FIELD EXPERIMENTS IN ECONOMICS: AN INTRODUCTION

Jeffrey P. Carpenter, Glenn W. Harrison
and John A. List

Experimental economists are leaving the reservation. They are recruiting subjects in the field rather than in the classroom, using field goods rather than induced valuations, and using field context rather than abstract terminology in instructions. We believe that there is something methodologically fundamental behind this trend. Field experiments differ from laboratory experiments in many ways. Although it is tempting to view field experiments as simply less controlled variants of laboratory experiments, this would be a serious mischaracterization. What passes for "control" in laboratory experiments might in fact be precisely the opposite if it is artificial to the subject or context of the task. We see field experiments as being methodologically complementary to traditional laboratory experiments.

In Section 1 we offer a taxonomy of field experiments in the literature from Harrison and List (2004). This taxonomy identifies the key characteristics defining the species. It also provides a terminology to better identify different types of field experiments, or more accurately to identify different characteristics of field experiments. We do not propose a bright line to define some experiments as field experiments and others as something else, but a set of criteria that one would expect to see in varying degrees in a field experiment. We propose five factors that can be used to determine the field context of an experiment: the nature of the subject pool, the nature of the information and experience that the subjects bring to the task, the nature of the commodity, the nature of the task or institutional rules applied, and the environment that the subjects operate in. In Section 2 we

Field Experiments in Economics
Research in Experimental Economics, Volume 10, 1–16

ISSN: 0193-2306/doi:10.1016/S0193-2306(04)10001-X

augment our taxonomy by discussing some reasons for conducting experiments in the field. In Section 3 we summarize the papers in this volume, placing them in the context of our taxonomy. In Section 4 we offer some general conclusions about the methodological contribution of field experiments.

This volume had it's origins in a conference that we organized in April 2003 at Middlebury College in Vermont. In addition, we put out a call for papers in the area. Each paper was refereed, typically by 3 or more experts, and all papers were reviewed by each co-editor. The resulting mix is a good reflection of the wide range of topics and methodological issues covered in field experiments.

Data files and computer programs to replicate statistical analyses are available for all papers. Each is listed as a project at the *ExLab* Digital Archive located at http://exlab.bus.ucf.edu. In each case the project name matches the title of the chapter. The editors are grateful to all authors for being willing to provide data and code.

1. DEFINING FIELD EXPERIMENTS

There are several ways to define words. One is to ascertain the formal definition by looking it up in the dictionary. Another is to identify what it is that you want the word-label to differentiate.

The *Oxford English Dictionary (Second Edition)* defines the word "field" in the following manner: "Used attributively to denote an investigation, study, etc., carried out in the natural environment of a given material, language, animal, etc., and not in the laboratory, study, or office." This orients us to think of the *natural environment* of the different components of an experiment.

It is important to identify what factors make up a field experiment so that we can functionally identify what factors drive results in different experiments. To give a direct example of the type of problem that motivated us, when List (2001) gets results in a field experiment that differ from the counterpart lab experiments of Cummings, Harrison and Osborne (1995) and Cummings and Taylor (1999), what explains the difference? Is it the use of data from a particular market whose participants have selected into the market instead of student subjects, the use of subjects with experience in related tasks, the use of private sports-cards as the underlying commodity instead of an environmental public good, the use of streamlined instructions, the less-intrusive experimental methods, or is it some combination of these and similar differences? We believe field experiments have matured to the point that some framework for addressing such differences in a systematic manner is necessary.

If we are to examine the role of "controls" in different experimental settings, it is appropriate that this word also be defined carefully. The *Oxford English Dictionary (Second Edition)* defines the verb "control" in the following manner: "To exercise restraint or direction upon the free action of; to hold sway over, exercise power or authority over; to dominate, command." So the word means something more active and interventionist than is suggested by it's colloquial clinical usage. Control can include such mundane things as ensuring sterile equipment in a chemistry lab, to restrain the free flow of germs and unwanted particles that might contaminate some test. But when controls are applied to human behavior, we are reminded that someone's behavior is being restrained to be something other than it would otherwise be if the person were free to act.

