

Experimental Methods for the General Economist: Five lessons from the Lab

Douglas Davis*

Abstract: I review a series five primary results from experimental economics that impact on the economics profession as a whole. These results regard the relative (un)importance of subject sophistication in laboratory markets, the importance of gender on economic outcomes, the propensity for humans to behave in less than fully rational ways, the importance of trading institutions on economic performance and the behavioral relevance of economic theory. In overview, I find that economics as a profession has benefited from the use of experimental methods by fostering a dialogue between theorists and empiricists, better informing policy and improving data collection techniques.

* Professor of Economics, Virginia Commonwealth University, Richmond VA 28284-4000. Email: dddavis@vcu.edu

A version of this paper was presented as the Presidential Address at the 85th Annual Meeting of the Southern Economic Association in New Orleans Louisiana. I am indebted to the good ideas and dedicated efforts of the many coauthors, whose joint work I cite in this paper.. In all cases, the usual disclaimer applies.

This work was supported in part by NSF Grant SES-1426980.

1. Introduction.

My interest in experimental methods dates from the end of my time as a graduate school at Indiana University. For my doctoral thesis I chose to work on the topic of ‘statistical discrimination’ or motives other than prejudice that an employer might underestimate the ability of a minority job candidate (Davis, 1987). My thought was that one possible reason for such misperceptions might be driven by a simple misuse of order statistics. Unlike a sample mean, which does not vary in expectation with the number of observations, the expected sample maximum is a function of the sample size. In contexts where employers focus on maximal rather than average quality, perhaps employers might incorrectly draw an inference about average quality from the sample maximum. Evaluating empirically such a hypothesis, particularly in light of the standard items in the toolkit of a budding economist in the early 1980’s was a formidable challenge. One simply cannot usefully ask people whether or not their apparently discriminatory activities are motivated by racial animus.

Arlington Williams, a then new assistant professor at Indiana University (and a student of Vernon Smith), thought that my problem was both interesting and testable. Not, however, with naturally occurring data on hiring decisions, but in a streamlined laboratory setting, where student participants were asked to make evaluations not of job candidates, but rather through something more directly related to my theory, specifically by having financially motivated participants take random draws from two urns in a context where many more observations came from one urn than the other and where the participants were financially motivated to focus on the highest number drawn. If my theory was correct, then students would errantly view the less intensively sampled urn as containing lower numbers on average.

That approach worked. Although at the time radical, this innovative method of data collection not only allowed me to finish my thesis, but it opened for me the door to a relatively small group of extremely talented economists, who were remarkably willing to talk with an interested young scholar. I am profoundly grateful to these economists not only for their openness and encouragement in these initial years of my career but for their contagious sense of excitement that they were on the cusp of something really important.

From those early days, when experimental economics was still quite a novelty, I would spend the lion’s share of my time in any seminar explaining to a usually polite but very often deeply

skeptical audience what might usefully be learned from games played by undergraduates for cash (sometimes with computers). Times have certainly changed. The use of experimental methods expanded, and then exploded following the award of the Nobel Prize in Economics to Vernon Smith and Daniel Kahneman in 2002. The explosion has continued unabated in more recent years with Nobel Awards to Eleanor Ostrom in 2007 and Alvin Roth in 2012.

The basic rationale for the experimental method has remained pretty much the same over the years: using the lab can overcome the difficulties of finding data in many naturally occurring settings, the lab has a high degree of “internal validity” because it allows the researcher to change one variable at a time. All of these advantages have contributed to its acceptance as a standard tool in the economist’s toolkit.

Along with acceptance of laboratory analysis as a legitimate research method, both techniques and research questions have changed. Within the lab, computerization has made both the implementation of experimental designs and the collection of data vastly easier. Outside the laboratory the development of field experiments has generated a remarkably useful complement to both laboratory methods and standard empirical methods. The use of field experiments outside the laboratory offers a remarkably useful complement to both laboratory methods and standard empirical methods as well.¹ Finally, an entirely new subfield of ‘behavioral economics’ has emerged.

Everyone appreciates that experimental methods now cast a large footprint on economics. The purpose of this paper is to consider the how the outlines of that footprint impact on the economics profession as a whole. In particular, I focus on the potential value of this large, sometimes overwhelming volume of activity for non-experimentalists. Although I certainly cannot even pretend to be comprehensive in a short paper (and couldn’t even, in the length of a sizable text),² I here humbly offer the following five ‘big lessons’ that I have learned from thirty years of experimental analysis. I won’t suggest that these are necessarily the five pinnacle

¹ See e.g., Harrison and List (2004) for an introduction to and a review of field experiments. John List provided an overview of field experiments in his Distinguished Guest Lecture at the 84th Annual Meeting of the Southern Economic Association in Atlanta Georgia.

² Comprehensive, albeit somewhat dated treatments of experimental economics include Davis and Holt (1993), and Kagel and Roth (1995). See also Kagel and Roth (2016).

contributions of laboratory methods to economics, but they certainly are lessons that first come to my mind as interesting and important.

