NBER WORKING PAPER SERIES

USING FIELD EXPERIMENTS IN ENVIRONMENTAL AND RESOURCE ECONOMICS

John A. List
Michael K. Price

Working Paper 19289
http://www.nber.org/papers/w19289

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2013

We thank the Editor, Charles Kolstad, for his tremendous patience with this manuscript and two anonymous reviewers for comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Using Field Experiments in Environmental and Resource Economics
John A. List and Michael K. Price
NBER Working Paper No. 19289
August 2013
JEL No. C9,C93,Q5

ABSTRACT

This study showcases the usefulness of field experiments to the study of environmental and resource economics. Our focus pertains to work related to field experiments in the area of 'behavioral' environmental and resource economics. Within this rubric, we discuss research in two areas: those that inform i) benefit cost analysis and ii) conservation of resources. Within each realm, we show how field experiments have been able to test the relevant theories, provide important parameters to construct new theories, and guide policymakers. We conclude with thoughts on how field experiments can be used to deepen our understanding of important areas within environmental and resource economics.

John A. List
Department of Economics
University of Chicago
1126 East 59th
Chicago, IL 60637
and NBER
jlist@uchicago.edu

Michael K. Price
Department of Economics
Andrew Young School of Policy Studies
Georgia State University
P.O. Box 3992
Atlanta, GA 30302-3992
and NBER
mprice25@gsu.edu

I. Introduction

Environmental and resource economics is a field that has never been short on empirical questions. The extent to which the field embraced controlled experimentation as a way to uncover causal relationships and develop policy relevant cost benefit estimates should thus come as no surprise. Until recently controlled experimentation relied largely on inference drawn from either lab subjects or natural experimentation.[1] However, the past decade has seen a growth in the prominence of field experimentation in environmental economics.

This article presents an overview of the use of field experiments in environmental and resource economics. We begin by reviewing a body of evidence exploring the stability and consistency of preferences. Within this rubric, we concentrate on two area of study; field experiments that (i) speak to the valuation of non-market goods and (ii) explore the origins of the WTA-WTP disparity.

We focus on these areas as they are central to environmental policy making. The ability to estimate the total value of non-market goods lays the groundwork for evaluating proposed regulations and represents the basis for damage assessment. Yet, there is a growing body of laboratory work suggesting that many individuals make choices violating the assumption that preferences are stable and consistent. For academics and policy-makers alike, these results are problematic and call into question the theoretical foundations of welfare economics. However, we review a series of field experiments highlighting that such concerns may be overblown. When investigated within a population familiar with the trading institution, behavior converges to neoclassical benchmarks – particularly as trading experience intensifies.

We conclude by reviewing a burgeoning literature exploring the effectiveness of dynamic pricing plans and non-pecuniary strategies such as normative appeals and tailored information as a means to manage the consumption of energy and water. Results from this literature suggest that both strategies are effective ways to manage demand. Moreover, these studies highlight complementarities between pecuniary and non-

[1] For an overview of lab experiments in Environmental and Resource Economics we refer the interested reader to the excellent survey articles by Cason (2010), Harrison (2006) or Sturm and Weimann (2006). For an overview of the use of quasi-experiments within this realm, the interested reader should see Greenstone and Gayer (2009).

pecuniary based policy measures. Whereas the former are most effective amongst lower income and low use households, the latter have greatest impact on high income households and larger user groups.

For academics, such studies are notable as they get to the heart of the externalities issue and foster a deeper understanding of the individual behaviors that generate public goods (bads). By elucidating the various influences that drive such actions, such studies highlight what models best predict behavior and outline directions for new theories. For policy makers, such studies are invaluable as they provide a blueprint that outlines ways to use insights from behavioral economics to promote policy goals.

Before proceeding, we would like to note that we have in no way attempted to review the voluminous literature to which field experiments are beginning to add. Rather we limit attention to studies that speak directly to the design and evaluation of environmental policy, particularly as viewed through the lens of the individual agent or consumer. Further, within these topics we discuss a limited number of papers that underscore what we view to be a central advantage of field experiments – the ability to examine behavior in naturally occurring settings with self-selected agents that vary in both experience and familiarity with the underlying trading institution.

The remainder of our study proceeds as follows. Section II defines the various field experiment types, and how they represent an interesting middling ground between the lab and observation data. Section III discusses work that affects benefit cost analysis, with a specific focus on preference elicitation and testing prospect theory. Section IV focuses on incentive schemes to promote the conservation of scarce resources. Section V concludes with directions for future use of field experiments in environmental and resource economics.

II. What is a Field Experiment?

Since we have defined the species 'field experiment' several times previously (see, e.g., Harrison and List, 2004 or List and Rasul, 2009), we draw upon that work heavily here to provide a quick sketch outlining how field experiments differ from more traditional means of measuring empirical relations. For our well-versed readership, we would advise skipping directly to Section III.

A fundamental challenge facing researchers who wish to estimate the causal effect of some action or policy is the construction of the correct counterfactual. The action of interest is either taken or it is not. The researcher is thus unable to observe what would have happened in the absence of treatment or if another action had been taken. Yet, it is possible to observe outcomes for similar others – or control group – who were not treated. Field experiments build upon the experimental model of the physical sciences as a means to create valid control groups. They provide a bridge between laboratory and naturally-occurring data in that they use randomization in naturally-occurring settings as an instrument to facilitate causal identification. In this regard, field experiments are a complement to laboratory and quasi-experimental approaches.

The Classification of Field Experiments: A Simple Taxonomy

Harrison and List (2004) propose six factors that can be used to determine the field context of an experiment and use these factors to classify field experiments into three categories: artefactual, framed, and natural. Figure 1 shows how these three types of field experiments compare and contrast with laboratory experiments and the analysis of naturally occurring data. On the far left in Figure 1 are laboratory experiments, which make use of randomization to identify a treatment effect of interest. The right-most part of the empirical spectrum in Figure 1 includes examples of empirical approaches such as instrumental variables, regression discontinuities, and propensity score matching that require making identification assumptions to identify treatment effects from naturally-occurring data. Between these endpoints are field experiments.

The most minor departure from the typical laboratory experiment is the "artefactual" field experiment, which mimics a lab experiment except that it uses "non-standard" subjects, typically experimental participants from the market of interest. Early contributions within the area of environmental economics in this genre include Bohm's (1972) seminal work comparing how willingness to pay for a sneak preview of a Swedish television show differs when the activity is purely hypothetical versus when the payment and sneak preview will actually occur.

Moving closer to how naturally-occurring data are generated, Harrison and List (2004) denote a "framed field experiment" as a field experiment that incorporates important elements of the context of the naturally occurring environment with respect to

the commodity, task, stakes, and information set of the subjects. Subjects in such experiments often know about the randomization and/or are aware of the study via a survey that is used to generate information for policy purposes.[2]

Both artefactual and framed field experiments are conducted in an environment in which subjects are keenly aware that their behavior is being monitored, recorded, and subsequently scrutinized. Decades of research within psychology highlight the power of the role obligations of being an experimental subject, the power of the experimenter herself, and the experimental situation (see Orne, 1962). This leads to our final field experiment type—"natural field experiments".

Natural field experiments are those experiments conducted in environments where subjects naturally undertake the desired task and do not know that they are participants in an experiment. Therefore, they neither know that they are being randomized into treatment nor that their behavior is subsequently scrutinized. While we restrict attention to studies that focus on topics germane to environmental and resource economics, natural field experiments have been used to answer a broad range of topics in economics.[3]

Further Considerations and the Limits of Field Experiments: Some Important Caveats

Considering the differences between field experimentation and other empirical methods, it is important to discuss some potential obstacles that arise when conducting field experiments. An important shortcoming of field experiments vis-à-vis laboratory experiments is the relative difficulty of replication. A fundamental advantage of the experimental approach is the ability of others to reproduce the study and independently verify its results.