We take care with these terms, since it is common for experimenters to think of the difference between lab experiments and field experiments as being synonymous with the trade-off between "internal validity" and "external validity." If the controls in the lab do their job, and do not artificially constrain behavior, then the lab affords more control almost by definition. But the premiss here is not obviously correct: there are many settings in which the controls of the lab can elicit artefactual behavior that is poorly correlated with naturally-occurring behavior. We simply argue that one cannot make this determination *a priori* on the basis of whether the experiment is conducted in the lab or the field. There is much more to the evaluation of an experiment than that. First we need to identify what criteria differentiates field experiments, and then one needs to decide if the experiment (lab or field) corresponds to the theory being tested.

1.1. Criteria that Define Field Experiments

We propose five factors that can be used to determine the field context of an experiment:

- the nature of the subject pool,
- the nature of the information and experience that the subjects bring to the task,
- the nature of the commodity,
- the nature of the task or institutional rules applied,
- the nature of the environment that the subject operates in.

The taxonomy that results will be important, we believe, as comparisons between lab and field experimental results become more common.

Student subjects can be viewed as the standard subject pool used by experimenters, simply because they are a convenience sample for academics. Thus when one goes "outdoors" and uses field subjects, they should be viewed as

non-standard in this sense. But we argue that the use of non-standard subjects should not *automatically* qualify the experiment as a field experiment. The experiments of Cummings, Harrison and Rutström (1995), for example, used individuals recruited from churches in order to obtain a wider range of demographic characteristics than one would obtain in the standard college setting. The importance of a non-standard subject pool varies from experiment to experiment: in this case it simply provided a less concentrated set of socio-demographic characteristics with respect to age and education level, which turned out to be important when developing statistical models to adjust for hypothetical bias (Blackburn et al., 1994). Alternatively, the subject pool can be designed to represent the national population, so that one can make better inferences about the general population (Harrison et al., 2002).

In addition, non-standard subject pools might bring experience with the commodity or the task to the experiment, quite apart from their wider array of demographic characteristics. In the field, subjects may be endowed with experiences that are more directly relevant for the question that motivates the research. For example, Cardenas (2003) collects experimental data from participants that have direct, field experience extracting from a common pool resource. Similarly, Carpenter, Daniere and Takahashi (2003) conduct social dilemma experiments with urban slum dwellers who face daily coordination and collective action problems, such as access to clean water and solid waste disposal.

The commodity itself can be an important part of the field. Recent years have seen a growth of experiments concerned with eliciting valuations over actual goods, rather than using induced valuations over virtual goods. The distinction here is between physical goods or actual services and abstractly defined goods. The latter have been the staple of experimental economics since Chamberlin (1948) and Smith (1962), but imposes an artificiality that *could* be a factor influencing behavior.[1] Such influences are actually of great interest, or should be. If the nature of the commodity itself affects behavior, in a way that is not accounted for by the theory being applied, then the theory has at best a limited domain of applicability that we should know about, and at worse is simply false. In either case, one can know the limitations of the generality of theory only if one tests for it, by considering physical goods and services.

Again, however, just having one field characteristic, in this case a physical good, does not constitute a field experiment in any fundamental sense. Rutström (1998) sold lots and lots of chocolate truffles in a laboratory study of different auction institutions designed to elicit values truthfully, but hers was very much a lab experiment despite the tastiness of the commodity. Similarly, Bateman et al. (1997) elicited valuations over pizza and dessert vouchers for a local restaurant. While these commodities were not actual pizza or dessert themselves,

but vouchers entitling the subject to obtain them, they were not abstract. There are many other examples in the experimental literature of designs involving physical commodities.[2]

The nature of the task that the subject is being asked to undertake is an important component of a field experiment, since one would expect that field experience could play a major role in helping individuals develop heuristics for specific tasks. The lab experiments of Kagel and Levin (1999) illustrate this point, with "super-experienced" subjects behaving differently than inexperienced subjects in terms of their propensity to fall prey to the winners' curse. An important question is whether the successful heuristics that evolve in *certain* field settings "travel" to other field and lab settings (Harrison & List, 2003). Another aspect of the task is the specific parameterization that is adopted in the experiment. One can conduct a lab experiment with parameter values estimated from field data, so as to study lab behavior in a "field-relevant" domain. Since theory is often domain-specific, and behavior can always be, this is an important component of the interplay between lab and field. Early illustrations of the value of this approach include Grether, Isaac and Plott (1981, 1989), Grether and Plott (1984) and Hong and Plott (1982).

