MONASH BUSINESS SCHOOL

DEPARTMENT OF ECONOMICS

ISSN 1441-5429

DISCUSSION PAPER 34/16

Behavioral Insights from Field Experiments in Environmental Economics

Daniel A. Brent*[a], Lana Friesen[b], Lata Gangadharan[c], and Andreas Leibbrandt[c]

Abstract:

Many environmental decisions are based on intrinsic motivations in addition to traditional economic incentives. Field experiments allow researchers to isolate a specific causal mechanism which can help advance our understanding of consumer and firm behavior in environmental markets. This article summarizes the literature on the use of field experiments in environmental economics, focusing on framed and artefactual field experiments as well as natural experiments targeting municipal energy and water demand. We set out a theoretical framework to improve the interpretation of results from field experiments in environmental economics. In addition to providing an overview of experimental methods and findings we also lay out a set of challenges for researchers interested in running a field experiment in environmental economics.

Key words: field experiments; environmental economics, intrinsic incentives, extrinsic incentives, behavioral economics
JEL classification: C93, Q50

*Funding from the Cooperative Research Centre for Water Sensitive Cities (CRC grant number 20110044) is acknowledged.
[a]Department of Economics, Louisiana State University
[b]School of Economics, University of Queensland
[c]Department of Economics, Monash University

1 Introduction

The use of experiments has grown exponentially in many areas of economics including environmental economics. Experiments provide powerful and reasonably low cost ways to analyze contemporary environmental policy issues. During the last three decades laboratory experiments were propelled by the fact that high quality data required for assessment of policy initiatives in the environmental area were often not available. Several important insights from this body of work were generated, in diverse areas such as compliance with environmental regulations and the importance of specific trading features in environmental markets. Multiple survey papers summarize the influence of laboratory experiments in addressing environmental problems (Friesen and Gangadharan, 2013; Cason, 2010; Normann and Ricciuti, 2009; Cherry et al., 2007; Sturm and Weimann, 2006; Bohm, 2003; Cummings et al., 2001; Muller and Mestelman, 1998).

In recent years field experiments have been used as complements to laboratory methods in the environmental area. In contrast to laboratory experiments, field experiments are conducted in the participant's natural everyday environment and often without their knowledge. These experiments range from artefactual experiments to framed experiments and natural experiments, including randomized control trials. Field experiments have two important desirable features: they help reach conclusions that account for the specific environment in which environmental choices are made and still allow for causal inferences. This level of specificity, however, sometimes makes it difficult to extrapolate findings from one setting to another. Thus, to get a comprehensive understanding of which environmental policies are likely to "work" in many different environments and why, it is important to compare and combine the insights generated from the many new field experiments in environmental economics.

This review summarizes recent research using field experiments in environmental economics and outlines the theoretical background for a burgeoning set of experiments that incorporate both extrinsic (financial) and intrinsic (pro-social) incentives. We highlight

how the different types of experiments complement each other by honing in on the causal mechanism. Our review builds on and complements existing work studying field experiments in environmental economics by John List and Michael Price (List and Price, 2013; Price, 2014). In particular, we focus our review around the theme of intrinsic and extrinsic incentives, which is an extremely active research area in environmental economics. We also incorporate leading theoretical models of incentives in order to ground the empirical findings in economic theory and discuss the challenge of extrapolating the results of field experiments. While List and Price (2013) and Price (2014) focus on natural field experiments, and field experiments involving stated preference methods, we highlight the significant literature using artefactual field experiments to understand resource management. Lastly, we also update the set of articles in this rapidly expanding field.

The link between theory and the field is a particularly important consideration in environmental economics where consumers respond to traditional economic incentives, but also make choices based on contributions to the public good of environmental protection. Depending on the context and environment, firms and individuals generate different amounts of pollution, use excessive energy or water, or mis-report on their usage/emissions. For example, theoretical frameworks predict that the degree of environmental consciousness, social pressure and social conformity can play an important role for public good contributions. Therefore, it is important to account for these variables to understand differences in empirical findings from field experiments conducted in different settings. Viewing the field experimental findings in light of theory is of crucial importance for policymakers and industry leaders who want to understand and anticipate how a certain policy intervention will work for their constituents and consumers. By incorporating the theoretical background with empirical evidence from popular field experiments in environmental economics we shed light on the mechanisms and their functioning.

The paper is structured as follows. The next section provides a description of the experimental methodology. Section 3 summarizes the recent literature on framed and artefactual field experiments with local environmental users. Section 4 describes recent

experiments in residential water and energy consumption in natural settings. A theoretical background for these results is outlined in Section 5. Section 6 presents challenges in linking theory to field experiments and in running field experiments in practice.

2 Experimental Methods

The experimental method has played a crucial role in environmental protection. In the 1970s, a group of scientists including Paul J. Crutzen, Mario J. Molina, and F. Sherwood Rowland (1995 Nobel Prize Winners in Chemistry) proved that chlorofluorocarbon (CFC) gases used in refrigerators, plastic foams and spray bottles damage the ozone layer. While their insights were based on the theoretical understanding of chemical processes and modeling techniques, only experiments could conclusively identify and measure the causal impact of CFC on ozone layer depletion (Parson, 2003).

The causal understanding generated from experiments is not constrained to understanding chemical processes. We also rely on experiments involving humans to develop new medicines and better theories on decision-making. More recently, environmental economists have employed the experimental method to generate knowledge on various behaviors affecting our environment, such as energy conservation, common pool extraction, and pollution.

In this section, we describe the experimental method, particularly emphasizing the role of field experiments and contrasting this with standard laboratory experiments. In addition, we summarize the different types of field experiments (Harrison and List, 2004). Researchers, who want to estimate how policies or other actions have affected environmental behaviors, typically face the challenge of finding the right counterfactual or, in the language of experimentalists, control group. If policies were uniformly implemented, often the only alternative is to compare environmental behaviors pre- and post-intervention and abstract from, or attempt to account for, relevant changes that occur during the time span that is analyzed. If policies were locally implemented, researchers often compare locations where they were implemented to areas where policies were not implemented,

assuming there are no differences in these locations or using some form of propensity score matching (Rosenbaum and Rubin, 1983; Abadie and Imbens, 2006). In both scenarios the causal interpretation of policy impacts on environmental behaviors can be severely misleading as unobservable factors, that may or may not be induced by a particular policy, can be responsible for an observed policy-environmental behavioral relationship.

Experiments have a crucial advantage over other empirical methods because the counterfactual, or control group, is randomly assigned. Thus, unobservable factors are assumed to be equally distributed across the treatment and control groups and cannot drive group differences. For example, using the experimental method to test the impact of a policy on environmental behavior would imply that only a random selection of resource users are treated with the policy. In this manner, the researcher can compare those resource users to the other resource users who did not face the policy and conclude that the difference is the result of the policy.

2.1 Types of Field Experiments

Field experiments have become a popular method that complements other empirical methods and laboratory experiments. They serve as a bridge between naturally occurring and laboratory data because they use randomization into treatments and take part in the natural environment of the behavior that we want to predict. While the large majority of experiments in the natural and social sciences are conducted in the laboratory, field experiments have proved to be particularly useful in improving our understanding of environmental behaviors.

Field experiments can be categorized into three main types: artefactual, framed, and natural (Harrison and List, 2004). All three types can be viewed as methodologies that lie between laboratory experiments and naturally occurring data. While not all researchers agree on this categorization and the distinctions between these three types are sometimes difficult to draw, they differ in the extent to which they include elements of the field and reduce elements of the laboratory.

Artefactual field experiments include some aspects of the field but still carry most ele-

ments of the laboratory. They are laboratory experiments that use experimental subjects from the environment of interest instead of the typical subject pool (i.e. university students). This distinction can be crucial as there may be important differences in subject pools, and incentives for students in a classic laboratory experiment are very different than for individuals whose livelihoods depend on environmental choices and natural resources. In addition, environmental variables may affect individuals and if we want to understand this relationship we need to investigate subjects who were and were not exposed to a particular variable. Recent examples (also discussed in Section 3) are Leibbrandt et al. (2013); Gneezy et al. (2015) who study fishermen at a lake and compare them to fishermen at the sea to show how the environment has formed their cooperative and competitive inclinations.

Framed field experiments incorporate field context with respect to the commodity, task, stakes, and information set of the subjects. They typically study subjects from the environment of interest. For example, Noussair et al. (2015) study recreational fishermen and their effort to catch fish in a social dilemma experiment. A key feature of both artefactual and framed field experiments is that subjects are aware that they participate in an experiment and that their behavior is scrutinized.

Natural field experiments, in contrast, are conducted without the knowledge of the subjects. Thus, one main advantage of natural field experiments is that the observed behavior cannot be the result of experimenter/interviewer demand effects (Orne, 1962). Natural field experiments are conducted in the natural environment in which the behavior of interest takes place using the subject pool from the environment of interest. Schultz et al. (2007), for example, use natural field experiments to test the use of normative and injunctive messages for promoting household energy conservation.

3 Framed and Artefactual Field Experiments

In this section, we describe the insights from research that has taken the Lab to the Field and conducted framed and artefactual field experiments. Researchers have used

the natural settings of actual decision makers and invited them to be participants in their experiments, and this innovative approach has allowed them to examine behavioral patterns of environmental and natural resource users in the field. This growing body of research helps us identify social norms, financial incentives and ecological and institutional frameworks that can impact resource use.

Many environmental goods and services, such as fisheries, rainforests and resource management, have features associated with common pool resources and public goods. While pure public goods are characterized by non-excludability and non-rivalry in consumption, several environmental public goods can be considered impure public goods. These goods may not be directly rivalrous, but the difficulty of excluding people from using these resources, and the fact that consumption of the resource by one person reduces the amount available to others, imply that these resource users face a typical social dilemma. Social dilemmas are group interactions in which an individual maximizes her own payoff when she does not cooperate, but attaining the social optimum requires cooperation. Several approaches have been suggested to mitigate these problems, some focus on external intervention by regulators and others emphasize the importance of community led approaches. Researchers, such as Nobel prize-winner Elinor Ostrom, have argued that if communities are able to design their own usage schemes, organize themselves and enforce the rules they design, then collective action and self-governance can be successful in reducing the impact of these social dilemmas (Baland and Platteau (1996); Ostrom (1990)).

An experimental paradigm commonly used to study social dilemmas is the voluntary contribution mechanism (VCM) (Isaac and Walker, 1988). In this game, each member of a group receives an endowment of money and then the members simultaneously decide to contribute some portion of their endowment to the group account. These contributions benefit all members of the group. The endowment left over goes to the individual's private account. The dominant strategy in these games is to place all of one's endowment in the private account, however for the social optimum to be attained all individuals need to

contribute their entire endowment to the group account. Researchers have therefore interpreted the percentage of the endowment placed in the group account as a measure of cooperation.

Similarly, in the common pool resource game (CPR) (Walker et al., 1990; Ostrom et al., 1992), the amount of resources not extracted from the group account, or the amount individuals chose to leave in the group account can be thought of as an indicator of their cooperativeness. Both the public good and the common pool resource game have been used extensively in the field to study environmental conservation, extraction of natural resources, and public good provision such as environmental management of resources and in this section, we will refer to these games often. In addition to the obvious connection of these games with the natural environment, it is relatively easy for researchers to introduce these games to community groups as most people relate to the differential incentives (individual versus group) embedded in these games. A majority of the research in this literature on artefactual and framed field experiments focuses on natural resource management in developing countries and while social dilemmas in such settings exist the world over, the preponderance of studies in poorer countries suggests that the combination of low incomes and bad governance makes the problem more acute.

