as follows. First, the researcher decides upon an objective for the experiment—a theory he would like to test, an anomaly he would like to explore, or a policy he would like to testbed. He designs the experiment and materials, including instructions for the subjects, keeping in mind the experiment's objective,[15] previous literature in the area (both empirical and theoretical), and some general rules of the road. The researcher will often show his instructions to colleagues and other interested parties for their comments. Most universities have committees to monitor experiments and to ensure that no harm is being done to the subjects. Researchers must obtain the permission of this committee before running their experiment.[16]

If the experiment is complex, a pilot experiment may be run. In a pilot, volunteers work through the experiment in a dry run setting and provide feedback regarding the clarity of the instructions, the logic of the procedures, and so forth. Once the materials and procedures are honed, the researcher recruits the subjects. Often these subjects are students at the researcher's university, although other subject pools may be used depending on the objective of the experiment. Once this recruitment is complete, the experiment can begin.

At the experimental session, subjects check in and are given instructions or explanatory materials.[17] Sometimes the materials are read aloud to create common knowledge, and other times they are read silently.

14. This categorization is based on Alvin E. Roth, *Laboratory Experimentation in Economics*, 2 ECON. & PHIL. 245 (1986).

15. *See infra* Part III.

16. This committee is often called a human subjects committee. Researchers need the permission of this committee in order to run their experiments and must follow whatever procedures it may require.

17. Due to recent technological developments, experiments can be conducted on the Internet, which allows subjects to log onto a particular web page from anywhere to begin the experiment.

Subjects are typically given a chance to ask questions either privately or publicly.

During the experiment, subjects make the decisions that face them. For experiments that are run by hand, this may involve checking a box indicating a choice or playing a game with another subject and recording the choices. For computerized experiments, subjects enter their decisions into the keyboard. As decisions are being made, the appropriate feedback is provided to the subjects.

Upon completing the experiment, subjects are compensated privately based on their earnings,[18] sign a receipt, fill out any remaining paperwork, and are dismissed. Occasionally, subjects are debriefed by the researcher and told the purpose and design of the experiment.

The details of these steps: the experimental design, the subject pool used, the experimental implementation (computer or face-to-face), and the manner in which they are paid, depend critically on the objective of the experiment being run. The next section details some specific types of experiments and offers guidance on how to think about these details for each type.

III. Types of Experiments

Experiments are categorized in various ways. For purposes of this article, I will rely on the categorizations of Professor Roth.[19] Roth identifies three types of experiments: those designed to address economic theories (speaking to theorists), those designed to investigate previously identified anomalies (searching for facts), and those designed to inform policy (whispering in the ears of princes).[20]

A. *Experiments That Address Theory*

Experiments in this category are explicitly designed to address theories. These experiments have an important place in the dialectic of the

18. *See* Vernon L. Smith, *Experimental Economics: Induced Value Theory*, 66 AM. ECON. REV. 274, 274–79 (1976).

19. *See generally* Roth, *supra* note 14, at 245–46. In addition to the way Roth categorizes experiments, Davis and Holt identify three types of experiments: those that test the behavioral hypotheses, those that stress-test theory, and those that search for empirical regularities. *See* DAVIS & HOLT, *supra* note 13, at 18–20. Hoffman and Spitzer identify five types of experiments: those that test theory, those that explore the usefulness of theory (here, these two are discussed under experiments that address theory), those that suggest the development of new theory (here, this is discussed under experiments that investigate anomalies), those that aid in the designs of institutions, and other policy-relevant experiments (here, these last two are discussed under experiments that testbed policies). *See generally* Hoffman & Spitzer, *supra* note 4, at 1003–23. Plott identifies four types of experiments: measurement, theory first (here, experiments that address theory), data first (here, experiments that investigate anomalies), and design and testbed approaches (here, experiments that testbed policies). *See generally* Charles R. Plott, *Policy and the Use of Laboratory Experimental Methodology in Economics*, *in* UNCERTAIN DECISIONS: BRIDGING THEORY AND EXPERIMENTS 303–07 (Luigi Luini ed., 1999).

20. *See generally* Roth, *supra* note 14, at 245–46.

scientific method. First, a theory is constructed or described and predictions are generated from it deductively. Then, an experiment is run to test the theory. If the data from the experiment are not consistent with the predictions of the theory, the theory is refined, new predictions generated, and the process continues.

While this dialectic is critical to the progress of science, it should be noted that experiments can neither conclusively prove or disprove theories. Simply because an experiment is consistent with a theory's predictions does not mean the theory is true; an alternative theory could be developed that is also consistent with the experiment's outcomes. Thus, a theory is not proven true by observation. This is the reasoning behind Karl Popper's *falsification hypothesis*, which states that empirical research can only falsify theories, not prove them.[21] In fact, Popper argues, the ability to be falsified is exactly what makes a theory scientific rather than circular.[22]

Similarly, theorists often argue that observations do not disprove theories. Theories' truths are judged in their internal consistency—if the conclusions follow from the premises, then the theory is true. Experiments can, however, provide evidence that the theory does not describe reality or organize the observed data well enough to be useful. Typically, this is seen as a critique of the theory's premise.

That said, experiments can be run to directly test theories (or their premises). For example, Professors Hoffman and Spitzer conducted an experiment to test the validity of the assumptions of the Coase Theorem.[23] In these experiments, subjects bargain in a setting with externalities, perfect information, and have no transaction costs. The Coase Theorem predicts that all individuals in these situations will agree upon efficient outcomes—this type of prediction is sometimes referred to as a point prediction because the prediction is that 100% of the groups will reach efficient solutions. In their observation of 114 negotiations, the researchers found that 89.5% of them resulted in efficient outcomes. Most researchers interpret this data as support for the Coase Theorem, even though the point prediction of 100% efficiency was not reached.

However, because of this property of the theory's prediction, almost any statistical test rejects the null hypothesis that the theory is true, that all outcomes are efficient. One way to avoid this problem in experimental design is to test not the point predictions of the theory (e.g., that 100% of the outcomes will be efficient), but rather a *comparative static* prediction (e.g., that more outcomes will be efficient when transaction costs are low than when they are high). Later experiments by these and other researchers follow this approach.

21. KARL R. POPPER, THE LOGIC OF SCIENTIFIC DISCOVERY 32–34 (1968).
22. *Id*. at 40–42.
23. Elizabeth Hoffman & Matthew L. Spitzer, *The Coase Theorem: Some Experimental Tests*, 25 J.L. & ECON. 73, 73–77 (1982).