The environment of the experiment can also influence behavior. The environment can provide context to suggest strategies and heuristics that a lab setting might not. Lab experimenters have always worried that the use of classrooms might engender role-playing behavior, and indeed this is one of the reasons that experimental economists are generally suspicious of experiments without salient monetary rewards. Even with salient rewards, however, environmental effects could remain. Rather than view them a uncontrolled effects, we see them as worthy of controlled study.

1.2. A Proposed Taxonomy

Any taxonomy of field experiments runs the risk of missing important combinations of the factors that differentiate field experiments from conventional lab experiments. However, there is some value in having broad terms to differentiate what we see as the key differences. Harrison and List (2004) therefore propose the following terminology:

- a *conventional lab experiment* is one that employs a standard subject pool of students, an abstract framing, and an imposed[3] set of rules;
- an *artefactual field experiment* is the same as a conventional lab experiment but with a non-standard subject pool;[4]

- a *framed field experiment* is the same as a artefactual field experiment but with field context in either the commodity, task, or information set that the subjects can use;[5]
- a *natural field experiment* is the same as a framed field experiment but where the environment is one where the subjects naturally undertake these tasks and where the subjects do not know that they are in an experiment.[6]

We recognize that any such taxonomy leaves gaps.

Moreover, it is often appropriate to conduct several types of experiments in order to identify the issue of interest. For example, Harrison and List (2003) conduct artefactual field experiments and framed field experiments with the same subject pool, precisely to identify how well the heuristics that might apply naturally in the latter setting "travel" to less context-ridden environments found in the former setting. And List (2004) conducts artefactual, framed and natural experiments to investigate the nature and extent of discrimination in the sportscard maketplace.

1.3. Other Types of Experiments

Apart from lab and field experiments, Harrison and List (2004) discuss three other types of experiments that economists conduct:

- *social experiments* entail some change in government policy, with the intent of observing if the change has an effect relative to some baseline or control treatment;
- *natural experiments* involve some exogenous change in economic circumstances that mimics a controlled field or social experiment, but in which the subjects do not know that they are being studied and in which the subjects are not deceived, and in which the researchers typically have no say in what treatments are imposed; and
- *thought experiments* are simply experiments without the benefit of implementation.

Each has strengths and weaknesses relative to lab and field experiments. Social experiments are often conducted on a scale that makes them directly relevant to policy, but suffer from a "rational expectations" inferential problem if the subjects being studied are aware of the exercise. Natural experiments avoid this pitfall, but typically only occur by chance. Thought experiments can be cheap, but you get what you pay for: *a priori* assumptions substituting for actual behavior.

Just as we see lab and field experiments as methodological complements, we also view social, natural and thought experiments as just different analytical tools in the economists' arsenal.

2. WHY CONDUCT EXPERIMENTS IN THE FIELD?

The conventional lab is comfortable. Students are relatively easy to recruit as participants, they are used to abstract reasoning, they can actually undertake abstract reasoning on a good day, and they provide a reasonably broad cross-section of the population on some important socio-economic dimensions. In addition, the computer lab is relatively sterile. It is now easy to write code for experiments[7] and isolate one terminal from another. And the coffee machine is usually right around the corner. So why should researchers give up this comfort to enter the field where experiments usually become much more messy?

We offer a few thoughts on this topic, but begin with a few words of caution based on our experiences in both the lab and the field. Properly conducted field experiments really are messy. There is often much more planning involved. One has to devote a lot of thought to identify which population of participants to target, and even more thought to figure out how to gain access to the target population. The opportunity cost of time for non-student populations is often much higher. This factor alone means the procedures often need to be streamlined to minimize the participants' commitment of time. But it also means that more thought must be put into these procedures, since researchers often have only one chance with the population. Therefore it is critical that the procedures run efficiently and gather the information that is important. In short, one way to differentiate field experiments from conventional lab experiments is that field experimentalists do their research "without a net."

So why walk the high-wire without a net? One obvious reason is to easily silence one of the most often leveled criticisms of lab experiments – the lack of external validity.[8] Any lab experimental study presented at a seminar in a location not frequented by other experimenters is bound to receive the standard external validity question: "Yes, interesting results, but who's to say 'real' people would behave this way?" Going to the field allows one to examine whether student results can be extrapolated to the population. The influential market research conducted by Vernon Smith and his collaborators was taken much more seriously when others were able to show that career traders often exhibited the same (or more severe) biases present in the student trader population.[9] Now the circle has come all the way around, with students of Wall Street relying on insights from the lab (e.g. Miller, 2001). Moreover, there is simply no way to answer the critically important development policy question posed in the title of Henrich and McElreath (2002), "Are Peasants Risk-Averse Decision Makers?" without going into the field to some extent.

