

What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?

Steven D. Levitt and John A. List

Economists have increasingly turned to the experimental model of the physical sciences as a method to understand human behavior. Peer-reviewed articles using the methodology of experimental economics were almost nonexistent until the mid-1960s and surpassed 50 annually for the first time in 1982; and by 1998, the number of experimental papers published per year exceeded 200 (Holt, 2006). Lab experiments allow the investigator to influence the set of prices, budget sets, information sets, and actions available to actors, and thus measure the impact of these factors on behavior within the context of the laboratory. The allure of the laboratory experimental method in economics is that, in principle, it provides *ceteris paribus* observations of individual economic agents, which are otherwise difficult to obtain.

A critical assumption underlying the interpretation of data from many laboratory experiments is that the insights gained in the lab can be extrapolated to the world beyond, a principle we denote as generalizability. For physical laws and processes like gravity, photosynthesis, and mitosis, the evidence supports the idea that what happens in the lab is equally valid in the broader world. The American astronomer Harlow Shapley (1964, p. 43), for instance, noted that “as far as we can tell, the same physical laws prevail everywhere.” In this manner, astronomers are able to infer the quantity of certain gases in the Sunflower galaxy, for example, from observations of signature wavelengths of light emitted from that galaxy.

■ *Steven D. Levitt is the Alvin H. Baum Professor of Economics and John A. List is Professor of Economics, University of Chicago, Chicago, Illinois. Levitt and List are both Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their e-mail addresses are <s-levitt@uchicago.edu> and <jlist@uchicago.edu>, respectively.*

The basic strategy underlying laboratory experiments in the physical sciences and economics is similar, but the fact that humans are the object of study in the latter raises special questions about the ability to extrapolate experimental findings beyond the lab, questions that do not arise in the physical sciences. While few scientists would argue that observation influences whether Uranium₂₃₉ would emit beta particles and turn into Neptunium, human behavior may be sensitive to a variety of factors that systematically vary between the lab and the outside world. In particular, we argue, based on decades of research in psychology and recent findings in experimental economics, that behavior in the lab is influenced not just by monetary calculations, but also by at least five other factors: 1) the presence of moral and ethical considerations; 2) the nature *and* extent of scrutiny of one's actions by others; 3) the context in which the decision is embedded; 4) self-selection of the individuals making the decisions; and 5) the stakes of the game. The remainder of this paper is devoted to examining how each of these factors influences decision making and the extent to which the environment constructed in the lab does or does not conform to real-world interactions on these various dimensions.¹

We begin by presenting a simple model in which utility maximization is influenced not only by wealth maximization, but also by an individual's desire to "do the right thing" or make the "moral" choice. We then discuss the empirical evidence concerning the role of the five factors (above) in laboratory experiments.² Although our arguments apply more generally (Levitt and List, 2006), we focus the bulk of the discussion on the class of experiments that is believed to measure pro-social preferences. We provide a summary of the most popular games of this type in Table 1. We next discuss the extent to which the five factors systematically differ between laboratory experiments and naturally occurring environments, and explore how these differences affect the generalizability of experimental results outside the lab. We conclude that, just as is the case with naturally-occurring data, great caution is required when attempting to generalize lab results out of sample: both to other populations and to other situations. Interpreting laboratory findings through the lens of theory helps us to understand the observed pattern of results and facilitates extrapolation of lab results to other environments. Field

¹ There are instances where generalizability might not be of first-rate importance. For example, when testing a general theory, generalizability might not be a concern. In fact, as a first test of theory, an experimenter might wish to create an artificial environment for its own purpose: to create a clean test of the theory. Another example would be using the lab for methodological purposes—that is, to inform field designs by abstracting from naturally-occurring confounds.

² This list certainly does not exhaust the set of reasons that lab experiments may not provide direct guidance with respect to behavior outside the lab. For instance, subjects tend to have less experience with the games they play in the lab, and there is no opportunity to seek advice from friends or experts in the lab. Also of potential importance is the fact that outside the lab, the design of institutions may be driven by sophisticated agents who seek ways to exploit the anomalous tendencies of those with whom they interact (Glaeser, 2004); this force is not at work inside the lab. For further discussion of these issues, see Harrison and List (2004) and Levitt and List (2006).

Table 1
Summary of Experimental Games Used to Measure Social Preferences

<i>Name of game</i>	<i>Summary</i>	<i>Typical finding</i>	<i>Social preference interpretation</i>
Ultimatum game ^a	A two-stage game where two people, a proposer and a responder, bargain over a fixed amount of money. In the first stage, the proposer offers a split of the money, and in the second stage, the responder decides to accept or reject the offer. If accepted, each player receives money according to the offer; if rejected, each player receives nothing.	<i>Proposer:</i> Majority of offers in the range of 25–50% of fixed amount. Few offers below 5%. <i>Responder:</i> Frequently reject offers below 20% of fixed amount.	<i>Proposer:</i> Fairness <i>Responder:</i> Punish unfair offers: negative reciprocity, fairness preferences, such as inequity aversion
Dictator game ^b	A variant of the ultimatum game: strategic concerns are absent as the proposer simply states what the split will be and the proposer has no veto power, rendering the proposed split as effective.	Usually more than 60% of subjects pass a positive amount of money, with the mean transfer roughly 20% of the endowment.	Altruism; fairness preferences, such as inequity aversion.
Trust game ^c	A sequential prisoner's dilemma game wherein the first mover decides how much money to pass to the second mover. All money passed is increased by a factor, $f > 1$, and the second mover then decides how much money to return to the first mover. In this light, the second mover is a dictator who has been given his endowment by the first mover.	<i>Proposer:</i> Average transfer of roughly 50% of endowment. <i>Responder:</i> Repayment is increasing in transfer. Average repayment rate is nearly 50% of transfer.	<i>Proposer:</i> Trust; foresee positive reciprocity <i>Responder:</i> Trustworthiness, positive reciprocity
Gift exchange game ^d	Similar to the trust game, but the money passed by the first mover (often labeled the "wage" or "price" offer), is not increased by a factor, rather it represents a pure lump-sum transfer. Also, the first mover requests a desired effort, or quality, level in return for the "wage" or "price" offer. The second mover then chooses an effort or quality level that is costly to provide, but increases the first mover's payoff.	<i>Proposer:</i> "Wage" or "price" offer is typically greater than the minimum allowed. <i>Responder:</i> Effort or quality increases in "wage" or "price" offer.	<i>Proposer:</i> Trust; foresee positive reciprocity <i>Responder:</i> Trustworthiness, positive reciprocity
Public goods game ^e	Generalization of the prisoner's dilemma game in that n group members decide simultaneously how much to invest in the public good. The payoff function is given by $P_i = e - g_i + \beta \sum_n g_j$, where e represents initial endowment; g_i is the level of tokens that subject i places in the group account; β is the marginal payoff of the public good; and $\sum_n g_j$ is the sum of the n individual contributions to the public good. By making $0 < \beta < 1 < n\beta$, the dilemma follows.	Players' contribution to public good is roughly 50% of endowment in one-shot games. Many players' contributions unravel to approach 0% in latter rounds of multi-period games	Altruism; fairness preferences, conditional reciprocity

^a See Roth (1995) for a discussion of ultimatum and dictator games. This game was first proposed in the economics literature by Guth, Schmittberger, and Schwarze (1982).

^b This game was first proposed in the economics literature by Kahneman, Knetsch, and Thaler (1986). A related game is the "punishment game," whereby an observer can, for a cost, punish the first mover by subtracting a portion of the first mover's payoff.

^c This game was first proposed in the economics literature by Berg, Dickhaut, and McCabe (1995).