2. Five Lessons from Laboratory Economics

2.1 *College Sophomores Are Real People.* One of the biggest sources of initial skepticism regarding the use of laboratory methods was that the typical subject pool consisted not of ‘reals’ but of a convenience sample of university undergraduates. When testing economic theories, particularly propositions regarding interactions between firms or financial exchanges, the frequent use of college students has – appropriately – raised the criticism that ‘reals’ such as bankers or businessmen are far more sophisticated than the subject pool used to evaluate a theory. The important news here is that in repeated comparison testing between college students and ‘real’ actors there is no important difference in behavior. In a recent paper Fréchette (2015a) evaluates thirteen studies conducted between 1964 and 2006 that contrast the behavior of college students with ‘real’ agents from a context pertinent to the theory being examined. A summary of his results appears as Figure 1. In only one instance did the pertinent professional in any sense out-perform the college students^{3,4} In another nine instances Fréchette reports no important differences across groups.⁵ In another two instances the college students actually conformed more closely to the predictions of the theory than did the relevant professionals.⁶

What explains this general absence of real versus undergraduate student subject pool effects? Most prominently, laboratory experiments are most frequently ‘theory tests,’ conducted in extremely streamlined contexts as close to the domain of the theory as possible. In such environments basic cognitive processes rather than occupation specific skills drive predictions and, as surprising as it may seem to many of those of us who regularly teach college

³ Namely, professional soccer players more nearly engaged in randomizing behavior in an asymmetric game with no pure strategy equilibrium.

⁴ The 13 studies Fréchette (2015a) analyzes are only a subset of economics experiments that evaluate decisions made by actors with experience in a natural context pertinent to the market or decision context being investigated. Fréchette confines attention to those studies that used both ‘convenience sample’ college undergraduates and ‘reals’ with pertinent experience. Other studies related have appeared more recently. A very nice current example is the recent study on the external validity of laboratory tax compliance experiments by Alm, Bloomquist and McKee (2015). As with the majority of the studies assessed by Fréchette the decisions of college undergraduates did not differ significantly from the decisions of taxpayers.

⁵ These instances included contexts involving bidding in sealed bid auctions, common value actions, option pricing and gaming against managerial incentive schemes.

⁶ One such instance involved oral auctions. The other involved free riding behavior in a voluntary contributions game.

undergraduates, there is no basis for thinking that the cognitive process of students differs from that of ‘reals’. Further, the occupation specific skills of ‘reals’ may actually hinder performance in a laboratory context. That college students sometimes out-perform the relevant ‘real’ actors occurs because those with experience in a pertinent natural context may invoke their own mental scripts in an attempt to fill in the blanks of experimental instructions. For this reason they may attempt to extract information from features of the laboratory environment that, while useful in the natural context, are meaningless in the laboratory environment.

Performance of Undergraduates vs. 'Reals' in Laboratory Experiments

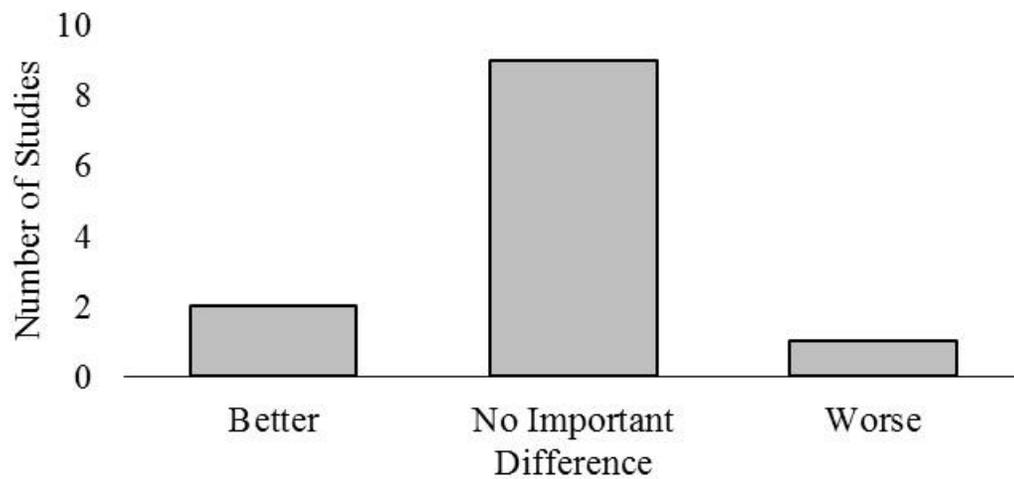


Figure 1. Relative Performance of Undergraduates Relative to ‘Reals’ in Thirteen Laboratory Experiments. Source Fréchette (2015a).

This is not to say that the choice of subject pool is universally a matter of indifference. To the contrary, an entire branch of behavioral economics has developed in the last decade concerned with measuring individual characteristics such as risk posture, time discount rate, altruism, etc. These measures can differ substantially across populations of ‘reals’ including groups with different ages, races, cultures and sexes.⁷ However, for the narrow purpose of testing theory, the use of a convenience sample rather than real agents does not importantly matter and people with pertinent expertise in a particular do not behave in a consistently different way.

⁷ Fréchette (2105b) reviews the literature that addresses the behavioral consequences of non-standard subject pools including non-human animals, people living in token economies, children, the elderly, demographically varied samples, and professionals.

The importance of individual characteristics on performance brings me to a second major lesson.