The second most often criticism leveled at experimental work is, "Yes, interesting results, but who's to say behavior would not change at 'real' stakes?" From a practical point of view, the fact that a few dollars or euros is a much

bigger fraction of one's monthly budget in many areas of the world outside of North American and Europe provides ample opportunity to examine the effect of stakes on behavior. Cameron (1999) is one of the most cited paper on the effect of stakes. She showed that first mover behavior in the ultimatum bargaining experiment was unaffected when the stakes of the game were raised to a level of three months expenditures by Indonesian students. In the wake of this experiment, it is now conventional to see stakes of a day's wage in field experiments in both industrialized and unindustrialized settings.

One reason to conduct experiments in general, discussed in Plott (1982) and Smith (1994), is particularly salient in the field: experiments in the field allow policy makers to examine the effect of changing or implementing new institutions on a small scale before fully implementing a project with potentially large consequences. A nice example, on a small scale, comes from Gneezy and Rustichini (2000) who examine the effect of fining parents who are late picking up their children from Israeli daycare centers. Conventional wisdom says that imposing a fine will reduce the likelihood that parents will be late. However, they showed that parents treat the fine as a price for being late that parents were willing to pay. As a result, the frequency of tardiness actually increased and most importantly, when the fines were removed, parents continued to be more likely to be late when gathering their children. The punchline, for our purposes, is that imposing a fine on a large scale would have put the daycare system on an alternative path that would have been worse than the status quo from the point of view of the people in charge of the system. Furthermore, this path change could not have been reversed.

3. SUMMARY OF THE PAPERS IN THIS VOLUME

Not only have economists begun leaving the reservation, they are doing so with increasing frequency. However, they are still spending most of their time in the neighborhood. Using our taxonomy, artefactual field experiments (lab experiments with non-standard participants) have become relatively common recently, but framed field experiments (that add a naturally occurring frame) are still relatively rare, and there are just a few natural field experiments (where the task is also familiar). The chapters of this book reflect the current distribution of field experiments. Leaving aside Chaps 1–4 and 9 for now, since they are more methodological, we have compiled three artefactual field experiments and one framed field experiment.

Chapters 5, 6, and 8 are excellent examples of artefactual field experiments. In each case standard laboratory experiments are conducted with participants that range from grade school children in Ohio (Chap. 8) to the working poor

in the Montreal metropolitan area (Chap. 6) to a cross-section of the Danish population (Chap. 5). In Chap. 8, Robert Slonim and Eric Bettinger illustrate how artefactual field experiments can be used to inform policy disputes like the effect of educational vouchers on student attitudes and performance. In this case, they take advantage of the fact that for four years a private foundation in Ohio used a lottery to allocate educational vouchers for children to attend private school. The random assignment of these vouchers allows them to identify their effect on self-confidence, a factor that has been claimed to have an effect on educational attainment. Self-confidence is measured using an experiment, and the results show that there is no robust difference that can be attributed to winning the voucher lottery in the larger populations. However, among the African American sub-population, lottery winners are significantly less *over*-confident.

In Chap. 6, Catherine Eckel, Cathleen Johnson and Claude Montmarquette use experiments to measure the time preferences of the working poor in Montreal. Along with showing that the discount rates (measured in intervals) for these individuals can be predicted by a mixture of experimental variables and individual characteristics (e.g. the investment period, the rate of return, age, and sex), they illustrate the phenomenon of *present-biased* time preferences in which people prefer an earlier payoff more strongly the closer this payoff is to the present. Twenty-three percent of the experimental population act in accordance with this bias in their task frame. Most interestingly, however, they find a correlation between their measure of discount rate and financial decisions that have real financial consequences. Specifically, the authors show that the time preferences of the participants, elicited at modest stakes, can be used to predict whether one is more likely to take cash over a substantial amount of money (targeted for one's retirement). These results illustrate how field experiments can be used to inform policy interventions that target poverty reduction. Using experimental procedures from the older literature, they find extremely high discount rates for short-term horizons (mean of 290% p.a.) that are consistent with the older literature reviewed in Coller and Williams (1999).[10] On the other hand, their elicited discount rates for longer-term horizons are much more consistent with the recent literature (mean of 32% p.a.). They find reasonably high risk aversion (mean CRRA = 0.78) that is consistent with other findings from the lab and field, but this is a deliberately specialized population of policy interest that would be expected to be slight more risk averse on average.