Often, the point predictions of a theory can be falsified, while the comparative static predictions are supported. For instance, in my previous work with Professor Johnston, we tested a theory of signaling in pretrial bargaining, comparing the impact of different legal regimes.[24] As with the Coase Theorem, this theory predicts that each bargaining pair should reach a particular outcome (a point prediction), but the existence of randomness in individual decisions makes it easy to reject those hypotheses statistically. However, the *comparative statics* of the theory are strongly supported. Under the regime where parties are predicted to go to court, i.e., not to settle, the experimentally observed likelihood of going to court is significantly higher than it is in other regimes where parties are predicted to settle out of court. Comparative-static tests involve comparing the theory's predictions of how various treatments will differ, while point-prediction tests compare outcomes within one treatment to those that the theory predicts.

A second type of experiment can distinguish between competing theories.[25] For example, Professors Coursey, Isaac, and Smith compare Demetz's contestable markets hypothesis with the competing theory of collusive pricing.[26] Demetz's theory predicts that in settings of natural monopoly (decreasing marginal costs; returns to scale) and two firms, instituting an auction can control collusive pricing. The competing theory of collusive pricing predicts that the auction format will not affect the prices set. The experiment compared auctions run with one seller and with two sellers. Results indicate support for the contestable markets hypothesis—the auction with two sellers resulted in significantly lower prices than similar auctions with monopolists. Notice that this experimental design tested the comparative static predictions of the two competing theories. Collusive pricing predicts that auctions with two sellers will have lower prices than auctions with one seller, while contestable markets predicts that the prices will be the same. By comparing the two treatments and seeing how they are different, the authors distinguished between the theories.

A third type of theory-addressing experiment is a *stress test*. When a theory has been supported empirically or in the lab, like the Coase Theorem, a new experiment may be conducted to push the limits of the theory to see if and where it breaks down. Professors Hoffman and Spitzer did just that with the Coase Theorem.[27] They ran bargaining ex-

24. *See generally* Rachel Croson & Jason S. Johnston, *Experimental Results on Bargaining Under Alternative Property Rights Regimes*, 16 J.L. ECON. & ORG. 50 (2000).

25. This distinction is partly captured by Plott, who describes theory first tests (experiments that test theory point predictions) and data first tests (experiments that compare multiple theories). *See generally* Plott, *supra* note 19, 303–05.

26. *See generally* Don L. Coursey et al., *Natural Monopoly and Contested Markets: Some Experimental Results*, 27 J.L. & ECON. 91 (1984).

27. *See generally* Elizabeth Hoffman & Matthew L. Spitzer, *Experimental Tests of the Coase Theorem with Large Bargaining Groups*, 15 J. LEGAL STUD. 149 (1986).

periments with groups of ten people and found continued support for the efficiency predictions of the theorem. They also ran an experiment with limited information where subjects knew their own payoffs, but not that of the other party, and again found support for the efficiency predictions of the theorem.[28] Another test added more realism to the situation by instituting a real property right—the right to avoid drinking a bad tasting, but harmless, substance in the lab—and relaxed the symmetry of the problem, but the outcomes remained largely efficient.[29] Stress tests such as these are not designed to prove or disprove theories, but instead are designed to see how robust a theory is and where its boundaries lie.

Finally, experiments can be used to estimate parameters of theories.[30] For example, much of economic theory suggests that individuals are risk averse—experiments can estimate the degree of risk aversion individuals exhibit.[31]

How to Conduct an Experiment That Addresses Theory

Experiments designed to address theories need to have a high degree of *internal validity*. This requires a deep understanding of the theory in order to construct a lab situation that exactly captures its assumptions. If the experiment is not internally valid, the data produced is flawed. For instance, if the theory a researcher is trying to test assumes that a game is infinitely repeated, the experiment run must implement such a game.[32] Similarly, one cannot test one-shot theories using experimentally repeated interaction. Repeated interactions are often used in experiments to glean more data out of subjects by allowing them to make decisions multiple times, but if subjects are paired with the same counterpart repeatedly then the experiment is no longer a test of the one-shot theory.[33]

Another aspect of internal validity has to do with the payoffs of the subjects. Typically, theories specify the payoffs from taking one action

28. *See generally* Hoffman & Spitzer, *supra* note 23, at 91–95.

29. *See generally* Don L. Coursey et al., *Fear and Loathing in the Coase Theorem: Experimental Tests Involving Physical Discomfort*, 16 J. LEGAL STUD. 217 (1987). *But see* Stewart Schwab, *A Coasean Experiment on Contract Presumptions*, 17 J. LEGAL STUD. 237 (1988) (adding different types of realism to the Coase Theorem and finding less efficiency than in previous experiments).

30. Plott refers to this category as measurement experiments. *See* Plott, *supra* note 19, at 124–26.

31. *See*, *e.g.*, Steven J. Kachelmeier & Mohamed Shehata, *Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence from the People's Republic of China*, 82 AM. ECON. REV. 1120, 1129–31 (1992); Haim Levy, A*bsolute and Relative Risk Aversion: An Experimental Study*, 8 J. Risk & Uncertainty 289, 295–300 (1994).

32. The way to implement an infinitely repeated game in the lab is to derive the theory's predictions under a discount rate of δ, then implement the discount rate by making the probability of the game ending in any given period equal to δ.

33. The way to implement a one-shot game is through a strangers, or 'zipper,' design where each subject meets each other subject once, at the most, during the experiment. *See*, *e.g.*, Ulrich Kamecke, Note, *Rotations: Matching Schemes That Efficiently Preserve the Best Reply Structure of a One Shot Game*, 26 INT'L J. GAME THEORY 409 (1997) (illustrating that an absence of repercussion is necessary to achieve the best reply structure).

or another. It is critical for theory testing that the subjects actually face the payoffs assumed by the theory. This has led to the practice of *induced valuation* in experimental economics, where a subject's compensation is responsive to the choices they make that are consistent with the theory being tested. This practice replaces the flat fee payments more common in psychology experiments.[34] In an experiment on pretrial bargaining, for example, the subject would be faced with the choice of going to court and earning an amount determined by a lottery with known odds, or settling and earning the amount agreed on by the parties. In this setting, it is critical that the financial decision being faced by the subject is the same as assumed by the theory.[35] Typically, payments are made in cash directly after the experiment, however, other forms of payment can be used provided they are consistent with the theory being tested.[36]

A second methodological issue has to do with subject pool selection. Most theories are designed to describe behavior generally, thus, no special selection is required. However, occasionally, a theory may be tested that distinguishes between types of subjects based on gender, expertise, etc. A subject pool designed to test these types of theories should be chosen to adequately test the theory.