3.1 The Institutional and Ecological Framework

The institutional setting that individuals operate under can have a critical impact on behavior in the field, and on their natural resource use. Carpenter and Seki (2006, 2011), conduct public good experiments with a fishing community in Japan to explore this impact. While all the experimental participants were part of the same fishing community, interestingly they were engaged in one of three different jobs. One group fished in teams, another group ran the cooperative shop in the community, and the third group were traders who wholesale the fish to retail outlets. Due to the nature of their job, they were exposed to different levels of on-the job competition and the authors test if this can explain their cooperative behavior in the experiment. They also examine if on-the-job competition can predict the rate of norm enforcement amongst the workers. To do

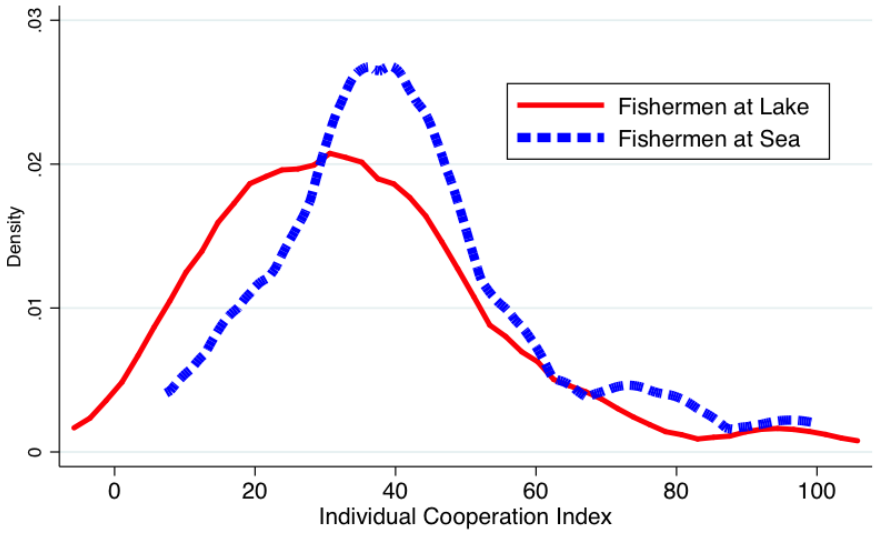
this, they allow participants to send disapproval signs (unhappy faces) to each other at a monetary cost to the sender but not to the receiver, after the first five periods of the experiment. The workers who are less exposed to competition use the disapproval signs more, though the use of these signs is less than the authors predicted. In terms of cooperation, their results are very clear. Incentives that promote competition on the job, and the resulting perceptions of competition by the employees, are associated with lower levels of cooperation between the workers. This matters for productivity as well, as the crews that exhibit higher levels of conditional cooperation are also more productive, in terms of fish caught per unit of effort.

Likewise, Gneezy et al. (2015) observe different kinds of behavior of fishermen depending on ecological and institutional factors. Their participants are also in the same occupation and live in similar environments, yet they are exposed to very different workplace organization that is shown to have a critical effect on their behavior in different experimental games. The organization of the workplace where group activities are encouraged influences norms of cooperation. In their setting in Brazil, fishermen located by a lake fish on their own and cooperate less, whereas fishermen located by the sea who fish in groups show higher cooperation levels in the field experiment. Sea fishermen also trust more, return more money in the trust game, propose more equal offers in the ultimatum game, donate more to a charity and try more often to coordinate in a stag-hunt coordination game than lake fishermen.[1]

[1]The Trust game (Berg et al., 1995) provides a measure of trust and trustworthiness. Participants are randomly assigned to the role of trustor or trustee. Both types receive an endowment. In the first stage of the game, the trustor chose how many monetary units to send to the trustee. The experimenter triples each unit sent. In the second stage, the trustee decides how many units to return. The first stage measures trust and the second stage is an indicator of trustworthiness. In the Ultimatum game (Güth et al., 1982) participants were assigned the role of proposers or responders. The proposer receives an endowment of units and has to decide how to divide this amount between him and the responder. The responder then has the opportunity to accept or reject the offer. If accepted, the proposed share is implemented, if rejected, both participants receive zero units. The donation game (Eckel and Grossman, 1996) is used to measure pro-social behavior towards individuals outside the participants' societies. All participants received ten units and can decide how many of these units to donate to a charity. The stag-hunt game is a simultaneous-move coordination game in which two participants can either choose to hunt a stag or hunt a hare. For example, if both choose to hunt a stag, each gets 10 units, but if one chooses to hunt a stag while the other chooses to hunt a hare, the one choosing to stag earns zero units. If one chooses to hare, his payoffs are seven units regardless of the other participant's decision. There are two Nash equilibria (both players choose hare or both choose stag), but only one of these is

The authors constructed for each participant an index composed of their behavior in the four experiments, with equal weights on each (back-transfers in trust, offers in ultimatum game, donations in donation game, contributions in public goods game). This index is an indication of overall cooperativeness of the participants and Figure 1 shows that the dashed line is shifted to the right of the solid line, which demonstrates that the extent of cooperation is significantly higher for fishermen at the sea in the various games.

Figure 1: Individual Cooperation



Notes: The figure from Gneezy et al. (2015) shows different levels of cooperation among fisherman who fish in groups (Sea) compared to those who fish alone (Lake).

Prediger et al. (2011), explore the impact of culture and ecology on cooperation levels in common pool resource games. The experiments were conducted in Namibia and South Africa, in communities which share the same traditions and culture, but where they have to cope with different ecological environments. In South Africa, there is a lack of reliable and easily observable markers of ecological change, whereas in Namibia the indicators of degradation are clearly visible in terms of bare soil patches and hence evoke cooperation from resource users. Namibians have a longer history of cooperative grazing, whereas in South Africa, internal conflict and corruption have mitigated cooperation historically. The authors find these institutional differences reflected in their field experiments as well,

Pareto-efficient (both players choose stag). If the players coordinate they can reach the Pareto-efficient equilibrium.

with cooperation levels much higher in Namibia.

3.2 Intrinsic and Extrinsic motivations: Crowding in or Crowding out?

Policy makers at the central and community level use several approaches to improve resource use. Some of the resource usage rules focus on a formal or regulatory approach, and employ financial incentives (subsidies, taxes, audits, penalties) to motivate individuals to use resources in a socially optimal manner. Other approaches aim to evoke intrinsic motivations of resource users. The interaction between such formal and informal schemes is interesting to examine and the field setting can be an ideal environment to explore the complementarity and substitutability between these two approaches, as both the financial and intrinsic motives are unmistakably clear in the field. Lopez et al. (2012), investigate the relative effectiveness of formal regulatory pressure and informal methods such as pro-social emotions in promoting cooperative actions in a public good field experiment in Colombia. The experiments were framed as a situation in which each fisherman decides whether to help clean the beaches and wharves. Keeping the beaches clean can prevent the migration of lobster and other fish, so it benefits everyone. In the treatments with formal regulations, the participants were audited and penalized if found non-compliant. The informal treatments focused on guilt and shame. In the Guilt treatment individuals were privately reminded of their action and its impact on others, while information about an audited individual's contribution decision was publicly revealed to the group in the Shame treatment. While the Guilt treatment was not effective in improving outcomes, Lopez et al. (2012) find that the Shame treatment, in which they reveal identity and contribution of one randomly picked member out of a five member group, significantly increases contributions and this approach is more effective than the regulatory approach.

D'Adda (2011) examines the impact of non-monetary and non-regulatory incentives on pro-social behavior relating to environmental conservation through experiments with users of forest and groundwater resources in Bolivia. Participants were in one of four

treatments and were asked to make decisions on contributions to a reforestation project from an endowment given to them. In the control treatment, called the private treatment, the contribution was anonymous, and the author suggests that this treatment captures the intrinsic valuation of the project by the participants. The other three treatments, add external incentives. In the information treatment, subjects were primed with information about past contributions of other subjects, to create a social norm about appropriate behavior. In the public treatment, subjects announced their contributions to others in the group, and this observability by others can measure the effect of social pressure on conformity with the norm. The punishment treatment allowed subjects to punish others whose contribution behavior they deemed inappropriate. Participants in all treatments also answered a survey question on civic engagement (number of days of community work performed over the previous year) and questions on individualism (e.g. whether independence is the one of the most important qualities to teach to a child). More socially engaged individuals contribute more, and the experimental treatments have a negative impact on socially engaged participants' contributions and a positive impact on the individualistic participant's contributions. Non-monetary and non-regulatory incentives therefore generate different kinds of motivation crowding out effects. Contributions were in general higher in the private treatment as compared to the other external incentives treatments.

Bouma et al. (2008) use experiments and household surveys in rural India to understand the relationship between social capital (community level trust) and the provision of a local semi-public good which specialized in investments in soil and water conservation. They find mixed evidence. While the average amount sent in the trust game is correlated with participation in community resource management, this is only observed in the case of public goods where the costs of effort is fully borne by the household that undertakes these conservation activities. When public good provision is subsidized by government agencies, this correlation is not observed. Introducing financial compensation hence makes private motivations dominant and likely crowds out pro-sociality.

Kerr et al. (2012) examine the impact of the introduction of financial incentive payments on collective action. They focus on the effects of different types of payments on management of a common resource in Mexico and Tanzania. Villagers were invited to participate in pro-social activities such as collecting litter from streets or planting trees. The payment for these activities were at the group (towards the school in the village) or the individual level. The key results from this research are that in the presence of social norms that favor communal work, the compensation regime does not matter. In situations where such social norms did not exist, overall financial payments (both individual and for group) help increase participation in communal activities, though group payments are less favored in villages where corruption is perceived to be high. Similarly, Narloch et al. (2012) examine the extent to which external rewards (such as payments for ecosystem services) crowd-in or crowd-out existing social norms and their effectiveness in increasing conservation levels, in Bolivia and Peru using a public good game with quinoa farmers. The financial rewards interact in complex ways with the social norms and are in general, observed to be beneficial when used as a complement.

3.3 Capturing Heterogeneity of Resource Users

Several researchers collect information about individual and social characteristics of subjects with the aim of understanding how these characteristics may influence their decision making. Non-student subjects allow for a richer, more heterogeneous, and more context-specific subject pool. As they are fully grounded in the environmental backdrop being explored, they may use information not only from the incentives embedded in the experimental treatments, but also bring in knowledge about the actual context and their personal preferences about the context. These multiple layers of information have been used effectively for example, in Cardenas and Ostrom (2004), where the authors conduct common pool resource experiments in rural Colombia and find that these different kinds of information are useful in understanding subject decisions in the experiment. Castillo et al. (2011), use a similar approach in the fishery game they conduct in Colombia and Thailand. They utilize information from experiments, from a collective role playing game

and from surveys. In the experiment, participants have to choose a location (out of two options) to fish and how much effort to put in to fish. They can choose between three resource management rules to solve the social dilemma and these consist of random access (where participants are randomly assigned to a location); rotation of access (in which a location is closed in some periods and inspected) and; a quota on effort (in which only a small fixed amount of effort can be put in each location). Participants vote on these rules. Their results show low levels of cooperation and low efficiency of the management rules. The role-playing games that took into account the local context, showed that fishermen behaved in an individualistic manner in both countries and the surveys suggest that fishermen do not trust local leaders and governmental organizations. Resource sustainability could hence depend on several factors, trust and cooperation amongst fishermen and also their trust in the local institutions. In a companion paper, Janssen et al. (2013) examine the decisions made in a social dilemma framed as a forestry experiment in Colombia and Thailand, and find that after participants voted for three possible forestry management rules, groups harvested less from the common resource. Again social norms, the local context and trust were important in understanding behavior.