In Chap. 5, Glenn Harrison, Morten Lau, Elisabet Rutström and Melonie Williams also gather data on individual risk and time preferences. However, this study examines a broad cross-section of Danish adults instead of the working poor in Canada. This study is important, not only for its estimate of discount rates and risk preferences among the 253 Danes who participated, but because of it's

contribution to the discussion of field methodology. In addition to showing that Danes exhibit slight risk aversion (mean CRRA = 0.33), have a mean individual discount rate in artefactual experimental frames that is equivalent to a really bad credit card (mean rate = 23%),[11] and that individual characteristics do a slightly better job predicting risk attitudes than time preferences (here only indicators for old age and living Copenhagen are significant), they extensively discuss the pitfalls of conducting this sort of research. For example, they discuss a new variant of the *multiple price list* method for eliciting subject responses in which participants pick one option at a time while moving down a list that helps to minimize the amount of confused responses by participants who flip back and forth between columns and, therefore, display inconsistent or imprecise preferences. They also address ways to quantify the possibility of a framing problem in which participants might have a natural tendency to flip between columns in the middle of the table of choices irrespective of the cost of doing so.

Chapter 7 by Jeffrey Carpenter, Stephen Burks and Eric Verhoogen is an example of a framed field experiment. They conduct ultimatum and dictator games at high stakes ($100) with people who work at a distribution center in Kansas City in addition to two control groups: traditional students at Middlebury College and non-traditional students at Kansas City Kansas Community College (KCKCC). What makes this a framed field experiment is the fact that each experiment was conducted in the natural environment of the subject population. The warehouse worker sessions were conducted in the breakroom of the warehouse and the student experiments were conducted in classrooms at the two locations. The point of having two control groups is to triangulate the effect of demographic characteristics separately from the effect of the natural setting. Comparing the two student groups allows one to test for the effect of demographic differences because the KCKCC resemble the warehouse workers demographically but have the same field setting as the Middlebury students. Similarly, comparing the KCKCC students to the warehouse workers allows one to examine the effect of the natural frame (school versus workplace). The results indicate that both demographics and framing matter. In the ultimatum game, demographic factors increase the offers made in Kansas City, but the workplace frame reduces them slightly so that offers can be ordered from lowest to highest: Middlebury, Warehouse, KCKCC. In the dictator game, only the framing of the situation has a robust effect on the altruism demonstrated by the participants. Workers are more generous than students in either setting. If one believes that phenomena like altruism are regulated by social norms, then this last result illustrates that norms can be endogenous with respect to framing and the nature of interactions.

The remaining chapters are oriented towards methodology and the existing literature. In Chap. 2, Glenn Harrison addresses a common myth among

experimentalists and other economists that field experiments must necessarily trade off control for relevance. A main theme of this chapter is that the artificial and sterile nature of many lab experiments constitutes a potential loss of control because participants have no clues that tell them which (highly relevant) heuristic rules of thumb to apply. Harrison systematically discusses the problem of control in natural and field experiments, in addition to the problems associated with the sterile framing of many lab experiments.

In Chap. 3 Andreas Ortmann expands on the issue of control by being critical of many of the field experiments that have been conducted in the past. Ortmann points out that going to the field is particularly onerous, because it is difficult to control factors that are taken for granted in the lab with students (e.g. literacy). However, he also points out that these difficulties are not automatically acceptable reasons for a lack of control. This chapter is a particularly useful balance to many of the other papers in this volume that emphasize the benefits of conducting experiment in the field.

