What Do Laboratory Experiments Tell Us About the Real World?

Steven D. Levitt and John A List*
University of Chicago and NBER

27 June 2006

Abstract

An important question facing experimental economists is whether behavior inside the laboratory is a good indicator of behavior outside the laboratory. We begin with a model that assumes the choices that individuals make depend not just on financial implications, but also on the nature and extent of scrutiny by others, the particular context in which a decision is embedded, and the manner in which participants are selected. To the extent that lab and naturally-occurring environments systematically differ on any of these dimensions, the results obtained inside and outside the lab need not correspond. Based on theory and empirical evidence, we argue that lab experiments are a useful tool for generating qualitative insights, but are not well-suited for obtaining deep structural parameter estimates. We conclude that the sharp dichotomy sometimes drawn between lab experiments and data generated in natural settings is a false one. Each approach has strengths and weaknesses, and a combination of the two is likely to provide deeper insights than either in isolation.

JEL: C9 (Design of Experiments), C93 (Field Experiments)

Key words: generalizability; laboratory experiments; field experiments; markets

*Thanks to seminar participants at the 2005 International Meetings of the ESA for useful suggestions. Excellent suggestions from Nicholas Bardsley, Gary Becker, Dan Gilbert, Uri Gneezy, Glenn Harrison, Reid Hastie, Dean Karlan, Jayson Lusk, 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 came from the National Science Foundation and the Sherman Shapiro Research Fund.

Nearly 400 years ago, Galileo performed the first recorded laboratory experiment, timing balls as they rolled down an inclined plane to test his theory of acceleration (Settle, 1961). Since that time, laboratory experiments have been a cornerstone of the scientific method. Feynman (1963) illustrates this fact when noting that "The principle of science, the definition almost, is the following: The test of all knowledge is experiment. Experiment is the sole judge of scientific 'truth.'"

Increasingly, economists have turned to the experimental model of the physical sciences as a method to understand human behavior. Holt (2005) documents that experimental economics is a "boom industry," showing that publications using the methodology were almost non-existent until the mid-1960s, surpassed 50 annually for the first time in 1982, and by 1998 there were more than 200 experimental papers published per year. The allure of the laboratory experimental method in economics is that, in principle, it provides *ceteris paribus* observations of motivated individual economic agents, which are otherwise exceptionally difficult to obtain using conventional econometric techniques. Lab experiments provide the investigator with a means to directly influence the set of prices, budget sets, information set, and actions available to actors, and thus measure the impact of these factors on behavior within the context of the laboratory.[1]

A critical maintained assumption underlying many laboratory experiments is that the insights gained in the lab can be extrapolated to the world beyond, a principle we denote as

[1] This explosion of research in experimental economics has had an important influence on the profession. For example, in the case of modeling preferences, a series of experiments and surveys revealing a disparity between subjects' willingness to pay and willingness to accept compensation, has spurred an entire body of theoretical research into the underlying mechanism at work (e.g., Kahneman and Tversky, 1979; Loomes and Sugden, 1982; Heiner, 1983; Hanemann, 1991; Tversky and Kahneman, 1991). Likewise, the finding that subjects seemingly have interdependent utilities in various environments has led to a theoretical exploration of the economic consequences of social preferences (Rabin, 1993; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Andreoni and Miller, 2002; Charness and Rabin, 2002). Indeed, the foundations of behavioral economics are firmly anchored in the observation that human decision making in the laboratory often deviates from the predictions of standard economic models (Rabin, 1998; Mullainathan and Thaler, 2000). Another indication of the impact of experimental economics on the discipline is Vernon Smith's 2002 Nobel Prize in Economics for his pioneering work in the field. The Nobel Prize Committee described Smith's contribution as "having established laboratory experiments as a tool in empirical economic analysis, especially in the study of alternative market mechanisms" (2002 Nobel Prize Announcement).

generalizability.[2] For physical laws and processes (e.g. gravity, photosynthesis, mitosis), the evidence to date supports the idea that what happens in the lab is equally valid in the broader world. Shapley (1964, p. 43), for instance, noted that "as far as we can tell, the same physical laws prevail everywhere."

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 fundamental questions about the ability to extrapolate experimental findings beyond the lab that do not arise in the physical sciences. While few would question whether $Uranium_{239}$ would emit beta particles and turn into Neptunium in the presence or absence of scientists, human behavior may be sensitive to a variety of factors that systematically vary between the lab and the outside world. In particular, we argue that human decisions are influenced not just by simple monetary calculations, but also by at least three other considerations:

1) The nature and extent to which one's actions are scrutinized by others,

[2] Many different types of phrases have been used to depict the relationship between the lab and the field, with "parallelism," "external validity," and "ecological validity" being the most popular. Parallelism, which we traced to Shapley (1964), is said to be established if the results found in the laboratory hold in other, particularly real-world, situations under *ceteris paribus* conditions (see, e.g., Wilde, 1981; Smith, 1982). Campbell and Stanley (1963) introduced external validity as follows: "external validity asks the question of generalizability: To what populations, settings, treatment variables, and measurement variables can this effect be generalized?" (p. 5). Ecological validity has taken on a multifarious set of meanings, including the notion that a study is ecologically valid if "one can generalize from observed behavior in the laboratory to natural behavior in the world" (Schmuckler, 2001, p. 419). But, confusion arises because it is clear that Egon Brunswik coined the term *ecological validity* to indicate the degree of correlation between a proximal (e.g., retinal) cue and the distal (e.g., object) variable to which it is related (see, e.g., Brunswik, 1955, 1956).

We take as given that the issue of generalizability is important (see also Pruitt and Kimmel (1977) who regard the problem of external validity as "one of the biggest problems of this [research] tradition" when critiquing laboratory experimental gaming). Alternatively, some renowned scholars are entirely unwilling to extrapolate from the lab to the field because "the same laws [do not] govern both the events in the laboratory and those of the cosmos" (Rapoport, 1970, p. 40). In addition, some experimental economists argue that there are instances where generalizability might not be of first rate importance. For example, Mook (1983) and Schram (2005) detail some arguments of cases where they believe that it is not important. Schram (2005) notes that the "classic" view of experimentation is one of theory-testing, arguing that "external validity is not an issue" in these instances. Writings of Plott (e.g., 1982) and Smith (e.g., 1982) reinforce this point (see Ortmann, 2003, for a recent review of Plott's arguments; for a skeptical dissent of some of these views, see Bardsley (2005b). Another example includes using the lab for methodological purposes—i.e., to inform field designs by abstracting from naturally-occurring confounds.

2) The particular context and process by which a decision is embedded, and

3) Self-selection of the individuals making the decisions.

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.[3] Yet, it should be clear that our arguments concerning the generalization of results from one setting to another are equally relevant to inference based on field data as well—whether the data are drawn from sumo wrestling matches or gathered from sportscard enthusiasts.[4]

We begin Section I by presenting a simple model that provides a convenient framework for understanding how, and why, these issues are important. The model stresses the fact that utility maximization is driven not only by wealth maximization, but potentially also by an individual's desire to "do the right thing," or make the "moral" choice. The weight that an individual places on "doing the right thing" is likely to increase when a subject is being watched, or the process by which choices and final allocations are reached is emphasized. Self-selection of individuals may be important if the weights given to moral versus wealth-maximizing considerations vary across individuals faced with the same problem, or people disagree as to

[3] Of course, 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 relative to situations that frequently arise in naturally occurring settings and typically the lab experience suppresses learning from peers. Likewise, many situations in the field are constructed endogenously. Moreover, experiments typically have short durations (minutes or hours), whereas many real life decisions (e.g. effort to exert in labor market settings) are made over much longer timeframes. We only briefly touch on these issues below. Other arguments exist as well (see, e.g., Kagel et al., (1979), Cross (1980), Starmer (1999a, 1999b), Bohm (2002), and Hertwig and Ortmann (2001)). The interested reader should also see the psychology literature, in particular the work of Gigerenzer, Cosmides, Tooby, and their colleagues who discuss the role of the environment in making inference from laboratory results.

[4] See Meyer (1995) on inference from natural and quasi-experimental data. Additionally, with the increasing popularity of micro-finance field experiments (see, e.g., Kremer, 2003; Ashraf et al., 2006), an important question in this literature relates to how the results scale-up and transport across countries, ethnicities, and time.

what the "correct" moral choice is in certain situations. Finally, as the stakes of the game rise, financial concerns are likely to become increasingly important relative to moral ones.