Cavalcanti et al. (2013), implement an environmental program in Brazil, where they provide fishing traps that reduce exploitation of the fisheries. They conduct experiments and also collect extensive survey evidence from the participants. They find that integration in a social network (measured by friends in the network) plays a role in improving cooperation in this environmental program. In addition those who participated in the development of this program are more cooperative. Reichhuber et al. (2009), apply different collective tax and subsidy mechanisms (Segerson and Miceli, 1998) that are popular in the area of non-point source pollution to a common pool resource problem in Ethiopia (harvesting of bee-hives). The Pigouvian tax is effective, but the subsidy leads to collusion and inefficient outcomes. Using a design in which all subjects participate in all treatments, they find that, independent of whichever mechanism is applied, there are some individual features that primarily drive decisions. Those who are compliers in one

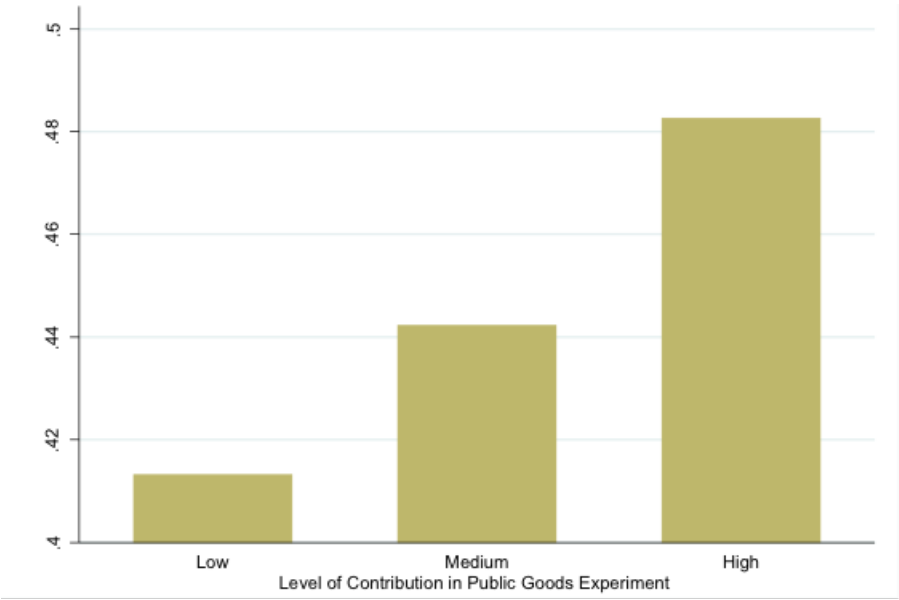
treatment also comply in other treatments.

Hayo and Vollan (2012) find that the socioeconomic heterogeneity of participants in their common pool resource experiment in South Africa and Namibia influenced group interactions to a large degree and can potentially explain why some policies are more effective in supporting cooperation on the commons than others. Voors et al. (2012) collected experimental and survey data in Sierra Leone and find that there is no single simple measure of pro-social preferences. They find little correlation between behavior as measured in a social invention public aid game and a conventional lab-in-the-field public good game. Heterogeneity along several dimensions such as moral and ethical considerations and attitudes towards conservation play a role in determining cooperative behavior.

Fehr and Leibbrandt (2011), also combine experimental and field data from fishermen in Brazil to study the role of cooperativeness and impatience in the fisheries. In addition to conducting a public good game and a time preference game with the fishermen, they collect data on the size of the holes in their shrimp traps. Using a nice overlay of experimental and field data they hypothesize that fishermen who contribute more in the public goods game use bigger holes in the traps, so that small immature shrimp can escape more easily. In a similar vein, fishermen who are impatient in the time preference game are expected to use traps with smaller holes.

They find that more cooperative fishermen use larger holes in their traps, report using fewer traps, and spend less time catching shrimp. As Figure 2 shows, fishermen with the lowest contributions have the smallest hole sizes, followed by fishermen with medium contributions. The fishermen with the highest contributions, who contribute at least half of their endowment, have substantially larger hole sizes than the others. Hence those who are pro-social in the public goods game are also less likely to use fishing instruments that exploit the fisheries. Impatience also has a strong predictive power and those who are more impatient in the time preference game are also more exploitative in their usage of fishing nets.

Figure 2: Contributions to Public Good



The y-axis in the figure from Fehr and Leibbrandt (2011) show different contributions to the public good based on observed differences in the size of holes in shrimp trap in centimeters. Larger holes trap less young shrimp, and thus smaller holes are associated with exploitation of the common pool resource.

3.4 Uncertainty and Resource Use

Decisions about the use of environmental resources and the strategies to be employed to mitigate harm are often intertwined with how individuals and communities perceive risk and uncertainty. Alpizar et al. (2011), conduct a framed field experiment in Costa Rica with coffee farmers and explore the relationship between farmer's adaptation to climate change and their preferences towards risk and ambiguity (preferences of known versus unknown risks). Due to the uncertainties related to climate change, both the status quo of no adaptation and the state of adaptation are associated with risk and ambiguity. The authors argue that ambiguity averse farmers are more likely to adapt to climate change in the presence of unknown risk than when the risk is known. They also explore the willingness of farmers to coordinate with each other such that the costs of adaptation become lower. Subjects participate in several rounds of decision making in the experiment, and in some rounds the risks are known while in others they are unknown. A 1% risk indicates that farmers may experience large losses due to climate change once every 100 years. The risk was varied across farmers. In some rounds, they have information about decisions made by neighboring farmers and they can communicate

with each other to lower adaptation costs. The authors find a high degree of risk aversion in their subject pool and that farmers are more cautious in the presence of ambiguity. Coordination was observed to lower adaptation costs, particularly in the presence of communication opportunities.

Blanco et al. (2015) focus on whether exogenous changes in the availability of a shared natural resource lead to more or less appropriation of the resource. The setting is one of a common pool resource, a social dilemma where individual appropriation and group welfare are negatively related. The experiment is run with peasants in rural Colombia, and the authors aim to test how resource size effects the behavior. They focus on three effects: (1) the magnitude effects of the reduction in resource size, (2) the order effect of intense versus progressive decreases in the size of the resource, and (3) the impact of having a rebound in the resource size after a small decrease. The authors find that the subjects increase approbation from the resource after experiencing a strong reduction in resource availability. Subjects who are in the rebound group, in which the resource returns to the initial size, appropriate more than those who did not experience any change in this resource. One explanation for their findings is that scarcity triggers competitive pressure amongst users. Alternatively, scarcity can impair cognitive reasoning and make subjects more myopic such that they only care about individual payoffs in a single round rather than the higher individual and social payoffs arising from collective action in the entire experiment.

3.5 The Context and Framing

The context of the experiment and the activities participants are engaged in can be an important component of decision making. Stoop et al. (2012) design a novel framed field experiment with recreational fishermen in the Netherlands to create a social dilemma and find that classical game theory has high predictive power in their experiment. Co-operation levels are very low in their public good game, much lower than in the standard laboratory public good game. They conduct several treatments to tease out the reasons for this and conclude that cooperativeness depends on the decision variable, the activity

that has to be modified in order to yield a benefit to the group. In their case recreational fishing is the behavior that needs to be modified, which is difficult since it is an activity they enjoy. Hence the benefits and costs of the action can be an important consideration for individuals.

Framing in terms of how choices are presented can also influence decision making. In Araña and León (2013), conference participants were asked to decide to pay for a carbon offset policy as a part of their travel expenses. Subjects registering for the conference were given an opportunity to pay an extra amount of money for compensating the carbon emissions resulting from their flight and their stay in the conference venue. The authors vary the way in which they request the extra price from the conference participants. In one treatment (opt-out treatment), the default option was that the extra price was already included as part of the conference fee and subjects could choose to opt-out of this. In the opt-in treatment, they had a choice of paying an explicit extra amount. The authors find more participants contribute towards carbon offsets in the opt-out treatment when it is framed as a default option, and argue that support and adoption of environmental policies could be dependent on how these policies are framed in public debate. Status quo and anchoring effects are some reasons for the observed differences in the two treatments.

Bouma et al. (2014), also examine the impact of framing, but find insignificant effects. They compare differences in legitimacy perceptions across villages based on the presence of water use associations to understand legitimacy's relationship to contributions to a public good, in a field experiment in rural India. They examine two aspects of legitimacy: perceived procedural justice and that based on shared vision. While both have some explanatory power, neither has a sufficiently robust effect on contributions.

In summary, research using the methodology of framed and artefactual field experiments has played an important role in identifying several important drivers of optimal resource use. The key lessons we have learnt so far indicate that first, environmental resource users are usually aware of the ecological and institutional constraints they operate under and this needs to be taken into consideration by policy makers to reinforce pro-

social institutions and circumvent the effect of anti-social institutions. Second, incentives, whether extrinsic or intrinsic, can interact in complex and sometimes unpredictable ways with existing norms. Hence deeper understanding of these interactions will help design more effective resource use policies. Third, one of the advantages of bringing the Lab to the Field, is that researchers can explore behavior of actual resource users without abandoning robust experimental procedures, which allow differences across treatments to be isolated without introducing additional confounding factors. Studies discussed in this section have shown that utilizing context specific information, framing, and individual specific information can improve the predictive power of the results and help explain decision making in the field. Fourth, while there is some evidence to the contrary, most studies have found that decisions taken by participants in the experimental games correspond to their day to day resource use, providing methodological validity to the experimental measures of behavior.

4 Natural Field Experiments: Energy and Water Conservation

While the previous section discussed artefactual and framed field experiments, predominantly in developing countries, that focus on natural resource management, this section summarizes a popular sector of predominantly natural field experiments that attempt to address environmental externalities associated with private energy and water consumption in developed countries. There are several features of municipal energy and water demand that attract environmental economists to study household behavior in these markets. Energy consumption produces local and global externalities in the form of air pollution and carbon emissions, while water use reduces water available for environmental flows and creates issues of scarcity for future generations. There is a strong role for economists to influence public policy since firms that supply energy and water to households are often public monopolies or heavily regulated.