Following List and Rasul (2009), we consider three levels at which replication can operate; (i) taking the actual data generated by an experiment and reanalyzing the data; (ii) running an experiment that follows a similar protocol but employs a new subject pool, and (iii) testing the hypotheses of the study using a new research design. Laboratory experiments lend themselves to replication in all three dimensions. While the same is true for many artefactual and framed field experiments, the second type of

[2] Social experiments and randomized control trials in the realm of development economics are prominent examples of framed field experiments. Over the past decade, such experiments have grown in importance and represent a very active area of research.

[3] We refer the interested reader to Harrison and List (2004), Levitt and List (2009), List and Rasul (2009), or Bandiera et al. (2011) for a more general discussion of field experiments and their use in economics.

replication is much more difficult when considering natural field experiments which are often opportunistic and require cooperation of outside entities. As such, it may be difficult to find opportunities to re-run the original study using a new pool of subjects.

A related concern is the 'external validity' of any given field experiment. When designed around the evaluation of a particular policy, field experiments are relatively easy for policy makers to understand. Yet, the simplicity of the approach often comes at a cost – they are designed to identify reduced form causal effects. This limits the extent to which such studies can be used to predict how outcomes will evolve over time or how similar interventions would impact other individuals or groups.

For example, Allcott and Mullainathan (2011) explore the generalizability of site-specific treatment effects and show the presence of a "partner selection bias" reflecting the selection of sites/firms willing to partner with researchers to implement large-scale field experiments. Using data from a series of field experiments designed to promote energy conservation, they show unexplained variation in treatment effects across sites that is both economically and statistically significant.

Finally, it is worth noting that while primary data collection is a key element of field experimentation, this raises the costs of entry and therefore limit the number of practitioners. Similarly, there are many instances where the nature of the research question is not amenable to field experimentation. In such cases, the controlled environment of the laboratory is an ideal starting point for inquiry.

III. Field Experiments that Inform Benefit Cost Analysis

One hallmark of public policy decision-making is a comparison of the benefits and costs associated with proposed regulations. A necessary condition underlying benefit-cost analysis is the ability to accurately estimate the total value of the affected goods and services. For commodities traded in the marketplace, prices provide a direct signal of value making the valuation task straightforward. Unfortunately, the task is more daunting for the practitioner interested in estimating the total benefits of nonmarket goods and services such as improved air or water quality. In such instances, policy makers frequently rely on stated preference methods to provide signals of value.[4] While

[4] Of course stated preference methods are not the only approach that one can use to value such changes. Other approaches based on revealed preferences – i.e., hedonic pricing or the measurement of expenditures

stated preference methods are literally the "only game in town" when it comes to measuring the total value of nonmarket goods, critics argue that contingent surveys are unreliable as the hypothetical nature of the approach allows respondents to distort statements of value without penalty.

Understanding whether and why people distort their actual preferences when asked a hypothetical question remains a fundamental issue facing environmental economists. Fortunately, a robust literature measuring the nature and extent of hypothetical bias and exploring methodologies to attenuate this tendency has emerged.[5] As a whole, this body of work highlights that statements of value are sensitive to both the mode of elicitation and the way in which survey questions are presented/implemented.

Although no single strategy is a panacea, the experimental literature makes clear that institutions matter. Respondents consider the costs and benefits of distorting their preferences when providing statements of value. Below we review of a body of literature that outlines conditions under which one would expect stated preference methods to elicit "true" preferences and provides practitioners a blueprint for mitigating strategic distortions.

Aligning Hypothetical and Real Statements of Value: The Role of Cheap Talk

Cummings et al. (1995) and Cummings and Taylor (1999) present evidence from laboratory experiments suggesting that hypothetical bias can be mitigated by utilizing an *ex ante* design they refer to as a "cheap talk" scheme. The underlying premise behind the "cheap talk" design is to induce truthful preference revelation by making hypothetical bias an integral part of survey questionnaires. Such scripts describe hypothetical bias, note its commonality in surveys, and discuss underlying reasons why it might occur. Moreover, the script asks subjects to consider this problem and adjust their response to the valuation questions.

---

on averting behaviors – have also been used to estimate such values. We refer the interested reader to Sugden (2005) for a nice discussion of when and why revealed preference approaches are preferable to stated preference methods.

[5] We refer the interested reader to Harrison (2006) who provides a critical review of laboratory experiments designed to assess stated preference methods and various strategies to mitigate or calibrate hypothetical bias. We agree with Harrison (2006) that one should carefully evaluate the inference drawn from this line of work and weigh the totality of the empirical evidence when designing stated preference studies to elicit homegrown values.

This section summarizes a number of empirical applications that extend the findings of Cummings et al. (1995) and Cummings and Taylor (1999) in field settings. The earliest such study is the framed field experiment of List (2001) comparing bids from a second-price auction for a 1982 *Topps Traded* Cal Ripken, Jr. baseball card across three treatments; hypothetical, hypothetical with cheap talk, and actual second-price auctions. All treatments were conducted on the floor of a sportscard show and employed actual market participants – either professional sporstcard dealers or ordinary consumers. For the sample of nondealers, List (2001) highlights an important difference in behavior across treatments. While the average hypothetical bid is statistically different than the average bid in either the hypothetical with cheap talk or actual auction treatment, there is no significant difference in average bids across the cheap talk and actual treatments.

Carlson et al. (2005) extend this line of inquiry using a novel framed field experiment to examine the impact of cheap talk on response in choice experiments (CE). Under the CE approach, hypothetical bias can occur at two levels: i) the decision to purchase and ii) the intra-buy decision (i.e., conditional on purchasing, the marginal value vector). Examining data from a sample of Swedish adults who received a choice experiment concerning the purchase of two goods – chicken and ground beef – the authors find evidence suggesting that cheap talk impacts marginal values. Of the ten attributes included in their study, seven are found to be valued significantly less amongst the subset of respondents randomly assigned a version of the survey containing a cheap talk script.[6]

Despite this evidence, the success of cheap talk in mitigating hypothetical bias is far from universal. A number of studies find that cheap talk is only effective amongst inexperienced subjects or those who are unfamiliar with the good being valued. For example, while Lusk (2003) finds that cheap talk eliminated bias amongst ordinary consumers, there is no such evidence when considering the sample of knowledgeable consumers. Similar insights are reported in a number of other framed field experiments, see e.g., List (2001), Aadland and Caplan (2003, 2006), and Blumenschein et al. (2008).

[6] List, Sinha and Taylor (2006) report similar results and find little difference in the estimated marginal values across respondents randomly assigned a cheap talk script and those facing actual purchase decisions.

Taken as a whole, these studies provide mixed support for the effectiveness of cheap talk. Yet, the observed data patterns highlight two factors that appear critically linked with the ultimate success of cheap talk scripts: (i) respondents' familiarity with the good being valued and (ii) the information content and length of the cheap talk script. As such, we believe this literature provides a playbook that outlines both conditions under which cheap talk may provide an effective method to overcome hypothetical bias and conditions under which the researcher should explore other alternatives. Importantly, such scripts hold promise when respondents are unfamiliar with the good being valued *and* the researcher can provide information on both the expected direction and magnitude of hypothetical bias.

Aligning Statements of Value: The Role of Consequentialism

While cheap talk has garnered much attention in the literature, scholars have explored other ways to attenuate hypothetical bias. In field settings, it is commonplace to present respondents with realistic scenarios. Accordingly, it is reasonable to assume that individuals place varying weight on the likelihood that their responses will influence public policy. Carson, Groves, and Machina (2000) suggest that such survey designs, which they denote as "consequential", will induce subjects to truthfully reveal preferences. Intuitively, if respondents believe that that their responses have the potential to influence policy measures, there is no incentive to distort behavior and misrepresent preferences.