A final methodological issue in experiments that test theory has to do with the level of context in the experimental instructions. The level of context can, of course, vary. Experimental economists, generally, prefer very little context when they are testing theories for three reasons. First, the theory being tested often does not rely on context, so the experiments should not either. Second, context often adds variance to the data. For example, if some subjects think that going to court is a good thing and others think it is a bad thing, then describing the experimental decision as 'going to court' as opposed to 'choosing option A' could change an individual's decisions. These changes might not affect the average or aggregate decision, but it can impact the variance of those decisions, reducing the likelihood of detecting statistically significant differences between treatments of the experiment. Finally, and most importantly, context can add systematic bias or demand effects.[37] For example, if subjects want to be seen as kind, gentle types by their professor, then describing

34. *See* Smith, *supra* note 18, at 275.

35. One criticism of experiments often involves the size of the payments made. In particular, the concern is that no experimenter will be able to match the size of financial payments seen by a typical plaintiff. In the realm of theory testing, however, this criticism is not as pressing. The theory does not specify that its predictions hold only when the dollar amounts under discussion are large, and if the theory does not work in the carefully controlled lab setting it is unlikely to work in the field. *See* Plott, *supra* note 19, at 296–97.

36. *See* Coursey et al, *supra* note 29, at 219–21 (paying subjects in dollars, after actually consuming the bitter-tasting drink); *see also* Rebecca R. Boyce et al., *An Experimental Examination of Intrinsic Values as a Source of the WTA-WTP Disparity*, 82 AM. ECON. REV. 1366, 1367 (1992) (mandating subjects either had to chop down or pay to preserve a baby tree in the lab); Daniel Kahneman et al., *Experimental Tests of the Endowment Effect and the Coase Theorem*, 98 J. POL. ECON. 1325, 1330–32 (1990) (giving incentives such as pens and mugs).

37. *See* MCGUIGAN, *supra* note 11, at 288–89.

the decision in terms of going to court might reduce everyone's likelihood of choosing that option. This would change the responses in a systematic way. Such systematic changes in the data will significantly change the conclusions reached, thus context should be avoided in these types of experiments. In theory testing experiments, there are only low costs associated with avoiding context.

B. Experiments That Investigate Anomalies

A second category of experiments investigate anomalies. An anomaly is an observed regularity that is not consistent with or predicted by current models. Professor Richard Thaler writes, "An experimental result qualifies as an anomaly if it is difficult to 'rationalize' or if implausible assumptions are necessary to explain it within the [self-interested, rational] paradigm."[38] These experiments demonstrate the importance of factors not captured by current models and, therefore, have a place in the more advanced stages of the scientific dialectic.

Professors Hoffman and Spitzer conducted an experiment to investigate an anomaly that occurred in their research of the Coase Theorem.[39] While the original tests and the stress tests both supported the efficiency predictions of the Coase Theorem, the distributional predictions of the theory were not supported by the data. In particular, subjects often agreed to negotiated outcomes that, while efficient, resulted in them earning less than they could have received if they refused to negotiate.

First, the researchers proposed an explanation for the anomaly. The explanation centers on the procedure they used in allocating the property right in the experiment: a coin flip. They hypothesized:

> that subjects behave in accord with a theory of distributive justice that says that flipping a coin is not a just way of allocating unequal property entitlements. Subjects perceived no morally justified difference between themselves, even though one 'legally' owned a substantial property entitlement and the other did not. Because they were 'morally equal' an equal split seemed to be the only fair allocation.[40]

38. This definition of anomalies and an excellent discussion of various specific ones can be found in RICHARD H. THALER, THE WINNER'S CURSE: PARADOXES AND ANOMALIES OF ECONOMIC LIFE 2 (1992) (this book collects many of Thaler's ongoing column in the *Journal of Economic Perspectives*).

39. *See generally* Elizabeth Hoffman & Matthew L. Spitzer, *Entitlements, Rights and Fairness: Some Experimental Evidence of Subjects' Concepts of Distributive Justice*, 14 J. LEGAL STUD. 259 (1985). A number of other experiments have been motivated by this same lab anomaly. *See*, *e.g.*, Glenn W. Harrison et al., *Coasian Solutions to the Externality Problem in Experimental Markets*, 97 ECON. J. 388, 388 (1987); Glenn W. Harrison & Michael McKee, *Experimental Evaluation of the Coase Theorem*, 28 J.L. & ECON. 653, 653–55 (1985).

40. *See* Hoffman & Spitzer, *supra* note 39, at 260.

Then, the researchers develop and discuss three competing theories of distributive justice that subjects might use: self-interested or utilitarian; egalitarian or an equal split; and natural law/desert or Lockian.[41]

Next, they design an experiment to test their explanation and to distinguish between these lay theories of distributive justice. In one treatment of the experiment, they replicate their previous data: Coasian bargaining with the property rights decided by a coin flip. In another treatment, they assign the property right according to a game of skill—Nim. A second manipulation, done independently, changes only the wording of the instructions to indicate that the subject had "earned" the property right rather than having been allocated it. If subjects' theories are self-interested, then those with property rights should demand more no matter what the treatment. If subjects' theories are egalitarian, then they should divide the money evenly no matter what the treatment. Only if the theories are Lockian should the behavior differ between the treatments. The results supported the Lockian theories. Subjects in the Nim and the earned treatments divided the money more unequally than those in the baseline treatment.[42] Again, notice that this experiment tests a comparative static prediction, comparing outcomes from two treatments under each theory rather than testing point predictions of what is observed in one treatment to what would be predicted by each theory.

This experiment was prompted by an anomalous observation from the lab—equal splits in bargaining settings—that was inconsistent with current theory. The researchers designed an experiment to replicate the effect and developed an explanation. The experiment then tested and, in this case, supported, their explanation.

This type of anomaly testing has occurred in the legal context as well. For example, legal theory assumes that all parties to a dispute share the same beliefs regarding probabilities and outcomes of a trial. These common beliefs, combined with risk aversion or costs of litigation, spur parties to settle before trial. However, not all lawsuits are settled out of court. This observation led Professor Babcock and colleagues to question the common belief assumption.[43]

They constructed an experiment in which individuals were bargaining in expectation of a trial.[44] Subjects were assigned the role of either plaintiff or defendant in a court case. All subjects read the same materials regarding the case. Before entering into pretrial negotiations, subjects estimated the likelihood that they would win if they went to court and the average judgment they could expect. Then the pairs of subjects negotiated the issues and tried to come to a settlement. Subjects who did

41. *Id.* at 261.
42. *Id.* at 280–83.
43. This work is reviewed in Linda Babcock & George Loewenstein, *Explaining Bargaining Impasse: The Role of Self-Serving Biases*, 11 J. ECON. PERSP. 109 (1997).
44. *Id.* at 111–16.

not settle "went to court" and were told of the decision made by a retired judge.