A key feature of both of these markets is that both intrinsic and extrinsic motivations

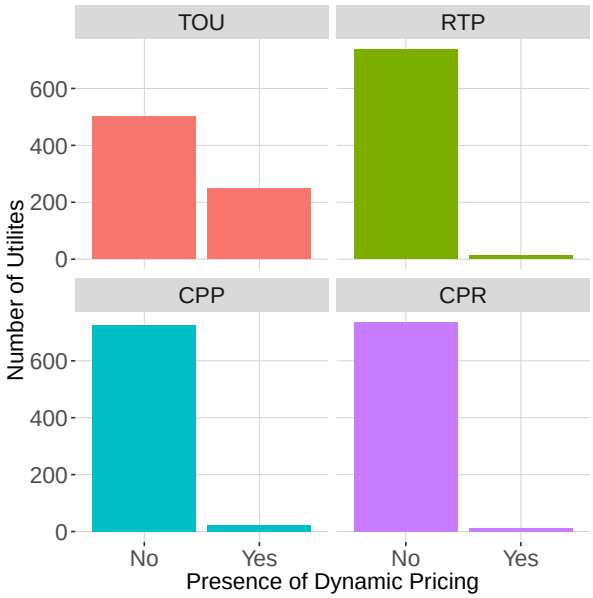
guide final consumption decisions. Households not only conserve energy and water in order to reduce their bills, but may also receive utility from warm glow or altruism associated with the public good of conservation. The dual set of motivations opens up policies beyond traditional economic incentives, such as prices and subsidies, to tap into social or moral motivations to reduce resource consumption. The presence of multiple incentives add complexity to the choice environment, positioning field experiments as an attractive method to disentangle causal mechanisms that drive consumer behavior. This section describes how environmental economists use field experiments to better understand how consumers respond to incentives in markets for municipal energy and water. We focus on field experiments relating to dynamic electricity pricing, utilizing information as a way to influence behavior, and investments in energy and water efficiency. We build on research that describes how field experiments and behavioral economics can be utilized by economists to promote conservation (Shogren and Taylor, 2008; Allcott and Mullainathan, 2010; List and Price, 2013; Price, 2014) by providing an updated review of experiments that focus on issues of intrinsic and extrinsic incentives in resource conservation.

4.1 Dynamic Electricity Pricing

Electricity generation suffers from a "peaking" problem. Supply and demand must be balanced in real time so electricity producers need to maintain infrastructure to supply electricity for peak demand. There are two types of peak demand: the daily peak demand and the annual peak demand, often described as the critical peak demand. Daily peak demand occurs either in the morning when customers wake up and begin using electricity or in the evening when customers return home. The critical peak is the highest 8-10 hours of electricity demand over the course of the year that correspond to the hottest or coldest days of the year when demand for heating or cooling spikes. The marginal cost is substantially higher for supplying peak demand partly due to the opportunity cost of owning and maintaining power plants that are infrequently utilized. These features create incentives for electric utilities to decrease peak demand beyond the concerns over

air pollution and climate change. There are three forms of dynamic pricing that help pass on the time-varying marginal cost of electricity generation to consumers: time of use pricing (TOU) varies prices over the hours of the day with higher prices during peak periods, critical peak pricing/rebates (CPP/CPR) increases prices or provides rebates for conservation during the projected critical peak hours, and real time prices (RTP) allow prices to vary dynamically with the marginal cost of electricity. Figure 3 shows how utilities in the United States are beginning to use various forms of dynamic pricing as of 2014 according to the Energy Information Administration (United States Department of Energy, Energy Information Administration, 2014). While TOU pricing is becoming more prevalent, RTP and incentives to address the critical peak are rare.

Figure 3: Forms of Dynamic Electricity Pricing in the United States (2014)



Note: TOU is time-of-use; RTP is real time pricing; CPP is critical peak pricing; CPR is critical peak rebate.

Dynamic pricing experiments offer some form of dynamic pricing to a treatment group and compare peak demand to a control group on the traditional pricing structure. Initial experiments testing dynamic pricing that began in the 1970s were able to shift consumption to non-peak hours (Caves et al., 1984). The next set of experiments took place from the mid-1990s to the mid-2000s, and offered a wider variety of incentives. Faruqui and Sergici (2010) summarize the results of 15 of these experiments and offer important

conclusions for more recent dynamic pricing policies. Two primary conclusions stand out: CPP is generally more effective at reducing peak demand than TOU, and enabling technology vastly increases the effectiveness of dynamic pricing. Enabling technology is defined as devices such as programmable thermostats or in home displays (IHD) that provide feedback on energy use and/or pricing information that increase the flexibility of response to time-varying prices (Faruqui and Sergici, 2010). Adding enabling technology increases average peak demand reductions from 4% to 26% for TOU and from 17% to 36% for CPP. Several caveats are needed in interpreting these results. First, CPP generally involves much larger price increases compared to TOU, and only takes place over a small number of days. For example in the California Statewide Pricing Pilot, which included both TOU and CPP treatment groups, the peak TOU rate was twice the control group rate while the CPP peak rate was over five times the control group rate. Additionally, as pointed out by Price (2014) many of the studies were opt-in and are therefore subject to selection bias.

Recent experiments aim to further isolate the role information plays in determining the efficacy of dynamic pricing of electricity. In a framed field experiment Jessoe and Rapson (2014) partner with an electric utility that assigns households into one of three groups: (1) a dynamic pricing group that increases prices 250% with a day ahead (DA) notification and 600% for thirty minute (TM) notification, (2) the same dynamic pricing with a real-time IHD, and (3) a control group. All treated customers experience both the DA and TM pricing events and the timing is equivalent across groups. For the DA pricing events the price-only treatment group did not experience statistically significant energy savings whereas the price+IHD group reduced usage by 10%. There was no statistically significant reduction in consumption for the TM treatment in either group. The results highlight that it is important that consumers are not only well informed about prices, but also have knowledge of their consumption in order to maximize the effectiveness of dynamic pricing.

In a working paper Ito et al. (2015) conduct a field experiment testing both social

norms (discussed below) and economic incentives to reduce critical peak demand. This is also a framed field experiment and utilizes the prior literature by providing all customers, including the control group, with an IHD. The economic incentive group receives day ahead and same day notifications that the price will increase by either 160% (40 cents/kWh), 340% (85 cents/kWh), or 420% (105 cents/kWh). The key finding is that while both moral suasion and economic incentives generated statistically significant savings, prices led to greater magnitude and persistence in conservation. Based on survey responses the savings from economic incentives were due to changes in habits as opposed to investments in the capital stock.

4.2 Non-pecuniary Interventions

Economists generally advocate for monetary incentives to correct market failure such as externalities, however, there is often a lack of political will or concerns over equity that limit the implementation of financial incentives to promote resource conservation. This political reality coincides with the growth of applied behavioral economics that uses changes in choice architecture or information nudges to modify consumer behavior without changing prices (Thaler and Sunstein, 2008). A particularly popular tool employed to reduce household energy and water consumption is the social comparison. Social comparisons consist of information sent to customers that compare their water or energy use to the consumption of a peer group. Work by Schultz et al. (2007) spawned a host of interventions utilizing social comparisons including Opower, a publicly listed company with over 95 electric utility companies.[2] Academic research analyzing Opower's early field experiments utilizing social comparisons found statistically significant savings of roughly 2%, with most of the savings coming from large users (Allcott, 2011; Ayres et al., 2012). In a large-scale intervention in a water utility in Cobb County, Georgia Ferraro and Price (2013) also found that social comparisons generated significant treatment effects that were more effective than other information-based treatments such as conservation tips

[2]Opower had 100+ customer based on the company website https://opower.com/our-clients/ as of May 2016.

and a generic moral appeal for conservation. Social comparisons for water conservation also generate more savings among high users (Ferraro and Miranda, 2013; Brent et al., 2015).

At this point there is a consensus in the literature that social comparisons are generally able to produce small reductions in household energy and water consumption. Our focus will be on the mechanisms through which social comparisons operate and the next generation of nudges for resource conservation. A social comparison plausibly generates both intrinsic and extrinsic incentives to conserve. Information on the consumption of a peer group may provide useful information to help a household re-optimize their resource use within their household budget with the goal of reducing monthly bills. Establishing a baseline level of consumption may be particularly useful since both energy and water suffer from an information-poor environment (Attari et al., 2010; Attari, 2014). An alternative motivation is that social comparisons raise the moral cost of resource consumption (Levitt and List, 2007).

There are several empirical studies that provide suggestive evidence that social comparisons target intrinsic incentives, by examining the type of households that respond to these interventions. These studies show that social comparisons generate higher treatment effects in populations who are plausibly more motivated by intrinsic incentives. Costa and Kahn (2013) find that social comparisons generate two to four times as much conservation among political liberals compared to conservatives, with the latter also more likely to opt out of receiving the reports. Political conservatives are less likely to believe that their energy use contributes to damages associated with climate change, indicating that removing the intrinsic motivation from social norms reduces their effectiveness. Similar results are reported by Bolsen et al. (2014); pro-social messages for water conservation were more effective for registered voters. The authors posit that households who contribute to the public good of voting are also more likely to contribute to the public good of resource conservation. Unlike Costa and Kahn (2013), Bolsen et al. (2014) do not find differences in water use between Republicans (political conservatives) and

Democrats (political liberals), which may be explained by the differences in how water and energy use are politicized in the United States.[3] While climate change has become a partisan issue, local water scarcity is likely less to be divided along party lines. Thus, it is important to consider the underlying public goods triggering the intrinsic motivations and the preference heterogeneity for the public good in the population of interest.

In addition to understanding the type of households that respond to social comparisons, the type of actions that consumers take also provide insight into the mechanisms through which norms generate resource conservation. While the data are scarce on specific actions that consumers take, the temporal pattern of treatment effects provides suggestive evidence on what consumers are doing, and consequently how to improve the efficacy of social comparisons. Allcott and Rogers (2014) analyze a rich dataset from Opower with high frequency daily electricity consumption where the social comparisons were randomly discontinued for some households. There is a pattern of "action and backsliding" where immediately after receiving a report there is a large reduction in energy use and then consumption gradually creeps up until the next report is received. Over time as consumers receive more reports this pattern attenuates. The intervention is relatively persistent; the treatment effect decays at 10-20% per year after reports are discontinued. The persistence greatly increases the cost effectiveness of the social comparisons. Remarkably, Bernedo et al. (2014) find the effects of a single letter containing social comparisons sent to encourage water conservation are detectable after six years, though the magnitude decreases by 50% after one year. The treatment effect is only statistically significant after six years among households that did not move. If the primary mechanism of the treatments is durable investments that remain in the home even after the homeowner who received the message moved, the treatment would also be effective when new residents move into homes with updated capital. Therefore, the results are consistent with consumers saving water through behavioral changes as opposed to investments in

[3]It should also be noted that Costa and Kahn (2013) use multiple measures of ideology besides voting including features of their community, paying for renewable energy, and donating to environmental groups.

water efficiency. Both Allcott and Rogers (2014) and Bernedo et al. (2014) claim that consumers form durable habits, consistent with a model of consumption cues as opposed to an immediate re-optimization. Ito et al. (2015) provide somewhat contradictory evidence that social comparisons are not persistent in reducing critical peak consumption. The treatment effect for social comparisons is statistically significant the first time in a season, but repeated mailers within a season are not statistically significant. While it is not appropriate to directly compare average household energy and water use with critical peak energy use, there clearly is variation in the persistence of social comparisons.