In section II, we investigate the degree to which lab environments correspond or diverge from the types of interactions encountered in the naturally-occurring world, paying particular attention to the implications for studies measuring deep structural parameters, such as pro-social preferences. We argue, based on decades of research in psychology and recent findings in experimental economics, that the lab is characterized by a special type of scrutiny that is unparalleled in the field, a particular type of self-selection of participants who are assigned roles and tasks exogenously, and small stakes. All of these forces potentially lead lab experiments to provide measures of social preferences that might not be broadly applicable to real-world situations, especially environments such as financial markets that are structured to provide anonymity, emphasize outcomes over processes, have substantial self-selection driven by talent and experience, and typically involve large stakes. Such estimates from the lab may, however, prove to be better suited for naturally-occurring settings in which there is a high degree of scrutiny of actions (e.g. employer-employee relationships, family interactions), or an emphasis on process (e.g. politics, judicial proceedings). Even in these settings, however, the nature of scrutiny, self-selection of participants, and low stakes *might* frustrate generalizability.

Section III considers the potential biases in games in which "doing the right thing" is not an issue. In situations where morality and wealth are not competing objectives, we expect that lab findings will be more readily generalizable, but a number of factors must be considered before one can make strong claims about behavior in the lab being indicative of behavior in naturally-occurring markets. First, in naturally-occurring environments there is systematic self-selection of the participants who are most skilled. Consistent with this conjecture, some

evidence suggests that individuals suffering from decision-making defects tend either to adapt, disappear from the market, never enter the market at all, or participate in markets that have developed securities to protect agents from falling prey of such defects. Second, in many cases experience reduces the prevalence of the sorts of decision-making defects that often arise in the lab, but experiments may not be of a sufficient duration or scope to allow this experience to arise.

Although most of the paper is focused on the shortcomings of laboratory experiments, Section IV highlights the many useful roles that such experiments can serve. In particular, there are large classes of experiments that are not subject to the criticisms we raise. Even for those experiments that are affected by our criticisms, it is quite likely that the qualitative findings of the lab are generalizable, even when the quantitative magnitudes are not. We draw attention to these and highlight the high intrinsic value and importance of laboratory experimentation in Section IV. Section V concludes with a summary.

I. Model

We begin by developing a model that makes precise our arguments regarding the potential factors that might influence individual decision-making. In building the model, we emphasize simplicity, sacrificing generality in the process. We do not claim originality in the ideas we are modeling. Indeed, there is a long history of economists—from Smith (1759) to Becker (1974)—emphasizing that decisions can have an impact on individual utility that goes beyond changes in wealth.[5] Rather, we view the model merely as a useful framework for organizing our discussion about the generalizability of lab experiments.

[5] Smith viewed decisions as a struggle between "passions" and an "impartial spectator," who was a "moral hector who, looking over the shoulder of the economic man, scrutinizes every move he makes" (Grampp, 1948, p. 317). For further discussion on this issue see the entertaining discussion of "Adam Smith, Behavioral Economist," in Ashraf et al. (2005). Becker (1974) introduced altruism as an explanation for why individuals give (or cooperate)

A utility-maximizing individual i is faced with a choice regarding a single action $a \in (0,1)$. The choice of action affects the agent's utility through two channels. The first effect is on the individual's wealth (denoted W_i). The higher the stakes or monetary value of the game, which we denote v, the greater the decision's impact on W_i.[6] The second effect is the non-pecuniary moral cost or benefit associated with action i, which we denote as M_i. If, for instance, an individual is altruistic, he will derive utility from charitable contributions. More generally, we have in mind that decisions which an individual views as immoral, anti-social, or at odds with his or her own identity (Akerlof and Kranton 2000; 2005) may impose important costs on the decision maker. For instance, in a dictator game, keeping a greater share for oneself increases an individual's wealth, but doing so may cause the agent moral disutility. This moral payoff might vary across people, religions, or societies.[7]

In practice, many factors influence the moral costs associated with an action across people, contexts, and societies, 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.

---

when it is money-maximizing not to give (or cooperate). Scores of economists in between have scribed of behavior being driven by factors beyond wealth (we point the interested reader to Stigler (1981) for a discussion).

Many other authors have addressed this problem in a much more thorough fashion than we attempt to do here. For example, in the recent charitable giving literature, an important alternative modeling approach to the pure altruism framework is to consider moral or group interested behavior (see, e.g., Laffont (1975), Sen (1977), Collard (1978), Margolis (1982), and Sugden (1984)). In Sugden, (1984), for instance, agents adhere to a "moral constraint," whereby they compare themselves to the least generous person when making their contributions. Moral concerns are also present in the work of Frey et al. (1996), Karni (1998), and Kotchen (2005), for example. Relatedly, in Bernheim's (1994) conformity model agents value status, and behavioral departures from the social norm impair status. Akerlof (1982) and Jones (1984) obtain similar conformity results by assuming deviations from social norms have direct utility consequences.

[6] In most experimental circumstances we have in mind, the wealth maximizing action is taken as the default prediction of economic theory, for example, playing the dominant strategy of zero contributions in a typical one-shot public goods game.

[7] Examples of the types of choices that might maximize M_i would be the Golden Rule of "Do unto others as you would be done by," Judaism's "What is hateful to you, do not to your fellow man. This is the law: all the rest is commentary," or Socrates' moral code "Do not do to others what would anger you if done to you by others." Formalizing the notion of morality in economic models is, as noted above, not novel to this study.

The greater is the negative impact of an action on others, the more negative the moral payoff M_i.[8] 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, extending one's middle finger vertically while the other fingers are in a fist is deemed extremely rude in American culture, but not in many other parts of the world. Although there is no financial externality borne by the recipient, there nonetheless are potential moral and social costs associated with such behavior. Likewise, the mere fact that an action is illegal (e.g., illicit drug use or smoking in restaurants), may impose an additional cost for partaking in such behavior. Mathematically, we denote these social norms against an action as n, with a greater value of n associated with a stronger norm against a behavior.[9]

Third, moral concerns will be greater when an individual's actions are more closely scrutinized (e.g., the act is being televised, is taking place in front of one's children, or is performed under the watchful eye of an experimenter), or the process by which a decision and final allocation are reached is emphasized (e.g., 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). While we view scrutiny as multi-dimensional, in our simple model we denote the effect of scrutiny as s, with higher levels of s associated with greater moral costs. We more thoroughly explore two dimensions of s that we find potentially important below.

[8] In Andreoni (1989; 1990), purely egoistic models assume that agents are not concerned with final outcomes.

[9] In the model social norms against an action do not have any direct financial impact, although one could imagine a generalization of the model involving repeated play in which such a financial impact would be present, e.g., if violators of social norms were punished by non-cooperation in the future. A good example of this is the ultimatum game. If there is a norm in favor of sharing, then the player who decides the initial allocation may be able to improve her payoff by being more generous because a high fraction of lopsided allocations will be rejected by the recipient, leaving both players with zero.

Focusing on the case in which utility is additively separable in the moral and wealth arguments, the utility function we have in mind when an individual *i* takes action *a* is

$$U_i(a,v,n,s) = M_i(a,v,n,s) + W_i(a,v) \qquad (1)$$

Solving this simple decision problem yields several predictions. First, when the wealth maximizing action has a moral cost associated with it, the agent will (weakly) deviate from that action towards one that imposes a lower moral cost. Second, the greater is the social norm against the wealth maximizing choice or the degree of scrutiny, the larger the deviation from that choice. In both cases, we envision the agent trading-off morality and wealth until an equilibrium is reached. Third, to the extent that individuals follow different moral codes (that is, $M_i \neq M_j$ for individuals *i* and *j*), they will generally make different choices when faced with the same decision problem. Fourth, in situations where there is no moral component (e.g., the choice between investing in a stock or bond index), the model reverts back to a standard wealth maximization problem.

Imposing a number of further restrictions on the utility function beyond those noted above yields crisper insights. Typically, we expect that as the stakes of the game rise, wealth concerns will increase in importance relative to issues of fairness, that is $|\partial M / \partial v| < |\partial W / \partial v|$.[10] Thus, while in a $5 dictator game a fair number of subjects might transfer $2.50 to their anonymous partner, in a $500 dictator game we would expect an increase in the level of money transferred, but not necessarily an equivalent proportion (i.e., more than $2.50 but less than $250 would be transferred). In such cases, a "split the money equally" norm proves too costly to

[10] Rabin's (1993) model of reciprocity makes the fairness component independent of the payoff scale. Thus, under either our model or Rabin's, for sufficiently high stakes W dominates M and there will be little concern (or no concern in Rabin's model) for others. Alternative models do exist, however; for example, Ledyard (1995) presents a model of voluntary contributions where altruism and selfishness are traded-off whereby an increase in the stakes has no influence on individual contributions.

implement, consistent with the empirical evidence from large stakes dictator games discussed below.