The results were inconsistent with the theoretical assumption of common beliefs. Subjects in the role of plaintiffs believed that they were significantly more likely to win at trial than those in the role of the defendants. Additionally, the plaintiff subjects believed they would be awarded a significantly larger amount than those in the defendant role. The experiment also demonstrated that this divergence of beliefs caused the lack of pretrial settlement. Pairs whose beliefs diverged the most were the least likely to settle.[45]

This experiment was prompted by an anomalous observation from the field—lack of pretrial settlement—that was inconsistent with current theory. The researchers designed an experiment to investigate the anomaly, to demonstrate its existence in a particular setting, and to examine its strength and dimensions. As a result of the experiments, the researchers were able to suggest some ways the current theory should be adapted to better explain the empirical and experimental results.

How to Conduct Experiments That Investigate Anomalies

The methods for experiments that investigate anomalies are related to, but are slightly different than those for other types of experiments. Like theory testing experiments, these experiments must have a high degree of internal validity. The conditions in the experiment should be such that the traditional theory can make a behavioral prediction. However, the experiment should be designed to create the anomaly as well.

Methods of experimentation differ based on what has prompted the inquiry. If the experiment was motivated by lab results, the previous lab experiment should be replicated to demonstrate that the anomalous results were not due to procedures or chance.[46] The experiment should involve little context and focus on demonstrating the implications and boundaries of the anomaly, along with what causes the anomaly to appear and disappear. If the experiment was motivated by field results, the field situations need to be implemented in the lab to see if the anomaly remains, and what causes it to appear and disappear. These experiments

45. Other related works include Linda Babcock et al., *Biased Judgments of Fairness in Bargaining*, 85 AM. ECON. REV. 1337 (1995) [hereinafter Babcock et al., *Biased*]; Linda Babcock et al., *Choosing the Wrong Pond: Social Comparisons in Negotiations That Reflect a Self-Serving Bias*, 111 Q. J. ECON. 1 (1996); George Loewenstein et al., *Self-Serving Assessments of Fairness and Pretrial Bargaining*, 22 J. LEGAL STUD. 135 (1993). Other experimental research on pretrial bargaining can be found in Don L. Coursey & Linda R. Stanley, *Pretrial Bargaining Behavior Within the Shadow of the Law: Theory and Experimental Evidence*, 8 INT'L REV. L. & ECON. 161 (1988); Linda R. Stanley & Don L. Coursey, *Empirical Evidence on the Selection Hypothesis and the Decision to Litigate or Settle*, 19 J. LEGAL STUD. 145 (1990).

46. *See, e.g.*, Hoffman & Spitzer, *supra* note 39, at 289–90.

tend to involve more context and focus on demonstrating the existence of the anomaly under controlled conditions.[47]

Once the anomaly is demonstrated, researchers engage in traditional theory testing to investigate why it is observed. In the Coase Theorem example, the researchers manipulated the source of the property right to investigate why behavior was different than predicted and what theory of distributive justice best predicts outcomes.[48] In the pretrial bargaining example, the researchers carefully ruled out alternative explanations for their results by providing subjects with identical information about the facts of the case.[49] The idea is to generate an explanation for the anomaly, and then engage in theory testing to demonstrate why and when it is observed.

As with experiments that test theory, induced valuation is critical in these experiments. Subjects may exhibit anomalous behavior when it costs little to do so, but behave quite differently when money is on the line. Similarly, the researcher must be careful to avoid *demand effects*—avoid suggesting the desired results to the subjects either explicitly or implicitly.

A postexperimental questionnaire that asks the subjects to describe what they did and why they did it is helpful in these types of experiments. These questionnaires, although hypothetical, often suggest new models to test and new experimental treatments to run.

Finally, these types of experiments suggest changes to existing theory that incorporate the anomalous result. In the Coase Theorem experiments, the results suggest that both property rights, and the manner in which they are acquired, matter. The results further suggest that these factors should be incorporated into models of legal bargaining and property rights regimes. In the pretrial bargaining experiment, the results suggest that pretrial beliefs of competing parties diverge substantially, even when they have the same information. Theories about pretrial settlement need to incorporate these divergent beliefs when predicting outcomes.[50]

These experiments not only suggest where current theory needs to be amended, but also where new theory needs to be developed. For example, the pretrial bargaining experiments suggest the need for a new theory explaining exactly how individuals form their expectations of what will happen in court.[51] This theory was not believed to be necessary when it was thought that all individuals believed the same thing about the distribution of possible outcomes.

47. *See, e.g.*, Babcock & Loewenstein, *supra* note 43, at 116–17.
48. *See* Hoffman & Spitzer, *supra* note 39, at 262–64.
49. *See* Babcock et al., *Biased*, *supra* note 45, at 1342.
50. *See id.* at 1341–42.
51. *See id.* at 1342.

C. Experiments That Testbed Policies

The final type of experiments involve those that testbed policies. As mentioned in the introduction, these experiments can provide a middle step between theorizing and implementation. When new policies are being considered, an experiment can be run to testbed the policy, investigate and illuminate any unintended consequences, and suggest parameters that policymakers might consider in their final implementation.

This type of experiment was used by Professors Hong and Plott to study a proposed policy change by the Interstate Commerce Commission (ICC).[52] The policy would require firms engaged in domestic, dry bulk commodity transportation on inland waterways to file proposed rate changes with the ICC fifteen days prior to their implementation.[53] The policy was described as a way to encourage competition, but opponents suggested that it might aid collusion among firms by signaling prices instead.

The experiment was designed to capture important elements of the real world situation, specifically grain traffic along the upper Mississippi and Illinois waterways. Parameters were chosen to represent the number and capacities of buyers and sellers in this market, the demand and supply elasticities, and the cyclical nature of the demand. Subjects in the experiment were placed in separate offices with their own phone, provided a phone directory of the other subjects playing the various roles, and told to negotiate prices.[54] Prices were negotiated via phone, as is common in this industry.