Cummings and Taylor (1998) provide the earliest experimental test of a "consequential" survey design. In their framed field experiment, subjects had the opportunity to vote in a referendum to finance the production and distribution of a Citizens Guide by the Southwest Research and Information Center.[7] Experimental treatments varied the probability that, if passed, the referendum would bind and require actual payment by the subjects. For treatments that employ low levels of probability ($p \leq 0.50$) to link voting behavior with actual economic commitment, respondents are significantly more likely to vote "Yes" than what is observed in a binding referendum.

[7] The Citizens Guide was distributed to low income, Hispanic families living in an area of Albuquerque, New Mexico where groundwater supplies had been contaminated by toxic substances. The purpose of the Citizen's Guide was to identify areas with contaminated groundwater, advise residents how to have their water tested at no cost, and outline different actions available to residents with contaminated wells.

But, at a higher probability level (p = 0.75), the authors are unable to distinguish voting behavior from that observed in the binding referendum.

Landry and List (2007) extend this analysis to systematically compare value statements obtained via "cheap talk" and "consequential" treatments. In the consequential treatment, subjects were informed that a coin flip would determine whether votes in the referendum would prove binding. Empirical results suggest the effectiveness of a "consequential" survey design – across all price levels the proportion of "Yes" votes in the "consequential" treatments are statistically indistinguishable from those in both the real and "cheap talk" treatments.

While the results of these studies are promising, the theory of consequentialism suggests that subjects should truthfully report preferences for probabilities as low as ε. A number of recent studies – e.g., Bulte et al. (2005), Carson et al. (2006), or Herriges et al. (2010) – find support for this invariance result. For example, Herriges et al. (2010) use data from the 2005 Iowa Lakes Survey to explore the causal impact on WTP of the perceived degree of consequentiality.[8] In their natural field experiment, a subset of individuals were randomly sent a magazine article indicating that results from previous surveys had influenced policy decisions at the state level. Noting that the receipt of information was positively correlated with perceived consequentiality, Herriges et al. (2010) estimate the "causal" impact of consequentiality and find support for the invariance result. Amongst those reporting the survey to be at least minimally consequential, there was no difference in underlying WTP distributions. Yet, WTP was significantly lower amongst those reporting the survey to be completely inconsequential.

Vossler et al. (2012) examine the import of consequentiality in the context of repeated choice experiments. They develop a game theoretic model providing conditions under which such surveys are incentive compatible and test the predictions of the model using a framed field experiment to elicit values for planting riparian buffers in agricultural areas of Quebec. Experimental results are largely consistent with theory and suggest the importance of consequentiality. Although the WTP distribution elicited via the stated preference treatment differed significantly from those elicited in real payment

[8] The Iowa Lakes Project was a four-year study designed to understand recreational use and the value of water quality for 130 lakes throughout Iowa. The 2005 survey included a supplemental question for 2000 households eliciting the extent to which they believed results of the study would affect public policy.

treatments, the observed differences are attenuated if one conditions on the belief that choices had more than a "weak" chance of influencing policy.

Combined this body of work suggests that individuals respond to incentives when formulating statements of value. When surveys are incentive compatible and perceived consequential, it appears as if respondents truthfully reveal preferences. However, when the response to a survey question is perceived inconsequential or has the possibility of affecting an outside, respondents may strategically distort preferences or fail to commit the cognitive resources required for a considered response. This suggests that practitioners should take great care in choosing the elicitation scheme when performing benefit cost analysis – institutions matter.

Aligning Statements of Value: Social Isolation and Interviewer Effects

The NOAA panel on contingent valuation (Arrow et al., 1993) recommends in-person interviews over either phone or mail surveys when implementing CV studies. While there are undoubtedly benefits to such an approach, there is ample evidence from the behavioral literature that individuals are more cooperative when interacting with others of a like social grouping (see, e.g., Devine, 1989; Fershtman and Gneezy, 2001; Andreoni and Petrie, 2008). Similarly, it is well documented that respondents may seek to distort answers to survey questions to please the interviewer or maintain consistency with societal norms (see, e.g., Atkin and Chaffee, 1972-1973; Campbell, 1981; Cotter et al., 1982; Finkel et al., 1991; Fisher, 1993; Davis, 1997; Krosnick, 1999). It is thus important to recognize that respondents in CV studies may be influenced by the presence and characteristics of the surveyor.

List et al. (2004) examine whether the manner in which contingent surveys are administered affects stated preferences. In their framed field experiment, nearly 300 subjects were randomly assigned to one of six treatment cells and asked to vote on whether to contribute $20 to provide start-up capital for the Center for Environmental Policy Analysis at the University of Central Florida. Experimental treatments vary subject anonymity and whether decisions are hypothetical or have real economic consequence.[9] Experimental results suggest that differences in actual voting decisions

[9] List et al. (2004) use a randomized response technique to promote anonymity and relax the degree of social pressure a subject faces when answering the stated preference question. As noted in Harrison (2006)

across treatments varying social isolation are similar in magnitude to those observed across hypothetical and real treatments.

Alpizar et al. (2008) extend this line of inquiry to examine the effect of anonymity on charitable donations in support of Poas National Park (PNP) in Costa Rica. Subjects in their natural field experiment were international tourists visiting the park that completed an interview and were asked to donate to the PNP. Experimental treatments varied whether contributions were made anonymously and placed in a ballot box or if they were registered by an interviewer. Empirical results highlight the important influence of social anonymity – average donations were approximately 25 percent higher when made in front of an interviewer.

A related line of inquiry examines the impact of interviewer effects on estimated statements of value for non-market goods. Leggett et al. (2003) assess interviewer bias in the context of face-to-face versus self-administered surveys. Using a split-sample contingent valuation survey of visitors to Fort Sumter National Monument, they highlight behavior consonant with social desirability bias – estimated WTP for a fort visit is approximately 23-29 percent lower when the survey is self-administered rather than conducted via an in-person interview.

A more recent set of framed field experiments set forth to decompose such effects by controlling various aspects of the interviewer-respondent interaction. For example, Bateman and Mawby (2004) examine the impact of interviewer appearance on stated willingness to pay and find that WTP was approximately 66.8 – 79.1 percent higher amongst respondents approached by an interviewer dressed in formal clothing. Loureiro and Lotade (2005) find that WTP for eco-labeled products are approximately 128.6 to 177.9 percent greater amongst subjects approached by an interviewer from a region that produces the products. Gong and Aadland (2009) find that monthly WTP for curbside recycling was approximately 7 – 8% higher amongst respondents interviewed by a Caucasian or a woman.

---

the use of such technique is not beyond reproach. While it preserves the secrecy provided by ballot boxes in the field, it can introduce ambiguity over the resolution of the referendum. To the extent that subjects are ambiguity averse, the randomized response technique could thus influence choice through channels other than its effect on anonymity. However, Alpizar et al. (2008) find stark differences in contributions made in front of an interviewer and those made via a ballot box. Such differences highlight the importance of anonymity and rule out ambiguity as the sole driver of treatment effects in the List et al. (2004) study.

As a whole, this literature highlights that statements of value are sensitive to the mode of elicitation and characteristics of those eliciting the value statement. Fortunately, there are a number of ways to control for and mitigate these effects. For instance, one can exploit variation in both the mode of elicitation and observable characteristics of the interviewer to identify and net out such effects. Alternately, one can *ex ante* attempt to minimize such influences through the use of a cheap talk script or consequential survey design.

Preference Anomalies – The Value Disparity

It has been more than four decades since researchers discovered that the WTP measure of value differed starkly from the WTA measure (see, e.g., Hammack and Brown, 1974). Initially, most economists believed that these results were a survey artifact and argued that WTA estimates should not be treated seriously (Kahneman, 1986). Despite these misgivings, Kahneman et al. (1990) provide strong evidence to reject the neoclassical postulate that preferences between two goods are independent of current entitlements.