There are several recent noteworthy innovations in social comparisons that explore the motivations behind behavioral changes. In two separate experiments that focus on student populations at the University of California Los Angeles (UCLA) researchers explore different framings and utilizations of social comparisons. Delmas and Lessem (2014) remove financial motivations in a field experiment in UCLA residence halls with students that do not pay electricity bills. In addition to the standard private social comparison there is a treatment that publicly displays which rooms on a floor are above the median consumption, introducing reputational incentives as a motivation for conservation. Delmas and Lessem (2014) find that private social norms do not have a statistically significant effect on conservation while public comparisons reduce energy use by 20%. The public information treatment effect exclusively stems from students that consume above the median and are thus subject to public censure. This is consistent with the results of a field experiment run by Yoeli et al. (2013) that finds public information increases the probability that households sign up for a program to reduce blackouts. A second experiment, designed by Asensio and Delmas (2015) in UCLA housing tailored for married students, where residents do pay their electricity bills, alters the framing of social comparisons in both monetary cost savings and public health benefits from using an "efficient level of energy, defined as the 10th percentile in the sample. The public health message was much more effective; households reduced energy use by 8% in the health group while consumption increased by a statistically insignificant 4% in the monetary framing.

The health treatment was particularly effective for households with children who reduced consumption by 19%. Another innovation in these studies is the ability to measure appliance level electricity. In the public versus private experiment the group with private information conserved by turning off lights, while savings from the public group came from reduced electricity for heating and cooling (Delmas and Lessem, 2014). Asensio and Delmas (2015) find that the health savings primarily reduced electricity from plug load and lighting, while the monetary group reduced lighting use, but increased electricity for heating.

4.3 Technology Adoption

It is unlikely that the social comparisons primarily work through investments in energy and water efficient technologies, as evidenced by the decay in the treatment effect over time. There is a large literature debating whether consumers under-invest in energy efficiency (Jaffe and Stavins, 1994; Allcott and Greenstone, 2012; Gerarden et al., 2015), which is partly explained by a model of internalities (Allcott et al., 2014). An internality is the opposite of an externality: consumers fail to internalize all the *private* costs of an activity. The potential welfare improvements from increased investments in energy efficiency, coupled with evidence demonstrating that social norm comparisons reduce consumption but do not encourage new investments, have spawned a new wave of field experiments focused on technology adoption.

Allcott and Taubinsky (2015) develop a novel approach in the context of choosing between incandescent and compact fluorescent lightbulbs (CFL). They use an incentivized artefactual field experiment to estimate the marginal internality bias by randomly providing hard information on the benefits of CFLs, which the authors term a "pure nudge". The information increases the market share of CFLs by 12 percentage points, justifying a $3 subsidy on CFLs based on their theoretical framework that is slightly larger than current subsidies in the United States. The artefactual field experiment is complemented by a natural field experiment in a home improvement retailer that does not find statistically significant effects of information. Taken in concert the results indicate that while a

subsidy for CFLs may be justified, a ban on incandescent bulbs decreases welfare.

Allcott and Sweeney (Forthcoming) examine a similar field experiment that randomly varies information and sales incentives for agents selling hot water systems. Both incentives and information increase the market share of energy efficient models when utilized together, but there is no statistically significant effect of either treatment in isolation. A novel feature of this design is that it targets sales agents that are selling energy durables as opposed to targeting consumers directly. Newell and Siikamäki (2014) also examine the role of information in the demand for energy efficiency in hot water systems by nesting a field experiment within a stated preference choice experiment. The information provided to respondents in the survey varies randomly within four primary categories: the currently used EnergyGuide, Energy Star labels, labels with CO_2 information, and an EU-style energy efficiency grade. The primary results indicate that the manner in which energy costs are communicated to consumers significantly impacts the willingness to pay (WTP) for energy efficiency. A version of the current EnergyGuide augmented with an Energy Star logo and the EU-style relative grade generated the highest WTP. These field experiments help decompose the energy efficiency gap and provide information to guide policies targeting investment in energy durables that can improve welfare from reducing both internalities and externalities.

5 Theory

According to the standard neoclassical model, individuals engage in environmentally improving pro-social behavior such as energy conservation, public good contributions, and lower harvesting of common pool resources, only if there is a private pecuniary or extrinsic motivation for doing so. Thus, they contribute zero to a public good, they ignore the effect of their resource extraction on others, and they are unaffected by information regarding the energy use of their peers. Substantial evidence however, including that described in previous sections, shows that additional motivations influence decisions. People exhibit altruism and behave cooperatively, and they are affected by framing and

social information, even in the absence of pecuniary consequences. Moreover, a substantive body of evidence demonstrates that monetary incentives often interact with other motivations in unexpected ways. In particular, the introduction of, or the increase in, explicit monetary incentives can crowd out non-pecuniary incentives even leading to a (net) reduction in the targeted pro-social behavior. In other settings monetary incentives are complementary, leading to crowding in of the desired behavior.[4]

These findings have led to the development of theoretical models that encapsulate motivations beyond just monetary incentives. In this section, we discuss theoretical models of pro-social behavior that model intrinsic and reputational motivations, in addition to the standard monetary incentives. These models provide a framework for interpreting the findings from the field experiments discussed earlier in this paper and facilitate generalization beyond the specific settings of each experiment. With this purpose in mind, we do not provide a comprehensive literature review but concentrate only on the most relevant models.[5] While the earlier models assume separability between extrinsic and other motivations so that crowding out (or in) cannot occur, more recent models describe mechanisms through which these interactions can occur.

5.1 Models Incorporating Intrinsic Motivations

Intrinsic motivations reflect motivation that is internal to each person and can arise for a number of reasons. First, individuals have personal moral codes that influence behavior, and they may experience a psychic cost (e.g. experience guilt) from not engaging in pro-social behavior, with this cost varying across individuals (Gordon, 1989). Importantly, this type of intrinsic motivation operates independently of what others are doing and whether your actions are observed; the benefits or costs are purely internal, and reflect individual heterogeneity in preferences and values. In the environmental context, this could reflect an individual's "green" preferences. Andreoni (1989, 1990) models the role

[4]Surveys of this evidence can be found in Frey and Jegen (2001); Fehr and Falk (2002); Gneezy et al. (2011); Bowles and Polanía-Reyes (2012)

[5] For example, we do not explore the theoretical literature on how social norms develop (and social equilibria) or inequality aversion (Fehr and Schmidt, 1999) as they are less relevant for the types of experiments we consider.

of altruism in public good contributions, separating pure from impure altruism, where people receive a personal benefit (a "warm glow") from giving in addition to helping others through providing the public good. As described above, these types of models typically include either a psychic cost from not choosing the moral option or a psychic benefit from doing so. Such motivation might arise from a desire to maintain one's own self-image or self-concept (Mazar et al., 2008) as a moral, cooperative, or fair person or as caring about the environment. Such motivations can explain the heterogeneity of individual behavior observed in some of the field experiments discussed in earlier sections.

A second category of intrinsic motivation arises with regard to the standard to which people compare their behavior. In addition to adhering to one's own moral code, people may also care about adhering to the social norm, experiencing a psychic benefit from adhering to the social norm or a cost to deviating from it (Myles and Naylor, 1996). These benefits or costs do not arise because of reputational effects, or because others observe one's behavior, rather they are internal. However, perceptions of the social norm can be affected by information about the behavior of others, as occurs in the energy and water field experiments, or the institutional context or framing of a situation can suggest cooperative social norms as in the framed common pool and public goods experiments.

The models described so far assume separability between extrinsic and other motivations and thus cannot generate crowding out (or in) effects. More recent theories provide mechanisms through which crowding might occur by modeling interactions between extrinsic and intrinsic motivations. For example, Brekke et al. (2003) present a model of moral (rather than social) norms for public good contributions where individuals want to maintain their own image as morally responsible. The "morally ideal action" is endogenous in their model and is affected by the economic incentives, among other factors such as the efficiency of public good provision. Specifically, symbolic incentives (which are not sufficient to provide the good) still leave you morally responsible but substantive ones do not, and so the latter crowds out voluntary contributions.

In a recent survey, Bowles and Polanía-Reyes (2012) provide a theoretical explanation

of results from 50 diverse experimental studies distinguishing between the effects of the presence of incentives (a categorical effect) from their magnitude (a marginal effect). They develop a model of state-dependent preferences reflecting that incentives convey information about the decision environment resulting in different states. Of particular relevance for common pool and public goods field experiments is that externally imposed penalties can change the decision frame, suggesting certain behavior is permissible thus causing "moral disengagement". In other settings, the presence of incentives suggests cooperation and thus generates crowding in, although such effects are less common.[6]

5.1.1 The Bowles and Polonia-Reyes model

Bowles and Polanía-Reyes (2012) describe two models. We describe their first model of incentives as signals in this subsection. Each individual contributes a to the public project, which benefits all citizens according to φA where A is the sum of all contributions and $\varphi > 0$ reflects the efficiency of the public project. Contributions are costly according to an increasing convex cost of $g(a)$ but maybe subsidized at rate s. Individuals also receive intrinsic benefits related to their social preference v. The individual decision problem is thus

$$\max_a \{\varphi A - g(a) + as + av\} \tag{1}$$

An important contribution of their model is to elaborate on the construction of an individual's social preference, v. Specifically

$$v = \lambda_0(1 + \mathbf{I}\{s > 0\}\lambda_c + s\lambda_m) \tag{2}$$

where λ_0 is the social preference in the absence of the monetary incentive, λ_c is the categorical effect of incentives, and λ_m the marginal effect.

The first order condition that defines the optimal choice a^* is given by

[6]Bowles and Polanía-Reyes (2012) identify three channels that might generate state-dependent preferences. Moral disengagement is the most relevant one for the field experiments we discuss. The two other channels are "bad news" (e.g. externally imposed penalties might suggest a lack of trust in the agent if imposed by the other player) and "control aversion" (incentives can undermine an individual's feeling of autonomy).

$$g'(a^*) = \varphi + s + \lambda_0(1 + \mathbf{I}\{s > 0\}\lambda_c + s\lambda_m) \tag{3}$$

As noted by Bowles and Polanía-Reyes (2012), introducing the incentive (subsidy) makes contributing more financially attractive, thus increasing the right side of (3) which they denote as θ. This leads to their key equation (equation 5 in their study) deriving how the benefits of contributing change when a subsidy is introduced:

$$\frac{\Delta\theta}{\Delta s} = 1 + \lambda_0(\frac{\lambda_c}{\Delta s} + \lambda_m) \tag{4}$$

Equation (4) shows how the direct effect of the subsidy (which is one) can be augmented or offset by social preferences. Notably, in the absence of baseline social preferences λ_0, crowding is not possible. Further, the larger the baseline social preference, the larger the magnitude of the crowding that occurs. If λ_c and λ_m are negative then crowding out will occur, and (4) will be less than one (the direct effect of the subsidy). "Strong crowding out" is also possible whereby (4) becomes negative if these effects are sufficiently large. This theoretical setup nicely demonstrates that a variety of cases are possible depending on the strength of the benefits of cooperation.

5.2 Models Incorporating Reputation Concerns

In addition to these internal or intrinsic motivations, people also care about maintaining their public image or social reputation, desiring social approval and wishing to avoid social disapproval. Individuals want to be seen as adhering to social norms and be perceived as being pro-social even if they are not. Thus, the second broad category of motivation that has been modeled is reputation effects. For these effects to operate actions must be observable by others, which has led to the development of a number of signaling models. Bernheim (1994), for example, develops a model where individuals care about both actions and status; i.e. how others perceive them. Actions provide signals about your unobservable underlying preferences, while status is related to the (endogenous) social norm.