2.2 Gender Matters. I am an admittedly reluctant convert to the notion that gender importantly affects behavior in relatively sterilized laboratory environments. While I am of course well aware of the bundle of differences in both biology and socialization that distinguish males and females, I was skeptical that these differences could meaningfully affect the way people played simple games. My conversion came fairly recently in a project regarding behavior in indefinitely repeated prisoner dilemma games (Davis, Ivanov and Korenok, 2016). It is well known that in such games subjects exhibit considerable heterogeneity in their propensities to cooperate and we were interested in identifying individual characteristics that affected cooperative propensities. So we did the following: We had individuals play a series of ‘matches’ or indefinitely repeated prisoner’s dilemma games. In each match participants were paired with a different, randomly selected, anonymous partner. Then, on a separate day, we had these same participants return to the lab and complete a battery of popular tests for eliciting individual characteristics, including risk postures, time discount rates, propensities for altruism, degrees of strategic sophistication and capacities to recognize of strategic dominance. Our intention was to see how these different characteristics correlated with individuals’ propensities to either cooperate or engage in behaviors that promote cooperation, such as leadership, leniency or forgiveness. The surprising, at least to us, result was that with one notable exception, these individual characteristics had little impact on either cooperativeness or behavioral aspects might affect cooperative. The exception was gender. By a large margin, females cooperated significantly less frequently than males. Importantly what was surprising here was not that we observed an identifiable difference between women and men. A variety of gender-based differences, have been well documented, such as for example, the higher degree of risk aversion among females.⁸ The surprise was the importance of the effect over a host of potential measures of individual characteristics.

Gender differences can have marked economic effects. Let me mention here two examples from laboratory experiments. The first regards ‘glass ceiling’ effects for women in the workplace

⁸ Women also tend to dislike competitive environments and respond negatively to competitive pressures. See the surveys by Croson and Gneezy (2009), Eckel and Grossman (2008) and Niederle and Vesterlund (20110).

(Niederlie and Vesterlund, 2008).⁹ These authors conducted a study where mixed groups of males and females were asked to make allocative decisions in a game in which earnings for all players would increase if exactly one player made a costly contribution to a group exchange prior to the expiration of a time limit. What they observed was that the males almost uniformly waited for females to make the contribution decision, and that females almost uniformly made it! The suggestion is that females (at least in mixed groups) feel compelled to make the costly decision for the group, a task that these two, very successful female economists considered analogous to volunteering to perform ‘non-promotable’ tasks in the workplace.

A second, equally provocative example studies gender effects in the context of asset markets Eckel and Fullbrunn (2015). A well-known area of behavioral research in financial economics regards the frequent (and almost pervasive) propensity for traders in a simple laboratory market to generate speculative pricing bubbles via the trade of assets with a commonly known payoff stream. Eckel and Fullbrunn divided traders by gender into separate groups, endowed the traders in each market with identical combinations of assets and cash and then observed trading behavior in the declining value design standard in this line of research. Figure 2 illustrates the median prices for six female-only and six male-only groups. As can be seen female traders were much less likely to generate pricing bubbles than comparable groups of male traders. Eckel and Fullbrunn then tried mixed gender groups, and again found that mixed gender groups engaged in significantly less speculative behavior than did the male only cohorts.

I don’t want to overstate the implications of these results. Recent work on gender effects in a very different asset market environment by Holt et al. (2015) suggest that it is preliminary to conclude that females may necessarily tame the wolves of Wall Street.¹⁰ Nevertheless, while the

⁹ Lise Vesterlund discussed glass ceiling effects in her 2014 Association Lecture to the 84th Annual Meeting of the Southern Economics Association in Atlanta Georgia.

¹⁰ Numerous differences distinguish the designs used in Eckel and Fullbrunn (2015) and Holt et al. (2015). In addition to a declining asset balance, trades in Eckel and Fullbrunn were conducted under double auction rules, traders were aware of the gender of other members of their cohort, and cash/asset endowments were heterogeneous across participants. In contrast in Holt et al. (2015), trades were conducted under call market rules, traders had homogeneous endowments, and relative to Eckel and Fullbrunn, traders were ‘cash rich’. As in Eckel and Fullbrunn, Holt et al. (2015) tested for risk aversion. Both studies report that females were relatively more risk averse.

interactions between genders and various features of laboratory designs remains unclear it is unquestionably the case that in some contexts gender prominently affects market outcomes.¹¹

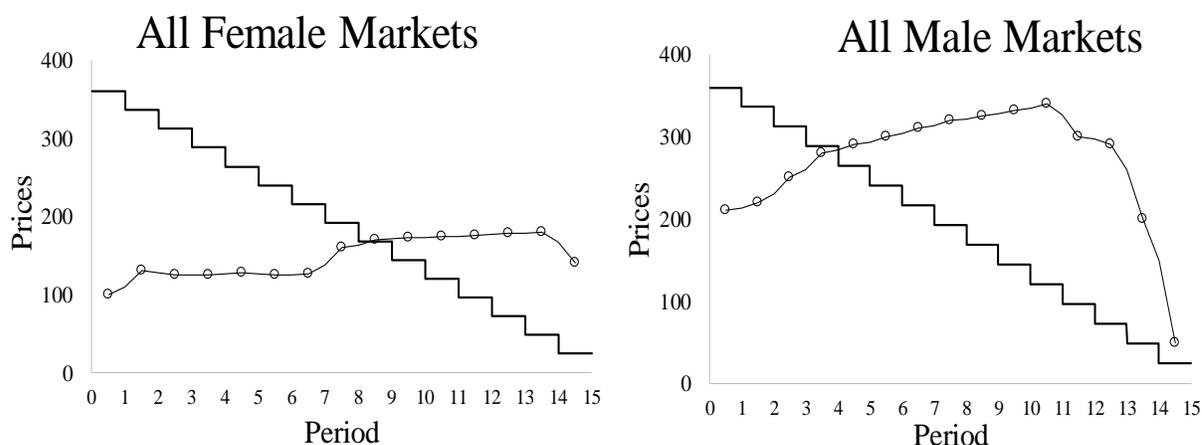


Figure 2. Median Contract Prices and Fundamental Prices for Asset Markets Composed of Female and Male Traders. Source: Adapted from Eckel and Fulbrunn (2015)

The presence and persistence of speculative trading bubbles suggests my third important lesson.