^d This game was first proposed in the economics literature by Fehr, Kirchsteiger, and Reidl (1993), and a related game is Camerer and Weigelt (1988). The payoff function description for the buyer is similar to Fehr, Gächter, and Kirchsteiger's (1997) S13-S16 treatments. In this case, the price represents a pure lump-sum transfer, which differs from the earlier joint profit equation (Fehr, Kirchsteiger, and Riedl, 1993), which was characterized by price increases leading to an increase in the sum of payoffs when $q < 1$.

^e See Ledyard (1995) for a discussion of these types of games. This is a generalization of the famous prisoner's dilemma game.

experiments, which exploit randomization in naturally-occurring settings, offer an attractive marriage of these competing empirical strategies.

A Model of Utility with Wealth and Morality

We begin by developing a model that makes precise our arguments regarding the potential factors that might influence individual decision-making. Many economists, dating back to Adam Smith, have emphasized that factors beyond wealth (for example, acting morally) enter into the utility function.³ We do not claim originality in the ideas we are modeling. Rather, we view the model merely as a useful framework for organizing our discussion about the generalizability of results from laboratory experiments.

A utility-maximizing individual i is faced with a choice regarding a single action a . The choice of action affects the agent's utility through two channels. The first effect is on the individual's wealth (denoted W). The higher the stakes or monetary value of the game, which we denote v , the greater the decision's impact on W . The second effect is the nonpecuniary moral cost or benefit associated with the action, which we denote as M . Decisions which an individual views as immoral, antisocial, or at odds with his or her own identity (Akerlof and Kranton, 2000, 2005) may impose important costs on the decision maker (see also Gazzaniga, 2005). This moral payoff might vary across people, religions, or societies.

In practice, many factors influence the moral costs associated with an action, but for modeling purposes, we focus on just three aspects of the moral determinant. The first of these is the financial externality that an action imposes on others. The greater is the negative impact of an action on others, the more negative the moral payoff M . We model the externality as being an increasing function of the stakes of the game v . The second factor that influences the moral choice is the set of social norms or legal rules that govern behavior in a particular society. For instance, the mere fact that an action is illegal (for example, illicit drug use or smoking in restaurants), may impose an additional cost for partaking in such behavior. We denote these social norms against an action as n , with a greater value of n associated with a stronger norm against a behavior. Third, moral concerns depend on the nature and extent of how an individual's actions are scrutinized—such as whether the act is being televised, is taking place in front of one's children, or is performed under the watchful eye of an experimenter—as well as the way in which the process for reaching the decision and final allocation is emphasized (for example, in bargaining between husbands and wives, it is not just the final allocation that matters, but also the nature of the discussion by which the decision is reached). We denote the effect of scrutiny as s , with higher levels of s associated with greater moral costs.

³ Smith viewed decisions as a struggle between “passions” and an “impartial spectator”—a “moral hector who, looking over the shoulder of the economic man, scrutinizes every move he makes” (Grampp, 1948, p. 317, as cited in Ashraf, Camerer, and Loewenstein, 2005). For formal models of such issues, see, for instance, Becker (1974), Akerlof (1982), and Bernheim (1994).

With these considerations in place, focusing on the case in which utility is additively separable in the moral and wealth arguments, the utility function for individual i is:

$$U_i(a, v, n, s) = M_i(a, v, n, s) + W_i(a, v).$$

Framing the problem of utility maximization in this way yields several predictions. For example, in situations without a moral component, like the choice between investing in a stock or bond index, the model reverts back to a standard wealth maximization problem. However, when the wealth-maximizing action has a moral cost associated with it, the agent will deviate from that action to some extent towards an action that imposes a lower moral cost. The greater is the social norm against the wealth maximizing choice, or the greater the degree of scrutiny when the wealth-maximizing action has a social cost, the larger the deviation from that choice. In both cases, we envision the agent trading-off morality and wealth. When individuals follow different moral codes, they will generally make different choices when faced with the same decision problem. Typically, we expect that as the stakes of the game rise, wealth concerns will increase in importance relative to fairness concerns, although this need not always be the case.⁴

We would also expect that these various determinants of moral costs can interact, although the extent of such interaction remains an open empirical issue. For instance, for any given social norm n , as stakes v rise, the moral penalty for violating a given norm will be greater. As an example, people frown on shoplifting, but are much more forgiving of shoplifting than of embezzling millions of dollars. Likewise, the moral cost of violating a social norm increases as scrutiny s rises. For instance, an individual likely faces a larger utility loss from a crime if his capture is broadcast on CNN rather than merely recorded in his rap sheet.

The relevant social norms and the amount of scrutiny are not necessarily exogenously determined, but can be influenced in the context of real-world situations. For instance, panhandlers often emphasize physical deformities or carry placards claiming veteran's status to elicit greater sympathy from potential givers. When churches use "open" rather than "closed" collection baskets, they are acting in a manner consistent with recognition of the importance of norms and scrutiny, as potential contributors can not only see the total amount already gathered, but neighbors can witness each others' contributions (Soetevent, 2005).

The utility function we describe has relevance for a wide variety of behavior. For instance, the model explains why out-of-town visitors to a restaurant will leave a tip, even though they never intend to dine there in the future. Although leaving a tip imposes a financial cost on the diner, tipping provides an offsetting nonpe-

⁴ In Rabin's (1993) model, for sufficiently high stakes there will be no concern for fairness. Under our model there will be less, but potentially some, concern for others as stakes increase. Alternative models do exist, however; for example, Ledyard (1995) presents a model of voluntary contributions in which altruism and selfishness are traded off in such a way that an increase in the stakes has no influence on individual contributions.

cuniary reward. This behavior holds true even if one is eating alone, but probably even more so when there is a higher degree of scrutiny such as when you are accompanied by business clients, on a first date, or when another diner is observing your actions. Conlin, Lynn, and O'Donoghue (2003) present results from an extensive data set on tipping that confirms many of these intuitions.

Our primary interest here lies in developing the model's implications for the generalizability of lab experiments to naturally occurring contexts. When a laboratory experiment diverges from the real-world environment on certain dimensions of interest, the model provides a framework for predicting in what direction behavior in the lab will deviate from that outside the lab.

Implications for Experiments Designed To Measure Social Preferences

The issues that we raise are relevant for a wide range of experimental results, but their bite is likely to be greatest for those games with potential for a strong moral component to behavior. Research on social preferences, one of the most influential areas in experimental economics in recent years, fits this bill. This broad class of games, described earlier in Table 1, includes dictator and ultimatum bargaining games, public goods games, as well as trust and gift exchange games. Results from these types of experiments have been used to argue that pro-social preferences are important in a wide range of real-world settings (for example, Fehr and Gächter, 2000; Camerer and Fehr, 2004)—an inference based on the assumption that the experimental findings are equally descriptive of the world at large.

In what follows, we examine the empirical evidence on possible complications arising when experimental findings are extrapolated to outside the lab. We are not denying that individuals have social preferences; indeed, our own model assumes that moral costs can be influenced by a concern for others as well as a concern for one's own appearance. Rather, we are interested in the extent to which the lab provides reasonable guidance as to the importance of such behavior in a wide range of naturally-occurring settings.

Scrutiny That Is Unparalleled in the Field

In the typical lab experiment, subjects enter an environment in which they are keenly aware that their behavior is being monitored, recorded, and subsequently scrutinized. Decades of research within psychology highlights the importance of the role obligations of being an experimental subject, the power of the experimenter herself, and the significance of the experimental situation. For instance, Orne (1962) wrote: "Just about any request which could conceivably be asked of the subject by a reputable investigator is legitimized by the quasi-magical phrase, 'This is an experiment,' and the shared assumption that a legitimate purpose will be served by the subject's behavior." Schultz (1969, p. 221) described the lab as having

a “superior–subordinate” relationship matched only by that of “parent and child, physician and patient, or drill sergeant and trainee.” Pierce (1908) warned of such effects almost a century ago:

It is to the highest degree probable that the subject[’s] . . . general attitude of mind is that of ready complacency and cheerful willingness to assist the investigator in every possible way by reporting to him those very things which he is most eager to find, and that the very questions of the experimenter . . . suggest the shade of reply expected. . . . Indeed . . . it seems too often as if the subject were now regarded as a stupid automaton.