While it remains an open empirical issue, we would also expect that the cross-partial derivatives between *v*, *n*, and *s* in the morality function are potentially important. For instance, for any given social norm, the incremental influence on the morality component associated with behavior that violates that norm increases as scrutiny rises, i.e. $\partial^2 M / \partial n \partial s < 0$. Likewise, as stakes 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 that crime than they are of bank robbery. Similar intuition holds for stakes and scrutiny, as an individual likely faces a larger utility loss of robbing a bank if his capture is broadcast on *CNN* rather than merely recorded in his rap sheet.

It is also worth recognizing that the relevant social norms and the amount of scrutiny are not necessarily exogenously determined, but rather, subject to influence by those who will be affected by the choices an agent makes. For instance, panhandlers often emphasize physical deformities or carry placards claiming veteran's status in order to elicit greater sympathy from potential givers. And, churches use of "open" rather than "closed" collection baskets is consistent with recognition of the importance of norms and scrutiny, as potential contributors can not only see the total amount already gathered, but direct neighbors can witness each others' contributions (see, e.g., Soetevent, 2005).

II. Implications of the model for experiments designed to measure phenomena such as social preferences.

The utility function we describe has relevance for a wide variety of behavior. In this model, for instance, it is easy to understand 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, there is an offsetting non-pecuniary reward for tipping. This is true if

one is eating alone, but probably even more so when there is a higher degree of scrutiny (e.g., you are with clients, on a first date, or when another diner is closely observing your actions).

Our primary interest, however, is developing the implications of the model for the generalizability of lab experiments to naturally occurring contexts. To address that question requires an understanding of how the lab environment and real-world settings may differ along the key dimensions of the model, namely the stakes of the game, the social norms at work, the nature and degree of scrutiny, and the type of individuals engaged in the activity.[11] The greater the extent that the lab environment mirrors the naturally occurring setting that it is modeling, the more confident one can be that the lab results will be generalizable. If the lab diverges from the environment of interest on certain dimensions, the model provides a framework for predicting in what direction behavior in the lab will deviate from that outside the lab.

While the concerns we raise are relevant to a wide range of experimental results, their bite is likely to be greatest for those games in which there is the potential for a strong moral component to behavior and for those experiments that use the lab for quantitative purposes, such as to measure deep structural parameters (or "physical constants"). One of the most influential areas of research in experimental economics in recent years precisely fits this bill: games that provide insights into social preferences.[12] This broad class of games includes bargaining games (Roth, 1995), public goods games (Ledyard, 1995), and trust, or gift exchange, games (e.g., Camerer and Weigelt, 1988; Fehr et al., 1993; Berg et al., 1995; Charness, 1996; Fehr et al.,

[11] Closely related to the sorting of individuals into the experiment is the considerable importance of allowing endogenous task (or institution) selection. For instance, much like we would suspect that panhandlers would receive less in gifts if passersby could easily "sort" themselves to the other side of the road to avoid interaction, if subjects could costlessly sort out of performing certain tasks in the lab, such as playing the dictator game, the measured level of kindness would decrease in the lab. In this spirit, our model has support from recent laboratory evidence that allows agents to avoid playing the dictator game (Lazear et al., 2004). Botelho et al. (2005) provide another example along the lines of institutional choice.

[12] While various definitions exist, we define an agent with social preferences as one who has preferences that are measured over her own and others' material payoffs. Such preferences might arise due to, for example, altruism, reciprocity, fairness, and inequality-aversion.

1997; Fehr and Falk, 1999; Fehr and Gachter, 2000; Gächter and Falk, 2002; Hannan et al., 2004; Brown et al., 2004). Findings from gift exchange games, for example, have been interpreted as providing strong evidence that many agents behave in a reciprocal manner even when the behavior is costly and yields neither present nor future material rewards. Further, the social preference results have been widely applied outside the laboratory, based on the assumption that the experimental findings are equally descriptive of the world at large—see the cited papers above and discussions in Sobel (2002) and Camerer (2003).

In what follows, we examine the empirical evidence regarding each of the possible complications to extrapolating the experimental findings outside the lab highlighted in our model. We should stress that we are not denying that individuals have social preferences. Indeed, in our own model we assume that M can be influenced both by a concern for others as well as a concern for one's own appearance. Rather, we are interested in whether the depth and scope of such behaviors measured in the lab are shared widely among individuals in the field. In this regard, comparisons of interest include observing behavior of (i) identical individuals across the lab and the field, (ii) agents drawn from the same population in the lab and field, where the lab selection rules might be different from the manner in which markets select individuals, and (iii) individuals drawn from different populations. Of course, for years psychologists have questioned cross-situational consistency of behavior (see, e.g., Mischel, 1968; Ross and Nisbett, 1991), but we are pursuing something different: whether important factors of the experimental environment, and associated experimental procedures, systematically influence behavior, and how we can use insights on such factors to provide more accurate predictions for comparisons (i)-(iii) above.

Scrutiny

We view scrutiny as inherently multi-dimensional, yet to clarify the discussion we focus only on two dimensions, which we denote as "lab" effects and "non-anonymity" effects. We view the former as a necessary characteristic of any laboratory study and by their very nature these effects are immutable given the current practices in experimental economics. The latter effects can be varied usefully in the lab and naturally vary in the field.[13]

A. Lab Effects

In the typical lab experiment, subjects are keenly aware that their behavior is being monitored, recorded, and subsequently scrutinized. While we concur with the view that the laboratory is a "real" room just like the multiple other "real" rooms in the extra lab world, there is little doubt that the lab is a special type of room. Decades of research within psychology highlight the power of the role obligations of being an experimental subject, the power of the experimenter herself, and the experimental situation (Orne 1959a, 1959b, 1962, Rosenthal 1967, 1976, 1991, Milgram 1963, Haney et al. 1973).[14] Such effects are discussed in Schultz (1969, p. 221), who observed the unique relationship in the lab as having "some of the characteristics of a superior-subordinate one…Perhaps the only other such one-sided relationships are those 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

[13] Some readers might recognize the similarities between our definitions and what experimental psychologists denote "obtrusive" and "unobtrusive" measurement.

[14] For instance, Orne (1962) writes "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." Relatedly, in physics the Heisenberg Uncertainty Principle reminds us that the act of measurement and observation alters that which is being measured and observed. And, the Hawthorne Effect is a famous example of an instance where people changed their behavior merely because they were being watched. The fact that some agents alter their behavior when observed by others has also been termed the "interpersonal self-fulfilling prophecy" and the "Pygmalion Effect" (see Harrison and List, 2004).

> 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

Indeed, the strength of such factors is so compelling that in medical drug trials researchers often go above and beyond simple placebo and treatment groups by keeping the administrators themselves in the dark about which patients receive the treatment. Relatedly, in psychology, experimenters commonly deceive subjects to control for demand-induced effects. In economics, deceptive practices are frowned upon, perhaps leading to important demand effects.[15]

Clearly, the nature of scrutiny in the lab environment is unparalleled, or at least rarely encountered, in the field. For example, the role obligations inherent in being an experimental subject are completely absent in any important markets that we are aware. As a consequence, it might be the case that behavior in the lab is more influenced by moral concerns and less aligned with wealth maximization than behavior in many naturally-occurring settings.

One approach to explore this hypothesis is to move systematically from a laboratory environment to a naturally occurring environment by building a bridge between the lab and the field, as was done in List (2005a). That study 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 using experienced sportscard traders as subjects. The results mirrored the typical findings with other subject pools. Still in a lab environment, List added context to the gift exchange game by making the good that was exchanged a baseball card and

[15] We are not taking a side on the debate about whether deception should be practiced in experimental economics (see, e.g., Hertwig and Ortmann, 2001 for a good discussion on deception), rather we are merely noting that without the ability to deceive the experimenter interested in measuring parameters such as social preferences seemingly has a much more difficult chore than one who can practice deception. This has been recognized for decades among psychologists, who routinely deceive their clientele to avoid some of the effects that we discuss.

further changing the environment to mirror the naturally occurring marketplace.[16] Since quality is difficult to detect in this market for untrained consumers, if social preferences play a role in this case the card's quality and the price offer should be positively correlated. Once the buying agents had purchased each of the cards from the dealers, List had every card professionally graded. The results using baseball cards as the trading object mirrored those obtained in the standard gift exchange game.

List then extended the analysis by exploring behavior outside of the lab, sending confederates as buying agents to approach dealers on the floor of a sportscard show, instructing them to offer different prices in return for sportscards of varying quality. Strikingly, in a market-based naturally occurring environment where scrutiny is low, when the dealer 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. In other words, the social preferences so routinely observed in the lab are significantly attenuated in the field, even though the same agents revealed strong evidence of having social preferences in laboratory experiments.[17]

While this area is nascent in experimental economics and it is premature to make strong inference from one study, other field generated data yield similar conclusions. For example, making use of personnel data from a leading United Kingdom based fruit farm, Bandiera et al.