Chapter 4 by Juan Camilo Cardenas and Jeffrey Carpenter begins by discussing how conducting field experiments may benefit the study of economic development. This first theme highlights the traditional reasons to conduct experiments (e.g. control, replication, and internal validity) and links this rationale to the study of behavioral factors in economic development. In their second theme, they stress a non-standard use of experiments to gather behavioral data that can be used to inform more directly relevant analyses. For example, they consider a possible link between norms of cooperation among slum dwellers in Southeast Asia and their living standard. In their final theme, they point out that experimentalists often forget that debriefing can be an important part of this type of research. Without a discussion of the experiment and its outcome, researchers often leave without communicating their purposes and results to the people who, in a field setting, might be best suited to use them.

The book is concluded by an example of why we must be careful in our interpretation of the results of experiments in both the field and the lab. In Chap. 9, Anabela Botelho, Glenn Harrison, Marc Hirsch and Elisabet Rutström draw an important distinction between culture and demographics. Using results from new experiments, as well as previously unused demographic control data from Roth, Prasnikar, Okuno-Fujiwara and Zamir (1991), Slonim and Roth (1998) and Cameron (1999), they illustrate that one cannot rely on standard practices of randomizing subjects into treatments when conducting experiments in many locations because the resulting demographic differences between the populations may be highly correlated with the location. The implication is that the variance in behavior previously attributed to location (or culture) can often be explained by the differential effect of demographics within locations. The punchline is that

there is no excuse not to collect demographic control data when conducting experiments under most circumstances and economists should be wary when presented uncontrolled results.

4. CONCLUSION

We avoid drawing a single, bright line between field experiments and lab experiments. One reason is that there are several dimensions to that line, and inevitably there will be some trade-offs between those. The extent of those trade-offs will depend on where researchers fall in terms of their agreement with the argument and issues we raise.

Another reason is that we disagree where the line would be drawn. One of us (Harrison), bred in the barren test-tube setting of classroom labs *sans* ferns, sees virtually any effort to get out of the classroom as constituting a field experiment to some useful degree. Another (List), raised in the wilds amidst naturally occurring sportscard geeks, would include only those experiments that used free-range subjects. And the last of us (Carpenter), who only seems to go to the field if there is good food involved, has decided that the line should probably be a plane, at least. Despite this disagreement on the boundaries between one category of experiments and another category, however, we agree on the characteristics that make a field experiment differ from a lab experiment.

The main conclusion we draw is that experimenters should be wary of the conventional wisdom that abstract, imposed treatments allow general inferences. In an attempt to ensure generality and control by gutting all instructions and procedures of field referents, the traditional lab experimenter has arguably lost control to the extent that subjects seek to provide their own field referents. The obvious solution is to conduct experiments both ways: with and without naturally occurring field referents and context. If there is a difference, then it should be studied. If there is no difference, one can conditionally conclude that the field behavior *in that context* travels to the lab environment.

NOTES

1. It is worth noting that Smith (1962) did not use real payoffs to motivate subjects in his experiments, although he does explain how that could be done and reports one experiment (Note 9, p. 121) in which monetary payoffs were employed.

2. We would exclude experiments in which the commodity was a gamble, since very few of those gambles take the form of naturally occurring lotteries.

3. The fact that the rules are imposed does not imply that the subjects would reject them, individually or socially, if allowed to.

4. To offer an early and a recent example, consider the risk aversion experiments conducted by Binswanger (1980, 1981) in India, and Harrison, Lau and Williams (2002), who took the lab experimental design of Coller and Williams (1999) into the field with a representative sample of the Danish population.

5. For example, the experiments of Bohm (1984b) to elicit valuations for public goods that occurred naturally in the environment of subjects, albeit with unconventional valuation methods; or the Vickrey auctions and "cheap talk" scripts that List (2001) conducted with sport card collectors, using sports cards as the commodity and at a show where they trade such commodities.

6. For example, the manipulation of betting markets by Camerer (1998), the solicitation of charitable contributions by List and Lucking-Reiley (2002), or the adjustment of work incentives in Nagin, Rebitzer, Sanders and Taylor (2002).

7. Many experiments can now be accessed and run as freeware on the web, such as the V*e*conlab maintained by Charles Holt at http://www.people.virginia.edu/~cah2k/programs.html. For a modest initial time commitment, one can program almost any conceivable experiment using Urs Fischbacher's *Z-Tree* software and templates available at http://www.iew.unizh.ch/home/fischbacher/.