The strength of such factors is so compelling that researchers in medical drug trials often go above and beyond using placebo and treatment groups by keeping the administrators themselves in the dark about which patients receive the treatment. In psychology experiments, to avoid demand-induced effects, subjects are often deceived about what exactly the investigator is measuring. In economics, however, deceptive practices are frowned upon. Clearly, the *nature* of scrutiny inherent in the lab is rarely encountered in the field, and represents an important aspect of the situation that needs to be accounted for when generalizing laboratory results.

Our theory suggests that such scrutiny will exaggerate the importance of pro-social behaviors relative to environments without such scrutiny. For example, List (2006) carries out gift exchange experiments in which buyers make price offers to sellers, and in return, sellers select the quality level of the good provided to the buyer. Higher quality goods are costlier for sellers to produce than lower quality goods, but are more highly valued by buyers. List began by running a standard gift exchange game in a laboratory context, but used experienced sports-card traders as subjects. The results mirrored the typical findings with other subject pools: strong evidence for social preferences was observed, in the sense that sellers offered higher quality levels to buyers who offered higher prices—although the sellers were not obligated by the rules of the game to do so. List then carried out a second lab experiment that maintained the central elements of the gift exchange game, but the goods exchanged in this lab treatment were actual baseball cards whose market values are heavily influenced by minor differences in condition that are difficult for untrained consumers to detect. If social preferences are present on the part of card sellers, then buyers who offer more money should be rewarded with higher-quality cards. When card sellers were brought into the lab to sell their cards, which were subsequently professionally graded, the results paralleled those obtained in the standard gift exchange game with student subjects.

List (2006) then moved from a lab environment (in which sellers knew their behavior was being scrutinized) to the sellers’ natural environment. Importantly, dealers in this treatment were unaware that their behavior was being recorded as part of an experiment. Confederates were sent as buying agents to approach sellers on the floor of a sports-card show, instructing them to offer different prices in

return for sports-cards of varying quality, just as in the lab treatment described above. When the dealers believed that consumers could not have the card graded or when there was likely to be little future interaction, little statistical relationship between price and quality emerged. Only when there were reputational consequences to a dealer (that is, when quality was verifiable and the potential for a long-term relationship existed), was high quality provided. The social preferences so routinely observed in the lab—even for this very same group of traders—were attenuated in the field.

Other field-generated data yield similar conclusions. For example, making use of personnel data from a leading United Kingdom–based fruit farm, Bandiera, Rasul, and Barankay (2005) find that behavior is consistent with a model of social preferences when workers can be monitored: when other workers can observe their productivity, workers internalize the negative externality that they impose on others under a relative compensation scheme. Yet this effect disappears when workers cannot monitor each other, which rules out pure altruism as the underlying cause of workers' behavior. Being monitored proves to be the critical factor influencing behavior in this study.

Relatedly, Benz and Meier (2006) compare how individuals behave in donation laboratory experiments with how the same individuals behave in the field. They find some evidence of correlation across situations, but find that subjects who have never contributed in the past to the charities gave 60 percent of their endowment to the charity in the lab experiment. Similarly, those who chose not to give to the charities in the two-year period after the experiment gave more than 50 percent of their experimental endowment to the charities in the lab experiment. Similar insights are reported in Laury and Taylor (forthcoming), who find little correlation between an “altruism parameter” estimated from a public goods lab experiment and actual contributions to a real public good (in this case, an urban tree-planting nonprofit organization).

In a “dining game,” Gneezy, Haruvy, and Yafe (2004) find that behavior in a social dilemma game in the laboratory exhibits a considerable level of cooperative behavior—in the lab, students showed great reluctance to impose negative externalities. Yet, in a framed field experiment that resembles the laboratory game—diners were taken to eat at a restaurant—they find no evidence of cooperative play, even though both experimental samples are drawn from the same student population. They speculate that unfamiliarity with the task and confusion are two reasons why negative externalities are influential in the lab but not in the field. Such results are consistent with our model.

Overall, these results are consistent with the wealth of psychological literature that suggests there is only weak evidence of cross-situational consistency of behavior (for example, Mischel, 1968; Ross and Nisbett, 1991). Long ago, Hartshorne and May (1928) discovered that people who cheat in one situation are not the people who cheat in another. If this result spills over to measurement of pro-social preferences, it means either that (a) there is not a general cross-situational trait

called “social preferences,” and/or (b) the subjects view one situation as relevant to social preferences and the other as irrelevant.

Anonymity in the Lab and the Field

Another element of scrutiny is the degree of anonymity conferred upon experimental participants. Anonymity in this case takes two forms. One aspect of anonymity is between experimenter and subject; in some research designs, the experimenter cannot determine what actions the subject takes. This aspect of anonymity is our primary focus. Additionally, there is the question of anonymity among subjects, an issue to which we devote less attention.

In the typical lab experiment, the identity of the subject can readily be linked to individual choices by the experimenter. Our theory predicts that the absence of anonymity will be associated with an increased level of pro-social behavior relative to settings in which individuals are more anonymous.

If the lack of anonymity between the experimenter and subject contributes to pro-social behaviors, then taking steps to reduce the extent to which subjects are knowingly observed should reduce the amount of such behavior.⁵ To accomplish this goal, Hoffman et al. (1994; 1996) used a “double-blind” approach wherein the experimental monitor could not infer individual subjects’ actions in a dictator game. Hoffman, McCabe, Shachat, and Smith (1994) find that 22 of 48 dictators (46 percent) donate at least \$3 of a \$10 pie under normal experimental conditions, but when subject–experimenter anonymity is added, only 12 of 77 dictators (16 percent) give at least \$3. Hoffman, McCabe, Shachat, and Smith (1994, p. 371) conclude that observed “behavior may be due not to a taste for ‘fairness’ (other-regarding preferences), but rather to a social concern for what others may think, and for being held in high regard by others.” Davis and Holt (1993, p. 269) note that these results “indicate that this apparent generosity is not altruism, but rather seems to arise largely from concerns about opinions of outside observers,” which not only highlights the power of anonymity but also the important interaction between lab and anonymity effects. Consistent with this interpretation, Andreoni and Bernheim (2006) report subjects are much more likely to split the pie 50–50 in dictator games as scrutiny increases.⁶

List, Berrens, Bohara, and Kerkvleit (2004) adopt a different approach to generating anonymity between the subject and experimenter (as well as among subjects) using a “randomized response” technique. In this approach, for instance, a subject is told to answer “no” if either (a) they chose not to contribute to a public good, or (b) their mother was born in the first six months of the year. The

⁵ We find the lab a fine tool to explore this type of scrutiny, yet manipulating the experimental environment in this manner may induce other difficulties in interpretation. For example, lessons learned from social psychologists teach us that such efforts to ensure anonymity might result in subjects inferring that the experimenter “demands” them to behave in a manner that might be deemed unacceptable (Loewenstein, 1999).

⁶ It should be noted, however, that Bolton, Zwick, and Katok (1998) and Laury and Taylor (1995) collect data that cast doubt on Hoffman, McCabe, Shachat, and Smith’s (1994) results.

experimenter therefore cannot determine with certainty whether the subject contributed to the public good or not. List, Berrens, Bohara, and Kervleit (2004) found that as decisions became less anonymous, a greater number of subjects opted to give to the public good in a one-shot decision. Both the degree of anonymity between the experimenter and subject, as well as anonymity among subjects, proved important.

Other dimensions of anonymity can also affect giving. For instance, Haley and Fessler (2005) find that giving in a dictator game significantly increases when a pair of eyes is shown on the computer screen where the dictator makes the allocation. This simple manipulation—meant to signal that the subjects' actions were being observed—increased the proportion of nonzero givers from 55 percent in the control treatment to 88 percent in the “eyespot” treatment. Likewise, Allen (1965) reports that increases in privacy reduce conformity. Individuals are also more likely to conform with the social norm of hand-washing when they are being observed (Harris and Munger, 1989).