[16] Harrison and List (2004) denote this type of exercise as a "framed" field experiment, in that it is identical to a typical laboratory experiment but one that makes use of a subject pool that has experience with the task, information set, stakes, etc., of the decision problem.

[17] Only among local dealers, selling goods that could be readily graded, who could reasonably expect to have repeat interactions with the buyer did behavior consistent with social preferences emerge. Such behavior, however, is equally consistent with a purely self-interested attempt on the part of local dealers to develop a reputation for honesty to foster long-term relationships with the buyers. One does not typically ascribe social preferences to corporations that give frequent buyer discounts to repeat customers, and it likewise appears to be a stretch to attribute social preferences to the sportscard traders in List's experiment.

(2005) find that behavior is consistent with a model of social preferences when workers can be monitored, but when workers cannot be monitored, pro-social behaviors disappear. Being monitored proves to be the critical factor influencing behavior in this study. Further, Benz and Meier (2005) combine insights gained from a controlled laboratory experiment and naturally occurring data to compare how individuals behave in donation laboratory experiments and how the same individuals behave in the field. Consistent with our theory, they find some evidence of correlation across situations, but find that subjects who have never contributed in the past to the charities gave 75 percent of their endowment to the charity in the lab experiment. Similarly, those who never gave to the charities subsequent to the lab experiment gave more than 50 percent of their experimental endowment to the charities in the lab experiment.[18]

Gneezy et al. (2004) find that while behavior in a social dilemma game in the laboratory exhibits a considerable level of cooperative behavior, in a framed field experiment that closely resembles the laboratory game 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 simple model.[19]

We do not infer from such evidence that laboratory experimentation is seriously "flawed" in some manner. Contrary to this position, we view this as an opportunity to combine two observations—one from the lab, one from the field—to create a deeper theory of human behavior

[18] Several different interpretations exist for these data. For example, perhaps there is not a general cross-situational trait called "charitability," and hence we would not expect correlations between lab and field, or between lab and lab for that matter. Or, the subjects saw one situation as relevant to "charitability" and one as irrelevant. This raises an important issue of framing effects, which clearly can influence behavior and cause similar situations to be interpreted differently by subjects. Thus, the frame alone can importantly frustrate proper inference.

[19] Similar results are found in Chen (2005), who uses a large data set drawn from the Australian Workplace Industrial Relations Survey to explore reciprocity in the workplace and finds little evidence consistent with positive or negative reciprocity. Yet, there are interesting naturally-occurring data that are consistent with several models, including a model of reciprocity (see, e.g., Krueger and Mas, 2004).

than either could alone. Additional lab experiments can then be used to learn *qualitative* insights about the types of variables that can cause, attenuate, or exacerbate pro-social behaviors. With an understanding of the proper interpretation of such lab results through the lens of a theoretical model, the researcher can then return to the field for further testing.

B. Non-Anonymity Effects

Closely tied to the lab effects discussed above is what we denote as non-anonymity effects, or how changes in the degree of confidentiality influence behavior. In the typical lab experiment subjects are purely anonymous in relation to other subjects, but the identity of the subject can readily be linked to individual choices by the experimenter. In this section, we focus attention on anonymity between the experimenter and the subjects, but now and again mention anonymity among the subjects.[20] In the field there are many market transactions that span the anonymity spectrum. Some market transactions, for instance, are completed with substantial anonymity and little concern about future analysis of one's behavior. The rise of the internet has further depersonalized some market transactions. Other naturally occurring settings, such as employee-employer relationships or choices made in front of one's children are under somewhat greater scrutiny. For example, the actions a firm takes towards an employee may be carefully watched by others who expect to be in a similar situation in the future. There are even situations where less anonymity is provided, such as when one knows that his actions are being televised.[21]

If the lack of anonymity between the experimenter and subject contributes to pro-social behaviors, then under our model taking steps to reduce the extent to which subjects are knowingly observed should reduce the amount of such behavior. One way of accomplishing this

[20] Anonymity can usefully be broken down to three configurations: relative to other experimental subjects, relative to the experimenter, and anonymity that rules out pecuniary gains through reputation formation.

[21] Levitt (2005) finds no evidence of discrimination towards Blacks or women by participants on the televised game show "The Weakest Link."

is to create an environment that provides anonymity for the subjects in the laboratory. Hoffman et al. (1994; 1996), for instance, used a "double-blind" approach whereby the experimental monitor could not infer individual subjects' actions in a simple dictator game. List et al. (2004) explored changes in anonymity between the subject/experimenter as well as among subjects using a "randomized response" technique to ensure that subjects would preserve their anonymity when asked to contribute money to a real public good. Both studies, and other related research (e.g., Eckel and Grossman, 1996; Masclet et al., 2003; Rege and Telle, 2004), find that generous actions and the degree of anonymity are crucially linked.

Hoffman et al. (1994), for example, find that 22 of 48 dictators (46%) 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%) give at least $3. Hoffman et al. (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.[22] Additionally, List et al. (2004) found that both

[22] Agreeing with this general point, in their template to guide inference from their experimental data, Andreoni and Miller (2002, p. 738) recognize that "social variables like the level of anonymity, the sex of one's opponent, or the framing of the decision….are known to affect the outcome." We should stress that some experimentalists do not consider this result a stylized fact, however. As Roth (1995, pp., 299-301) points out, there are a number of reasons to treat the conclusions from Hoffman et al. (1994) with caution. And, even though many studies find such effects, not all studies find similar data patterns. For instance, Bolton et al. (1998) and Laury et al. (1995) collect data that casts doubt on the Hoffman et al. (1994) results. One potential explanation for these results is that some subjects in such environments might not fully embrace the anonymity promise. We informally polled subjects who were part of a recent anonymous experiment at the University of Maryland on whether they believed it was truly anonymous and a non-trivial portion of subjects reported "that the experimenter could likely determine identities." Relatedly, our model also provides an explanation for the experimental results that show declining positive survey responses in favor of a proposed public project as the likelihood that the survey is executed 'for real' increases (see, e.g., Cummings and Taylor, 1998).

the degree of anonymity between the experimenter/subject as well as among subjects was critical: as decisions became less anonymous, a greater number of subjects opted to give to the public good in a one-shot decision.

In a related literature, other dimensions of anonymity are found to affect giving in a way that is consonant with our model (Allen, 1965; Cason and Mui, 1997; Bohnet and Frey, 1999a, 1999b; Frolich et al., 2001; Burnham, 2003; Haley and Fessler, 2005). 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 with which the dictator makes his allocation. This simple manipulation -- meant to signal that the subjects' actions were being observed -- increased the proportion of nonzero givers from 55% in the control treatment to 88% in the eyespot treatment. Likewise, Allen (1965) reports that increases in privacy reduce conformity. Research on hand washing behavior even suggests that individuals are more likely to conform with the social norm of washing when they are being observed (Harris and Munger, 1989).

Some of the experimental approaches described above, however, are subject to criticism. 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). Thus, differences in behavior observed across anonymous and non-anonymous settings might understate the true spectator effect in our model. Nevertheless, our broader point is that laboratory experiments are a good tool to explore the *qualitative* effects of anonymity. In this case, overall the received results on anonymity are consistent with our model; though we certainly accept the view that this is a ripe area for further study.

Context matters and it is not completely controlled by the experimenter.

Human behavior is heavily influenced by context. The actions people take are influenced by a dazzlingly complex set of relational situations, social norms, frames, past experiences, and the lessons gleaned from those experiences. Consequently, the experimental investigator might not have complete control over the full context within which the subject makes decisions.

Experimentalists are fully aware that context in their instructions, inducing "role" playing, framing, and the like can influence subject behavior (see, e.g., Roth 1995, Hertwig and Ortmann, 2001, Bohnet and Cooter, 2005). In a wide range of experimental settings, it has been shown that subtle manipulations can have drastic effects on play. 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); more generally, using terms like "opponents" versus "partners" influence play in a myriad of games (see Burnham et al., 2000, for an example), asking people to "contribute" or to "allocate" funds in a linear public goods game matters, and to "punish" or to "assign" points" to other agents in a punishment game can considerably influence play (see, e.g., Gintis, 2001).[23]

One subtle aspect of the context that the experimenter can control that has not been fully explored is the fact that restrictions on the available choice set affect behavior. Consider a dictator game, where one agent unilaterally determines how a prize should be split between himself and another player, and that decision determines the allocation. In the typical dictator game, the subject is given, say, $10 and asked what portion he would like to share with the other player. 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 a $10 prize and is told

[23] For instance, in the trust game of Burnham et al. (2000), instead of the term "counterpart" for referring to the person that an individual is matched, the authors use the word "partner" in one treatment, "opponent" in the other. Such differences produced a significant change in measures of trust and trustworthiness: trustworthiness with "partner" is over twice that for "opponent," and this reinforces trust. Such results have led some experimenters to consider using context to match the natural environment. This issue is discussed more fully below.

that he can give any portion of his prize away to the second player, or confiscate up to an additional $10 from the other player, subjects give little to the other player (see Bardsley, 2005a and List 2005b for studies in this spirit). Importantly, in real-world contexts, there is typically the option of both giving and receiving, which may help explain in part why, contrary to the lab environment, people rarely receive anonymous envelopes stuffed with cash.