Bénabou and Tirole (2006) model all three motivations: intrinsic, extrinsic and reputational. Importantly, they show how net crowding out is theoretically possible when people use pro-social behavior to signal their generosity. In such a situation, increased extrinsic incentives make it harder to demonstrate that your motives are pure or have a signal jamming effect. Donating to a public good when there are sanctions is not being pro-social but donating in the absence of monetary incentives is. Similarly, the signaling effect from installing solar panels falls as the economic benefits increase. They note that compared with the model of Bernheim (1994) "multidimensional uncertainty ... is essential to generating over-justification effects and net crowding out" (p.1655). Compared with Bernheim (1994) where the social norm is endogenous, in Bénabou and Tirole (2006) the social norm is given; individuals want to be perceived as being prosocial and not as greedy, independent of what others are doing.[7][8]

5.2.1 Benabou and Tirole Model

As mentioned before, in the Bénabou and Tirole (2006) model, pro-social choices have benefits that fall into three categories: extrinsic, intrinsic and reputational. Each individual chooses an action a (e.g. the amount to contribute to a public good), which costs $C(a)$ but yields pecuniary benefit ya. The individual values money according to v_y but also receives a non-monetary value of v_a. Individuals thus receive a direct benefit, comprised of the extrinsic and intrinsic motivation, from choosing a of

$$(v_a + v_y y)a - C(a) \tag{5}$$

[7]There are also signaling models of the principal-agent relationship where the principal's contract choice reveals information and can crowd out prosocial behavior. This mechanism was mentioned by Bowles and Polanía-Reyes (2012). See for example Ellingsen and Johannesson (2008) and Sliwka (2007).

[8]In addition to the model described in the previous subsection, where social preferences are fixed, Bowles and Polanía-Reyes (2012) develop a second model to study the development of preferences or social norms; i.e., "endogenous preferences". In their model of state-dependent preferences, the preference parameters are fixed. In this second model, preference parameters are affected by the incentives so that over time the proportion of social types in the population changes. Similar to the mechanism in the Bénabou and Tirole (2006) signaling model, incentives interfere with the inference one can draw from actions to traits. In this case, the presence of an incentive affects the accuracy of perceptions regarding the proportion of "civic types" in the population. In particular, the perceived frequency of pro-social types falls when monetary incentives are present. In contrast to Bénabou and Tirole (2006), here incentives have long run implications as well as affecting short run behavior.

As (5) makes clear, the intrinsic value component is simpler than in the Bowles and Polanía-Reyes (2012) model but an additional effect is added; reputation. While everyone knows their own type (v_a, v_y), this is private information and others only know the distribution of types. However observed choices of a are potentially informative about a person's type. The given social norm is that people want to be perceived as prosocial and not greedy. The reputational payoff is modeled as follows, where the parameters μ captures both visibility of actions and how much you care about reputation, and E is the posterior expectation:

$$R(a, y) \equiv \mu_a E(v_a|a, y) - \mu_y E(v_y|a, y) \tag{6}$$

Combining these equations gives the individual's decision problem

$$\max_{a \in A} \{(v_a + v_y y)a - C(a) + \mu_a E(v_a|a, y) - \mu_y E(v_y|a, y)\} \tag{7}$$

The first order condition is

$$C'(a) = v_a + v_y y + \mu_a \frac{\partial E(v_a|a, y)}{\partial a} - \mu_y \frac{\partial E(v_y|a, y)}{\partial a} \tag{8}$$

As equation (8) shows, choices of a reflect the combined effect of these three motivations and thus it is not straightforward to infer another person's type (or equivalently signal your own type), particularly when monetary incentives are present.

Bénabou and Tirole (2006) demonstrate the basic operation of the mechanism in a simple version of the model where they assume that v_a and v_y are independent random variables and that the μ parameters are fixed. In this case, if $y = 0$ (monetary incentives are absent) then the choice of a is "very informative" about the agents social preference.[9] The direct (standard) effect of introducing a monetary incentive is to increase contributions, by expanding the set of individuals who contributes; i.e. it is now worthwhile for

[9] As Bénabou and Tirole (2006) note, p.1659, "In the continuous case, prosocial orientation is learned perfectly; in the discrete case one learns whether it is above or below a known cutoff."

those with lower values of v_a to contribute. But the monetary incentive dilutes the reputation effect of contributing and some of those who contributed when $y = 0$ now reduce their contribution. The overall effect is therefore uncertain but clearly net crowding out is possible in certain cases.

6 Challenges with Field Experiments

Field experiments can provide important insights into the optimal use of environmental goods, which are difficult to obtain with existing empirical data or laboratory experiments. However, there are also several challenges with field experiments. Unlike laboratory experiments, for example, where the researcher can freely create and modify the environment, field experiments are embedded in the natural environment and therefore subject to its constraints. In this section, we describe some of the main challenges researchers encounter when they take their research questions to the field.

6.1 Testing Theories in the Field

It is probably fair to say that many theories, and in particular those that relax the standard self-interest assumption, have been predominantly experimentally tested in the laboratory. This is due to several reasons. Most prominent of these is that it is difficult to test pro-social motivations in the field as it is often impossible to rule out other confounding explanations. For example, often behavior can be driven by reputation concerns as shown in the theoretical models discussed in the previous section. A good illustration of this issue is that the adaption of environmentally friendly technologies may be driven by an intrinsic motivation to be environmentally friendly or rather a motivation to appear environmentally friendly. Cavalcanti et al. (2013), for example, find that fishermen who are more central in the social network are more willing to change to environmentally friendly shrimp traps than those at the periphery, which may be due to enhanced reputational concerns. However, this does not imply that pro-social motivations cannot be tested in natural field experiments. The role of pro-social motivations can still be tested as long as other variables such as reputational concerns are held constant across

treatments.

A particular challenge with the interpretation of findings from natural field experiments is that they are not always designed to test theory but as a means to optimize the output variable of interest. The focus on such optimization is particularly common whenever field experiments are designed and conducted jointly with practitioners. The problem with such practice is that the mechanisms behind treatment effects remain uncovered in the absence of theory testing. Thus, the value of such a field experiment may be constrained to the particular environment in which it took place.

6.2 Difficulty of Replication

Many field experiments can unfortunately be difficult to replicate. Replication, the possibility to independently reproduce a study, gives researchers and policy makers more confidence in findings and their robustness. While the replication of laboratory studies and their findings is straightforward and relatively easy (Camerer et al., 2016), replication of field experimental studies is much more complex, sometimes impossible, and typically costly. Replication operates on different levels. Three important levels are: (i) re-analysis of the data generated by a field experiment to test whether the analysis confirms the original findings, (ii) re-run the field experiment with a similar protocol but in a different environment and different subjects, and (iii) test hypothesis of original study with a different field experimental design List and Rasul (2011). In contrast to laboratory experiments and artefactual and framed field experiments, the replication of natural field experiments is significantly more challenging on these levels.

The easiest level of replication for natural field experiments is the re-analysis of the generated data. However, while this is usually possible there are also studies using confidential data sets, which are not accessible to other researchers. Re-running a natural field experiment with a different subject pool is often difficult and costly. First, researchers need to have access to a different field setting and subject pool. Second, even if they have such access, it is expensive to conduct a replication study of a natural field experiment as it will likely be difficult to find partners who will fund a replication study (instead of

an original study). Similar challenges are also present for new natural field experiments testing the hypotheses of the original study, albeit likely to a lesser extent as there is more freedom in choice of potential field settings. Many researchers are also often discouraged from running pure replications (whether in the lab or field) due to the pressure for publishing novel results.

6.3 Ethical Issues and Informed Consent

A key distinction between natural field experiments and other types of experiments is that subjects are unaware of taking part in an experiment. Thus, behavior in natural field experiments is less prone to experimenter demand effects and other behavioral reactions that may be the result of scrutiny (Levitt and List, 2007). However, this implies that we observe behavior from subjects who did not give consent to take part in an experiment, which raises ethical issues.

Informed consent is a cornerstone in human experimentation. The Nuremberg Code of 1947 prescribed the requirement of informed consent and provides guidelines on research ethics as a response to immoral experimentation by Nazi doctors. There is little disagreement that informed consent as the default is appropriate for medical trials as they often involve serious risk potential. However, opinions on informed consent differ for field experiments in the social sciences, where risks for the involved subjects are typically small or even nonexistent (Dingwall, 1980; Punch, 1986). Whatever the opinions are, it is important to acknowledge that natural field experiments are common practice among companies, which use us as experimental subjects without our knowledge. Perhaps more importantly, in some environments only natural field experiments can uncover undesirable behaviors and test how to curb these, as subjects would not display such behaviors knowing they are observed.

Another ethical issue arises as a product of field experimentation when subjects are randomized in different treatments. While randomization renders possible causal inferences, it leads to differential treatment of subjects. For example, Herberich et al. (2011) test the demand for compact fluorescent light bulbs (CFLs) by offering them to house-

holds under varying prices. Thus, some subjects had the option to buy CFLs for a lower price than others. Another example is Gans et al. (2013), which studies the impact of smart meters on energy consumption. Randomization in such environments can be justified by the fact that is often impossible to immediately roll out interventions (in this study smart meters) for all subjects. In such environments, the researcher can only randomly determine the order in which subjects are treated but not whether they are treated.

Although ethical concerns cannot simply be delegated, it is important to include Research Ethics Committees and Institutional Review Boards (IRBs) in the discussion and the assessment of research proposals. These committees and boards can assist researchers and policy makers in weighing the costs and benefits of a natural field experiment, and determine whether subjects are fairly treated in the field experiment and have selected into the environment on their own volition.

7 Discussion and Directions for Future Research

Field experiments incorporate elements of the naturally occurring environment to generate insights into environmental behaviors within this environment. Fitting the experiment to the environment has the advantage of capturing the key determinants in a given environment but renders it difficult to predict environmental behaviors in different environments, which may be characterized by different key determinants. For example, it is often unclear how findings will change over time or how similar policies will affect resource users in a different location. Extrapolation of findings - regardless of which empirical and experimental method was used to generate the findings - is always a challenge. For example, how confident can we be that the insights from a specific field experiment on energy conservation in Australia apply to resource conservation in New Zealand? There are at least two means that we can use to increase the extent to which field experimental findings can be extrapolated: consistency with theoretical predictions and similarity of key determinants.

Field experiments are often designed around a specific policy and it can be tempting

to test the status quo against an alternative policy instead of relying on theory to motivate treatments. However, if there is no clear theoretical prediction, or, if several policy variables change simultaneously, the interpretation of treatment effects (status quo vs. alternative policy) is speculative. In contrast, if we test the status quo against a policy that differs on one variable that is predicted to matter, then we can say that this variable is responsible for the treatment difference. To increase our confidence that field experimental findings can be extrapolated, we need to make sure that the variable that caused the treatment effect is also predicted to be relevant in different settings.