Environmental Economics may be the branch of economics most affected by this research. For example, when losses associated with changes in the status quo cost consumers significantly more than the gains associated with these changes, the decision on whether to use compensating or equivalent variation measures is of central import (Knetsch, 1990). More generally, the "WTA/WTP disparity" calls into question the applicability of Hicksian theory and the legitimacy of cost-benefit analysis. Moreover, the value disparity changes the procedure necessary to resolve damage disputes. Below, we focus on field experiments in this area.[10]

WTA/WTP Evidence from the Field

One early question for field experimentalists was "do experienced subjects display less WTA/WTP disparity than their inexperienced counterparts?" In a series of framed field experiments, List (2003; 2004a; 2004b) probes this very question using protocols similar to those of Knetsch (1989) and Kahneman et al. (1990). List's studies

[10] The WTA/WTP disparity and associated implications for Hicksian theory has motivated the experimental work of many scholars such as Knetsch (1989) and Bateman et al. (1997). For an excellent overview of the literature exploring the value disparity in the lab, we refer the interested reader to Plott and Zeiler (2005) who provide results suggesting that institutions and a subject's familiarity with the trading protocol are fundamentally related to the severity of the WTA/WTP gap.

can be split into four categories: i) examining *trading* patterns of "familiar" goods, ii) examining *trading* patterns of "unfamiliar" goods, iii) examining *bidding* patterns for "familiar" goods, and iv) examining *bidding* patterns for "unfamiliar" goods.

In the "familiar" goods trading experiments, subjects were randomly endowed with unique memorabilia and subsequent trading rates examined. In these situations, subjects are familiar with both the trading environment and the traded goods. Empirical results from these studies highlight an important caveat on the earlier literature on the WTP/WTA disparity – institutions and experience matter. Observed behavior becomes increasingly neoclassical as trading experience intensifies. Amongst the sample of professional dealers and experienced non-dealers, trading rates and final holdings are independent of initial endowment.

Although promising, the results from these studies raise a natural question - do the observed patterns of choice hold when the good is unfamiliar? To separate the role of experience in a market from experience with a good, subjects were endowed with an "unfamiliar" good – either a mug or a candy bar. Since psychological research suggests that transfer of learning across situations is quite weak, this exercise represents a particularly strict test of the role of market experience on shaping choices (Loewenstein, 1999). Results *again* highlight that individual behavior converges to the neoclassical prediction as trading experience intensifies.

Data on bidding patterns for "familiar" goods and "unfamiliar" goods were gathered in the same manner as the trading data. However, in these treatments, WTP and WTA measures were elicited using either a random *n*th price auction or a Becker-DeGroot-Marschak discrete-choice auction. For both types of goods the data suggest that individual behavior converges to the neoclassical prediction as trading experience intensifies. Decomposing this result by separately evaluating WTA and WTP measures, the data suggest a potential channel through which experience impacts the value disparity. Whereas there is no difference in WTP across consumer types, more experienced subjects state significantly *lower* WTA figures than inexperienced counterparts.

Viewed in their totality, these data suggest that perhaps the main effect of endowment is not to enhance the appeal of the good one owns but rather the "pain" of

giving it up (Loewenstein and Kahneman, 1991). *Ex ante,* agents may over-estimate the cost they will incur from giving up a good (and so state a high WTA). Through market interactions, agents may come to realize that the pain associated with a loss is not as great as initially imagined and learn to take advantage of arbitrage opportunities.[11] Here psychological effects explain both the economic anomaly and its attenuation.

Zhou and Kling (2001, 2004) provide an alternate rationale for the value disparity – the presence of commitment costs and asymmetric beliefs about market opportunities. For goods with uncertain value, WTP and WTA will reflect compensation for the fact that one could learn that good has a different value than what was initially believed at the time of purchase (sale) and the associated cost of reversing the initial transaction. Even slight asymmetries in the perceptions about these costs across buyers and sellers can lead to considerable divergence between WTA and WTP. Kling et al. (2010) report data from a series of framed field experiments that lend support to the commitment cost story. Subjects placed in the role of perspective buyer believe it more difficult to trade in the outside market than those placed in the role of a perspective seller – a difference that is mitigated as market experience intensifies.

As a whole, this literature highlights that concerns regarding the stability and consistency of preferences may be overblown. Behavior converges towards neoclassical predictions when investigated within a population of experienced agents familiar with the trading institution. For academics and practitioners alike, these results underscore an important caveat on the earlier literature on the WTP/WTA disparity – institutions and experience matter.[12] As such, the theoretical foundations of welfare economics may be more stable than some would surmise.

---

[11] This line of thought is consonant with recent findings in health and behavioral economics, where studies oftentimes report that individuals are better at adapting to tragic loss of a limb or divorce, for example, than they predicted *ex ante*.

[12] Although the thrust of List's results on experience have been broadly replicated in both the laboratory and field (see, e.g., Feng and Seasholes, 2005; Kermer et al., 2006; Dhar and Zhu, 2006; Munro and De Sousa, 2008; Greenwood and Nagel, 2009; Gachter et al., 2009; Choe and Eom, 2009; Engelmann and Hollard, 2010; Seru et al., 2010) one may be concerned with the endogeneity of market experience. Hence, while List's work attempts to parse treatment (market experience) from selection, the results rely on his modeling assumptions. List (2011) attempts to address this issue by exogenously inducing market experience and provides results that are qualitatively similar to those reported herein. While promising, we believe that more work is needed in this area before one can conclude that it is treatment rather than selection that drives the attenuation of the value disparity.

IV. Promoting Conservation Efforts

Most travelers have been confronted with a strategically placed card in a hotel washroom urging them to protect the environment by reusing their towels. Such efforts are consistent with a growing trend of employing norm based messages and social comparisons to influence individual decision-making. Such strategies build upon Festinger's (1954) social comparison theory which posits that individuals validate the appropriateness of an action through comparisons to others. In this section, we summarize a growing body of work that uses field experiments to examine the effectiveness of normative appeals and targeted information as a way to manage the consumption of energy/water and meet our climate policy goals.

A broad body of work within the social psychology literature examines the use of social-norm marketing, feedback, and tailored information campaigns to promote environmental conservation. Amongst this literature, the work by Schultz et al. (2007) has proven most influential it pilots an approach for promoting household energy conservation that was subsequently adoption by OPOWER.[13] Their study found that combining normative messages detailing the energy use of one's neighbors with injunctive messages – emoticons ☹ and ☺ - generated significant reductions in energy consumption while mitigating the so-called boomerang effect.[14]

Given the scope of OPOWER's operations and the increased popularity of Energy Efficiency Resource Standards – policies that require utilities to promote and document reductions in energy use – a recent body of literature evaluating the effectiveness of various OPOWER programs has emerged. Allcott (2011) evaluates data from seventeen natural field experiments targeting more than 600,000 residential households randomly assigned to either a treatment group, which received home energy reports, or a control group.[15] Point estimates for the Average Treatment Effect across the seventeen programs

[13] OPOWER is a company that helps utilities meet their efficiency goals through the use of targeted messages designed to promote reductions in household energy use

[14] The boomerang effect refers to the phenomenon whereby informing individuals of typical peer behavior inadvertently inspires those who have been under-estimating the prevalence of an activity to inadvertently increase undesired behavior.

[15] The Home Energy Report was a multiple page letter that included a Social Comparison Module detailing the household's electricity consumption over the past twelve months to both the mean and 20th percentile of its comparison group and an Action Steps Module that suggested ways in which the household could conserve energy.

suggest an approximate 1.4 to 3.3 percent reduction in average monthly energy consumption relative to the control households. The estimated effects imply that households in the treatment group conserved 0.62 kilowatt-hours of electricity per day – the equivalent of approximately 10.4 hours of 60-watt light bulb use.