The literature is replete with examples in which the experimenter intentionally varies context to analyze the sensitivity of respondent behavior.[24] Contextual factors that are beyond the control of the experimenter appear to have equally profound impacts on actions. Perhaps the most powerful evidence of this kind comes from Henrich et al. (2001; 2004). In the latter study the group of scholars conducted ultimatum, dictator, and public goods games in fifteen different small-scale communities in developing countries. They found enormous variation in behavior across communities, differences they are able to relate to interactional patterns of everyday life and the social norms operating in these various communities. For instance, as Henrich et al. (2004, 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 that are characterized by a much different ethnographic description, such as the Hadza of Tanzania, low offers and high rejection rates are observed in ultimatum games. As Henrich et al. note

[24] An interesting research program in the lab would be to reverse the typical economics experiment: along with treatments holding the environment constant and changing economic variables, one could hold the economic variables constant and change the environment. This would lead to comparative static comparisons of great interest (see, e.g., Cookson, 2000).

(2004, p. 33), these "contrasting behaviors seem to reflect their differing patterns of life, not any underlying logic of hunter-gatherer life ways."

Critically, in all of the experiments Heinrich et al. (2004) conducted, the context that the experimenter can control—the payoffs, the description of the way the game is played, etc.—was almost identical (this was apparently not the case in the Heinrich (2001) study, see Ortmann (2005) for a discussion). It was the part of the context that actors themselves bring to the game and experimenters cannot control (e.g. past experiences and internalized social norms) that proved centrally important in the outcome of play.

While there is good reason to be concerned about the possibility that the context that is beyond the control of the experimenter will affect any preferred interpretation, the results of Henrich et al. (2004) do not necessarily provide guidance as to the direction of the distortion that will be induced since the results suggest that it is context specific. There is, however, one particular example of this phenomenon that potentially influences measurement of pro-social behavior. The unique aspect of lab experiments for measuring pro-social behavior is that the experimenter can control for reputational concerns by using one-shot experimental designs. If effective, then this approach is quite useful because it can parse the "social preference reciprocity" explanation from the "repeated game reciprocity" explanation (i.e., I give today to maintain a good reputation so people will cooperate with me in the future). In naturally occurring data it is quite difficult to separate these competing explanations.[25]

[25] One might argue that the lab actually underestimates social preferences in the field because the lab is based on interactions of complete strangers, anonymity between subjects, absence of any social relations between subjects, and restricted communication channels between subjects. We concur that upon relaxing these aspects of the environment that agents might have stronger social preferences (such insights are consistent with List's (2005a) data). Yet, one must take care to properly account for the important underlying factors at work in such an analysis. In natural settings it is clear that agents might exhibit behaviors that suggest they have social preferences to enhance their competitive position, but if such behaviors are aimed at reputation building they should not be confused with social preferences. In this light, it is difficult to consider a Las Vegas Pit Boss giving a "free steak card" to a high roller she conversed with for hours at the craps table to be considered evidence of social preferences at work.

As various authors have pointed out (e.g., Hoffman et al., 1996; Ortmann and Hertwig, 2000; Harrison and Rutstrom, 2001), however, activities in the spirit of public good provision, dictator, ultimatum, trust, gift exchange games, and other social dilemmas, are typically not one-time encounters in the outside world, but rather repeated games. Ortmann and Hertwig (2000) go further and speculate that "there is a good chance that many of the experimental results for the classes of games discussed here [social dilemmas] are artifacts of experimental design."[26] Indeed, from the Henrich et al. (2001; 2004) studies we learn that to the extent that participants perceive these games in the laboratory as some form of social dilemma, they are likely to retrieve experiences and strategies that, unbeknownst to the experimenter, change the nature of the games (see also Binmore 1994). Effectively, personal experiences may cause the subjects to play these one-shot games as if they are repeated games, and the experimenter may have little or no ability to moderate this phenomenon. In repeated games, reputation plays an important role, leading wholly self-interested agents to engage in cooperative behavior. If the experimenter mistakenly assumes that the actor is treating the game as one-shot, reputation-building behavior can be misconstrued as social preferences.

While as researchers we might hope and expect that experimental subjects should be able to make clear strategic adjustments across one-shot and repeated contexts, the empirical evidence is mixed and seemingly depends on the game being played. 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 (2003) report that five studies find more cooperation among the randomly re-matched, six find more cooperation among the repeatedly paired, and

[26] As empirical evidence, Ortmann and Hertwig (2000) use the fact that these types of games (one-shot) are situations wherein the experimental subjects have to overcome powerful base rate experiences (from repeated interactions) that serve them well in everyday life.

four studies fail to find a difference between the two treatments. Relatedly, using an indefinitely repeated, 2-player Prisoner's Dilemma game, Duffy and Ochs (2005) compare fixed and random pairs matching protocol and report that "while there is no difference in the way inexperienced subjects first play these games, experience under the fixed protocol drives each group of subjects to a much higher level of cooperative play than is observed under any of the random pairings treatments."

Yet, Fehr and Fischbacher (2003) find that responders react strongly to the possibility of acquiring a reputation. Similarly, Gachter and Falk (2002) and Brown et al. (2004) find that students can perceive such differences in gift exchange games. Other studies, such as Seinen and Schram (2006) and Engelmann and Fischbacher (2002), also find increases in helping behavior if subjects can build a reputation versus when they cannot, suggesting that student subjects understand the strategic incentives inherent in the two environments.

This line of reasoning also might lend an explanation to certain data patterns observed in laboratory experiments. One example is the standard laboratory puzzle that contributions in linear public goods games begin high and plunge in the latter rounds of an experiment (Andreoni, 1995; Houser and Kurzban, 2002). One explanation proposed in the literature is "frustrated attempts at kindness" (Andreoni, 1995), another is confusion (Andreoni, 1995; Houser and Kurzban, 2002). Viewed from the lens of our model, this result can be explained by certain subjects bringing context learned in everyday life into the lab, not being rewarded for that behavior, and subsequently reducing contributions. Under this model, therefore, subjects are "cursed" by their previous everyday experiences, and if one were to bring back the same subjects and clearly summarize that they are playing the same linear public goods game they played earlier, our model predicts that contributions would start, and remain, relatively low throughout

all periods of play (but not necessarily be zero for everyone due to other factors in our model). We are unaware of any such experimental evidence, though the "restart" experiments in the public goods literature share similarities.

A final contextual concern is whether the games played in laboratory experiments are reflective of the real life situations they attempt to model and to which inference is ultimately made by the experimenter. And, if not, can the experimenter simply use context to "control" the perceived situation. In terms of the former question, from the experimental economics literature we know of little concrete, direct evidence beyond the insights gained from Henrich's (2001; 2004) work. But, a recent study does shed some light on the matter. Ferraro and Vossler (2006) make use of the archetypal linear public goods game to examine questions related to why agents contribute to a public good when it is a dominant strategy to free ride completely and give nothing. Most importantly for our purposes, upon examining their data they ask "Is it really possible that so many subjects are oblivious to the dilemma experimentalists are attempting to induce in the laboratory?" (Ferraro and Vossler, 2006, p. 29). To explore that question they use post-experimental questions and find that many players "view the game as an Assurance Game, rather than a linear Public Goods Game" (p. 32).

Clearly it is possible that subjects are playing a different game than believed by experimentalists (which in some cases is good outcome, provided the researcher understands such perceptions). One solution that has been discussed in the literature is to use context in the lab games themselves. For instance, Loewenstein (1999) argues that "the goal of external validity is served by creating a context that is similar to the one in which economic agents will actually operate, not by engaging in futile attempts to eliminate context (see also the arguments in Harrison and List, 2004). Yet, this might not always solve the problem either. In a recent

study, Eckel and Wilson (2004), had subjects play the standard trust game but in the instructions they framed the game as a "loan." After the experiment, Eckel and Wilson asked the subjects "what does this situation remind you of?" Interestingly, only 25/110 (36/108) of first (second) mover subjects mentioned "loan," and of those 20 (22) mentioned "loan to/from a friend" rather than "loan to/from a stranger," which is what it was. In this case, if subjects pay back loans to friends with greater frequency than they honor loans to strangers, then any inference about the scope of social preferences is likely to be exaggerated.