The researchers compared prices, volume, and efficiency in the markets under three conditions: status quo, with no information sharing and independently negotiated prices; the proposed policy, where the sellers prices are distributed to everyone in advance of negotiations; and an intermediate treatment, where the posted prices are distributed to buyers, but not the sellers. Results of the experiment suggested that the proposed policy increased prices, lowered volumes, and resulted in less efficient outcomes.[55] The policy also hurt small sellers and benefited large ones. This experiment concluded by recommending against the policy change.[56]

52. *See generally* James T. Hong & Charles R. Plott, *Rate Filing Policies for Inland Water Transportation: An Experimental Approach*, 13 BELL J. ECON. 1 (1982). Other experiments that explicitly testbed policies include David M. Grether et al., *The Allocation of Landing Rights by Unanimity Among Competitors*, 71 AM. ECON. REV. 166 (1981); David M. Grether & Charles R. Plott, *The Effects of Market Practices in Oligopolistic Markets: An Experimental Examination of the Ethyl Case*, 22 ECON. INQUIRY 479 (1984); Kemal Güler et al., *A Study of Zero-Out Auctions: Testbed Experiments of a Process of Allocating Private Rights to the Use of Public Property*, 4 ECON. THEORY 67 (1994).

53. *See* Hong & Plott, *supra* note 52, at 1.

54. *See id*. at 3.

55. *See id.* at 9.

56. *See id.* at 13.

Two other policy-relevant experiments by Professors Babcock and Pogarsky examine the impact of damage caps on lawsuits.[57] The experiments give subjects details on a personal injury case and allow them to bargain pretrial.[58] The treatments are distinguished by the existence and level of damage caps that subjects are told apply to their case. The results showed that when the damage cap is low relative to the amounts involved in the case, its existence increases the likelihood of settlement and lowers the settlement amounts.[59] On the other hand, when the damage cap is high relative to the amounts involved in the case, its existence reduces settlement rates and increases settlement amounts.[60] Conclusions from these experiments suggested that the impact of damage caps will be driven by the level at which the caps are set.

How to Conduct Experiments That Testbed Policies

Experiments that testbed policies focus on *external validity* rather than *internal validity*. The critical factor is to design experiments that appropriately capture the situation as it occurs in reality.[61] Of course, no lab situation can capture every nuance of a real world situation, so judgment is required to identify the most important aspects of the situation and of the policy being investigated. In the grain-shipping experiment, institutional details like the number and size of the buyers and sellers, the communication method used for the negotiations, and the cyclical nature of demand were all captured in the design.[62]

As with theory-testing, these types of experiments often compare multiple policies (comparative statics) rather than simply testing one policy proposal (point predictions). Often the policy question itself involves multiple options or a decision on whether to depart from the status quo. As with theory testing experiments, it is typically easier to conclude that one policy is more efficient than another, a comparative static prediction, than to conclude that a given policy is efficient in isolation, a point prediction. For example, in the damage cap experiments, the policy of having no cap is compared to the policy of high or low caps.[63]

The importance of external validity in these experiments leads to slightly different methods than in previous types. First, experiments that

57. Linda Babcock & Greg Pogarsky, *Damage Caps and Settlement: A Behavioral Approach*, 28 J. LEGAL STUD. 341 (1999) [hereinafter Babcock 1999]; Greg Pogarsky & Linda Babcock, *Damage Caps, Motivated Anchoring, and Bargaining Impasse*, 30 J. LEGAL STUD. 143 (2001) [hereinafter Pogarsky 2001].

58. *See* Babcock 1999, *supra* note 57, at 359–62; Pogarsky 2001, *supra* note 57, at 150–53.

59. *See* Babcock 1999, *supra* note 57, at 362–63.

60. *See* Pogarsky 2001, *supra* note 57, at 153–58.

61. *See* Charles R. Plott, *Market Architectures, Institutional Landscapes and Testbed Experiments*, 4 ECON. THEORY 3, 3–10 (1994) (additional methodological discussion of policy-testing experiments).

62. *See* Hong & Plott, *supra* note 52, at 6–9.

63. *See* Babcock 1999, *supra* note 57, at 362–63; Pogarsky 2001, *supra* note 57, at 153–58.

testbed policy often involve context. Subjects are not told that they are manufacturing a widget, but that they are setting a price for grain transportation. The previous warning about how context can affect decisions still applies, but because the context affects real world decisions and the goal of the experiment is to explain or investigate those decisions, context is necessary in these experiments.

For that same reason, in these experiments, researchers are more likely to use professionals in the field as subjects.[64] There are two reasons for this choice. First, if professionals in this area have particular sets of experiences or biases that have developed in the field, and these will exhibit themselves once the policy is in place, thus you want those experiences also to be exhibited in the lab. The second reason has to do with the increased context of these experiments. As context increases, experiments get more complicated. If the researcher used student subjects, they might never gain enough experience to understand the setting and make intelligent decisions. In contrast, professionals trained in the area can rely on their previous experience to quickly grasp the situation and make decisions based on an understanding of the policy being proposed.[65]

Once the researcher moves from students to professionals, the incentives used in the experiment become more important and more difficult. An undergraduate student can be induced to think hard about a problem if the difference between making the right decision and the wrong one is around twenty dollars. A professional whose income and opportunity costs of time are higher may require significantly more money to participate. More troublesome is that high earning professions may not be motivated by money at all, at least not on the scale most experimentalists can pay. For these subjects, participation can sometimes be induced in exchange for a learning experience—the researcher promises to share the results from the study and perhaps offer free consulting on how to improve the subject's own decision making. Appealing to either professional pride or their desire to discover the answer are alternative tactics used to motivate professionals.

64. *See, e.g.*, Douglas Dyer et al., *A Comparison of Naive and Experienced Bidders in Common Value Offer Auctions: A Laboratory Analysis*, 99 ECON. J. 108–15 (1989) [hereinafter *Comparison*] (comparing bidding behavior of experienced business executives in construction with that of student subjects in a lab experiment); *see also* Douglas Dyer & John H. Kagel, *Bidding in Common Value Auctions: How the Commercial Construction Industry Corrects for the Winner's Curse*, 42 MGMT. SCI. 1463 (1996) [hereinafter *Bidding*]; Charles R. Plott & David P. Porter, *Market Architectures and Institutional Testbedding: An Experiment with Space Station Pricing Policies*, 31 J. ECON. BEHAV. & ORG. 237, 249–50 (1996) (using NASA administrators with the responsibility of developing and implementing station resource allocation policies as subjects).

65. There are also risks to using professionals as subjects. For example, professionals may bring to the experiment incentives and institutions that exist in the world but not in the experimental design, they may have opinions about the implications of the experiment and try to influence the results to impact policy, and they may not prepare for the experiment as carefully as student participants (e.g., they may only skim instructions). I thank Charles Plott for suggesting these arguments.

D. Other Experiments

While the categorization above spans most types of research experiments, there are experiments designed for other purposes. The largest class of these are experiments designed for pedagogical purposes. These experiments are explicitly designed and selected to illustrate a theoretical or empirical result clearly and simply. For example, experiments designed for use in the classroom have often been pretested to ensure consistent, robust results. In this sense, they are very unlike research experiments, where the whole point is that the experimenter does not know what will happen.