Ayres et al. (2009) analyze data from two large-scale, natural field experiments conducted by OPOWER in conjunction with the Sacramento Municipal Utility District and Puget Sound Energy. Taken jointly, data from the two field studies provide evidence consonant with Allcott (2011); properly framed peer comparisons have the ability to affect energy conservation. However, such effects are more pronounced amongst the highest user groups and depend on the frequency of messaging.[16] Empirical results from the PSE experiment suggest that treatment primarily impacts day-to-day patterns of use rather than promoting investments in energy saving technologies – nearly 38 percent of the observed reductions in use are manifest on Sunday and Monday.

Costa and Kahn (2010) re-analyze the SMUD data from Ayres et al. (2009) and compare the effect of this information on consumption patterns for environmentally experienced subjects—those that had given money to environmental non-profits before—with the effect on non-college aged Republicans. They find that while environmentally experienced subjects reduced consumption versus the control group, Republicans actually increased their consumption. Such heterogeneity is noteworthy and suggests that one need to be careful when using normative messages to influence behavior – there is no one-size fits all approach. Effective messages should adjust the content of appeals to account for differences in ideology or norms across groups to minimize unintended behavioral response.

Ferraro and Price (2013) examine the effectiveness of normative messages as a means to manage residential water demand. In conjunction with the Cobb County Water System (CCWS), they implement a natural field experiment targeting more than 100,000 residential households. Experimental treatments implement one of three commonly employed conservation strategies; (i) the dissemination of information on behavioral and

[16] Allcott (2011) reports similar differences in the effectiveness of reports delivered monthly versus those delivered quarterly - the Average Treatment Effect for the monthly treatment group was approximately one-third greater than that observed amongst households receiving quarterly reports. Moreover, the relative treatment effect in the quarterly treatment group was statistically lower in the 2nd and 3rd months after receiving the report suggesting that the effect of such messages tends to wane over time.

technological modifications, (ii) appeals to pro-social preferences, and (iii) the provision of social comparisons to enhance appeals to pro-social preferences. Empirical results highlight that technical advice has but a small impact on water use – consumption falls by approximately 1 percent. However, augmenting technical advice to include pro-social appeals or social comparisons generate substantially larger reductions; particularly amongst high use households.

Allcott and Rogers (2012) extend this line of research to examine whether norm-based messages influence behavior in the long-run. Using data from an OPOWER program that has been running continuously since 2008, they explore both within month variation in electricity use across treatment and control households and the long-run persistence of treatment effects. The empirical results suggest a pattern of within month action and backsliding – households in the treatment group reduce use within days of receiving the home energy report but the response decays quickly. Over time, this pattern is attenuated as the immediate decrease in usage becomes smaller and the rate of date over the course of the month becomes indistinguishable from zero.

To explore the persistence of treatment effects, the authors exploit the fact that approximately 12,000 randomly selected households in the treatment group stopped receiving the home energy reports after two-years of intervention. Amongst the set of households that received reports throughout the entire sample, the estimated treatment effects grow throughout the four-year period. For the group whose reports were discontinued, the estimated treatment effects decays but does so at a rate that is orders of magnitude slower than that observed in the initial months of the program. Taken jointly, this suggests that households in the treatment group develop a "habit" for conservation in the sense of Becker and Murphy (1988) and that the resulting change in "capital" leads to persistent conservation efforts.[17]

Taken jointly, this body of literature highlights the importance of moral payoffs and norms on consumption decisions. Framing conservation as a normative behavior and providing salience to the norm by including comparisons to like others are powerful tools

[17] Ferraro et al. (2011) show similar effects. Using data from Ferraro and Price (2013) on initial treatment assignment, they explore post-treatment usage over the period 2007-2009 for households in the original CCWS experiment. Empirical results suggest that while appeals to pro-social preferences and social comparisons affect short-term patterns of use, only messages augmented with social comparisons have a lasting impact on water demand.

to manage residential demand for energy and water; especially amongst high user groups. In this regard, policies based on messages targeting the "why" and "how much" of conservation may prove a useful complement to pecuniary measures as they are most effective amongst those who are least sensitive to price changes. For practitioners, the lesson learned from this literature is clear – norms matter.

Managing the Peak-Load Problem: Dynamic Pricing Experiments

While policies based on normative appeals and social comparisons have garnered much attention in the literature, scholars have explored a number of other mechanisms to promote conservation efforts. For example, economists have long recognized the promise of dynamic pricing strategies such as "peak load" or "real-time" pricing as a means to manage to manage consumption during periods when the marginal cost of production is high. In this section we summarize a growing body of work that uses field experiments to examine the effectiveness of various dynamic pricing schemes.

Wolak (2006) evaluates data from a critical peak pricing experiment involving 123 residential consumers of the City of Anaheim Public Utilities (APU). Participants were randomly assigned to either a treatment or control group and received a "smart" meter that recorded consumption over 15-minute intervals. Control group customers were charged according to APU's prevailing increasing-block, fixed-price schedule. Customers in the treatment group paid the same tariff except during peak hours on critical peak pricing (CPP) days where they received a rebate of 35 cents/KWh for reductions in consumption relative to a reference level – the average of the three highest peak consumption levels for the consumer over all non-CPP days.

Empirical results highlight the promise of such pricing plans – households in the treatment group consume approximately 12 percent less electricity during peak hours on CPP days than counterparts in the control group. However, Wolak uncovers a perverse effect accounting for roughly half of the estimated treatment effect. Households in the treatment group significantly increase use during peak period of non-CPP days relative to counterparts in the control during similar periods.[18]

[18] Such differences suggest that households in the treatment group were attempting to distort their reference consumption level in order to obtain higher rebates during CPP days.

Wolak (2011) extends this earlier work to consider a broader array of dynamic pricing plans – hourly pricing, critical peak pricing, or critical peak pricing with a rebate – on electricity use for a representative sample of 1,245 residential consumers throughout Washington, DC.[19] Empirical results confirm prior findings on dynamic pricing plans – treated customers reduce electricity use during high-priced periods (peak events). However, the average treatment effect for the critical peak pricing treatment is significantly greater (13 versus 5.3 percent) than that for the critical peak pricing with rebate treatment. Although not discussed in Wolak (2011), the differences across theses these treatments are consonant with a growing behavioral literature highlighting that incentives framed as losses (penalties) loom larger than those framed as gains (rebates).

Allcott (2010) evaluates data from a real-time pricing experiment involving 693 households in and around Chicago. Relative to the same months of the previous year, Allcott (2010) finds that households facing hourly prices determined by day-ahead prices on the wholesale electricity market reduced consumption by about 10 percent; more than twice that observed amongst counterparts in the control group. Interestingly, conservation effects are significantly enhanced following High Price Alerts triggered whenever the day-ahead wholesale price exceeded 10 cents/KWh.

Pushing this line of inquiry further, Jessoe and Rapson (2012) use data from a real time pricing experiment to explore the effect of information feedback on the price elasticity of demand. In their study, a subset of households facing exogenous price changes during peak events was given an in-home display that provided real-time feedback on the price and quantity of electricity consumed. While households exposed to real-time prices reduced demand by up to seven percent, those provided real-time feedback on use demonstrated reductions in the range of 8 to 22 percent – a three standard deviation increase in the average treatment effect.

Ida et al. (2012) use data from a randomized field experiment to examine the relative effectiveness of economic and non-economic incentives on peak demand. Consumers in their study were randomly assigned to treatment groups that either faced a

[19] Households in the CPP treatment faced a reduced block schedule but paid an additional 78 cents/KWh during critical peak events. Households in the CPP with rebate treatment faced the prevailing block schedule but received rebates for reductions in consumption during CPP events. And those in the hourly pricing treatment were charged according to prices that tracked the day-ahead wholesale market for the District of Columbia.

dynamic pricing scheme that varied the marginal price of electricity during critical peak periods or a received an appeal calling for conservation during critical peak periods. Results from the study highlight an important asymmetry in the effect of the various incentive schemes – economic incentives have a greater impact amongst low income households whereas conservation warnings are most effective amongst higher-income groups.