In sum, the crux of our arguments in this subsection highlight the fact that context and relational situations heavily influence behavior. This fact presents a particularly vexing situation because the activity of *ceteris paribus* testing in and of itself might alter the phenomenon of interest. Unlike natural phenomena such as bumble bees, bacterial genes, and water, which are identifiable as such inside and outside of the laboratory, the phenomena of interest to many experimentalists—measuring social preferences for example—might not retain their identities without their relation to field referents. By isolating social phenomena to study such factors we might alter the very phenomena we purport to study, yielding the subjects' interpretation of the situation vastly different than the experimenters.

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 roughly consonant with this view. For example, in dictator games an increase in stakes generally leads to a less than proportionate increase in monies transferred (see, e.g., List and Cherry, 2005; Carpenter et al., 2005a). In Carpenter et al. (2005a), an increase in stakes from $10 to $100 caused the median offer to drop from 40% to 20% of the endowment. Similar

insights are gained from second mover play in ultimatum games: the acceptance rate is increasing in the amount offered, conditional on the share offered (i.e., a $1 offer in a $5 game is rejected more often than a $100 offer in a $500 game). For example, Slonim and Roth (1998) find that in each range of offers below 50%, the acceptance rate goes up as the level of stakes increase (from 60 to 1500 Slovak Koruna, respectively).[27]

While there certainly are many transactions over low stakes in the real-world, the stakes for which experiments are played are typically quite small relative to some important naturally-occurring environments that the experiments are seeking to reflect, such as labor negotiations, contract disputes, general worker/employee relations, and the like.[28] Consequently, we argue that if the analyst does not account properly for the differences in stakes across settings, the lab will likely yield inaccurate measures of pro-social preferences relative to real-world situations. The magnitude of such mis-measurement is a rich area for future research, and it would be interesting to compare such effects to the size of the effects of the factors discussed above.

Self-selection into the experimental subject pool

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."

[27] In other types of games, the existing evidence is mixed. For example, in terms of trust and gift exchange games (sequential prisoner's dilemma games), Fehr et al. (2002) report that fairness concerns play an important role for both low and high stakes games whereas Parco et al. (2002) find that raising financial incentives causes a breakdown in mutual trust in centipede games. These types of results have led commentators to conclude that "In the experimental literature there is no consensus on the relationship between co-operation and stakes in social dilemmas" (Clark and Sefton, 2001, p. 54).

[28] Although economic experiments, which use monetary rewards that are contingent on behavior, are likely to suffer less from this bias than psychology experiments that typically do not use pecuniary incentive schemes. A related issue is whether subjects behave differently over "house" money and own-(or earned) money. In simple dictator games, empirical evidence suggests that dictators give less when they are allocating earned money or have earned their position in the game (see, e.g., Hoffman et al., 1994; Rutstrom and Williams, 2000; List and Cherry, 2005).

Indeed, Rosenthal and Rosnow (1969) conclude that 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).[29] In contrast, market participants are likely to be a highly selected sample of individuals whose traits allow them to excel in the marketplace.

To the extent that the students participating in experiments have different moral codes, laboratory findings may not provide accurate guidance for behavior in naturally-occurring situations for which the experimenter is attempting to make inference. A priori, from strands within the social psychology literature cited above, one might suspect that the nature of the selection into the lab would yield exaggerated pro-social behavior relative to real-world markets. On the other hand, lab participants may have less social preferences than those who select into particular naturally-occurring environments, such as the clergy, or public defenders.

One approach to investigating biases from selection is to examine whether professionals, or other representative agents, and students behave similarly in laboratory experiments.[30] In order for these laboratory findings to be meaningful, however, it must be the case that the scope of lab and non-anonymity effects (e.g., the increase in emphasis on the moral action) are similar across experimental samples. Fehr and List (2004) examine experimentally how Chief

[29] For example, 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. 136-137). We should highlight, however, that 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 (see, e.g., Kagel et al, 1979). There is some evidence that volunteer subjects in an economics experiment have more interest in the subject than non-volunteers (Kagel et al, 1979), consistent with the social psychology literature. Their study, however, also finds that other important variables are not different across volunteers and non-volunteers. This is a clear example where much more research is needed in experimental economics.

[30] Harrison and List (2004) denote this type of exercise as an "artefactual" field experiment, in that it is identical to a typical laboratory experiment but one which makes use of a non-standard subject pool.

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 more trusting and exhibit more trustworthiness than students.[31] 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.

Even if strong insights could be gained about subject pool differences from these experiments, a related issue concerns the possibility that only certain types of participants—students or professionals—are willing to take part in the lab experiment. In this case, if the selection rules differ across subject pools, then valid inference might be frustrated. For example, as noted earlier, volunteers (both students and 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.

There exists some limited, but suggestive, data from field and lab experiments that provide some support for this argument about selection into laboratory gift exchange experiments. List (2005a) approached a number of sportscard sellers about participating in a laboratory experiment. Some sellers declined his invitation, but later and unbeknownst to them, participated as sellers in a gift exchange 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

[31] Consistent with Fehr and List (2004), in dictator games, Carpenter et al. (2005b) find that employees at a Kansas City distribution center are more generous than students. Yet, Harbaugh et al. (2003a), 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. Alternatively, in dictator games, the youngest children tend to make considerably smaller transfers than do older children and adults in Harbaugh et al. (2003b).

due to small sample sizes and therefore not statistically significant at conventional levels. Although not a gift exchange game, 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) in a series of dictator games. 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.

III. Implications of our model for experiments in which the moral and wealth-maximizing actions are not competing objectives

In Section II we emphasized how the lack of congruence between moral and wealth-maximizing actions can lead laboratory experiments to yield quantitative insights that may not be readily extrapolated to the outside world. In the large class of games where there is no inherent conflict between the moral choice and the wealth-maximizing choice—e.g., certain experiments exploring general economic theory, Bayesian updating, risk and uncertainty, psychological phenomena such as loss aversion, hyperbolic discounting, impersonal auctions, market experiments where the demand and cost functions are unknown—many of the above concerns, such as scrutiny effects, become inconsequential.

Potential distortions that remain within the model are the impact of stakes on actions and non-random selection of participants into the experiment. In addition, although not formally part of our static model, we raise four extensions to our model that highlight additional issues: (1) the amount of experience one has with a task is likely to be much greater in naturally occurring environments than in the lab, (2) real-world institutions may be endogenously designed by sophisticated agents to exploit behavioral defects of naive players, (3) experiments typically are completed over short durations (minutes or hours), whereas many real life decisions (e.g. effort to exert on the job) are made over much longer time-frames, and (4) there may be group

differences (e.g., male vs female) in how the lab affects behavior and in self-selection. These six considerations are addressed in turn, with much of the discussion centered on individual bidding behavior in auctions since this area represents a vast and expanding literature. All of these forces, we argue, can lead to more anomalous behavior in the lab than in naturally-occurring environments. Yet, there are certain cases where the opposite result may hold.

Stakes and cognitive costs

The simple model we presented earlier abstracts from any cognitive or effort costs associated with determining the wealth maximizing action. To the extent that such costs are present, theory suggests that the likelihood of the wealth maximizing choice being made is an increasing function of the gap between the stakes of the game and the costs of effort. Smith and Walker (1993) find evidence of the interaction between stakes and cognitive costs in a comprehensive review of thirty-one published laboratory experiments. Camerer and Hogarth (1999) extend Smith and Walker's survey by examining 74 experimental papers, and find evidence in favor of the cognitive-effort theory in that the variance of play decreases with stakes; they note that "higher levels of incentives have the largest effects in judgment and decision tasks." List and Lucking-Reiley (2000) explore these issues with auction experiments in the field, finding results that indicate that stakes matter—high-priced auctioned goods produced more of the theoretically predicted strategic behavior than did lower-priced goods.