Context Matters and Is Not Completely Controlled By the Experimenter

The actions people take are affected by a dazzlingly complex set of relational situations, social norms, frames, past experiences, and the lessons gleaned from those experiences. Consequently, the experimental investigator often lacks complete control over the full context within which the subject makes decisions (see also Harrison and List, 2004).

Experimentalists are fully aware that context in their instructions, inducing role-playing, framing, and the like can influence subject behavior (for example, Roth, 1995; Hertwig and Ortmann, 2001; Bohnet and Cooter, 2005). In a wide range of experimental settings, subtle manipulations have been shown to have drastic effects on actions. Rates of defection in prisoner dilemma games swing wildly depending on whether subjects are playing a “Community” or “Wall Street” game (Ross and Ward, 1996); using terms like “opponents” versus “partners” influences play in a myriad of games (Burnham, McCabe, and Smith, 2000, offer an example); asking people to “contribute” or to “allocate” funds in a linear public goods game matters, as does whether the game is framed as a positive externality or a negative one (Andreoni, 1995). Further, whether the agent “punishes” or “assigns” points to other agents can considerably influence play (for example, Gintis, 2001).

Contextual factors beyond the control of the experimenter appear to have equally profound impacts on actions. Henrich et al. (2005) provide powerful evidence of such effects. This group of scholars conducted one-shot ultimatum, dictator, and public goods games in 15 different small-scale communities in developing countries. They found enormous variation in behavior across communities, differences they were able to relate to patterns of everyday life and the social norms operating in these various communities. For instance, as Henrich et al. (2005, p. 31) note, the Orma readily recognized “that the public goods game was similar to the *harambee*, a locally-initiated contribution that Orma households make when

a community decides to construct a public good such as a road or school,” and subsequently gave quite generously. Likewise, among the whale-hunting Lamalera of Indonesia and the Ache in Paraguay, societies with strong norms of sharing, very generous ultimatum game offers are observed and very few offers are rejected. Alternatively, in small-scale foraging societies, such as the Hadza of Tanzania, low offers and high rejection rates are observed in ultimatum games. As Henrich et al. note (2005, p. 33) these “contrasting behaviors seem to reflect their differing patterns of life, not any underlying logic of hunter-gatherer life ways.”

In all of the experiments Henrich et al. (2005) conducted, the context that the experimenter can control—the payoffs, the description of the way the game is played, and so on—was almost identical. But the context that actors themselves brought to the game and that experimenters cannot control—like past experiences and internalized social norms—proved centrally important in the outcome of play.

These examples highlight that an aspect of the lab over which experimenters have incomplete control is that subjects may not be playing the game that the experimenter intends. For instance, lab experiments in economics often seek to eliminate concerns as to whether behavior is motivated by a desire to build a reputation by using one-shot experimental designs. The basis for this methodology is that in a one-shot game, subjects will only display cooperative or pro-social behavior out of “social preference reciprocity,” rather than because they are seeking to build and maintain a good reputation so other people will cooperate with them in the future. However, many real-world activities that have aspects of dictator, ultimatum, trust, or gift exchange games, public good provision, and other social dilemmas are typically not one-time encounters, but rather repeated games (for example, Hoffman, McCabe, and Smith, 1996; Ortmann and Hertwig, 2000; Harrison and Rutstrom, 2001). Effectively, personal experiences may cause the subjects to play these one-shot games as if they have some repetition, and the experimenter may have little or no ability to moderate this phenomenon. The Henrich et al. (2005) study of ultimatum games around the world showed that participants in laboratory games are likely to retrieve experiences and strategies that, unbeknownst to the experimenter, change the nature of the games. If an experimenter mistakenly assumes that the agent is treating the game as one-shot, reputation-building behavior can be misconstrued as social preferences.

While researchers might hope that experimental subjects will make clear strategic adjustments from repeated contexts to one-shot games, the empirical evidence is mixed. For instance, in a review of 15 studies that compare behavior across voluntary contribution games where subjects are randomly re-matched with new partners every round, as opposed to being paired with the same subjects over all rounds, Andreoni and Croson (forthcoming) report that five studies find more cooperation among the randomly re-matched, six find more cooperation among the repeatedly paired, and four studies fail to find a difference between the two treatments.

On the other hand, Fehr and Fischbacher (2003) find that responders react strongly to the possibility of acquiring a reputation; Andreoni and Miller (1993)

report similar insights using data drawn from a prisoner's dilemma game. However, even results that suggest that subjects have an ability to distinguish between situations that have different prospects for future interaction do not necessarily imply that subjects behave in a one-shot experimental situation as if *no* prospect exists for future interaction. The received results are entirely consistent with a model whereby subjects recognize the difference between games with and without an explicit future, but still hold some prospect for future interaction in games described as one-shot (Samuelson, 2005).

While we know of no evidence that suggests those who exhibit strong social preferences in the lab behave similarly outside the lab, we do not doubt that such evidence can be collected. Yet, even if such data are gathered, many simple manipulations in the lab experiment can yield drastically different measures of individual propensities. This result does not necessarily imply that preferences are labile. Rather, we view such data as evidence that when critical elements of the situation change, behavior will change in predictable ways.⁷

Stakes

Our model predicts that in games that have both a morality and wealth component, financial concerns will take on increasing prominence as the stakes rise. The evidence in the literature is only partially consistent with this view. In dictator games, a large increase in stakes generally leads to a less-than-proportionate increase in money transferred. For example, in Carpenter, Verhoogen, and Burks (2005), an increase in stakes from \$10 to \$100 caused the median offer to drop from 40 percent to 20 percent of the endowment. This result is much weaker for smaller changes in stakes: Cherry, Frykblom, and Shogren (2002) find no perceptible differences in offers across a \$10 and \$40 dictator game. Stakes effects have also been found in second-mover play in ultimatum games, in which the acceptance rate is generally increasing in the amount offered, conditional on the share offered—that is, a \$1 offer in a \$5 game is rejected more often than a \$100 offer in a \$500 game. Slonim and Roth (1998) find that in each range of offers below 50 percent, the acceptance rate goes up as the level of stakes increase (from 60 to 1500 Slovak koruna, the latter of which represents eight days of wages for the typical worker). In another type of game that involves some form of trust, the centipede game, Parco, Rapoport, and Stein (2002) similarly find that raising financial incentives causes a breakdown in mutual trust.⁸ Fehr, Fischbacher, and Tougareva (2002), however, report fairness concerns play an important role for both low and high stakes in trust and gift exchange games.

⁷ In this spirit, our arguments bear similarities to the Lucas critique.

⁸ The centipede game is an extensive form game that involves potentially several rounds of decisions. The game begins with player one deciding whether to take the payoff in the pot or to pass the decision to player two. If passed, player two then has a similar decision over a different payoff space. After each passing of the pot, the summation of payoffs is slightly increased, but the payoffs are arranged so that if one player passes and the opponent takes the pot, the player that passed receives less than if he or she had taken the pot.

We are not arguing that low stakes games in the lab have no market parallels; we take part in such transactions in well-functioning markets everyday. Our point is that if the analyst does not account properly for the differences in stakes across settings, inaccurate inference concerning the importance of pro-social preferences will likely result. The magnitude of such mismeasurement is a rich area for future research, and it would be interesting to compare the size of the low-stakes effect with that of the other factors discussed above.