2.3 *Humans Can Be Irrational.* That we individually all make foolish decisions is, from a social perspective, relatively uninteresting. Perhaps a natural corollary to the old saw that ‘a fool and his money are soon parted’ is that ‘most of us have less money than we might’. More interesting for us as economists are instances where people act in a *predictably* irrational way. Laboratory methods have been powerfully instrumental in identifying such instances.¹² It wasn’t so many years ago, for example, that the rational expectations hypothesis - the idea that asset prices conveyed the most accurate information available regarding the value of a firm, was regarded as an article of faith among financial economists. Work in behavioral finance, much of it stemming from the just mentioned research on speculative behavior in laboratory asset markets, initiated in Smith, Suchanck and Williams (1988), has led to a much broader appreciation by economists that asset prices may persistently deviate from fundamental value.¹³

¹¹ My comments in the text regard only the effects of gender on decisions in a laboratory context. A related and fascinating area of investigation regards manipulations of hiring procedures that counteract gender bias in employment. For a recent example, see Bohnet, Van Geen and Bazerman (2015), who show that making decisions collectively rather than individually tends to offset the effects of unconscious stereotypes in hiring decisions.

¹² For a more general review of predictable irrationality, see Airely (2008).

¹³ In this vein, observe that Robert Shiller, although not an experimental economist, received a share of the 2013 Nobel prize in economics for his work on behavioral finance.

A related example comes from the study of the convergence properties of oligopoly markets. The speed and completeness of convergence in markets with a limited number of sellers turns critically on the adjustment rules used by sellers. Industrial organization economists have long considered the consequences of adjustment rules that vary with respect to differences in sellers' assumptions about others' responses to previous outcomes. A seller, for example, may believe that others will repeat their immediately previous actions, or, alternatively that others will pick actions closer to the average of all previous plays. Importantly, however, in either case, the seller is presumed to best respond to this assumed choice. Laboratory sellers, however, do not uniformly engage in such optimization.

I illustrate with summary results of a differentiated product oligopoly experiment reported in Davis (2011). The design features a market structure with four sellers who have symmetrically differentiated demand and constant costs. Sellers made a sequence of action choices in a series of 40 consecutive 'trading periods'. In addition to action choices each period sellers also submitted forecasts of what other sellers would do in the coming period. The primary treatment was the basis of competitive interactions, with sellers interacting as price setting Bertrand competitors in one case and as quantity setting Cournot competitors in another. To further help sellers with their action choices they were also provided with a profit calculator that allowed them to identify the profit maximizing response to their forecasts.

Figure 3 illustrates the relative propensity of sellers to follow a best response rule over the final 20 trading periods. As is clear from the figure sellers in both environments made choices closest to optimizing responses only about 40% of the time. Further, while some randomness accompanied action choices, as a whole, responses could hardly be described as random. Rather, in addition to best responses, the sellers relied on two distinctly non-optimal alternatives: 'Herding' or picking a choice very close to their forecast of the next period's average choice, and 'Inertia' or selecting an action choice close to their own past period action. As seen in the middle and right columns of Figure 3, Sellers are closest to an 'inertia' rule about as often as best responses, and made choices closest to 'herding' in another 20% of instances.

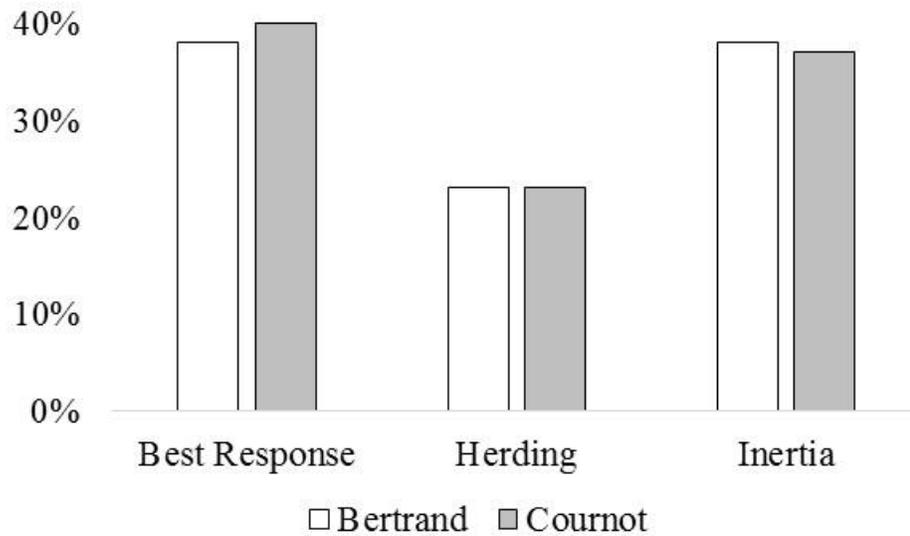


Figure 3. A Classification of Seller Adjustment Patterns in a Differentiated Product Oligopoly
Source: Davis (2011) .