8. We know what people think they mean by this expression, but we are not so clear. What is valid in an experiment depends on the theoretical framework that is being used to draw inferences from the observed behavior in the experiment. If we have a theory that (implicitly) says that hair color does not affect behavior, then any experiment that ignores hair color is valid from the perspective of that theory. But one cannot identify what factors make an experiment valid without some priors from a theoretical framework, which is crossing into the turf of "internal validity." Furthermore, the "theory" at issue here should include the assumptions required to undertake statistical inference with the experimental data (Ballinger & Wilcox, 1997).

9. In fact, Smith (1991, p. 157) recalls the reaction that academics had to his very first paper: "Whatever the exact genesis, I got up the courage to write a paper reporting on all the experiments I had done from 1956 to 1960. It wasn't easy. People had been skeptical that there was a trick, some simple reason why the experiments worked that had nothing to do with economics or theory or that overused, undefined thing that economists call the "real world."

10. Newer methods, such as employed by Coller and Williams (1999) and Harrison, Lau and Williams (2002) result in much lower discount rates.

11. This fact, given the number of people who use such credit cards, makes their results very plausible.

REFERENCES

Ballinger, T. P., & Wilcox, N. T. (1997). Decisions, error and heterogeneity. *Economic Journal*, *107*(July), 1090–1105.

Bateman, I., Munro, A., Rhodes, B., Starmer, C., & Sugden, R. (1997). Does part-whole bias exist? An experimental investigation. *Economic Journal*, *107*(March), 322–332.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural India. *American Journal of Agricultural Economics*, *62*(August), 395–407.

Binswanger, H. P. (1981). Attitudes toward risk: Theoretical implications of an experiment in rural India. *Economic Journal*, *91*(December), 867–890.

Blackburn, M., Harrison, G. W., & Rutström, E. E. (1994, December). Statistical bias functions and informative hypothetical surveys. *American Journal of Agricultural Economics*, *76*(5), 1084–1088.

Camerer, C., & Fehr, E. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. In: J. Henrich, R. Boyd, S. Bowles, H. Gintis, E. Fehr & C. Camerer (Eds), *Foundations of Human Sociality: Experimental and Ethnographic Evidence From 15 Small-scale Societies*. New York: Oxford University Press.

Camerer, C. F. (1998). Can asset markets be manipulated? A field experiment with racetrack betting. *Journal of Political Economy*, *106*(3), 457–482.

Cameron, L. A. (1999, January). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry*, *37*(1), 47–59.

Cardenas, J. C. (2003). Real wealth and experimental cooperation: Evidence from field experiments. *Journal of Development Economics*, *70*, 263–289.

Carpenter, J. (2002). Measuring social capital: Adding field experimental methods to the analytical toolbox. In: J. Isham, T. Kelly & S. Ramaswamy (Eds), *Social Capital and Economic Development: Well-Being in Developing Countries*. Northampton, UK: Elgar.

Carpenter, J., Daniere, A., & Takahashi, L. (2004). Cooperation, trust, and social capital in Southeast Asian urban slums. *Journal of Economic Behavior & Organization*, *54* (forthcoming).

Chamberlin, E. H. (1948). An experimental imperfect market. *Journal of Political Economy*, *56*, 95–108.

Coller, M., & Williams, M. B. (1999). Eliciting individual discount rates. *Experimental Economics*, *2*, 107–127.

Cummings, R. G., Harrison, G. W., & Osborne, L. L. (1995). Can the bias of contingent valuation be reduced? Evidence from the laboratory. Economics Working Paper B-95–03, Division of Research, College of Business Administration, University of South Carolina.

Cummings, R. G., Harrison, G. W., & Rutström, E. E. (1995, March). Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive compatible? *American Economic Review*, *85*(1), 260–266.

Cummings, R. G., & Taylor, L. O. (1999, June). Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. *American Economic Review*, *89*(3), 649–665.

Durlauf, S. N. (2002). On the empirics of social capital. *Economic Journal*, *112*, 459–479.

Ferber, R., & Hirsch, W. Z. (1978, December). Social experimentation and economic policy: A survey. *Journal of Economic Literature*, *16*(4), 1379–1414.

Ferber, R., & Hirsch, W. Z. (1982). *Social experimentation and economic policy*. New York: Cambridge University Press.

Gneezy, U., & Rustichini, A. (2000). A fine is a price. *Journal of Legal Studies*, *29*(1), 1–17.

