

The Methodological Promise of Experimental Economics

by

Glenn W. Harrison*

September 2010

Forthcoming, *Journal of Economic Methodology*

TEXT WORD COUNT: 2498

* Department of Risk Management and Insurance and CEAR, Robinson College of Business, Georgia State University. E-mail contact: GHARRISON@GSU.EDU. Thanks to the U.S. National Science Foundation for research support under grants NSF/HSD 0527675 and NSF/SES 0616746.

Experimental Economics: Rethinking the Rules by Bardsley, Cubitt, Loomes, Moffatt, Starmer and Sugden [2010] attempts forensic pathology on corpses that are still partying. And it is one of those noisy, alcohol-fueled parties on a hot night. In one corner there is a large argument that threatens to turn ugly, in another corner there are some near-sleepers that are still managing to mumble something, at the bar there are some earnest folk taking their appreciation of the alcohol a bit too seriously, and around the grill there are some carnivores balancing it all brilliantly. There are many people at the party that I do not know, although I do see many that are clearly not having as much fun as they used to in earlier times. Of course, there are a few traitors that I am happy to see quartered and gibbeted by our pathologists, even if they still show a pulse and the scalpels are dull.

The corpse, of course, is experimental economics, and correctly defined broadly to subsume behavioral economics. The party is the academic publishing environment in which the field manages to find itself. The bar-tenders are the editors of academic journals. But more important than the bar-tenders, who seem to serve without listening anymore, are the milling masses. The field of experimental economics evolves in the way it does because those that consume the stuff have generally become uncritical and undemanding.

But there is still extraordinary promise here. The intellectual division of labor in economics as a whole has gone far beyond the point at which there are benefits of specialization, and there is a risk of the same things happening to experimental economics. And yet, and yet ... maybe this is the place to try to regroup. My passion for the sub-field comes from the derived demands it makes on theory, econometrics, and common sense, the potential for policy insight, and the *lingua franca* it enforces on the whole intellectual conversation. It allows economists to realize the methodological promise of offering operationally meaningful statements that the rest of us can laugh at when appropriate. It shares with econometrics the functional utility of being a tool that is useable in virtually all areas of economics.

There is no question that economics has been improved by being reminded of the heterogeneity of economic behavior, and the fact that much of the behavior observed does not fit well with existing models. If one can get past the marketing hype of the behavioral economics tradition, then this contribution can and should be applauded. However, now begins the more serious task of restating, re-applying, and extending the tools of traditional economics.

This is where this valuable book fits in. It proposes a methodological assessment of experimental economics, but aims more broadly. And it succeeds for the very reason I thought it might not: the authors are among those at the party that seem to have balanced it all brilliantly over long careers, while those around them do not and show no signs of ever sobering up. It is hard to write at this level without occasionally grinding an axe, and yet it is only those with intellectual skin in the game that I want to hear from.

Chapter 1 lays out the agenda for the book, with some concrete examples that are revisited later in more detail and with subtlety. The best part of the chapter, §5, is where their product is differentiated from the professional methodologists that typically populate this journal. Many of the semantic distinctions that bother eunuchs do not make a difference in practice, such as the imaginary line between methods and methodology, and what constitutes explanation and causation. And, yes, we get a first, appropriate dosing from the Duhem-Quine Finger Wag that is a staple of any methodologist's mantra.

Chapter 2 examines what it means to test a theory with experiments. Apart from naive claims that a theory is meant to be universal, a useful distinction is drawn between the domain of behavior that the theory is intended to characterize and the domain that the theory can be appropriately tested on. For example, I have no idea how one could meaningfully claim to have tested all of the assumptions needed to simulate the effects of trade reform on the poor in Brazil, such as Harrison et al. [2004], but that would not give me an instant of hesitation in using that theory for that important purpose. And being

among the first to revise the simulation model if and when we know how to do so. The real bite of this distinction comes when we see how experimental behavior is ignored. When the editors of the *JPE* do not like a result, they claim that the latter domain is “the field” and not “the lab.” But they are not defending the theory, so much as just saying that they really don’t care that much about errant behavior in the lab. When I pointed out that the subjects in many canonical experiments were within ϵ of any optimum when measured in terms of expected income or utility, this was just an attack on claims that these are interesting or powerful tests of the theory. To a Mere Theorist, or the editors of the *QJE*, they are apparently interesting tests. Finally, when Hey and Orme [1994] examine choice behavior over pairs of lotteries *not* designed for their “trip-wire” fragility to small violations of standard theory, standard theory does pretty well. So the validity of these tests, as far as they go, is not the issue, just the fact that they do not go very far at all. This is a very different thing than claiming that the core theory is valid.