These experiments can be extremely useful. Rather than having students simply listen to lectures, they actively participate in experiments and discuss the results. Thus, experiments have the potential to transform a passive learning experience into an active one. Economics textbooks, complete with experimental instructions, lab reports for students to turn in, and discussion questions, have been written explicitly to achieve this purpose.[66]

Other experiments are not run only as academic research, but for private purposes. Companies and other organizations often sponsor lab and field experiments to investigate changes they are considering implementing.[67]

Experiments take many shapes and forms, the major categories reviewed in this article are those that address theory, those that investigate anomalies, and those that testbed policies. The methodologies used are both general in nature and specific to the purpose of the experiment. Experiments that test theory tend to involve abstract situations and student subjects, while those that testbed policies include context and use professionals. The next section gives some general advice regarding experiment design, implementation, and presentation for those who want to conduct experiments themselves.

IV. RUNNING EXPERIMENTS

Learning a new methodology is always a challenging task. In addition to the explicit advice one can find about running experiments,[68] each

66. *See, e.g.*, THEODORE C. BERGSTROM & JOHN H. MILLER, EXPERIMENTS WITH ECONOMIC PRINCIPLES (1997); *see also* Denise Hazlett, *An Experimental Education Market with Positive Externalities*, 31 J. ECON. EDUC. 44 (2000); Charles A. Holt, *Teaching Economics with Classroom Experiments: A Symposium*, 65 S. ECON. J. 603 (1999); Maureen A. Kilkenny, *A Classroom Experiment About Tradable Permits*, 22 REV. AGRIC. ECON. 586 (2000); Jane N. Leuthold, *A Public Goods Experiment for the Classroom*, 18 J. ECON. EDUC. 58 (1987).

67. *See, e.g.*, *Interfaces*, INST. MGMT. SCI. (Gary Bolton & Anthony Kwasnica eds., forthcoming 2002) (special issue devoted to economics experiments that were designated to apply to corporate decisions).

68. The best book on experimental economics methodology is DANIEL FRIEDMAN & SHYAM SUNDER, EXPERIMENTAL METHODS: A PRIMER FOR ECONOMISTS (1994).

field has a body of implicit "rules of the game." In this section, I discuss a few of these rules, focusing on mistakes that novice experimenters tend to make.

A. *Experimental Design and Preparation*

By far the most common mistake made by novice researchers in any area is to bite off more than they can chew. Graduate school is essentially training in finding manageable research questions. In experimental economics, this mistake reveals itself in extremely complicated experimental designs.

A design can be complicated on a number of different dimensions. First, within a given experimental condition the experiment can be overly complicated. To run a successful experiment, the researcher must identify the important features of the theory or empirical situation and include them alone. This is particularly challenging for experiments that inform policy, where the researcher is tempted to incorporate more and more institutional detail in a quest for external validity. However, extraneous information will only confuse the subjects and lead to uninformative results. Instructions should be at most two to three pages long and should be pretested for comprehension. This usually means giving the instructions to students, colleagues, or other readers who are unfamiliar with the research project, having them read the instructions, and asking them questions about the task. Researchers particularly want to ensure that subjects know what decision they should be making, how they should make that decision, where and how they should record it, and the various possible outcomes resulting from that decision. If the pretesters cannot answer those basic questions, the instructions need to be simplified.

Many experimenters include a quiz that presents subjects with some hypothetical decisions, and asks them to calculate the outcomes that would result. This quiz method can be quite useful, but it contains the danger of creating a demand effect in which subjects interpret the examples presented in the quiz as suggested behavior. Thus, it is important, if the instructions include quizzes, to make them as unbiased as possible. This can be done by using different scales in the quiz and in the experiment, e.g., if in the experiment subjects choose quantities to produce between one and ten, use examples like twenty-five for the quiz; by making the quiz abstract rather than concrete, e.g., tell subjects they choose quantity X and their counterpart chooses quantity Y, then ask them for the expression describing their profits; or by 'balancing' the quiz to present the spectrum of possible decisions.

A second complexity dimension has to do with the number of treatments the experimental design includes. Veteran experimenters tend toward designs with between three and six treatments, with four being a common choice. Experimental designs with more than six treatments are often better presented as multiple studies. No one experiment can investigate everything, manipulate every possible cause of behavior, or test for all parameters, so a researcher should not design one that tries to encompass everything. Subject time and attention is limited; experimental designs with too many treatments inevitably suffer from too few observations in each. A rough rule of thumb is to aim for twenty to thirty independent observations for each treatment in the experiment. Thirty observations is generally enough to run parametric rather than nonparametric statistical tests on the outcomes.

This leads to the related question of confounding designs. Researchers want to set up experimental designs so that each treatment changes one factor at a time. If treatments differ on two dimensions, and indeed result in different behavior, then the researcher cannot identify what dimension is causing the change in behavior. Therefore, it is important to compare treatments with only one changed element to determine what causes behavioral changes.

A final experimental design issue to be addressed is whether to run a within- or between-subject design. In a within-subject design, the same subject experiences more than one treatment in the experiment. In a between-subject design, each subject engages in only one treatment. Within-subject designs collect more information, because the researcher gets multiple observations from each subject. This gives the researcher the ability to compare behavior within an individual and to use stronger statistical or paired tests.

However, these designs have dangers. Primarily, the outcomes in one treatment may affect decisions in the next. Subjects may feel they must change their decision when faced with a different treatment, or they must remain consistent. The solution to these problems is to *counterbalance* the order of treatments shown to the subjects in the experiment, so some of the subjects see one treatment first and others see a different one. However, counterbalancing provides its own challenges, particularly in experiments where groups of people interact. The alternative is a between-subject design where each subject faces only one treatment. This is the method generally preferred by experimental economists, although important exceptions exist.

B. Experimental Implementation

Consistent with the idea that only one element should be changed between treatments in the experimental design is the notion that it is important to randomly assign subjects to treatments. For instance, if a researcher is using students as subjects, do not assign the students in the morning class to one treatment and those in the afternoon class to a separate treatment. Systematic differences in the type of student who enrolls in the morning class compared to those who enroll in the afternoon class may generate a spurious difference between the treatments and lead to an erroneous conclusion. The researcher may conclude that the manipulated factors changed the behavior when, in fact, it was simply the differences in the subject pools. Assigning subjects randomly is a critical way to control for unknown individual differences in the subject pool.