Selection into the Experiment

If participants in laboratory studies differ in systematic ways from the actors engaged in the targeted real-world settings, attempts to generalize lab results directly might be frustrated. Most laboratory experiments have been conducted using students who self-select into the experiments. As Doty and Silverthorne (1975, p. 139) note, volunteers in human research “typically have more education, higher occupational status, earlier birth position, lower chronological age, higher need for approval and lower authoritarianism than non-volunteers.” Indeed, Rosenthal and Rosnow (1969) conclude that social experimentation is largely the science of “punctual college sophomore” volunteers, and have further argued that subjects are more likely to be “scientific do-gooders,” interested in the research, or students who readily cooperate with the experimenter and seek social approval (see also Orne, 1962).⁹

In contrast, market participants are likely to be a selected sample of individuals whose traits allow them to excel in the marketplace. If such markets select agents who place a higher (or lower) value of W (or M) on decision tasks than student subjects, then one might suspect that the nature of the student selection into lab experiments might yield exaggerated pro-social behavior relative to such markets. On the other hand, lab participants may have less pro-social preferences than those who select into particular naturally-occurring environments, such as the clergy or public defenders.

One approach to investigating subject pool biases is to examine whether professionals, or other representative agents, and students behave similarly in laboratory experiments. Fehr and List (2004) examine experimentally how chief executive officers (CEOs) in Costa Rica behave in trust games and compare their behavior with that of Costa Rican students. They find that CEOs are considerably

⁹ When experimentally naïve high school students were asked, “How do you think the typical human subject is expected to behave in a psychology experiment?” over 70 percent circled characteristics labeled “cooperative” and “alert” (Rosenthal and Rosnow, 1973, pp. 136–7). However, these discussions typically revolve around social psychology experiments. Since economic experiments involve different subject matter and involve monetary payments, such arguments might not generalize across disciplines. Kagel, Battalio, and Walker (1979) offer some evidence that volunteer subjects in an economics experiment have more interest in the subject than nonvolunteers, but other important variables are not different across volunteers and nonvolunteers.

more trusting and exhibit more trustworthiness than students.¹⁰ These differences in behavior may mean that CEOs are more trusting in everyday life, or it may be that CEOs are more sensitive to the lab and non-anonymity effects discussed above, or that the stakes are so low for the CEOs that the sacrifice to wealth of making the moral choice is infinitesimal.

A related issue concerns the possibility that only certain types of participants—students or professionals—are willing to take part in the experiment. For example, volunteers, whether students or CEOs, who have social preferences or who readily cooperate with the experimenter and seek social approval *might* be those who are most likely to participate in the experiment. In this case, games that purport to measure pro-social behaviors will yield upper bound estimates on the propensities of the target population.

Some limited but suggestive data from field and lab experiments supports this argument about selection into laboratory gift exchange experiments. When List (2006) approached a number of sports-card sellers about participating in the laboratory experiment described earlier, some sellers declined his invitation. But later and unbeknownst to them, these same sellers participated in the parallel field experiment. Those who declined to participate in the lab portion of the experiment were less pro-social in the field compared to dealers who agreed to participate in the lab experiment, although the differences were imprecisely measured due to small sample sizes and therefore not statistically significant at conventional levels. In a series of dictator games, Eckel and Grossman (2000) compare volunteers (those who select into the lab for an experiment) and pseudo-volunteers (those who are part of a class that is asked to participate during class time). Besides finding observable differences across the subject pools, they find that pseudo-volunteers give more than volunteers, but also that volunteers behave in a less extreme manner than pseudo-volunteers.

Artificial Restrictions on Choice Sets and Time Horizons

Another issue closely related to those that we raise in the model is that in experiments, the researcher creates a set of rules governing the interactions, chooses the wording of instructions, and defines the set of actions the subject is allowed to take. In stark contrast, in naturally occurring environments, the choice set often is almost limitless and institutions arise endogenously.

Even among those who choose to participate in lab experiments, restrictions on the available choice set can affect observed behavior. For example, pro-social behavior might be observed less frequently in markets merely because people can avoid situations where they must make costly contributions to signal their generosity. This idea is illustrated in Lazear, Malmendier, and Weber (2006), who in an

¹⁰ Harbaugh, Krause, Liday, and Vesterlund (2003) conducted a set of trust experiments with students in third, sixth, ninth, and twelfth grade and found little variation across the participants in terms of trust and trustworthiness. However, in dictator games, the youngest children tend to make considerably smaller transfers than do older children and adults.

experiment allowed agents an opportunity to pay to opt out of playing the dictator game. They find that “the majority of subjects share without really wanting to, as evidenced by their willingness to avoid the dictator game and to even pay for avoiding it.” Such forces are readily observable in the field as well—panhandlers receive less in gifts if passersby can easily “sort” themselves to the other side of the road to avoid interaction.

Another example of how the available choice set influences play in the dictator game can be found in Bardsley (2005) and List (forthcoming). In the typical dictator game, the subject is given, say, \$10 and asked what portion the subject would like to share with the other player who received less than \$10. The experiment is framed such that “giving nothing” is the least generous act, and substantial sums of money are given away. If instead, the subject is given \$10 and is told that the rules allow giving any portion of this money away to the second player, or confiscating up to an additional \$10 from the other player, subjects give little to the other player. Likewise, Andreoni, Brown, and Vesterlund (2002) make use of a sequential public goods game with an asymmetric equilibrium and find results consistent with the data in Bardsley and List. Real-world contexts typically offer the option of both giving and receiving, which may help explain in part why, contrary to the lab environment, people rarely receive anonymous envelopes with cash inside.

These examples also highlight that laboratory experiments often restrict the response mode to a single dimension, whereas real-world settings almost always involve multiple response modes. Consider again the act of giving within a dictator game. An agent who is inclined to help others might give money in the dictator game in the lab. In the field, this same agent might give nothing, instead using other more efficient means to express generosity, such as volunteering time to help others. In this example, the laboratory evidence is consistent with some type of broader preference, but that preference might be expressed through a different activity in the field. Thus, when making comparisons across domains, one should take care to include all relevant dimensions.

Related to the choice set is the nature and temporal aspect of the task. Laboratory experiments usually consist of at most a few hours of fairly passive activities. For example, in a standard trust, or gift exchange, games in the laboratory, student subjects typically play several rounds of the game by choosing an effort or wage level (by circling or jotting down a number) in response to pecuniary incentive structures. The experiment usually lasts about an hour and a result often observed is that effort levels and wages are positively correlated. Such results are often interpreted as providing support for the received labor market predictions of Akerlof (1982) that the employer–employee relationship contains elements of gift exchange.

Such inference raises at least two relevant issues. First, is real-world, on-the-job effort different in nature from that required in lab tasks? Second, does the effect that we observe in the lab manifest itself over longer time periods? The evidence is sparse within the experimental economics literature on these issues, but studies are beginning to emerge. Using data gathered from a test of the gift exchange

hypothesis in an actual labor market, Gneezy and List (2006) find that worker effort in the first few hours on the job is considerably higher in a “gift” treatment than in a “non-gift” treatment. However, after the initial few hours, no difference in outcomes was observed over the ensuing days of work. The notion that positive wage shocks do not invoke long-run effects in effort levels is also consistent with data reported in Al-Ubaydli, Steffen, Gneezy, and List (2006), Hennig-Schmidt, Rockenbach, and Sadrieh’s (2006) field experiment, and Kube, Maréchal, and Puppe (2006). These results suggest that great care should be taken before making inference from short-run laboratory experiments to long-run field environments.¹¹

Such insights are consonant with results from the psychology literature in that important behavioral differences exist between short-run (“hot”) and long-run (“cold”) decision making. In the hot phase, visceral factors and emotions might prove quite important, whereas in the cold phase, immediate reactions may be suppressed. Loewenstein (2005) reviews some of the empirical evidence on behavioral differences across cold and hot states.

Generalizing the Findings of Laboratory Experiments to Actual Markets

We believe that several features of the laboratory setting need to be carefully considered before generalizing results from experiments that measure pro-social behaviors to market settings they purport to describe. The model that we advance provides a framework to begin a discussion of the relevant economic and psychological factors that might influence behavior. Such factors include both the representativeness of the situation as well as the representativeness of the population: the nature and extent of scrutiny, the emphasis on the process by which decisions are made, the artificial limits placed on the action space, the imposition of task, the selection rules into the environments, and the stakes typically at risk.