Experimentalists have identified a host of such predictable ‘non-rational’ regularities in the way people process and respond to information, ranging from the ameliorative effects of behavioral ‘nudges’ to factors affecting the human propensity to ‘trust.’ These developments, all fascinating in their own right, are more important collectively as they have led to the incorporation of psychology into economics. The emerging interdisciplinary field of behavioral economics contributes in fundamental ways to both economics and psychology, offering for psychologists an analytic foundation and forcing economists to consider seriously the essential elements of a parsimonious utility specification as well as the ways that human limitations, such as bounds on rationality, can be incorporated into economic models.¹⁴

This comparison of decision rules in differentiated product Cournot and Bertrand environments motivates a fourth lesson.

2.4 *Institutions Matter.* Consider again the just-described differentiated product Cournot and Bertrand market experiment. That sellers followed the same adjustment rules in relative terms does not imply that the two market structures converge on Nash predictions with equal speed or completeness. To the contrary, as can be seen in the mean transaction price paths for two representative markets shown in Figure 4, performance varies markedly with the change in

¹⁴ See Camerer (2003) for an excellent overview of behavioral economics.

strategic interactions. In stark contrast to the immediate collapse on Nash predictions in the Bertrand market, outcomes in the comparable Cournot market vacillate both persistently and wildly.

The difference in outcomes in the two market structures is driven by the very different gaps between forecasts and best responses that changing the basis of interactions induces. In the Bertrand market, where actions are strategic complements, optimization, herding and inertia are all very close together. In contrast, in the Cournot market, where actions are strategic substitutes, the best response to a forecast that deviates from the Nash prediction is a deviation of almost equal magnitude on the other side of the equilibrium. Thus, while the behavior of agents does not importantly vary across institutional structures, the way the institution interacts with these propensities can yield starkly different outcomes.

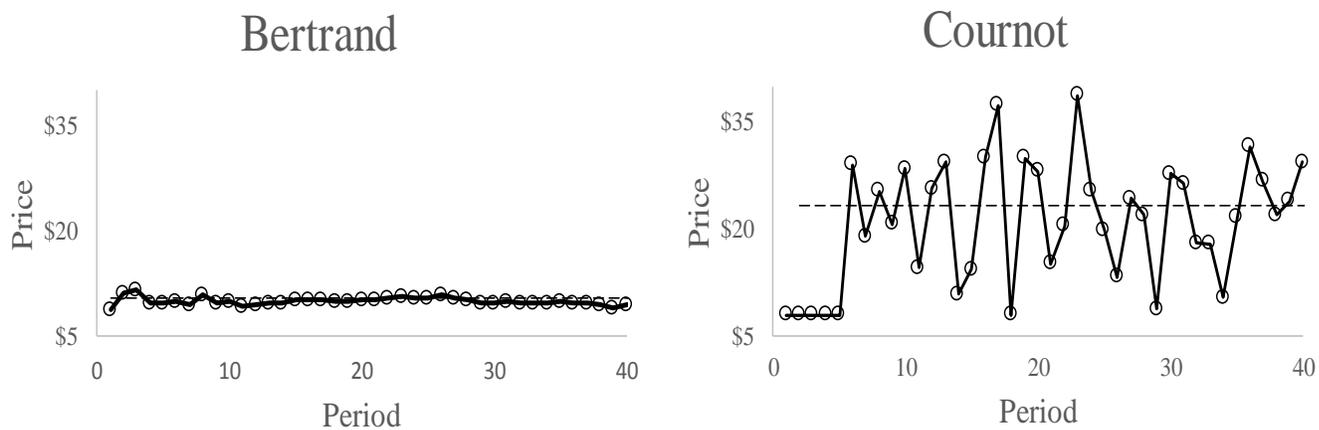


Figure 4. Mean Transactions Price Paths for Representative Differentiated Product Bertrand and Cournot Markets. Source: Davis (2011).

Historically, the notion that institutions affect market performance was a primary motivation for the use of laboratory methods to investigate allocative questions in economics. In the late 1940's a residual uncertainty regarding the capacity of markets to allocate goods efficiently festered among some economists. This post-depression/post WWII skepticism, combined with an utter incapacity to observe directly either the supply or the demand curves necessary to evaluate competitive predictions motivated Edward Chamberlin's famous experiment evaluating the performance of a classroom market organized like the 'pit' on the Chicago mercantile exchange (Chamberlin, 1948). Chamberlin's experiment, which has become a staple exercise for introductory economics classes, yielded something less than fully efficient behavior. Smith

(1962), however, subsequently showed that centralizing bids and asks with an ‘auctioneer’ (a task performed, for example, by the specialist on the floor of the New York Stock Exchange) generates robustly efficient competitive outcomes.

Although the direct observation that ‘markets can work’ was certainly a fundamental insight in economics, more important was the offshoot – that changes in the way markets are organized fundamentally affects their performance. The performance difference of markets organized under Chamberlin’s ‘negotiated price’ rules and Smith’s ‘double auction’ trading rules is only one illustration. A more pronounced example arises from the comparison of the posted price trading rules typical of retail type exchange with double auction markets. Given the same underlying supply and demand conditions, not only is it the case that a ‘posted offer’ market extracts less of the available surplus than a ‘double auction’, but further that posted offer markets are very markedly less responsive to changes in underlying conditions.