One of the most stringent rules in experimental economics is that the researcher may not deceive the subjects of an experiment. This prohibition on deception includes deception about the purpose of the experiment, the payoffs the subjects will earn, or the characterization of the subjects' counterparts. In economics, the validity of the experiment rests on the link between behavior and expected payoffs. If subjects are deceived about that link, the validity of their decisions is called into doubt. A second reason deception is disfavored has to do with the public-goods nature of trust in the experimenter. For instance, if subjects are routinely deceived in experiments by being told they will take home their earnings in the game, and then actually receiving a five dollar flat fee for their participation, they will begin to distrust the experimenter. This lack of trust could lead the subjects to change their behavior in the future. This strong antideception philosophy is one of the distinguishing characteristics of experimental economics, and will be discussed again below.[69]

Similarly, it is important to respect subjects' privacy. Most human subjects committees at universities require the researcher to report only aggregated data and, if an individual subject's decisions must be reported separately, they must be anonymized. This is important not just in the publication but also in the experimental implementation. Never identify a subject during an experiment, and under no circumstances should the researcher discuss one subject's performance or decisions with other subjects. This last rule is sometimes difficult to follow at the end of the experiment, when many subjects want to know how they did relative to the others.

Another concern about experimental instructions and implementation is avoiding demand effects. A demand effect is when a subject acts in a particular way to please the experimenter.[70] The experimenter is in a position of authority relative to the subject, and sometimes is even their professor, so many students want to do as the experimenter wishes. It is critical when writing instructions, designing quizzes, and answering questions to avoid leading your subjects to the answer you want them to provide. The fact that people will act in a particular manner when prompted is not of much interest to researchers—the point of running experiments is to see what individuals do when unprompted.

69. *See infra* Part IV.E.
70. *See* MCGUIGAN, *supra* note 11, at 287–91.

C. Analyzing and Presenting Data

Once the experiment is run and the data has been recorded, it is time for analysis. First, it is important to always report *all* the data collected, even if it does not support the theory or researcher's personal predisposition. It is through surprises and anomalistic results that the field moves forward. This is not to say that no data will ever be excluded. If there is independent evidence that a subject in the experiment was confused, e.g., if they did not pass the quiz included in the instructions; or that an entire experimental session needs to be excluded, e.g., there was inappropriate communication, collusion among the subjects, or a mistake was made in the experimental implementation, like incorrect feedback given to the subjects; then the data should clearly be excluded. However, the decision to exclude data in these cases should be made *before* the data is analyzed, as soon as the mistake is discovered. Excluding data because the researcher does not like the results is particularly troubling.

Second, do not data mine. Data mining is a troublesome procedure when a researcher runs many statistical tests and reports only those that produce significant results. As when excluding data, the researcher has to exercise judgment regarding which tests to report. It is entirely natural to run some first level analyses, then to refine them as the data start to speak. However, running 100 tests, and reporting five that are significant at the five percent level leads to conclusions that are no more likely than chance to be correct. This is a problem not just in experiments, but in all empirical research. One solution is to report all the tests that were run and let the reader judge for himself. Other statistical tests adjust significance levels to account for multiple comparisons.

Third, keep the data analysis and presentation clean. One joy of running experiments is that data is much cleaner than the data one gets from empirical data, since experiments directly manipulate the effect of one variable. Such data should demonstrate the impact of that manipulation directly without need for complex analysis, transformed variables, and other tricks of the econometrics trade. Many experimental economists apply the "inter-ocular trauma test" to their data; the differences between treatments should jump up and hit the reader in the eye. To facilitate this test, almost all experimental papers include figures as well as tables for the presentation of their data.

Researchers may need to analyze their data using nonparametric statistics.[71] Although the data generated by experiments is often cleaner than empirical data, there is also less of it. Because data sets from experiments tend to be small and non-normally distributed, researchers often use nonparametric statistics to analyze their data. These analyses do

71. An extremely useful book on nonparametric statistics is SIDNEY SIEGEL, NONPARAMETRIC STATISTICS FOR THE BEHAVIORAL SCIENCES (1956).

not rely on properties of the distribution of the data, and are thus appropriate for experimental data.

A final point has to do with the difference between ex ante and ex post explanations of the experiments' results. Ideally, researchers should develop hypotheses of what will occur in the lab before the experiment is run. The hypotheses will either be supported or not by the data. Theorizing that happens as a *result* of looking at the data is ex post, and should not be mistaken for hypotheses generation. Instead, it should be identified as ex post theorizing and belongs in the discussion or conclusion section, not in the introduction and motivation sections.

D. Presenting Experimental Work

Now that the data analysis is done, hypotheses are tested, and results written, it is time to present the experimental work. Whenever experiments are presented, particularly to audiences unfamiliar with the methodology, a set of stock questions/challenges routinely appear. These questions occur particularly when the experiment does not support the theory's (or the researcher's) predictions.

The first objection involves external validity. The questioner might object that this research is just a game, it is not the real world, and thus the experiment cannot tell anything about reality. There are multiple possible responses to this objection, depending on the type of experiment the researcher is presenting. The simplest is the response for a theory testing experiment. The theory being tested typically does not specify that it only holds true in particular circumstances. The theory predicts behavior whenever its assumptions are met, whether in the laboratory or the real world.[72] A similar answer can be used for experiments that investigate anomalies motivated by previous lab results. For these experiments, the objective is to explain behavior previously observed in the lab—where better than the lab to investigate this question?

For experiments investigating field anomalies and for those addressing policy questions, the answer is more methodological. For these topics, the lab provides a controlled environment to untangle competing explanations or predictions of behavior. The goal of these experiments is to provide a controlled test of the anomaly/policy under discussion. Certainly if it is observed/successful in the lab, it may (or may not) also be observed/successful in the real world, but if it does not work in the lab, it is unlikely to work in reality.[73]

A second standard objection has to do with the subject pool. Challengers often argue that the subjects in many experiments are students, and do not make decisions in the same way as "real people." There are a number of responses to this objection as well. First, students *are* real

72. *See* Plott, *supra* note 7, at 1521.
73. *Id*. at 1520.

people, only younger. There is no reason to think, in general, that students make decisions differently than professionals—in fact, many studies demonstrate that, if anything, students are less biased than professionals.[74] This is highlighted by the fact that the students of today are the professionals of tomorrow. A second response relies on the experimental design. Ideally, the experiment was not designed to rely on any institution-specific knowledge of the situation at hand, so a student's lack of knowledge would not adversely affect their decisions.