Taken jointly, this body of literature highlights the promise of dynamic pricing plans as a means to promote energy conservation and the reduction of greenhouse gases. However, such effects are dampened by incomplete information – consumers rarely observe the amount of energy consumed at any point in time and are thus uncertain about the associated marginal costs/benefits of their actions. Hence, providing consumers real time feedback on prices and use is an effective way to increase the sensitivity to prices. In this regard, policies designed to promote the uptake of in-home electricity displays and/or disseminate information on real-time energy use should be viewed as complements to pecuniary strategies that rely upon financial incentives to influence demand; salience matters.

V. Lessons Learned and Next Steps

Before outlining what we view as fruitful avenues for future research, we would like to summarize the important lessons learned from the extant literature. Within the context of non-market valuation we believe the most important lesson learned is that institutions matter. Practitioners should thus take care in choosing an elicitation scheme as statements of value are sensitive to both the mode of elicitation and characteristics of those eliciting the value.

Within the context of the value disparity and the resulting applicability of cost-benefit analysis, the most important lesson learned is that experience and institutions matter. When investigated within a population of experienced agents familiar with the underlying trading institution, behavior converges to neoclassical predictions. Hence, concerns regarding the stability of preferences and the relevance of Hicksian theory are overblown.

Within the context of energy and water conservation, we believe the most important lessons learned are that norms and saliency matter. Targeted messages that

frame conservation as a normative behavior through pro-social appeals or comparisons with like others are powerful tools to manage residential consumption. Likewise dynamic pricing plans that increase the costs of consumption during periods of high demand are promising ways to influence consumption. However, the effects of price based policies are dampened due to incomplete. Making consumption (and its cost) salient by providing consumers real-time feedback on use therefore serves to increase the sensitivity to prices and enhance the effectiveness of pecuniary based strategies.

Given these lessons learned, what do we view as fruitful avenues for future research in the area of environmental and resource economics? Methodologically, we envision that artefactual, framed, and natural field experiments designed to identify one type of behavior can be used to explain or predict non-experimental outcomes. For example, it might be efficient to explore how individual decisions in simple experimental games relate to land use and land use change. Perhaps agents who behave most selfishly in common pool resource games tend to make aggressive harvesting decisions in the field? If so, it can be instructive to devise mechanisms in artefactual and framed field experiments that garner cooperation from these types. One can then take these mechanisms to the field.

An early example of this is Carpenter and Seki (2006), who use an artefactual field experiment to explore the determinants of individual contributions in a standard public goods game among workers within the fishing industry of one particular Japanese community. They report that individual contributions in the public goods games are higher for those individuals who face less on-the-job competition in their workplace.

Another methodological advance lies in the use of field experiments as a means to test bed the design of new environmental markets or regulatory policies. To date, experimentalists have relied almost exclusively upon laboratory settings for such studies (see, e.g., Cason, 1995; Cason and Plott, 1996; Cason and Gangadharan, 2005; Murphy and Strandlund, 2007; Suter et al., 2010). Yet, as noted in Cason (2010), field experiments may provide qualitatively different insights regarding market design and the relative performance of different trading institutions. In this regard, we view the work of Poe, Suter and Vossler (2010) comparing the behavior of student subjects and agricultural professionals in an ambient-tax experiment a model for future work.

In terms of topic areas, the nonmarket valuation discussion highlights that there is still much to learn about how people formulate values when responding to CV questions. For example, future work should focus on furthering our understanding of the underlying causes of observed biases. A theoretical model of such biases would be most welcome, not only to place the results into perspective but to guide future field experiments. In addition, exploring new mechanisms to minimize hypothetical bias (and other biases) represents valuable research. In this manner, the inferred valuation approach of Lusk and Norwood (2009) holds promise.

Within the area of resource economics, we see tremendous opportunities to explore mechanisms revolving around land use and land use change issues. What are the best incentive schemes for the policymaker to utilize to achieve her goals? Are there more cost effective methods? We can begin to understand these, and related questions, by running field experiments. In this regard, we view the framed field experiment of Jack (2011) comparing alternate mechanisms to allocate tree planting contracts in Malawi as an important first step.

Similarly, we see significant promise for continued work at the intersection of development, health, and environmental economics. In particular, we view randomized field trials designed to promote investments in safe drinking water (Kremer et al., 2011) or reduce individual exposure to contaminated groundwater (Bennear et al., 2010) as natural complements to the research outlined in Section IV.

Another important area of future research is making better use of behavioral economics to promote our policy goals. For example, consider the power of defaults. Economists have found dramatic effects of using a default option: most people stick to the default rather than choosing other available options. Lofgren et al (2009) use an arefactual field experiment to explore the power of defaults within the context of an important question for environmental economists. They report that individual experience with environmental questions changes the sensitivity of subjects to a default when deciding whether or not to offset the CO_2 emissions from their air transit.

A related use of behavioral economics as a means to promote our policy objectives is to explore the use of goal setting as a way to reduce energy consumption amongst present-biased consumers with reference-dependent preferences. In this regard,

we view the natural field experiment of Harding and Hsiaw (2012) an important foundation upon which to build.

Finally, we see promise in studies that explore the use of social comparisons and normative appeals as a means to promote the adoption of green technologies. As noted in Section IV, such strategies have proven an effective way to manage residential water/energy consumption. Yet, whether and to what extent such programs can be used to overcome the "energy paradox" and promote the adoption of green technologies remains an open empirical question. In this regard, we view the natural field experiment of Herberich et al. (2011) comparing the relative impact of price reductions and normative appeals on the decision to purchase compact fluorescent light bulbs as an important first step.

References

Aadlan, David and Arthur J. Caplan, "Willingness to Pay for Curbside Recycling with Detection and Mitigation of Hypothetical Bias," *American Journal of Agricultural Economics*, 85 (2003), pp. 492 – 502.

Aadlan, David and Arthur J. Caplan, "Cheap Talk Reconsidered: New Evidence from CVM," *Journal of Economic Behavior and Organization*, 60 (2006), pp. 562 – 578.

Allcott, Hunt, "Social Norms and Energy Conservation," *Journal of Public Economics*, 95 (2011), pp. 1082-1095.

Allcott, Hunt, "Rethinking Real-Time Electricity Pricing," (2010) MIT Working Paper.

Allcott, Hunt and Sendhil Mullainathan, "External Validity and Partner Selection Bias," (2011) Working Paper, New York University.

Allcott, Hunt and Todd Rogers, "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation," (2012) NBER Working Paper 18492.

Alpizar, Francisco, Frederik Carlsson, and Olof Johansson-Stenman, "Anonymity, Reciprocity, and Conformity: Evidence from Voluntary Contributions to a National Park in Costa Rica," *Journal of Public Economics*, 92 (2008), pp. 1047 – 1060.

Andreoni, James and Ragan Petrie, "Beauty, Gender, and Stereotypes: Evidence from Laboratory Experiments," *Journal of Economic Psychology*, 29 (2008), pp. 73 – 93.

Arrow, Kenneth, Robert Solow, Edward Leamer, Paul Portney, Roy Radner, and Howard Schuman, "Natural Resource Damage Assessments under the Oil Pollution Act of 1990," Federal Register, 58 (1993), pp. 4601 – 4614.

Atkin, B. A. and S.H. Chaffee, "Instrumental Response Strategies in Opinion Interviews," *Public Opinion Quarterly*, 36 (1972-1973), pp. 69 – 79.