With this empirical evidence in hand, the next step is to discuss how the lab might differ from the field across these important dimensions. As noted earlier, the stakes in lab experiments are typically much smaller than in many real-world settings of interest. On the other hand, the cognitive costs associated with forming an optimal strategy might be higher or lower in the lab than in the field. Recall in the typical lab experiment an abstract task with little context is the

rule rather than the exception. In the field, referents, cues, and experience-related tasks often ease the burden of selecting an action. An example should suffice. In lab auction games, many times students are presented with complex tasks. Consider students facing the task of computing an optimal bid in a "simple" first-price auction, assuming no resale opportunities and *n* symmetric, risk-averse players. This calculation leads to the following differential equation defining an optimal bidding function $b_N(x)$ (see Lange et al., 2005, for details)

$$\left(\frac{1}{\rho(b_N(x))} F(x)^{n-1} \right)' = \frac{\left(F(x)^{n-1} \right)'}{K_N(x)}. \qquad (2)$$

Without even defining the variables in (2), it is evident that this is a hard problem, perhaps not worth solving carefully for an expected increase in payoff of nickels (see Harrison (1989) on this point). While equally daunting tasks certainly arise in the field as well, such players typically have more at stake, more resources to draw from, and more than a few minutes to compute an optimal bid. Of course, multiple rounds of the same game can help to refine play, but such paths of play might not be representative of "real world" play where experts discuss strategies and rules of thumb and heuristics are developed over years of experience.

Non-random selection of participants

If participants in laboratory studies differ in systematic ways from the actors engaged in the targeted real-world settings, lab results can be misleading. This concern is particularly acute in settings related to environments where (1) the real-world stakes are both high, and (2) it is frequently groups of individuals, or firms, rather than individuals making such choices. In these cases, actual decision makers are likely to be a highly select group of individuals with expertise and experience. Lab participants lacking such experience and expertise might not provide accurate guidance as to how real decision-making occurs. The selection issue is perhaps best

illustrated by Crack (2000, pp. 201-3), who notes that some Wall Street firms use the Monty Hall problem as a screening device for evaluating job candidates. In such cases, those less versed in applying Bayes' law, or perhaps those who are cognitively-challenged more generally, might never have a chance in these firms but would be readily accepted into the laboratory subject pool. Similar insights are gained from a 1991 deposition of John Mack, the CEO of Morgan Stanley, who openly recognizes that traders with loss averse preferences are not welcome employees at his firm: *"One of the critical criteria I use in judging my traders is their ability to take a loss. If they can't take a loss, they can't trade."*

One piece of laboratory evidence that suggests selection effects of this sort may be important comes from the case of the Winner's Curse in common value auctions (Bazerman and Samuelson 1983, Thaler, 1992). The Winner's Curse (WC) represents a disequilibrium behavior in which bidders systematically overbid and thus earn a negative payoff upon winning. The phenomenon arises because bidders fail to take into account the fact that *if they win* then they may have over-estimated the value of the object, and correct their bids for that fact. Cox et al. (2001), who use a laboratory experiment with free entry and exit, show that only a subset of potential bidders elected to bid in auctions with a WC element possible. The worst offenders chose to stay out, largely mitigating the WC problem; a result consonant with our arguments concerning social preference measurement when tasks are endogenous.[32]

[32] Selection effects have been shown elsewhere as well. In a study measuring individual risk propensities, Harrison et al. (2005) find that the use of show-up fees to increase experimental participation generates a sample of more risk averse participants. A recent example of the importance of selection effects is highlighted in Casari et al. (2005). Comparing bidding of various types of agents in common value auctions, the authors (p. 1) report that "there are strong selection effects in bid estimates for both experienced and inexperienced subjects…..ignoring these selection effects is most misleading for inexperienced bidders. A related point on selection that pervades not only this literature but the experimental literature more broadly is that the lab might yield a biased treatment effect if there is a treatment-demographic interaction and the lab misses the important demographic. Anderssen et al. (2005) illustrate this point by considering preferences over risk and time. They find that there (p. 1) "are also differences in treatment effects measured in the lab and the field that can be traced to interactions between treatment and demographic effects. These can only be detected and controlled for properly in the field data." Similar selection effects are found

The role of experience

In markets, firms that earn high returns tend to grow and prosper, whereas those that systematically make mistakes shrink or disappear. Such competitive forces, in most cases, work towards eliminating decision-making defects.[33] Of course, this intuition is not critical of lab experiments *per se*, rather their typical implementation. For example, it is not difficult to gather a pool of sophisticated and well-practiced agents as experimental participants. Yet, it is important to understand that subjects enter the lab with rules of behavior learned in the outside world. Depending on whether the specific experimental design in the lab rewards or punishes these rules of thumb, radically divergent results can be obtained.[34]

Harrison and List (2005), for instance, examine the behavior of professional bidders in their naturally occurring environments. They find that in their real-world bidding, they do not fall prey to the WC. When the expert bidders are placed in unfamiliar roles, however, they often fall prey to the WC, just as happens in the lab. This result, coupled with those of Cox et al.

---

in Rutström (1998), who explored the role of recruitment fees in securing experimental participants and found some evidence of group differences (male versus female, white versus non-white) in bidding in auctions in some of the treatments and no differences in other treatments.

[33] There are other cases when the anomalies observed in the lab might not be expected to disappear in the field. One prominent example is Bayesian decisionmaking, where there is evidence consistent with the hypothesis that experience does not considerably induce individuals to become "better Bayesians" (see, e.g., Camerer, 1987; 1992; Alevy et al., 2006). In Alevy et al. (2006), market professionals from the Chicago Board of Trade are not more Bayesian than student subjects in simple information cascade laboratory experiments. Likewise, in other cases such as in everyday consumption or worker choices, feedback mechanisms might be too weak, or non-existent, to signal to the individual that his behavior is irrational. An individual missing out on a small level of additional utility due to irrational choices might not realize such losses for a period of time because important information is never made apparent to him.

[34] This result is one of the important insights in Burns (1985), who reports that students do better than businessmen (wool traders) in progressive oral auctions. This result occurs because the wool traders used their related experience in the actual wool markets with which they operate, even in cases where such heuristics were inappropriate. Burns (1985, p. 150) summarizes this insight as follows: "In general, where the rules of the experimental market conflicted with the market rules that they were used to, the buyers instinctively chose the latter." More generally, students tend to outperform experts in more cognitively challenging and abstract tasks. This result is illustrated in Cooper et al. (1999), who argue that students appear to outperform experts because they are more likely to have the "test-taking" skills such as accurately computing cognitively challenging calculations, following abstract chains of logic, and the like, that is required in their laboratory game.

(2001), is consonant with the notion that agents with decisionmaking defects tend either to adapt, disappear from the market, or never enter the market at all.[35]

Similarly, there is lab and field experimental evidence that is consistent with the hypothesis that the endowment effect (and the WTA/WTP disparity) shrinks with market experience (e.g., Knez et al., 1985; Coursey et al., 1987; Brookshire and Coursey, 1987; Myagkov and Plott, 1997; List, 2003; 2004). For example, examining individual trading and valuation decisions from subjects drawn from quite diverse subject pools—from the typical student subject to agents in a collector pin market—List (2003) reports evidence that is consistent with the view that individual behavior converges to the neoclassical prediction as trading experience increases. His evidence includes experimental data collected from the same individuals over time, thus permitting learning and selection to be separately identified. While selection is found to be an important issue, many subjects who gained market experience over time "learned" to behave in a manner consistent with neoclassical theory.

Studies using non-experimental data report similar insights. For example, using a unique housing market data set drawn from downtown Boston, Genesove and Mayer (2001) find that seller behavior across investors and owner-occupants is different: owner-occupants exhibit about *twice* the degree of loss aversion compared to investors. Perhaps providing a cleaner result

[35] Another means for the WC to be attenuated in naturally-occurring markets is for the market to develop securities to protect bidders from such defects. One example of this can be found in Dyer and Kagel (1996), who review how executives in the commercial construction industry appear to avoid the winner's curse in the field (p. 1464): "Two broad conclusions are reached. One is that the executives have learned a set of situation-specific rules of thumb which help them to avoid the winner's curse in the field, but which could not be applied in the laboratory markets. The second is that the bidding environment created in the laboratory and the theory underlying it are not fully representative of the field environment. Rather, the latter has developed escape mechanisms for avoiding the winner's curse that are mutually beneficial to both buyers and sellers and which have not been incorporated into the standard one-shot auction theory literature." This passage highlights that creating an artificial environment and executing a standard laboratory common value auction using the executive experts as subjects will likely lead to students outperforming the experts because the situation-specific rules of thumb that the experts have developed will not help them avoid the WC. Furthermore, even if the WC was rampant in this environment, it would not influence market outcomes because the market has endogenously created "escape mechanisms" to allow bidders to avoid repercussions of the WC.

for the purposes herein, Shapira and Venezia (2005) analyze investment patterns of a large number of Israeli investors and report that professionals exhibit considerably less loss averting behavior compared to independent investors. Finally, studying trade histories for professional floor traders, Locke and Mann (2005) present evidence that suggests certain classes of "successful" traders exhibit less loss averting behavior than their less-successful rivals.