A final objection to experimental work has to do with the size of the payoffs. Nonexperimentalists often argue that the stakes in the experiment are too small, normally ten to twenty dollars, to induce optimal behavior. There are two possible rebuttals to this objection. First, empirically, many experiments have been run with both high and low stakes and comparing the data indicates very few differences in behavior.[75] Second, and perhaps more troubling, is that it is not always obvious that increasing the stakes will increase the likelihood of equilibrium behavior. If true, this suggests that increasing the size of the stakes will not necessarily bring behavior closer to predicted behavior.[76]

One final word about presenting experimental work—it is positively infectious. Each audience member will inevitably have a suggestion about how the researcher should have examined this treatment or tested for that effect. I routinely leave each presentation with five or six suggestions for other experiments I should have run. Some of these suggestions are good, others have one of the methodological problems mentioned above. Treat them not as criticisms (even though they are often presented that way), but as constructive suggestions for future work.

E. Psychology and Economics Compared

Before concluding, I would like to spend a moment comparing psychology experiments and economics experiments. Much of the method-

74. *See Comparison*, *supra* note 64, at 113–15; *Bidding*, *supra* note 64.

75. *See, e.g.*, Jane Beattie & Graham Loomes, *The Impact of Incentives upon Risky Choice Experiments*, 14 J. RISK & UNCERTAINTY 155 (1997); Colin F. Camerer & Robin M. Hogarth, *The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework*, 19 J. RISK & UNCERTAINTY 7 (1999); Vernon L. Smith & James M. Walker, *Monetary Rewards and Decision Cost in Experimental Economics*, 31 ECON. INQUIRY 245 (1993).

76. For example, imagine an ultimatum game where one player offers another some amount of a pie of fixed size. The second player can accept or reject the offer. If the second player rejects, both players earn nothing, but if the second player accepts, the pie is divided as proposed. The subgame-perfect equilibrium of this game is for the first player to offer one penny and for the second player to accept—the earnings of one penny if the second player accepts is more than the earnings of zero if they reject—regardless of the size of the pie. Now imagine a subject's likely behavior, either as the first or second player, if the pie is ten dollars or if the pie is 100 dollars. While we observe behavior that deviates from the equilibrium prediction when the pie is ten dollars, this deviation is not alleviated when the pie is 100 dollars. Intuitively, offering one penny out of a pie of 100 dollars is even *more* unlikely to be observed than offering one penny out of a pie of ten dollars. Increasing the stakes does not thus move behavior closer to equilibrium. *See* Elizabeth Hoffman et al., *On Expectations and the Monetary Stakes in Ultimatum Games*, 25 INT'L J. GAME THEORY 289, 289 (1996).

ology of experimental economics comes from experimental psychology, and many of the procedures used and the questions asked by the two fields are similar. For example, both fields often use undergraduate students as subjects, both elicit decisions from the subjects, and use those decisions to learn about the world. However, the *objectives* of psychology and economics experiments are often different. Economics experiments are designed to address economic theories; psychology experiments are designed to address psychological theories. This distinction may seem obvious at first, but it has important and often unforeseen implications for methodological differences between the two fields. Those implications arise from the differences in theories being tested.

Economic theories are primarily about choices. Individual behavior, market prices and quantities, and policy implications are the focal points. As a result, economists are interested in experimental decisions—what subjects in the experiment choose when faced with a situation that captures either the assumptions of a model or a real world situation. In contrast, psychological theories are primarily about internal states like thought processes (cognitive psychology) or attitudes about social interactions (social psychology). Thus, psychologists are interested in the internal states subjects experience when faced with a situation. This is not to say that economists are not interested in internal states, or that psychologists are not interested in choices, just that the two disciplines highlight different things.

Because economists are interested in choices, it is critical to economics experiments that subjects are paid according to their decisions.[77] In experimental economics, actions speak much louder than words. What subjects claim they are doing may be very different than what they actually do. Since psychologists are interested in eliciting and understanding internal states, subjects can be paid a flat fee or nothing at all, and are often asked to introspect about their motives and reasoning.[78] Thus, the differences in methodologies of how subjects are paid flows directly from what the researchers in question are studying.

A second difference has to do with deception of subjects. For example, psychology experiments that attempt to differentiate psychological theories about reasoning compare thought processes when subjects believe one thing or another. Thus, the practice of deceiving the subjects about the state of the world, the existence of other subjects, or other features of the experiment is a valuable tool that can be used to differentiate the theories. In contrast, experiments that differentiate economic theories do not need deception to answer the questions they are asking. This deception issue is one of great concern to both camps.[79] The use of de-

77. *See* Smith, *supra* note 18, at 275.
78. *See* MCGUIGAN, *supra* note 11, at 276.
79. For an ongoing discussion of deception in the literature, see Shane Bonetti, *Experimental Economics and Deception*, 19 J. ECON. PSYCHOL. 377 (1998); Shane Bonetti, *Response: Reply to Hey*

ception imposes a negative externality on other researchers—if subjects do not believe that the experiment is about what you say it is about, they may act in unpredictable ways. Because economists are more concerned with outcomes, this unpredictability is extremely costly. It is less obvious, however, how subjects' disbelief in the purpose of the experiment would impact their self-reported attitudes.

Therefore, deception is cheaper for economists to forgo and more expensive for them to engage in than psychologists. Consistent with the law of demand, deception is not used by economists but is used by psychologists. The norm against deception is quite strong in experimental economics. Generally, an economic journal will not publish an experiment in which deception was used, while in psychology journals deception is commonplace and accepted.

V. CONCLUSION

The goal of this article was to describe and illustrate methodological concerns about the use of experiments in law and to provide a guide to those who either conduct or consume experimental economics research. I have described three different types of economics experiments, mentioned the methodological considerations that accompany each type, and provided illustrative examples from law and economics. I also have given some advice for running, analyzing, and presenting experiments.

Experiments can accomplish objectives that other forms of analyses cannot. They can offer clean tests of economic theories by constructing experiments that meet the assumptions of the theories, and observing the outcomes. They can investigate alternative causes of observed anomalies, either from the lab or from the field, by independently manipulating factors that in reality are confounded. They can provide a testing ground for potential policies, and offer suggestions to social planners that can increase efficiency and equity. Many legal experiments have been run that do these things, and more should be used in the future. Experiments are an important addition to the researcher's toolbox that can help achieve our goal of better understanding and explaining our world.

---

and Stamer & McDaniel, 19 J. ECON. PSYCHOL. 411–14 (1998); John D. Hey, *Experimental Economics and Deception: A Comment*, 19 J. ECON. PSYCHOL. 397 (1998); Tanga McDaniel & Chris Starmer, *Experimental Economics and Deception: A Comment*, 19 J. ECON. PSYCHOL. 403 (1998).