Ayres, Ian, Sophie Raseman, and Alice Shih, "Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage," (2009) NBER working paper 15386.

Bandiera, Oriana, Iwan Baranky, and Imran Rasul, "Field Experiments with Firms," *Journal of Economic Perspectives*, 25 (2011), pp. 63-82.

Bateman, Ian J. and James Mawby, "First Impressions Count: Interviewer Appearance and Information Effects in Stated Preference Studies," *Ecological Economics*, 49 (2004), pp. 47 – 55.

Bateman, I.J., Munro, A., Rhodes, B., Starmer, C. and Sugden, R. "A test of the theory of reference-dependent preferences," *Quarterly Journal of Economics*, 112 (1997), pp. 479-505.

Becker, Gary and Kevin Murphy, "A Theory of Rational Addiction" Journal of Political Economy, 96 (1988), pp. 675-700.

Bennear, Lori, Alessandro Tarozzi, Alexander Pfaff, H.B. Soumya, Kazi M. Ahmed, and Alexander van Geen, "Bright Lines, Risk Beliefs, and Risk Avoidance: Evidence from a Randomized Intervention in Bangladesh," (2010) working paper Duke University.

Blumenschein, Karen, Glenn C. Blomquist, Magnus Johannesson, Nancy Horn, and Patricia Freeman, "Eliciting Willingness to Pay Without Bias: Evidence from a Field Experiment," *The Economic Journal*, 118 (2008), pp. 114 – 137.

Bohm, P., "Estimating the Demand for Public Goods: An Experiment," *European Economic Review*, 3 (1972), pp. 111-130.

Bulte, Erwin, Shelby Gerking, John A. List, and Aart de Zeeuw, "The Effect of Varying the Causes of Environmental Problems on Stated WTP Values: Evidence from a Field Study," *Journal of Environmental Economics and Management*, 49 (2005), pp. 330 – 342.

Campbell, B.A., "Race-of-Interviewer Effects Among Southern Adolescents," *Public Opinion Quarterly*, 45 (1981), pp. 231 – 244.

Carlsson, Fredrik, Peter Frykblom, and Carl Johan Lagerkvist, "Using Cheap Talk as a Test of Validity in Choice Experiments," *Economics Letters*, 89 (2005), pp. 147 – 152.

Carpenter, J. P. and E. Seki, "Competitive Work Environments and Social Preferences:

Field Experimental Evidence from a Japanese Fishing Community," *The B.E. Journal of Economic Analysis & Policy*, Berkeley Electronic Press, vol. 0(2) (2006).

Carson, Richard T., Theodore Groves, and Mark J. Machina, (2000), "Incentive and Informational Properties of Preference Questions," Paper presented at the Kobe Conference on Theory and Application of Environmental Valuation, Kobe: Kobe University, January.

Carson, Richard T., Theodore Groves, and John A. List, (2006), "Probabilistic Influence and Supplemental Benefits: A Field Test of the Two Key Assumptions Behind Using Stated Preferences," working paper University of California, San Diego.

Cason, Timothy N., "An Experimental Investigation of the Seller Incentives in EPA's Emission Trading Auction," *American Economic Review*, 85 (1995), pp. 905-922.

Cason, Timothy N., "What Can Laboratory Experiments Teach Us About Emissions Permit Market Design?," *Agricultural and Resource Economics Review*, 39 (April 2010), pp. 151-161.

Cason, Timothy N. and Lata Gangadharan, "A Laboratory Comparison of Uniform and Discriminative Price Auctions for Reducing Non-Point Source Pollution," *Land Economics*, 81 (2005), pp. 51-70.

Cason, Timothy N. and Charles Plott, "EPA's New Emissions Trading Mechanism: A Laboratory Evaluation," *Journal of Environmental Economics and Management*, 30 (1996), pp. 133-160.

Costa, Dora L. and Matthew E. Kahn, "Energy Conservation 'Nudges' and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment," (2010) NBER Workign Paper No. 15939.

Cotter, P., J. Cohen, and P.B. Coulter, "Race of Interviewer Effects in Telephone Interviews," *Public Opinion Quarterly*, 46 (1982), pp. 278 – 294.

Cummings, Ronald G., Glenn W. Harrison, and Laura Osborne, "Can the Bias of Contingent Valuation be Reduced? Evidence from the Laboratory," Economics Working Paper B-95-03, Dividing of Research, College of Business Administration, University of South Carolina, 1995.

Cummings, Ronald G. and Laura O. Taylor, "Does Realism Matter in Contingent Valuation Surveys?" *Land Economics*, 74 (1998), pp. 203 – 215.

Cummings, Ronald G. and Laura O. Taylor, "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method," *American Economic Review*, 89 (1999), pp. 649 – 665.

Davis, D.W., "Nonrandom Measurement Error and Race of Interviewer Effects among African Americans," *Public Opinion Quarterly*, 61 (1997), pp. 183 – 207.

Devine, P., "Stereotypes and Prejudice: Their Automatic and Controlled Components," *Journal of Personality and Social Psychology*, 56 (1989), pp. 5 – 18.

Ferraro, Paul and Michael K. Price, "Using Non-Pecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment," *Review of Economics and Statistics*, 95 (2013), pp. 64-73.

Ferraro, Paul J., Juan Jose Miranda, and Michael K. Price, "The Persistence of Treatment Effects with Norm-Based Policy Instruments: Evidence from a Randomized Environmental Policy Experiment," *American Economic Review Papers and Proceedings*, 101 (2011): pp. 318-322.

Fershtman, C. and U. Gneezy, "Discrimination in a Segmented Society: An Experimental Approach," *Quarterly Journal of Economics*, 116 (2001), pp. 351-377.

Festinger, L., "A Theory of Social Comparison Processes," *Human Relations*, 7 (1954), pp. 117 – 140.

Finkel, Steven E., Thomas Guterbock and Marian J. Borg, "Race-of-Interviewer Effects in a Pre-election Poll Virginia 1989," *Public Opinion Quarterly*, 55 (1991), pp. 313 – 330.

Fisher, R. J., "Social Desirability Bias and the Validity of Indirect Questioning," *Journal of Consumer Research*, 20 (1993), pp. 303 – 315.

Gong, Min and David Aadland, "Interviewer Effects in an Environmental Valuation Telephone Survey," (2009) working paper University of Wyoming.

Greenstone, Michael and Ted Gayer, "Quasi-Experimental and Experimental Approaches to Environmental Economics," *Journal of Environmental Economics and Management*, 57 (2009), pp. 21-44.

Hammack, Judd and Gardner Brown, *Waterfowl and Wetlands: Toward Bio-Economic Analysis*, (1974), Baltimore: Johns Hopkins University Press.

Harding, Matthew and Alice Hsiaw, "Goal Setting and Energy Efficiency." (2012) working paper, Stanford University.

Harrison, Glenn W., "Experimental Evidence on Alternative Environmental Valuation Methods," *Environmental and Resource Economics*, 34 (2006), pp. 125-162.

Harrison, Glenn W. and John A. List, "Field Experiments," *Journal of Economic Literature*, 42 (2004), pp. 1009 – 1055.

Herberich, David, John A. List, and Michael K. Price, "How Many Economists Does it Take To Change a Light Bulb? A Natural Field Experiment on Technology Adoption," Working Paper, University of Chicago, Department of Economics, (2011).

Herriges, Joseph, Catherine Kling, Chih-Chen Liu, and Justin Tobias, "What Are the Consequences of Consequentiality?" *Journal of Environmental Economics and Management*, 59 (2010), pp. 67 – 81.

Ida, Takanori, Koichiro Ito, and Makoto Tanaka, "Using Dynamic Electricity Pricing to Address Energy Crises: Evidence from Randomized Field Experiments," Working Paper, Stanford Institute for Economic Policy Research, (2012).