In contrast to the lab, many real-world markets operate in ways that make pro-social behavior much less likely. In financial markets, for instance, the stakes are large, actors are highly anonymous, and little concern seems to exist about future analysis of one’s behavior. Individuals with strong social preferences are likely to self-select away from these markets, instead hiring agents who lack such preferences to handle their financial dealings. Thus, one must take great care when claiming that patterns measured in the experimental economics laboratory are

¹¹ Naturally occurring data concerning the effects of pay shocks on work effort is mixed. Chen (2005), who uses a large data set drawn from the Australian Workplace Industrial Relations Survey to explore reciprocity in the workplace, finds little evidence consistent with reciprocity. Lee and Rupp (2006) examine the effort responses of U.S. commercial airline pilots following recent pay cuts, and find that such effects are very short-lived, consistent with Gneezy and List (2006). In the first week after a pay cut, frequent and longer flight delays are observed, but after the first week, airline flight performance reverts to previous levels. On the other hand, Krueger and Mas (2004) provide evidence consistent with negative reciprocity on the part of disgruntled Firestone employees, and Mas (forthcoming) documents persistent adverse effects on police performance following arbitration decisions in favor of the municipality.

shared broadly by agents in certain real-world markets. It seems highly unlikely, for instance, that at the end of a day's trading, a successful trader would seek out the party that was on the wrong side of a market move and donate a substantial fraction of the day's profits to the person who lost—even though parallel behavior is routine in certain experiments. In addition, there is some trend in retail transactions away from an environment that fosters pro-social behavior towards one that does not, because of the rise of Internet sales and large retail chains.

In some naturally occurring settings, however, lab findings may *understate* the extent of pro-social actions. The degree of scrutiny present when making choices in front of one's children, or when one's actions are being televised, may far outstrip that in the lab. Thus, Levitt (2005) finds no evidence of discrimination towards blacks or women by participants on the televised game show "The Weakest Link." Also, inference from lab experiments measuring social preferences is typically based on interactions of complete strangers, anonymity between subjects, an absence of any social relations between subjects, and restricted communication channels between subjects. To the extent that such factors are not introduced into the lab environment by experimental subjects (yet, see Eckel and Wilson, 2004, footnote 15; Samuelson, 2005), such factors in the real world could induce a greater level of social preferences. For instance, one expects to find a great deal of altruism amongst family members, close friends, and comrades-in-arms. It is important to stress, however, that in settings with repeated interactions, it is difficult to distinguish between pro-social preferences and strategic actions taken with the goal of reputation building. Purely selfishly motivated individuals may forego short-term private gains, for instance, to support a cooperative equilibrium in an infinitely repeated prisoner's dilemma. When a firm treats an employee in a manner that is consistent with social preferences, the firm may simply be pursuing profit maximization. More careful empirical work in this area is warranted.

In addition, other important forces that are at work in naturally occurring markets can be absent in the lab. As Glaeser (2004) notes, it may be in the best interests of sophisticated agents to design institutions in such a way as to exploit the anomalous tendencies of others with whom they interact. Della Vigna and Malmendier (2006) provide an excellent example in the manner in which health clubs structure fees. Levitt (2004) similarly shows that bookmakers set lines that take advantage of the inherent biases of bettors. If certain markets are arranged whereby entrepreneurs must raise the prevalence of social behaviors to maximize their own profits, then the lab evidence might underestimate the importance of social preferences in comparison to such markets.¹²

¹² In the long run, the impact of endogenously generated institutions on the amount of pro-social behavior is ambiguous. For instance, we learn from evolutionary biology that selection pressures can work against organisms that overextract from their hosts. In this case, firms that implement such policies can be displaced by firms that extract less from consumers. Even without such evolutionary competition, or in cases where incumbency advantages are large, if such institutions significantly raise the cost of faulty decision making, learning might occur more quickly and to a greater degree in markets than in the lab. However, if feedback mechanisms are weak in the field, such effects may generally not be observed.

Concluding Remarks

Perhaps the most fundamental question in experimental economics is whether findings from the lab are likely to provide reliable inferences outside of the laboratory. In this paper, we argue that the choices that individuals make depend not just on financial implications, but also on the nature and degree of others' scrutiny, the particular context in which a decision is embedded, and the manner in which participants are selected to participate. Because the lab systematically differs from most naturally occurring environments on these dimensions, experiments may not always yield results that are readily generalizable. Specifically, we argue that lab experiments generally exhibit a special type of scrutiny, a context that places extreme emphasis on the process by which decisions and allocations are reached, and a particular selection mechanism for participants. In contrast, many real-world markets are typified by a different type of scrutiny, little focus on process, and very different forms of self-selection of participants.

The points we make concerning generalizability of lab data apply with equal force to generalizing from data generated from naturally occurring environments. Empirical economists understand that studies of sumo wrestlers or sports-card traders cannot be seamlessly extrapolated to other economic settings. Any empirical estimate requires an appropriate theory for proper inference—and this lesson holds whether the data are obtained in the lab, from coin collector shows, or from government surveys. We envision similar practices among experimental economists: just as economists would want a model of firm and consumer behavior to tell what parameter we are estimating when we regress quantities on prices, we need a model of laboratory behavior to describe the data-generating process, and how it is related to other contexts. Theory is the tool that permits us to take results from one environment to predict in another, and generalizability of laboratory evidence should be no exception.

The discussion in this paper suggests three important conclusions regarding research design and interpretation. First, combining laboratory analysis with a model of decision-making, such as the model we present in this paper, expands the potential role of lab experiments. By anticipating the types of biases common to the lab, experiments can be designed to minimize such biases. Further, knowing the sign and plausible magnitude of any biases induced by the lab, one can extract useful information from a study, even if the results cannot be seamlessly extrapolated outside the lab. In this sense, even in cases where lab results are believed to have little generalizability, some number from a laboratory estimate is better than no number, provided that a theoretical model is used to make appropriate inference.

Second, by adopting experimental designs that recognize the potential weaknesses of the lab, the usefulness of lab studies can be enhanced. For instance, one approach is to “nest” laboratory experiments one within another and then examine the different results of the related experiments. This approach may serve to “net out” laboratory effects and thus reveal more about deep structural parameters than running a simple, more traditional, experimental design. Additionally, lab exper-

iments that focus on *qualitative* insights can provide a crucial first understanding and suggest underlying mechanisms that might be at work when certain data patterns are observed. Indeed, many of the arguments that we put forth in this study can be usefully explored using a laboratory experiment. Further, in the area of social dilemmas, laboratory experiments might help to illuminate whether punishing those who defect from pro-social behavior is a more powerful force than rewarding those who practice pro-social behavior.

Finally, recognizing that shortcomings exist in both lab-generated data and data from natural settings, an empirical approach that combines the best of each is appealing. A well-designed field experiment, incorporating the virtues of true randomization, but in a setting more representative of the behavior about which economists are seeking to learn, can serve as a bridge connecting these two empirical approaches.