Another threat to the extrapolation of findings from natural field experiments can stem from a partner selection bias. Allcott (2015), for example, use data from 14 energy conservation field experiments in the U.S. to identify a statistically and economically significant site-specific treatment effect. He then shows a selection of sites that are willing to conduct natural field experiments.

Allcott (2015) shows that the Opower results, and perhaps other field experiments in environmental economics, suffer from a site selection bias. This is the notion that even though individuals do not select into treatment in a natural field experiment, the experimenter must often cooperate with a self-selected partner. Allcott (2015) find that the first 10 Opower sites overpredict the mean average treatment effect by roughly one half of a percentage point, which is significant in the context of Opower's claim of generating 2% energy savings. The results indicate that initial treatment effects from field experiment may be biased upwards since the first movers are the partners with the largest expected treatment effects. While studies such as Delmas and Lessem (2014) and Asensio and Delmas (2015) advance the literature by isolating specific behavioral mechanisms and developing new framings, more research needs to be undertaken to understand how results from student apartments with zero or low energy bills can generalize to the broader population.

Field experiments do not exist in a vacuum, and it is important to understand how field experiments interact with existing conservation policies. For example, Allcott and

Rogers (2014) and Brent et al. (2015) both find that social comparisons generate statistically significant increases in utility program participation such as efficiency rebates.[10] Thus one area of future research is to embrace the interconnections with utility rate structures and subsidy programs in order to uncover more generalizable features of energy and water demand. One example of this is Kahn and Wolak (2013), who randomly administer an energy cost tutorial that teaches consumers about the increasing block rates structure. This field experiment is nested within an active area of research that examines how consumers respond to increasing block rates Ito (2014); Wichman (2014). Additionally, LaRiviere et al. (2015) combine both subsidies and social comparisons in multiple framings to increase participation in a utility-wide home energy audit and subsidy program. They find an asymmetric effect of nudges: while they increase energy audits the nudges decrease (both unconditionally and conditional on receiving an audit) the probability of purchasing energy efficient durables.

While social comparisons are shown to be effective at reducing energy use there is relatively little research that assesses the welfare effect of these nudges. If social comparisons work by raising the psychic cost of consumption then economists need to account for the utility cost of this emotional tax in welfare analysis. Allcott and Kessler (2015) combine an incentive compatible elicitation of willingness to pay for nudges with a field experiment of social comparisons on energy use. They find that there is a substantial welfare cost of the nudges such that previous welfare estimates significantly overestimate the benefits of social comparisons. Furthermore, a significant portion of consumers are made worse off by the social comparison. They develop an algorithm to target welfare gains as opposed to energy conservation in order to optimize the welfare impacts of nudges. This is an important development to consider the utility cost of nudges.

The vast majority of academic research utilizing field experiments for energy and water conservation takes place in the United States; thus another area of future research that pertains to external validity is applying field experiments in other countries. Ito

[10]While there are statistically significant increases in utility program participation the increased rates do not explain a substantial proportion of the average treatment effect.

et al. (2015); Pellerano et al. (2015); Dolan and Metcalfe (2015) are examples of field experiments outside of the United States that use either social comparisons, financial incentives, or both for energy conservation. In particular Pellerano et al. (2015), which is set in Quito, Ecuador, is an example of applying field experiments for resource conservation in developing countries.

The interaction of financial and social interventions with intrinsic norms of behavior is another area of research that is fruitful. While substantial work has been done to understand the different motivations behind conservation, there is still much to learn about the interaction between these effects. For example, we do not know whether price increases in water consumption reduce resource users' voluntary restraint to waste water. Isolating the different drivers of behavior in the field and examining how influencing one would impact the others requires building rigorous theoretical intuition and designing experimental treatments carefully to tease out subtle behavioral characteristics. Fortunately, identifying causal mechanisms and tying them to theoretical intuition is an advantage of the experimental approach.

References

Abadie, Alberto and Guido W Imbens, "Large sample properties of matching estimators for average treatment effects," *Econometrica*, 2006, *74* (1), 235–267.

Allcott, Hunt, "Social norms and energy conservation," *Journal of Public Economics*, 2011, *95* (9-10), 1082–1095.

_ , "Site Selection Bias in Program Evaluation," *The Quarterly Journal of Economics*, 2015, *130* (3), 1117–1165.

_ **and Dmitry Taubinsky**, "Evaluating Behaviorally-Motivated Policy: Experimental Evidence from the Lightbulb Market," *American Economic Review*, 2015, pp. 1–80.

_ **and Judd B. Kessler**, "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons," *NBER Working Paper Series*, 2015, *21671.*

_ **and Michael Greenstone**, "Is There an Energy Efficiency Gap?," *Journal of Economic Perspectives*, 2012, *26* (1), 3–28.

_ **and Richard Sweeney**, "The Role of Sales Agents in Information Disclosure: Evidence from a Field Experiment," *Management Science*, Forthcoming, pp. 1–52.

_ **and Sendhil Mullainathan**, "Behavioral Science and Energy Policy," *Science*, 2010, *327* (6970), 1204–1205.

_ **and Todd Rogers**, "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation," *American Economic Review*, 2014, *104* (10), 3003–3037.

_ **, Sendhil Mullainathan, and Dmitry Taubinsky**, "Energy policy with externalities and internalities," *Journal of Public Economics*, apr 2014, *112*, 72–88.

Alpizar, Francisco, Fredrik Carlsson, and Maria A. Naranjo, "The effect of ambiguous risk, and coordination on farmers' adaptation to climate change - A framed field experiment," *Ecological Economics*, 2011, *70* (12), 2317–2326.

Andreoni, James, "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy*, 1989, *97* (6), 1447–1458.

_ , "Impure Altruism and Donations To Public Goods: a Theory of Warm-Glow Giving.," *Economic Journal*, 1990, *100* (401), 464–477.

Araña, Jorge E. and Carmelo J. León, "Can Defaults Save the Climate? Evidence from a Field Experiment on Carbon Offsetting Programs," *Environmental and Resource Economics*, 2013, *54* (4), 613–626.

Asensio, Omar I and Magali A Delmas, "Nonprice incentives and energy conservation," *Proceedings of the National Academy of Sciences*, 2015, *112* (6), E510–E515.

Attari, Shahzeen Z, "Perceptions of water use," *Proceedings of the National Academy of Sciences*, 2014, *111* (14), 5129–5134.

_ , **Michael L DeKay, Cliff I Davidson, and Wändi Bruine De Bruin**, "Public perceptions of energy consumption and savings," *Proceedings of the National Academy of sciences*, 2010, *107* (37), 16054–16059.

Ayres, Ian, Sophie Raseman, and Alice Shih, "Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage," *Journal of Law, Economics, and Organization*, August 2012, *29* (5), 992–1022.

Baland, Jean-Marie and Jean-Philippe Platteau, *Halting degradation of natural resources: is there a role for rural communities?*, Food & Agriculture Organization of the United Nations, 1996.

Bénabou, Roland and Jean Tirole, "Incentives and prosocial behavior," *American Economic Review*, 2006, *96* (5), 1652–1678.

Berg, Joyce, John Dickhaut, and Kevin McCabe, "Trust, reciprocity, and social history," *Games and economic behavior*, 1995, *10* (1), 122–142.

Bernedo, Maria, Paul Ferraro, and Michael Price, "The Persistent Impacts of Norm-based Messaging and their Implications for Water Conservation," *Journal of Consumer Policy*, 2014, *37*, 437–452.

Bernheim, B Douglas, "A Theory of Conformity," *Journal of Political Economy*, 1994, pp. 841–877.

Blanco, Esther, Maria Claudia Lopez, and Sergio Villamayor-Tomas, "Exogenous degradation in the commons: Field experimental evidence," *Ecological Economics*, 2015, *120*, 430–439.

Bohm, Peter, "Experimental evaluations of policy instruments," *Handbook of environmental economics*, 2003, *1*, 437–460.

Bolsen, Toby, Paul J. Ferraro, and Juan Jose Miranda, "Are Voters More Likely to Contribute to Other Public Goods? Evidence from a Large-Scale Randomized Policy Experiment," *American Journal of Political Science*, jan 2014, *58* (1), 17–30.

Bouma, Jetske A., K. J. Joy, Suhas Paranjape, and Erik Ansink, "The influence of legitimacy perceptions on cooperation - A framed field experiment," *World Development*, 2014, *57*, 127–137.

Bouma, Jetske, Erwin Bulte, and Daan Van Soest, "Trust and cooperation: Social capital and community resource management," *Journal of Environmental Economics and Management*, 2008, *56* (2), 155–166.

Bowles, Samuel and Sandra Polanía-Reyes, "Economic incentives and social preferences: Substitutes or complements?," *Journal of Economic Literature*, 2012, *50* (2), 368–425.

Brekke, Kjell Arne, Snorre Kverndokk, and Karine Nyborg, "An economic model of moral motivation," *Journal of Public Economics*, 2003, *87* (9-10), 1967–1983.

Brent, Daniel A., Joseph H. Cook, and Skylar Olsen, "Social Comparisons, Household Water Use, and Participation in Utility Conservation Programs: Evidence from Three Randomized Trials," *Journal of the Association of Environmental and Resource Economists*, 2015, *2* (4), 597–627.

Camerer, Colin F, Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan et al., "Evaluating replicability of laboratory experiments in economics," *Science*, 2016, *351* (6280), 1433–1436.

Cardenas, Juan Camilo and Elinor Ostrom, "What do people bring into the game? Experiments in the field about cooperation in the commons," *Agricultural Systems*, 2004, *82* (3), 307–326.

Carpenter, Jeffrey and Erika Seki, "Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama Bay," *Economic Inquiry*, 2011, *49* (2), 612–630.

Carpenter, Jeffrey P and Erika Seki, "Competitive Work Environments and Social Preferences : Field Experimental Evidence from a Japanese Fishing Community," *Contributions to Economic Analysis & Policy*, 2006, *5* (2), 1–23.

Cason, Timothy N, "What can laboratory experiments teach us about emissions permit market design?," *Agricultural and Resource Economics Review*, 2010, *39* (2), 151–161.

Castillo, Daniel, Francois Bousquet, Marco A. Janssen, Kobchai Worrapimphong, and Juan Camillo Cardenas, "Context matters to explain field experiments: Results from Colombian and Thai fishing villages," *Ecological Economics*, 2011, *70* (9), 1609–1620.

Cavalcanti, Carina, Stefanie Engel, and Andreas Leibbrandt, "Social integration, participation, and community resource management," *Journal of Environmental Economics and Management*, 2013, *65* (2), 262–276.

Caves, Douglas W., Laurits R. Christensen, and Joseph A. Herriges, "Consistency of Residential Customer Response in Time-of-Use Electricity Pricing Experiments," *Journal of Econometrics*, 1984, *26*, 179–203.

Cherry, Todd L, Stephan Kroll, Jason Shogren et al., *Environmental economics, experimental methods*, Routledge, 2007.

Costa, Dora L. and Matthew E. Kahn, "Energy Conservation Nudges and Environmentalist Ideology: Evidence From a Randomized Residential Electricity Field Experiment," *Journal of the European Economic Association*, jun 2013, *11* (3), 680–702.

Cummings, R, Michael McKee, and Laura O Taylor, *To whisper in the ears of princes: laboratory economic experiments and environmental policy*, London: Elgar Publishers, 2001.