Chapter 3 is an extended application of the Duhum-Quine Finger Wag. OK, we got it in chapter 1, and if we are going to have scotch I prefer it to be single-malt and served neat (Ross [2005]). The application here does involve some subtle and important discussions in §4 of the extent to which we can infer preferences from choices in strategic settings the way we do in non-strategic settings. Real game theorists, who don’t forget that the defining treatise was called *Theory of Games and Economic Behavior* after all, get this easily. But they have the convenient excuse that they want to analyse the game given utility numbers as payoffs that are consistent with revealed preferences. As Binmore [2007; p. 12] tartly puts it, “Critics who think that human beings are basically altruistic therefore go astray when they accuse game theorists of using the wrong analysis of Prisoners’ Dilemma. They ought to be accusing us of having correctly analyzed the wrong game.” Testing game theory is really hard when you try to do it right. Even the specification of social preferences is tricky: should the arguments be monetary payoffs, the utility of the money payoffs, the subjectively perceived utility of the money payoff, a normative concept of utility;

should we consider the experimenter as a player in the experimental game, since he literally is in the experiment if not in the game being modelled; and is there some heterogeneous mixture of these considerations across agents? One could add concerns about common knowledge, and the havoc that relaxations of the Reduction of Compound Lotteries axioms do to strategic concepts. Experimental economists have barely begun to evaluate strategic games carefully.

Chapter 4 is one of the most important, and considers the manner in which experiments have evolved from being tasked with evaluating deductive theories to generating the regularities, or irregularities, that demand theoretical attention. In short, the shift towards the generation of “stylized facts” that may or may not have an immediate theoretical home. In large part the early days of experimental economics were driven by a visceral and oft-voiced terror at being branded a-theoretical, and banished to the lower tiers of academic journals. This is a clear memory from those of us in the second-generation of experimental economists that attended the earliest professional meetings. It was only reinforced as theorists pounded away in the anonymous confessional called refereeing. One important experimental project on auctions was hammered at all of the best journals for testing single-shot theory with experiments in which the bidders were partners in repeated games, since that was the “anything can happen” flavor of the month in theory. So there was no highbrow methodological stance at work, just a new field struggling for respect.

But there is something very dark about the way in which experiments are now used to generate stylized facts. In this chapter the concept of the “experiment as exhibit” is nicely developed. From a metaphysical perspective, an experiment is an object, whether or not it was conducted by an economist or a psychologist, whether it fits one theory, several or none, and even whether it was conducted well or not. It can have a blanket of knowledge claims attached to it, but the procedures and behavior are what they are. All this is well developed, and the appropriate methodological cautions advanced.

The abuse of experimental methods and results here is not due to subtle methodological error, but just plain bad scholarship that is uncritically tolerated. The sad state of neuroeconomics is just the latest, striking example (Harrison [2008a][2008b]). One common *modus operandi* in behavioral economics is to cite some casual, hypothetical study of behavior that finds some odd and startling things, and then briefly mention that the same results are obtained when one uses real incentives (e.g., the literature on discounting anomalies, such as Loewenstein and Prelec [1992; §II]). The problems here are not fixed by thinking more carefully from a methodological perspective, but by simply calling a spade a spade.

Chapter 5 takes on the issue of external validity, and the now-tense relationship between lab and field experiments. In some applications the lab environment is presented as *necessarily* unreliable (Levitt and List [2007]). The complementarity between the two, in general, is clear: see Harrison and List [2004] and Harrison, Lau and Rutström [2010] for extended discussion. In particular, the popular taxonomy of Harrison and List [2004] does not reflect their conclusion (§10) that there are many combinations of criteria that can be used to define degrees of external validity, and that *no* bright lines should be imposed. One should not confuse that taxonomy as *aide memoire* with the arguments behind it.

While it is true that lab experiments traditionally use artefactual and stylized tasks that are free of field cues, in order to generate the type of control that is seen as essential to hypothesis testing, field experiments have other weaknesses that *a priori* are equally important to recognize (Harrison [2005]). Most importantly, the ability to implement necessary controls on experimental conditions in the field is much more limited than in the lab, as is the ability to implement many counterfactual scenarios. In addition, recruitment is often done in such a way that it is difficult to avoid and control for sample selection effects; indeed, in many instances the *natural* process of selection provides the very treatment of interest. However, that means that one must take the sample with *all* of the unobservables that it might have selected on, and just assume that they did not interact with the behavior being measured.

Finally, the cost of generating observational data can be quite significant in the field, at least in comparison to the lab. Of course, we simply need more data points in which the lab and the field are directly compared (e.g., Andersen, Harrison, Lau and Rutström [2010]).

There is a constructive middle-ground here, apart from simply conducting parallel lab and field experiments and hoping for the best (or worst, depending on your disposition). Fiore, Harrison, Hughes and Rutström [2009] propose a new experimental environment, the Virtual Experiment (VX), that has the potential of generating both the internal validity of lab experiments and the external validity of field experiments. A VX is an experiment set in a controlled lab-like environment, using either typical lab or field participants, that generates synthetic field cues using Virtual Reality technology. The experiment can be taken to typical field samples, such as experts in some decision domain, or to typical lab samples, such as student participants. The VX environment can generate internal validity since it is able to closely mimic explicit and implicit assumptions of theoretical models, and thus provide tight tests of theory; it is also able to replicate conditions in past experiments for robustness tests of auxiliary assumptions or empirically generated hypotheses. The VX environment can generate external validity because observations can be made in environments with cues mimicking those occurring in the field. Dynamic scenarios can be presented in a realistic and physically consistent manner, making the interaction seem natural for the participant. Thus the VX builds a bridge between the lab and the field, allowing the researcher to smoothly go from one to the other and see what features of each change behavior. VX is a methodological frontier enabling new levels of understanding via integration of laboratory and field research in ways not previously possible.