Consider for example, double auction and posted offer performance in a ‘trend demand’ design that features a series of first ‘inflationary then deflationary demand shifts (Davis, et al, 1993) The mean contract price paths, shown as Figure 5 illustrate the results of three double auction and three posted offer markets. The double auction markets shown in the left panel, tracked the changes in equilibrium quite well, with the closing contract in each period (the hollow dots) typically providing very accurate information regarding the underlying market price. In contrast, in the posted offer markets shown in the right panel, prices responded only very slowly to an inflationary change in underlying conditions (as sellers realize only that they’re selling out their inventories each period), and then fell apart altogether in the subsequent deflationary regime when demand fell sufficiently to bring the equilibrium price below current market prices, leaving sellers suddenly with neither any sales nor any information about the prices that might again generate sales.

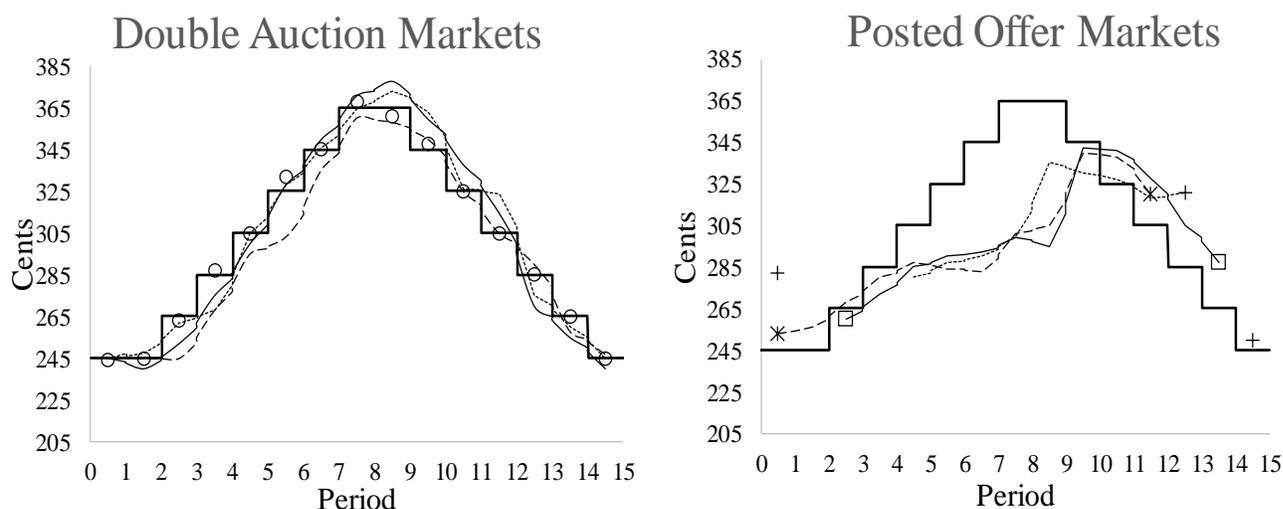


Figure 5. Mean Contract Price Paths from Three Double Auction and Three Posted Offer Markets in a Trend Demand Design. Source: Left panel, Davis et al. (1993), Right panel Davis and Holt (1997).

Given the pervasiveness of retail type exchange in industrialized economics, the poor dynamic performance of posted offer markets has implications for the performance of economies a whole. For example, Davis and Korenok (2011) reports an experiment conducted to isolate the relative effects of ‘sticky prices’ and ‘sticky information’ in driving imperfect market responses to fully anticipated changes in price levels – an issue of longstanding concern among macroeconomics. In a monopolistically competitive market where sellers post prices each period, we found it was neither – rather the posted offer rules themselves induced an exceedingly slow adjustment to an inflationary increase in both prices and outputs.

The capacity of institutional alterations to fix market imperfections suggests that new ‘hybrid’ institutions can be constructed to allow efficient allocations in contexts where markets previously did not or could not exist. Examples include not only computer assisted markets where bidders demands for one property or asset is contingent on their ability to purchase a related property or asset, but also new market structures like government radio spectrum auctions, (e.g., Plott and Salmon, 2004) markets for sodium dioxide emissions (e.g., Burtaw et al. 2010) and electricity distribution (e.g., Rassenti, Smith and Wilson, 2003). Also relevant here are allocative mechanisms where for reasons of equity or practicality monetary bids cannot for used, such as the National Medical Residency Assignment Program, the Kidney exchanges and mechanisms for determining school assignments (See e.g., Roth 2008).

The success of these various market and non-market allocative institutions raises my fifth and final lesson.

2.5 *On the Whole Economic Theory Works Remarkably Well.* When Vernon Smith first established that competitive predictions organized market behavior, his initial response was to ‘stress test’ this result, examining behavior in limiting conditions, including markets with few sellers on either side of the market. What he observed was that competitive predictions organized market performance under incredibly robust limiting conditions. Conditions so general that he actually proposed double auction trading rules as a mechanism for monopoly restraint (Smith, 1981)!

From the perspective of theory the double auction is problematic in that it is difficult to tractably model the price determination process. Other trading institutions, however, are far more amenable to theoretical analysis, and in cases where theory offers predictions, those predictions often organize the data quite well. I offer some examples from my own research.