■ *Thanks to seminar participants at the 2005 International Meetings of the ESA for useful suggestions. Excellent suggestions from James Andreoni, Nicholas Bardsley, Gary Becker, Gary Charness, David Cooper, Dan Gilbert, Uri Gneezy, Hays Golden, Glenn Harrison, Reid Hastie, Dean Karlan, Dan Levin, Jayson Lusk, Ulrike Malmendier, Ted McConnell, Kevin Murphy, Andreas Ortmann, Charles Plott, Jesse Shapiro, Andrei Shleifer, Robert Slonim, and Richard Thaler greatly improved the study. Colin Camerer, Ernst Fehr, and Alvin Roth provided detailed comments and made suggestions that have resulted in many improvements, although not as many as they would have wished. Seminar participants at Brigham Young University, University of Chicago, Laval University, McMaster University, Oberlin College, and the University of Nevada, Reno also provided useful feedback on bits and pieces of this research. Financial support for Levitt came from the National Science Foundation and the Sherman Shapiro Research Fund. An earlier version of this paper that discussed the generalizability of a much wider class of experiments circulated under the title “What Do Laboratory Experiments Tell Us about the Real World?”*

References

- Akerlof, George A. 1982. “Labor Contracts as Partial Gift Exchange.” *Quarterly Journal of Economics*, November, 97(4): 543–69.
- Akerlof, George A., and Rachel E. Kranton. 2000. “Economics and Identity.” *Quarterly Journal of Economics*, August, 115(3): 715–53.
- Akerlof, George A., and Rachel E. Kranton. 2005. “Identity and the Economics of Organizations.” *Journal of Economic Perspectives*, Winter, 19(1): 9–32.
- Allen, Vernon. 1965. “Situational Factors in Conformity.” In *Advances in Experimental and Social Psychology*, vol. 2, ed. Leonard Berkowitz, 133–76. New York: Academic Press.
- Al-Ubaydli, Omar, Steffen Andersen, Uri Gneezy, and John A. List. “Incentive Schemes to Promote Optimal Work Performance: Evidence from a Multi-Tasking Field Experiment.” Unpublished paper, University of Chicago.
- Andreoni, James. 1995. “Cooperation in Public Goods Experiments: Kindness or Confusion?” *American Economic Review*, 85(4): 891–904.
- Andreoni, James, and B. Douglas Bernheim. 2006. “Social Image and the 50–50 Norm: Theory

and Experimental Evidence." <http://www.hss.caltech.edu/media/seminar-papers/bernheim.pdf>.

Andreoni, James, Paul Brown, and Lise Vestlund. 2002. "What Makes an Allocation Fair? Some Experimental Evidence." *Games and Economic Behavior*, 40(1): 1–24.

Andreoni, James, and Rachel Croson. Forthcoming. "Partners versus Strangers: Random Rematching in Public Goods Experiments." *Handbook of Experimental Economic Results*, ed. C. Plott and V. Smith.

Andreoni, James, and John Miller. 1993. "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence." *Economic Journal*, 103(418): 570–85.

Ashraf, Nava, Colin F. Camerer, and George Loewenstein. 2005. "Adam Smith, Behavioral Economist." *Journal of Economic Perspectives*, 19(3): 131–45.

Bandiera, Oriana, Iwan Rasul, and Imran Barankay. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data." *Quarterly Journal of Economics*, 120(3): 917–62.

Bardsley, Nicholas. 2005. "Altruism or Artifice? A Note on Dictator Game Giving." Center for Decision Research and Experimental Economics Discussion Paper 2005-10.

Becker, Gary S. 1974. "A Theory of Social Interactions." *Journal of Political Economy*, 82(6): 1063–93.

Benz, Matthias, and Stephan Meier. 2006. "Do People Behave in Experiments as in Real Life? Evidence from Donations." Institute for Empirical Research in Economics, University of Zurich, Working Paper 248.

Berg, Joyce, John W. Dickhaut, and Kevin A. McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior*, July 10(1): 122–42.

Bernheim, B.D. 1994. "A Theory of Conformity." *Journal of Political Economy*, 102(5): 841–77.

Bohnet, Iris, and Robert D. Cooter. 2005. "Expressive Law: Framing or Equilibrium Selection?" John F. Kennedy School of Government Faculty Research Working Paper RWP03-046.

Bolton, Gary E., Rami Zwick, and Elena Katok. 1998. "Dictator Game Giving: Rules of Fairness Versus Acts of Kindness." *International Journal of Game Theory*, 27(2): 269–99.

Burnham, Terence, Kevin McCabe, and Vernon L. Smith. 2000. "Friend-or-Foe Intentionality Priming in an Extensive Form Trust Game." *Journal of Economic Behavior and Organization*, 43(1):57–73.

Camerer, Colin F., and Ernst Fehr. 2004 "Mea-

suring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists." In *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, eds. Joseph Henrich et al., 55–95. Oxford: Oxford University Press.

Camerer, Colin F., and Keith Weigelt. 1988. "Experimental Tests of a Sequential Equilibrium Reputation Model." *Econometrica*, 56(1): 1–36.

Carpenter, Jeffrey, Eric Verhoogen, and Stephen Burks. 2005. "The Effect of Stakes in Distribution Experiments." *Economics Letters*, March, 86(3): 393–8.

Chen, Paul. 2005. "Reciprocity at the Workplace: Do Fair Wages Lead to Higher Effort, Productivity, and Profitability?" http://gemini.econ.umd.edu/cgi-bin/conference/download.cgi?db_name=esam06&paper_id=222.

Cherry, Todd, Peter Frykblom, and Jason F. Shogren. 2002. "Hardnose the Dictator." *American Economic Review*, 92(4): 1218–21.

Conlin, Michael, Michael Lynn, and Ted O'Donoghue. 2003. "The Norm of Restaurant Tipping." *Journal of Economic Behavior and Organization*, 52(3): 297–321.

Davis, Douglas D., and Charles Holt. 1993. *Experimental Economics*. Princeton University Press.

DellaVigna, Stefano, and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review*, 96(3): 694–719.

Doty, Richard L., and Colin Silverthorne. 1975. "Influence of Menstrual Cycle on Volunteering Behavior." *Nature*, (March, 13), 254(5496): 139–40.

Eckel, Catherine C., and Phillip Grossman. 1996. "Altruism in Anonymous Dictator Games." *Games and Economic Behavior*, 16(2): 181–191.

Eckel, Catherine C., and Philip J. Grossman. 2000. "Volunteers and Pseudo-Volunteers: The Effect of Recruitment Method in Dictator Experiments." *Experimental Economics*, 3(2): 107–120.

Eckel, Catherine C., and Rick K. Wilson. 2004. "Is Trust a Risky Decision?" *Journal of Economic Behavior and Organization*, December, 55(4): 447–65.

Fehr, Ernst, and Urs Fischbacher. 2003. "The Nature of Human Altruism." *Nature*, 425(6960): 785–91.

Fehr, Ernst, Urs Fischbacher, and E. Tougareva. 2002. "Do High Stakes and Competition Undermine Fairness? Evidence from Russia." Institute for Empirical Research in Economics, University of Zürich, Working Paper 120.

Fehr, Ernst, and Simon Gächter. 2000. "Fairness and Retaliation: The Economics of Reci-

procuity." *Journal of Economic Perspectives*, 14(3): 159–81.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger. 1997. "Reciprocity as a Contract Enforcement Device: Experimental Evidence." *Econometrica*, 65(4): 833–60.

Fehr, Ernst, George Kirchsteiger, and Arno Riedl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics*, 108(2): 437–59.

Fehr, Ernst, and John A. List. 2004. "The Hidden Costs and Returns of Incentives—Trust and Trustworthiness among CEOs." *Journal of the European Economic Association*, 2(5): 743–71.

Gazzaniga, Michael S. 2005. *The Ethical Brain*. Dana Press.

Gintis, Herbert. 2001. "The Contribution of Game Theory to Experimental Design in the Behavioral Sciences." *Behavioral and Brain Sciences*, 24(3): 411–12.

Glaeser, Edward. 2004. "Psychology and the Market." *American Economic Association Papers and Proceedings*, 94(2): 408–13.

Gneezy, Uri, Ernan Haruvy, and H. Yafe. 2004. "The Inefficiency of Splitting the Bill: A Lesson in Institution Design." *The Economic Journal*, April, 114(495): 265–80.

Gneezy, Uri, and John A. List. 2006. "Putting Behavioral Economics to Work: Field Evidence of Gift Exchange." *Econometrica*, September, 74(5): 1365–84.