Chapter 6 takes on the topic of incentives, and argues the silly, contrarian perspective that hypothetical tasks have been given a bum rap. Here one has to object that there have been decades of mischievous debate that has diverted resources inefficiently; and I stress that this is of immediate policy

significance in the area of environmental damage assessment (Harrison [2006]), and not just a procedural issue for backwater debate between experimenters. By all means do a hypothetical study to scope out an issue, if you must and are lazy, but don't leave it there if something interesting is discovered. Don't misquote assertions that one gets the same results: in *general*, and that is the key, one does not (Harrison and Rutström [2008]). And don't make nonsensical claims about how I asserted that "payoff dominance" is "essential for subjects to take the task seriously." (p. 255) Read [2005] is a wonderful, blunt satire of the arguments in favor of using hypothetical rewards, although in this chapter he is read as seriously advocating the positions he so brilliantly undermines.

Chapter 7 considers the topic of "noise," and makes the important point that one cannot safely divorce the jobs of theorist, experimenter and econometrician. I agree completely, although without the bias in favor of the random preference model that one always expected these authors to display: see Harrison, Lau and Rutström [2010] for a contrary view of the issue, consistent with the core point about the critical role of experiments in waking up theorists and econometricians.

This is a fine book, for all my disagreements. Despite the incoherent sub-title, I will add it to my (background) reading list for classes in experimental economics and behavioral econometrics. It significantly advances the debate on the methodological contribution of experimental economics. I'll happily get all of these authors the next round at the party, off in that quiet corner untouched by the zombies.

References

- Andersen, Steffen; Harrison, Glenn W.; Lau, Morten I., and Rutström, E. Elisabet, "Preference Heterogeneity in Experiments: Comparing the Lab and Field," *Journal of Economic Behavior & Organization*, 74, 2010, 209-224.
- Bardsley, Nicholas; Cubitt, Robin; Loomes, Graham; Moffatt, Peter; Starmer, Chris, and Sugden, Robert, *Experimental Economics: Rethinking the Rules* (Princeton, NJ: Princeton University Press, 2010).
- Binmore, Ken, *Playing for Real: A Text on Game Theory* (New York: Oxford University Press, 2007).
- Fiore, Stephen M.; Harrison, Glenn W.; Hughes, Charles E., and Rutström, Elisabet, "Virtual Experiments and Environmental Policy," *Journal of Environmental Economics & Management*, 57(1), January 2009, 65-86.
- Harrison, Glenn W., "Field Experiments and Control," in J. Carpenter, G.W. Harrison and J.A. List (eds.), *Field Experiments in Economics* (Greenwich, CT: JAI Press, Research in Experimental Economics, Volume 10, 2005, 17-50).
- Harrison, Glenn W., "Making Choice Studies Incentive Compatible," in B. Kanninen (ed.), *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Guide to Theory and Practice* (Boston: Kluwer, 2006, 65-108).
- Harrison, Glenn W., "Neuroeconomics: A Critical Reconsideration," *Economics & Philosophy*, 24(3), 2008a 303-344.
- Harrison, Glenn W., "Neuroeconomics: Rejoinder," *Economics & Philosophy*, 24(3), 2008b, 433-444.
- Harrison, Glenn W.; Lau, Morten I., and Rutström, E. Elisabet, "Theory, Experimental Design and Econometrics Are Complementary (And So Are Lab and Field Experiments)" in G. Frechette and A. Schotter (eds.), *The Methods of Modern Experimental Economics* (New York: Oxford University Press, 2010 forthcoming).
- Harrison, Glenn W.; Rutherford, Thomas F.; Tarr, David G, and Gurgel, Antonio, "Trade Policy and Poverty Reduction in Brazil," *World Bank Economic Review*, 18(3), 2004, 289-317.
- Harrison, Glenn W., and Rutström, E. Elisabet, "Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Experiments," in C.R. Plott and V.L. Smith (eds.), *Handbook of Experimental Economics Results* (New York: Elsevier Press, 2008).
- Hey, John D., and Orme, Chris, "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62(6), November 1994, 1291-1326.
- Levitt, Steven D., and List, John A., "What do Laboratory Experiments Measuring Social Preferences Reveal About the Real World," *Journal of Economic Perspectives*, 21(2), 2007, 153-174.

Loewenstein, George, and Prelec, Drazen, "Anomalies in Intertemporal Choice: Evidence and Interpretation," *Quarterly Journal of Economics*, 107, 1992, 573-592.

Read, Daniel, "Monetary Incentives, What Are They Good For?" *Journal of Economic Methodology*, 12(2), 2005, 265-276.

Ross, Don, *Economic Theory and Cognitive Science: Microexplanation* (Cambridge, MA: MIT Press, 2005).