The discrete unit version of the posted offer trading institution that was traditionally compared to double auction markets may be analyzed as a model of Bertrand Edgeworth competition. Holt and Solis-Soberon (1992) develop a general method for identifying equilibria in discrete unit models of Bertrand-Edgeworth competition. Subsequently, Davis and Holt (1994) extended these results to examine the tradeoffs between ‘numbers’ (concentration) and predicted market power. ‘Market power’ (in the sense that the competitive equilibrium is not a Nash equilibrium for the market analyzed as a stage game) was the clear winner. Similarly, Davis and Wilson (2000) investigate the relationship between synergies and market power in a model of Bertrand Edgeworth competition, and found – precisely as subsequently stated in the 2011 revision of the DOJ/FTC Horizontal Merger Guidelines – that synergies mitigate market power concerns only if they affect the costs of marginal units.

These, of course, are only a few very specialized examples. More generally, experiments have created a dialogue between experimentalists and theorists that allows for a refinement of theory in ways that make theory behaviorally relevant. A pertinent example comes from game theory. In the late 1990’s a number of ‘basic contradictions’ had emerged as stylized facts, casting doubt on the capacity of game theory to organize human behavior in strategic situations. Adapting the decision making process to include noise, Goeree and Holt (2001) shows that these

behavioral inconsistencies could be both coherently explained and in fact predicted by allowing for a measure of noise in the decision making process.

My examples are certainly far from comprehensive. Useful interchanges between theorists and experimentalists have arisen in diverse areas including game theory, auctions, public goods, finance and labor economics, among others. My point is that traditional neoclassical economic theory provides a very useful basis for organizing economics outcomes, and that experimental methods are helping to both refine theory and make it more broadly useful.¹⁵

3. Conclusion

The above ‘lessons’ reflect my excitement for this subfield, and by illustrating what has been learned I hope offer some insight into the promise of this subfield for what may be learned in the future. In closing, I offer a few words of how laboratory tools fit into the economist’s tool box—that is, what the use of experimental methods are and are not.

Let’s start with what experimental methods can add to economic analysis. Laboratory methods are perhaps most frequently used for testing the predictions of a theory. It is difficult to say much from the laboratory about whether a theory organizes behavior well in some pertinent natural context. If, however, the theory fails to organize data in a streamlined ‘best shot’ environment, it is difficult to imagine how it could plausibly organize data in a much more complex natural context. Importantly, even ‘theory tests’ can be policy relevant. Policies, after all, are based on theories, and the failure of a theory on (or very close to) its domain raises very justifiable concerns about the value of that theory as a basis for policy.

Also, laboratory experiments can be used to isolate in streamlined conditions regularities in human behavior that deviate from standard ‘rational actor’ assumptions. I discussed several pertinent examples of such experiments above, including those involving gender effects and collective irrationality. Here, I stress that laboratory results are, by themselves, only suggestive of behavior in natural contexts. Such experiments, however, can complement and allow exploration of parallel observations from less controlled studies in natural contexts.

¹⁵ In his introduction to a volume discussing the role of experimental methods in economics Roth (1987) offers what has become perhaps the most standard categorization of experiment types: ‘Speaking to Theorists’ or testing and refining theory, ‘Searching for Facts’ or attempting to evaluate individual preferences or the limitations of rationality and ‘Whispering in the Ears of Princes,’ or using experiments to evaluate policy alternatives. Of course, as Roth is careful to observe, many experimental investigations fail to fit neatly in one of these groups.

Experimental methods also can and have been used to improve data collection. Over twenty five years ago, Vernon Smith (at the time boldly) argued that economics differed from meteorology in that economists were not restricted to learn about the economy by watching events pass (Smith, 1994). In addition to running experiments in the laboratory, however, Smith suggested that we could do a better job collecting data. The development and acceptance of experimental methods, particularly experiments done in the field, have helped considerably in this effort. The mindset of randomized trials, along with the variety of methods of eliciting individual characteristics has unquestionably improved our methods of collecting data in naturally occurring contexts.¹⁶

Finally, consider what the use of laboratory methods will not be. Experimental methods will neither turn theory on its head, converting economic analysis into psychology nor replace more conventional empirical analysis. Rather, they provide an important complement to these existing methods, and in the process they will, I contribute to making economic analysis a more useful and predictive science.

¹⁶ The wide variety of field experiments investigate diverse topics including measuring preferences, identifying the effects of institutions on behavior, and the effects of racial and gender bias on market outcomes.. Harrison and List (2004) and Levitt and List (2007) provide excellent reviews. List and Levitt (2009) offers an interesting, and more general overview of the longer term history of field experiments