Grether, D. M., Isaac, R. M., & Plott, C. R. (1981). The allocation of landing rights by unanimity among competitors. *American Economic Review (Papers & Proceedings)*, *71*(May), 166–171.

Grether, D. M., Isaac, R. M., & Plott, C. R. (1989). *The allocation of scarce resources: Experimental economics and the problem of allocating airport slots*. Boulder: Westview Press.

Grether, D. M., & Plott, C. R. (1984). The effects of market practices in oligopolistic markets: An experimental examination of the ethyl case. *Economic Inquiry*, *22*(October), 479–507.

Harrison, G. W., Lau, M. I., & Williams, M. B. (2002, December). Estimating individual discount rates for Denmark: A field experiment. *American Economic Review*, *92*(5), 1606–1617.

Harrison, G. W., & List, J. A. (2003). Naturally occurring markets and exogenous laboratory experiments: A case study of the winner's curse. Working Paper 3–14, Department of Economics, College of Business Administration, University of Central Florida.

Harrison, G. W., & List, J. A. (2004, December). Field experiments. *Journal of Economic Literature*, *42*(4), 1013–1059.

Henrich, J., & McElreath, R. (2002, February). Are peasants risk-averse decision makers? *Current Anthropology*, *43*(1), 172–181.

Hoffman, E., McCabe, K. A., & Smith, V. L. (1996). On expectations and the monetary stakes in ultimatum games. *International Journal of Game Theory*, *25*(3), 289–301.

Holt, C. A., & Laury, S. K. (2002, December). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Hong, J. T., & Plott, C. R. (1982). Rate filing policies for inland water transportation: An experimental approach. *Bell Journal of Economics*, *13*(Spring), 1–19.

Kagel, J. H., Battalio, R. C., & Walker, J. M. (1979). Volunteer artifacts in experiments in economics: Specification of the problem and some initial data from a small-scale field experiment. In: V. L. Smith (Ed.), *Research in Experimental Economics* (Vol. 1). Greenwich, CT: JAI Press.

Kagel, J. H., & Levin, D. (1999, September). Common value auctions with insider information. *Econometrica*, *67*(5), 1219–1238.

List, J. A. (2001, December). Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards. *American Economic Review*, *91*(4), 1498–1507.

List, J. A. (2004). The nature and extent of discrimination in the marketplace: Evidence from the field. *Quarterly Journal of Economics*, *119*(1), 49–89.

List, J. A., & Lucking-Reiley, D. (2002). The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign. *Journal of Political Economy*, *110*(1), 215–233.

Miller, R. M. (2002). *Paving wall street: Experimental economics and the quest for the perfect market*. New York: Wiley.

Nagin, D. S., Rebitzer, J. B., Sanders, S., & Taylor, L. J. (2002, September). Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *American Economic Review*, *92*(4), 850–873.

Plott, C. R. (1982). Industrial organization theory and experimental economics. *Journal of Economic Literature*, *20*(December), 1485–1527.

Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991, December). Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *American Economic Review*, *81*(5), 1068–1095.

Rutström, E. E. (1998). Home-grown values and the design of incentive compatible auctions. *International Journal of Game Theory*, *27*(3), 427–441.

Slonim, R., & Roth, A. E. (1998, May). Learning in high stakes ultimatum games: An experiment in the Slovak Republic. *Econometrica*, *66*(3), 569–596.

Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of Political Economy*, *70*, 111–137.

Smith, V. L. (1991). Experimental economics at Purdue. In: V. L. Smith (Ed.), *Papers in Experimental Economics*. New York: Cambridge University Press.

Smith, V. L. (1994). Economics in the laboratory. *Journal of Economic Perspectives*, *8*(Winter), 113–131.

Warner, J. T., & Pleeter, S. (2001, March). The personal discount rate: Evidence from military downsizing programs. *American Economic Review*, *91*(1), 33–53.

Uncited references

References cited in the text must appear in the reference list; conversely, each entry in the reference list must be cited in the text... The author must make certain that each source referenced appears in both places and that the text citation and reference list entry are identical in spelling and year.

Camerer & Fehr (2004), Carpenter (2002), Carpenter et al. (2004), Durlauf (2002), Ferber & Hirsch (1978), Ferber & Hirsch (1982), Hoffman et al. (1996), Holt & Laury (2002), Hong & Plott (1982), Kagel et al. (1979), Miller (2002) and Warner & Pleeter (2001).