Grapp, William. 1948. "Adam Smith and the Economic Man." *The Journal of Political Economy*, August, 56(4): 315–36.

Guth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization*, 3(2): 367–88.

Haley, Kevin J., and Daniel M. T. Fessler. 2005. "Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game." *Evolution of Human Behavior*, 26(3): 245–56.

Harbaugh, William T., Kate Krause, Steven G. Liday, and Lise Vesterlund. 2003. "Trust in Children." In *Trust, Reciprocity and Gains from Association: Interdisciplinary Lessons from Experimental Research*, ed. Elinor Ostrom and James Walker, 303–22. New York City, NY: Russell Sage Foundation.

Harris, S. J., and K. Munger. 1989. "Effects of an Observer on Hand Washing in Public Restrooms." *Perceptual and Motor Skills*, 69, pp. 733–5.

Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature*, December, 42(4): 1009–55.

Harrison, Glenn W., and Elisabet Rutstrom.

2001. "Doing It Both Ways—Experimental Practice and Heuristic Context." *Behavioral and Brain Sciences*, 24(3): 413–4.

Hartshone, Hugh, and Mark A. May. 1928. *Studies in Deceit*. New York: Macmillan.

Hennig-Schmidt, Heike, Bettina Rockenbach, and Abdolkarim Sadrieh. 2006. "In Search of Workers' Real Effort Reciprocity—A Field and a Laboratory Experiment." Governance and the Efficiency of Economic Systems Working Paper 55. Free University of Berlin, Humboldt University of Berlin.

Henrich, Joseph., et al. 2005. "Economic Man' in Cross-Cultural Perspective: Ethnography and Experiments from 15 Small-Scale Societies." *Behavioral and Brain Sciences*. 28(6): 795–815.

Hertwig, Ralph, and Andreas Ortmann. 2001. "Experimental Practices in Economics: A Challenge for Psychologists?" *Behavioral and Brain Sciences*, 24(4): 383–451.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. 1994. "Preferences, Property Rights, and Anonymity in Bargaining Games." *Games and Economic Behavior*, 7(3): 346–80.

Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith. 1996. "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review*, June, 86(3): 653–60.

Holt, Charles A. 2006. *Markets, Games, and Strategic Behavior: Recipes for Interactive Learning*. Addison-Wesley.

Kagel, John H., Raymond C. Battalio, and James M. Walker. 1979. "Volunteer Artifacts in Experiments in Economics: Specification of the Problem and Some Initial Data from a Small-Scale Field Experiment." *Research in Experimental Economics*, ed. Vernon L. Smith, 169–97. JAI Press.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1986. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review*, 76(4): 728–41.

Krueger, Alan B., and Alexandre Mas. 2004. "Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires." *Journal of Political Economy*, 112(2): 253–89.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe. 2006. "Putting Reciprocity to Work: Positive versus Negative Responses in the Field." University of St. Gallen Department of Economics Discussion Paper 2006-27.

Laury, Susan K., James M. Walker, and Arlington W. Williams. 1995. "Anonymity and the Vol-

untary Provision of Public Goods." *Journal of Economic Behavior and Organization*, 27(3): 365–80.

Laury, Susan K., and Laura O. Taylor. Forthcoming. "Altruism Spillovers: Are Behaviors in Context-free Experiments Predictive of Altruism toward a Naturally Occurring Public Good?" *Journal of Economic Behavior and Organization*.

Lazear, Edward P., Ulrike Malmendier, and Roberto A. Weber. 2006. "Sorting in Experiments." National Bureau of Economic Research Working Paper 12041.

Ledyard, John O. 1995. "Public Goods: A Survey of Experimental Research." In *Handbook of Experimental Economics*, ed. J. Kagel and A. Roth, chap. 2. NJ: Princeton University Press.

Lee, Darin, and Nicholas G. Rupp. 2006. "Retracting a Gift: How Does Employee Effort Respond to Wage Reductions?" <http://www.ecu.edu/econ/wp/05/ecu0523.pdf>.

Levitt, Steven D. 2004. "Why Are Gambling Markets Organized So Differently from Financial Markets?" *Economic Journal*, 114(495): 223–46

Levitt, Steven D. 2005. "Testing Theories of Discrimination: Evidence from the Weakest Link." *Journal of Law & Economics*, 47(2): 431–52.

Levitt, Steven D., and John A. List. 2006. "What Do Laboratory Experiments Tell Us about the Real World?" <http://pricetheory.uchicago.edu/levitt/Papers/jep%20revision%20Levitt%20&%20List.pdf>.

List, John A. 2006. "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions." *Journal of Political Economy*, 114(1): 1–37.

List, John A. Forthcoming. "On the Interpretation of Giving in Dictator Games." *Journal of Political Economy*.

List, John A., Robert Berrens, Alok Bohara, and Joe Kerkvliet. 2004. "Examining the Role of Social Isolation on Stated Preferences." *American Economic Review*, 94(3), pp. 741–52.

Loewenstein, George. 1999. "Experimental Economics from the Vantage-Point of Behavioural Economics." *Economic Journal*, February 109(453): F23–34

Loewenstein, George. 2005. "Hot-cold Empathy Gaps and Medical Decision-Making." *Health Psychology*, 24(4): S49–S56.

Mas, Alexandre. 2006. "Pay, Reference Points, and Police Performance." *Quarterly Journal of Economics*, 121(3): 783–821.

Mischel, Walter. 1968. *Personality and Assessment*. New York; Wiley.

Orne, Martin T. 1962. "On the Social Psycho-

logical Experiment: With Particular Reference to Demand Characteristics and Their Implications." *American Psychologist*, 17(10): 776–83.

Ortmann, Andreas, and Ralph Hertwig. 2000. "Why Anomalies Cluster in Experimental Tests of One-Shot and/or Finitely Repeated Games: Some Evidence from Psychology." <http://home.cerge.cuni.cz/Ortmann/Papers/10baseratesii06142000.pdf>.

Parco, James E., Amnon Rapoport, and William E. Stein. 2002. "Effects of Financial Incentives on the Breakdown of Mutual Trust." *Psychological Science*, 13(3): 292–7.

Pierce, A. H. 1908. "The Subconscious Again." *Journal of Philosophy, Psychology, & Scientific Methods*, 5(10): 264–71.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, December, 83(5): 1281–1302.

Rosenthal, Robert W., and Ralph L. Rosnow. 1969. *Artifact in Behavioral Research*. New York: Academic Press.

Rosenthal, Robert W., and Ralph L. Rosnow. 1973. *The Volunteer Subject*. New York: John Wiley and Sons.

Ross, Lee, and Andrew Ward. 1996. "Naive Realism: Implications for Social Conflict and Misunderstanding." In *Values and Knowledge*, ed. T. Brown, E. Reed, and E. Turiel, 103–35. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ross, Lee, and Richard E. Nisbett. 1991. *The Person and the Situation: Perspectives of Social Psychology*. New York: McGraw-Hill.

Roth, Alvin E. 1995. "Bargaining Experiments." In *The Handbook of Experimental Economics*, ed. J. H. Kagel and E. R. Alvin, 253–342. Princeton, NJ: Princeton University Press.

Samuelson, Larry. 2005. "Economic Theory and Experimental Economics." *Journal of Economic Literature*, March, 43(1): 65–107.

Schultz, Duane P. 1969. "The Human Subject in Psychological Research." *Psychological Bulletin*, 72(3): 214–28.

Shapley, Harlow. 1964. *Of Stars and Men: Human Response to an Expanding Universe*. Westport CT: Greenwood Press.

Slonim, Robert and Alvin E. Roth. 1998. "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic." *Econometrica*, 66(3), pp. 569–96.

Soetevent, Adriaan R. 2005. "Anonymity in Giving in a Natural Context—A Field Experiment in Thirty Churches." *Journal of Public Economics*, 89(11–12): 2301–23, Table 1.