References

- Airely, Daniel. 2008. *Predictably irrational: The hidden forces that shape our decisions*. New York: Harper Collins.
- Alm, James, Kim M. Bloomquist and Michael McKee. 2015. On the external validity of laboratory tax compliance experiments. *Economic Inquiry* 53:1170-86
- Bohnet, Iris, Alexandra Van Geen and Max Bazerman. 2015. When performance trumps gender bias: Joint versus separate evaluation” *Management Science*, (forthcoming)
- Burtaw, Dallas, Jacob K. Goeree, Charles A./Holt, Karen Palmer, Erica Myers and William Schobe. 2010. “An experimental analysis of auctioning emission allowances under a loose cap. *Agricultural and Resource Economics Review* 39: 162-75.
- Camerer, Colin F. 2003. *Behavioral game theory: Experiments in strategic interaction*. Princeton NJ: Princeton University Press.
- Chamberlin, Edward. H. 1948. An experimental imperfect market. *Journal of Political Economy* 56:95-108.
- Croson, Rachel T. A., and Uri Gneezy. 2009. Gender differences in preferences. *Journal of Economic Literature* 47: 448–74.
- Davis, Douglas D. 1987. Maximal quality selection and discrimination in employment. *Journal of Economic Behavior and Organization* 8:97-112.
- Davis, Douglas D. 2011. Behavioral convergence properties of Cournot and Bertrand markets: An experimental analysis. *Journal of Economic Behavior and Organization* 80: 443-58.
- Davis, Douglas D., Glenn W. Harrison and Arlington W. Williams. 1993. The effects of nonstationarities on the convergence to competitive equilibria. *Journal of Economic Behavior and Organization* 20: 1-22.
- Davis Douglas D. and Charles A. Holt. 1993, *Experimental Economics*. Princeton NJ: Princeton University Press.
- Davis, Douglas D and Charles A. Holt. 1994. "Market power and mergers in laboratory markets with posted prices. *RAND Journal of Economics* 25:467-87.
- Davis, Douglas D and Charles A. Holt. 1997. "Price rigidities and institutional variations in markets with posted prices. *Economic Theory*, 9:63-80.
- Davis, Douglas D. Asen Ivanov and Oleg Korenok. 2016. Individual characteristics and behavior in repeated games: An experimental study. *Experimental Economics* (forthcoming)
- Davis, Douglas D. and Oleg Korenok. 2011. Nominal price shocks in monopolistically competitive markets: An experimental analysis. *Journal of Monetary Economics* 58: 578-89.

- Davis, Douglas D. and Bart J. Wilson. 2000. Cost savings and market power exercise. *Economic Theory* 16: 545:65.
- Eckel, Catherine C. and Sascha C. Fullbrunn. 2015. “Thar “SHE” blows? Gender, competition, and bubbles in experimental asset markets. *American Economic Review* 105: 906-20.
- Eckel, Catherine C., and Philip J. Grossman. 2008. Differences in the economic decisions of men and women: Experimental evidence. In *Handbook of Experimental Economics Results*, edited by Charles R. Plott and Vernon L. Smith. New York: Elsevier. 509–19
- Fréchette Guillaume R. (2015a) “Laboratory Experiments: Professionals versus Students.” In *Handbook of Experimental Economic Methodology*. Edited by Guillaume R. Fréchette and Andrew Schotter Oxford UK: Oxford University Press. 360-90.
- Fréchette Guillaume R. (2015b) “Experimental Economics Across Subject Populations,” in the *Handbook of Experimental Economics*, vol. 2. Edited by John H. Kagel and Alvin E. Roth. Princeton NJ: Princeton University Press, forthcoming.
- Goeree, Jacob K and Charles A. Holt. 2001. Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review* 90: 1402-22.
- Harrison, Glenn W. and John A. List. 2004. Field experiments. *Journal of Economic Literature* 42:1009–55.
- Holt, Charles A. Megan Porzio and Michelle Yingze Song (2015). Price Bubbles and Expectations in Experimental Asset Markets: Gender and Risk Aversion. Manuscript, University of Virginia.
- Holt, Charles A and Francisco Solis-Soberon (1992) “The calculation of equilibrium mixed strategies in posted-offer auctions” in *Research in Experimental Economics*, vol.4 edited by R. Mark.Isaac. Greenwich CT: JAI Press 189-229.
- Kagel, John H. and Alvin E. Roth. 1995. *The Handbook of Experimental Economics*. Princeton NJ: Princeton University Press.
- Kagel, John H. and Alvin E. Roth. 2016. *The Handbook of Experimental Economics vol. 2*. Princeton NJ: Princeton University Press. (forthcoming)
- Levitt, Steven D. and, John A. List. 2007. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* 21:153–74.
- List John A. and Steven D. Levitt. 2009. Field experiments in economics: The past, the present, and the future. *European Economic Review* 53: 1-18.
- Niederle, Muriel and Lise Vesterlund. 2008. Gender differences in competition. *Negotiation Journal* 24: 337-463.

- Niederle, Muriel and Lise Vesterlund. 2011. Gender and competition. *Annual Review of Economics* 3: 601–30.
- Plott, Charles R. and Timothy C. Salmon. 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* 53:353-83.
- Rassenti, Stephen J., Bart J. Wilson, and Vernon L. Smith. 2003. Controlling market power and price spikes in electricity networks: Demand-side bidding. *Proceedings of the National Academy of Sciences*, 100(5),
- Roth, Alvin E. 1987. Introduction and Overview .In *Laboratory Experimentation in Economics: Six Points of View* edited by Alvin E. Roth. Cambridge: Cambridge University Press 1-13.
- Roth, Alvin E. 2008. What have we learned from market design? *Economic Journal* 118: 285–310.
- Smith, Vernon L. Gary Suchanek and Arlington W. Williams. 1988. Endogenous expectations in experimental spot asset markets. *Econometrica* 56: 1119-51.
- Smith, Vernon L. 1962. An experimental study of competitive market behavior. *Journal of Political Economy*, 70: 111-137.
- Smith, Vernon L. 1981. An empirical study of decentralized institutions of monopoly restraint. In *Essays in Contemporary Fields of Economics in Honor of E.T. Weiler, 1914-1979*, edited by James Quirk and George Horwich. West Lafayette IN: Purdue University Press. 83-106.
- Smith, Vernon L. 1994. Economics in the laboratory. *Journal of Economic Perspectives* 8:113-31.