D'Adda, Giovanna, "Motivation crowding in environmental protection: Evidence from an artefactual field experiment," *Ecological Economics*, 2011, *70* (11), 2083–2097.

Delmas, Magali A. and Neil Lessem, "Saving power to conserve your reputation? The effectiveness of private versus public information," *Journal of Environmental Economics and Management*, 2014, pp. 1–18.

Dingwall, Robert, "Ethics and Ethnography," *The Sociological Review*, 1980, *28* (4), 871–891.

Dolan, Paul and Robert Metcalfe, "Neighbors, Knowledge, and Nuggets: Two Natural Field Experiments on the Role of Incentives on Energy Conservation," 2015.

Eckel, Catherine C and Philip J Grossman, "Altruism in anonymous dictator games," *Games and economic behavior*, 1996, *16*, 181.

Ellingsen, Tore and Magnus Johannesson, "Pride and Prejudice : The Human Side of Incentive Theory," *American Economic Review*, 2008, *98* (3), 990–1008.

Faruqui, Ahmad and Sanem Sergici, "Household response to dynamic pricing of electricity: a survey of 15 experiments," *Journal of Regulatory Economics*, aug 2010, *38* (2), 193–225.

Fehr, Ernst and Andreas Leibbrandt, "A field study on cooperativeness and impatience in the Tragedy of the Commons," *Journal of Public Economics*, 2011, *95* (9-10), 1144–1155.

_ **and Armin Falk**, "Psychological foundations of incentives," *European economic review*, 2002, *46* (4), 687–724.

_ **and Klaus M Schmidt**, "A theory of fairness, competition, and cooperation," *Quarterly journal of Economics*, 1999, pp. 817–868.

Ferraro, Paul J. and Juan José Miranda, "Heterogeneous Treatment Effects and Mechanisms in Information-Based Environmental Policies: Evidence from a Large-Scale Field Experiment," *Resource and Energy Economics*, 2013, *35* (3), 356–379.

Ferraro, PJ and MK Price, "Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment," *Review of Economics and Statistics*, 2013, *95* (1), 64–73.

Frey, Bruno S. and Reto Jegen, "Motivation Crowding Theory," *Journal of Economic Surveys*, 2001, *15* (5), 589–611.

Friesen, Lana and Lata Gangadharan, "Environmental markets: What do we learn from the lab?," *Journal of Economic Surveys*, 2013, *27* (3), 515–535.

Gans, Will, Anna Alberini, and Alberto Longo, "Smart meter devices and the effect of feedback on residential electricity consumption: Evidence from a natural experiment in Northern Ireland," *Energy Economics*, 2013, *36*, 729–743.

Gerarden, By Todd, Richard G Newell, and Robert N Stavins, "Deconstructing the Energy-Efficiency Gap : Conceptual Frameworks and Evidence," *American Economic Review*, 2015, *105* (5), 183–186.

Gneezy, Uri, Andreas Leibbrandt, and John A. List, "Ode To The Sea: Workplace Organizations And Norms Of Cooperation," *The Economic Journal*, 2015.

_ **, Stephan Meier, and Pedro Rey-Biel**, "When and Why Incentives (Don't) Work to Modify Behavior," *Journal of Economic Perspectives*, 2011, *25* (4), 191–210.

Gordon, James P F, "Individual morality and reputation costs as deterrents to tax evasion," *European Economic Review*, 1989, *33*, 797–805.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze, "An experimental analysis of ultimatum bargaining," *Journal of economic behavior & organization*, 1982, *3* (4), 367–388.

Harrison, Glenn W and John A List, "Field Experiments," *Journal of Economic Literature*, 2004, *42* (4), 1009–1055.

Hayo, Bernd and Bjrn Vollan, "Group interaction, heterogeneity, rules, and cooperative behaviour: Evidence from a common-pool resource experiment in South Africa and Namibia," *Journal of Economic Behavior and Organization*, 2012, *81* (1), 9–28.

Herberich, David, J List, and Michael Price, "How many economists does it take to change a light bulb? A natural field experiment on technology adoption," 2011.

Isaac, R Mark and James M Walker, "Group size effects in public goods provision: The voluntary contributions mechanism," *The Quarterly Journal of Economics*, 1988, pp. 179–199.

Ito, Koichiro, "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing," *American Economic Review*, February 2014, *104* (2), 537–563.

_ **, Takanori Ida, and Makoto Tanaka**, "The Persistence of Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand," *E2e Project Working Paper Series*, 2015, *017*.

Jaffe, Adam B. and Robert N. Stavins, "The energy-efficiency gap What does it mean?," *Energy Policy*, 1994, *22* (10), 804–810.

Janssen, Marco A., Francois Bousquet, Juan Camilo Cardenas, Daniel Castillo, and Kobchai Worrapimphong, "Breaking the elected rules in a field experiment on forestry resources," *Ecological Economics*, 2013, *90*, 132–139.

Jessoe, Katrina and David Rapson, "Knowledge is (Less) Power: Experimental Evidence from Residential Energy Use," *American Economic Review*, apr 2014, *104* (4), 1417–1438.

Kahn, Matthew E. and Frank A Wolak, "Using Information to Improve the Effectiveness of Nonlinear Pricing: Evidence from a Field Experiment," 2013.

Kerr, John, Mamta Vardhan, and Rohit Jindal, "Prosocial behavior and incentives: Evidence from field experiments in rural Mexico and Tanzania," *Ecological Economics*, 2012, *73*, 220–227.

LaRiviere, Jacob, Scott Holladay, David Novgorodsky, and Michael Price, "Asymmetric Effects of Non-Pecuniary Signals on Search and Purchase Behavior for Energy-Efficient Durable Goods," 2015.

Leibbrandt, Andreas, Uri Gneezy, and John A List, "Rise and fall of competitiveness in individualistic and collectivistic societies," *Proceedings of the National Academy of Sciences*, 2013, *110* (23), 9305–9308.

Levitt, Steven D. and John A. List, "What do laboratory experiments measuring social preferences reveal about the real world?," *The Journal of Economic Perspectives*, 2007, *21* (2), 153–174.

List, John A and Imran Rasul, "Field Experiments in Labor Economics," *Handbook of Labor Economics*, 2011, *4*, 103–228.

_ **and Michael K Price**, "Using Field Experiments in Environmental and Resource Economics," *NBER Working Paper Series*, 2013, (19289), 35.

Lopez, Maria Claudia, James J Murphy, John M Spraggon, and John K Stranlund, "Comparing the Effectiveness of Regulation and Pro-Social Emotions to Enhance Cooperation: Experimental Evidence from Fishing Communities in Colombia," *Economic Inquiry*, 2012, *50* (1), 131–142.

Mazar, Nina, On Amir, and Dan Ariely, "The dishonesty of honest people: A theory of self-concept maintenance," *Journal of marketing research*, 2008, *45* (6), 633–644.

Muller, R. Andrew and Stuart Mestelman, "What have we learned from emissions trading experiments?," *Managerial and Decision Economics*, 1998, *19* (4-5), 225–238.

Myles, Gareth D. and Robin A. Naylor, "A model of tax evasion with group conformity and social customs," *European Journal of Political Economy*, 1996, *12* (1), 49–66.

Narloch, Ulf, Unai Pascual, and Adam G. Drucker, "Collective Action Dynamics under External Rewards: Experimental Insights from Andean Farming Communities," *World Development*, 2012, *40* (10), 2096–2107.

Newell, Richard G and Juha Siikamäki, "Nudging Energy Efficiency Behavior: The Role of Information Labels," *Journal of the Association of Environmental and Resource Economists*, 2014, *1* (4), 555–598.

Normann, Hans-Theo and Roberto Ricciuti, "Laboratory experiments for economic policy making," *Journal of Economic Surveys*, 2009, *23* (3), 407–432.

Noussair, Charles N, Daan van Soest, Jan Stoop et al., "Cooperation in a dynamic fishing game: A framed field experiment," *American Economic Review*, 2015, *105* (5), 408–13.

Orne, Martin T, "On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications.," *American psychologist*, 1962, *17* (11), 776.

Ostrom, Elinor, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press, 1990.

_ , James Walker, and Roy Gardner, "Covenants with and without a Sword: Self-governance Is Possible.," *American political science Review*, 1992, *86* (02), 404–417.

Pellerano, José A, Michael K Price, Steven L Puller, and Gonzalo E Sánchez, "Do Extrinsic Incentives Undermine Social Norms? Evidence from a Field Experiment in Energy Conservation," 2015.

Prediger, Sebastian, Björn Vollan, and Markus Frölich, "The impact of culture and ecology on cooperation in a common-pool resource experiment," *Ecological Economics*, 2011, *70* (9), 1599–1608.

Price, Michael K, "Using field experiments to address environmental externalities and resource scarcity: major lessons learned and new directions for future research," *Oxford Review of Economic Policy*, 2014, *30* (4), 621–638.

Punch, M., *The politics and ethics of fieldwork* Qualitative research methods, Sage Publications, 1986.

Reichhuber, Anke, Eva Camacho, and Till Requate, "A framed field experiment on collective enforcement mechanisms with Ethiopian farmers," *Environment and Development Economics*, 2009, *14* (5), 641–663.

Rosenbaum, Paul R and Donald B Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 1983, *70* (1), 41–55.

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*, 2007, *18* (5), 429.

Segerson, Kathleen and Thomas J Miceli, "Voluntary environmental agreements: good or bad news for environmental protection?," *Journal of environmental economics and management*, 1998, *36* (2), 109–130.

Shogren, Jason F. and Laura O. Taylor, "On Behavioral-Environmental Economics," *Review of Environmental Economics and Policy*, apr 2008, *2* (1), 26–44.

Stoop, Jan, Charles N Noussair, and Daan van Soest, "From the Lab to the Field: Cooperation among Fishermen," *Journal of Political Economy*, 2012, *120* (6), 1027–1056.

Sturm, Bodo and Joachim Weimann, "Experiments in environmental economics and some close relatives," *Journal of Economic Surveys*, 2006, *20* (3), 419–457.

Thaler, Richard H. and Cass R. Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*, New Haven, CT: Yale University Press, 2008.

United States Department of Energy, Energy Information Administration, "Survey Form EIA-826, Monthly Electric Utility Sales and Revenue Report with State Distributions," https://www.eia.gov/survey/#eia-826 2014. Accessed: 2016-02-27.

Voors, Maarten, Ty Turley, Andreas Kontoleon, Erwin Bulte, and John A List, "Exploring whether behavior in context-free experiments is predictive of behavior in the field: Evidence from lab and field experiments in rural Sierra Leone," *Economics Letters*, 2012, *114* (3), 308–311.

Walker, James M, Roy Gardner, and Elinor Ostrom, "Rent dissipation in a limited-access common-pool resource: Experimental evidence," *Journal of Environmental Economics and Management*, 1990, *19* (3), 203–211.

Wichman, Casey J., "Perceived price in residential water demand: Evidence from a natural experiment," *Journal of Economic Behavior & Organization*, 2014, pp. 1–16.

Yoeli, Erez, Moshe Hoffman, David G Rand, and Martin A Nowak, "Powering up with indirect reciprocity in a large-scale field experiment," *Proceedings of the National Academy of Sciences*, 2013, *110* (Supplement 2), 10424–10429.