Jack, B. Kelsey, "Allocation in Environmental Markets: A Field Experiment with Tree Planting Contracts in Malawi," Working Paper, Tufts University (2011).

Jessoe, Katrina and David Rapson, "Knowledge is (Less) Power: Experimental Evidence from Residential Energy Use," (2012) Working Paper, University of California, Davis.

Kahneman, Daniel, "Comments," in *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*, edited by Ronald G. Cummings, David S. Brookshire and William D. Schulze. Totow, NJ: Rowman and Allanheld, 1986.

Kahneman, D., Knetsch, J.L., and R.H. Thaler, "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, 98 (1990), pp. 1325-1348.

Kling, Catherine L., John A. List, and Jinhua Zhao, "A Dynamic Explanation of the Willingness to Pay and Willingness to Accept Disparity," (2010) NBER Working Paper No. 16483.

Knetsch, Jack L., "The Endowment Effect and Evidence of Nonreversible Indifference Curves," *American Economic Review*, 79 (1989), pp. 1277-84.

Knetsch, Jack L., "Environmental Policy Implications of Disparities between Willingness to Pay and Compensation Demanded Measures of Values," *Journal of Environmental Economics and Management*, 18 (1990), pp. 227-237.

Kremer, Michael, Jessica Leino, Edward Miqueal, and Alix P. Zwane, "Spring Cleaning: Rural Water Impacts, Valuation and Property Rights Institutions," *Quarterly Journal of Economics*, 126 (2011), pp. 145-205.

Krosnick, J.A., "Maximizing Measurement Quality: Principles of Good Questionnaire Design," in *Measures of Political Attitudes*, ed. J.P. Robinson, P.R. Shaver, and L.S. Wrightsman. (1999) New York: Academic Press.

Landry, Craig E. and John A. List, "Using Ex Ante Approaches to Obtain Credible Signals for Value in Contingent Markets: Evidence from the Field," *American Journal of Agricultural Economics*, 89 (2007), pp. 420-429.

Leggett, Christopher G.,Naomi S. Kleckner, Kevin J. Boyle, John W. Duffield, and Robert Cameron Mitchell, "Social Desirability Bias in Contingent Valuation Surveys Administered through In-Person Interviews," *Land Economics*, 79 (2003), pp. 561 – 575.

Levitt, Steven D. and John A. List, "Field Experiments in Economics: The Past, The Present, and The Future," *European Economic Review*, 53 (2009), pp. 1-18.

List, John A., "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards," *American Economic Review*, 91 (2001), pp. 1498 – 1507.

List, John A., "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, 118 (2003), pp. 41-71

List, John A., "Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace," *Econometrica*, 72 (2004a), pp. 615-625

List, John A., "Substitutability, experience, and the value disparity: evidence from the marketplace," *Journal of Environmental Economics and Management,* (2004b), 47(3), pp. 486-509.

List, John A., Robert P. Berrens, Alok K. Bohara, and Joe Kerkvliet, "Examining the Role of Social Isolation on Stated Preferences," *American Economic Review*, 94 (2004), pp. 741 – 752.

List, John A. and Imran Rasul, "Field Experiments in Labor Economics," *Handbook of Labor Economics*, Volume 4, edited by Orley Ashenfelter and David Card, Elsevier, (2011), pp. 104-228.

List, John A., Paramita Sinha, and Michael H. Taylor, "Using Choice Experiments to Value Non-Market Goods and Services: Evidence from Field Experiments," *The B.E. Journal of Economic Analysis and Policy*, Advances, Vol. 6: Issue 2 (2006), Article 2.

Loewenstein, G. "Experimental economics from the vantage-point of behavioural econmics." *Economic Journal,* 109 (1999), F25-F34.

Loewenstein, G. and Kahneman, D, "Explaining the Endowment Effect," 1991 working paper, Department of Social and Decision Sciences, Carnegie-Mellon University.

Löfgren, Å., P. Martinsson, M. Hennlock, and T. Sterner, "Does experience eliminate the effect of a default option? - A field experiment on CO2-offsetting for air transport," Working Papers in Economics 391 (2009), Göteborg University, Department of Economics.

Loureiro, Maria L. and Justus Lotade, "Interviewer Effects on the Valuation of Goods with Ethical and Environmental Attributes," *Environmental and Resource Economics*, 30 (2005), pp. 49 – 72.

Lusk, Jayson L., "Willingness-to-Pay for Golden Rice," *American Journal of Agricultural Economics*, 85 (2003), pp. 840 – 856.

Lusk, Jayson L., and Bailey Norwood," Bridging the Gap between Laboratory Experiments and Naturally Occurring Markets: An Inferred Valuation Method," *Journal of Environmental Economics and Management*, 58 (2009), pp. 236-250.

Murphy, James and John Stranlund, "A Laboratory Investigation of Compliance Behavior under Tradable Emissions Rights: Implications for Targeted

Enforcement," *Journal of Environmental Economics and Management*, 53 (2007), pp. 196-212.

Orne, Martin T., "On the Social Psychology of the Psychology Experiment: With Particular Reference to Demand Characteristics and Their Implications," *American Psychologist*, 17 (1962), pp. 776-783.

Plott, Charles R. and Kathryn Zeiler, "The Willingness to Pay – Willingness to Accept Gap, the 'Endowment Effect', Subject Misconceptions, and Experimental Procedures for Eliciting Values," *American Economic Review*, 95 (2005), pp. 530-545.

Poe, Gregory L., Jordan F. Suter, and Christian A. Vossler, "External Validity of Ambient-Based Pollution Control Experiments: A Comparison of Student Participants and Agricultural Professionals," 2009 Working Paper, Cornell University.

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius, "The Constructive, Destructive, and Reconstructive Power of Social Norms," *Psychological Science*, 18 (2007), pp. 429 – 434.

Sturm, Bodo and Joachim Weimann, "Experiments in Environmental Economics and Some Close Relatives," *Journal of Economic Surveys*, 20 (2006), pp. 419-457.

Sugden, Robert, "Anomalies and Stated Preference Techniques: A Framework for a Discussion of Coping Strategies," *Environmental and Resource Economics*, 32 (2005), pp. 1-12.

Suter, Jordan F., Kathleen Segerson, Christian A. Vossler, and Gregory L. Poe, "Voluntary-Threat Approaches to Reduce Ambient Water Pollution," *American Journal of Agricultural Economics*, 92 (2010), pp. 1195-1213.

Vossler, Christian A., Maurice Doyon, and Daniel Rondeau, "Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Field Experiments," *American Economic Journal: Microeconomics*, 4 (2012), pp. 145-171.

Wolak, Frank A., "Residential Customer Response to Real-Time Pricing: The Anaheim Critical-Peak Pricing Experiment," (2006) Stanford University Working Paper.

Wolak, Frank A., "Do Residential Customers Respond to Hourly Price? Evidence from a Dynamic Pricing Experiment," *American Economic Review: Papers & Proceedings*, 101 (2011), pp. 83-87.

Zhao, Jinhua and Catherine L. Kling, "A New Explanation for the WTP/WTA Disparity," *Economics Letters*, 73 (2001), pp. 293-300.

Zhao, Jinhua and Catherine L. Kling, "Willingness to Pay, Compensating Variation, and the Cost of Commitment," *Economic Inquiry*, 42 (2004), pp. 503-517.

Figure 1: A Field Experiment Bridge

Controlled Data			Naturally-Occurring Data
Lab	AFE	FFE	NFE, NE, PSM, IV, STR

- Lab: Lab experiment
- AFE: Artefactual field experiment
- FFE: Framed field experiment
- NFE: Natural field experiment
- NE: Natural experiment
- PSM: Propensity score estimation
- IV: Instrumental variables estimation
- STR: Structural modeling