Such results have led even noted skeptics of neoclassical theory to conclude that market experience can eliminate disparities between WTA and WTP (see, e.g., Loomes et al, 2005). One interesting data pattern found in the field experiments is that behavior among the least experienced agents is similar to individual behavior observed in the lab with student subjects. Thus, the lab is seemingly able to provide an accurate indication of data patterns generated by the least experienced agents in this case.[36]

Other similar data patterns exist across the lab and the field for other types of games as well. For instance, even though money illusion is a powerful force for naïve experimental subjects, almost all people eventually learn to play the equilibrium (see, e.g., Fehr and Tyran, 2001). Further, from a field experiment List and Lucking-Reiley (2000) learn that more experienced auction bidders exhibit a greater tendency to behave according to strategic theories than did lesser experienced bidders. In sum, experience influences behavior, leading us to conclude that if real market experience is important for the case at hand, then student behavior might not provide an adequate ideal for extrapolation purposes.

[36] There are two standard counter-arguments to the points we make above. First, that allowing subjects to engage in multiple rounds of the same game or activity provides them with the chance to adapt to the environment. While for some tasks laboratory learning might be adequate, the data in List (2003) suggest that useful cognitive capital builds up slowly, over days or years, rather than in the short-run of an experiment (see also Camerer and Hogarth (1999) on this point). Second, that the use of experts rather than students in lab experiments will solve this problem. As we noted earlier, the design and nature of the experimental environment strongly influence the relative performance of experts and students.

Again, we view such results as highlighting an attractive setting when laboratory experimentation can be quite useful. For example, lab experiments can be used to explore the types of variables that can influence the magnitude of the endowment effect. As our model suggests, lab experiments are not well suited to inform us of how large the endowment effect is in Boston real estate markets, or how it influences equilibrium outcomes in any naturally-occurring settings for that matter. Rather, their power is in creating an artificial environment to determine what can happen, not what usually does happen in natural environments. In this way, *qualitative* insights, such as factors that strengthen or alleviate the endowment effect, can be importantly measured in a controlled setting and subsequently explored in the field.

The endogenous design of institutions to exploit behavioral defects

In naturally occurring environments, institutions arise endogenously. 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 (Glaeser, 2004). For instance, Della Vigna and Malmendier (2005) provide evidence that health clubs structure fees in a manner that capitalizes the overly optimistic expectations of potential customers. Levitt (2004) similarly shows that bookmakers set lines that take advantage of the inherent biases of bettors.

Certain factors suggest that we might expect that less anomalous behavior will be observed in the field than in the lab due to endogenously generated institutions. For instance, we learn from evolutionary biology that selection pressures can work against organisms that over-extract from their hosts. In this case, under a very simple model firms which implement such policies will 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.[37] Relatedly, in important cases third-parties emerge in naturally occurring markets, when individuals face situations that are difficult or costly for novices to solve (e.g. lawyers, mortgage brokers, money managers). In this spirit, individual experiments in the traditional sense might not be well suited for analyzing markets characterized by the presence of such third-party agents.

Yet, it is possible that anomalous behavior is even greater in the field than in the lab due to endogenously generated institutions. First, since the latter environment is chosen by the experimenter, whose interests and goals are not necessarily similar to those of entrepreneurs in the field who endogenously generate profit opportunities, the lab might yield less anomalous behavior. This effect is reinforced if one considers that there is a steady supply of arbitrageurs selected into tasks by the market who work towards increasing the prevalence of such behaviors in markets by creating extractive institutions, whereas such pressures are usually absent in the lab.[38]

Relatedly, we should stress that there are also circumstances in which the field may be more "behavioral" than the lab. Haltiwanger and Waldman (1985; 1989) demonstrate that if the actions of the naïve and sophisticated agents are strategic complements, sophisticated agents have an incentive to mimic the behavior of the naïve types, causing the aggregrate market outcome to look more "behavioral" than the lab outcomes. The intuition is that in this case the naïve types are foregoing relatively little utility and therefore have less of a reason to adapt their behavior or drop out of the market. Fehr and Tyran (2006) present a nice discussion of factors

[37] As alluded to earlier, one must be cognizant of the fact that if feedback mechanisms are weak in the field such effects will generally not be observed.

[38] Relatedly, there is strong evidence that there are contextual effects everywhere, yet in the field the context (naturally part of the strategic environment) is oftentimes chosen by the agent who has market power. This is known by all agents and is reflected in their actions, which is quite different from the typical lab setting. Below we discuss other examples where the field might, in some sense, be "more behavioral" than the lab.

affecting the aggregation of individual decisions to the market level and highlight the various literatures that are influenced by such reasoning: from the efficient market hypothesis debate (see, e.g., DeLong et al., 1990a; 1990b; 1991; Shleifer and Vishny, 1997; Shleifer, 2000) to the importance of money illusion (see, e.g., Fehr and Tyran, 2001).

Short-run elasticity versus long-run elasticity

In most cases laboratory experiments are designed to last no more than a few hours. Yet, inference is oftentimes made over much longer time periods. Consider 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. The literature has taken such results as providing support for the received labor market predictions of Akerlof (1982) on gift exchange.

One result from the psychology literature is that there are important behavioral differences between short run (*hot*) and long (*cold*) run decision making. In the hot phase, visceral factors and emotions might prove quite important, whereas in the cold phase immediate reactions are more carefully suppressed. In this sense, the hot/cold settings can lead to much different behaviors (see, e.g., Loewenstein and Schkade, 1999). Loewenstein (2005) reviews some of the empirical evidence on behavioral differences across cold and hot states.[39]

The evidence is sparse within the experimental economics literature on this issue, but there is one study that provides a first 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 the gift treatment than in the non-gift treatment, but after the initial few hours no difference in outcomes was observed. We interpret these findings as suggesting that

[39] Likewise, whether real effort decisions are important remains an open empirical question.

great care should be taken before making inference from short-run laboratory experiments, which might be deemed as "hot" decision making, to long-run field environments, which typically revolve around "cold" decision making.

Group differences

Our model also has implications for laboratory studies that report differences across groups of race, sex, and age (see Croson and Gneezy, 2005, for a review). In particular, the lab *may* exaggerate group differences through a number of channels. First, as we have stressed throughout this paper, scrutiny might lead subjects to increase the weight they place on the "right" behavior relative to the wealth-maximizing behavior. If the design of the experiment signals to the subjects that the experimenter desires that minority groups underperform the dominant group, for example, members of both groups may respond (see Steele 1997 and Steele and Aronson, 1995).

Another channel that may induce systematic gaps across groups is differential self-selection of subjects into experiments. For instance, some experimental evidence suggests that women are more pro-social than men. Other studies have shown that women experience increases in elation and activity near the time of ovulation, whereas premenstrual and menstrual periods increase tension, irritability, depression, anxiety, and fatigue (Moos et al., 1969; Parlee, 1973; Sutherland and Stewart, 1965; De Marchi and Tong, 1972). Interestingly, Doty and Silverthorne (1975) find that most of the female volunteers for their experiment were in the ovulatory phase, whereas most of the female non-participants were in the postovulatory, premenstrual, and menstrual phases. If similar selection effects occur in economics experiments, then one cannot be sure that the gender results over social preferences are due to selection or natural gender differences.

How important might such selection effects be for experiments that do not trade-off morality and wealth? Chen et al. (2005) provide some insights in their study of bidding behavior in sealed bid auctions with independent and private valuations in a laboratory setting. They find that women bid significantly higher and earn significantly less money than men in first-price auctions (they find no evidence of a gender difference in the likelihood of dominant strategy play in the second-price auction). Importantly for our purposes, in the first-price auction, women who are menstruating do not bid differently from men. Thus, the gender difference in the first-price auction is driven by women during other phases of the menstrual cycle when they have higher estrogen levels.

IV. The Virtues of Lab Experiments

Thus far we have focused almost exclusively on concerns regarding the limitations of laboratory experiments, only pausing to explicate their strengths sporadically. In this section, we summarize the value of such experiments.

First, it is important to recognize that our model predicts that a wide class of laboratory results should be directly applicable to the field. In particular, when moral concerns are absent, the computational demands on participants are small, non-random selection of participants is not an important factor, experience is unimportant or quickly learned, and the experimenter has created a lab context that mirrors the important aspects of the real-world problem, then quantitative results from the lab are likely to be a closer guide to real-world behavior.

Moreover, except in rare circumstances, laboratory experimentation is likely to be a useful tool for providing *qualitative* evidence, even when the generalizability of deep structural parameters is suspect. For example, consider the notion of diminishing marginal value, which has had a profound impact on the development of neoclassical theory. Horowitz et al. (2006) use

a series of simple price and exchange institutions to show strong evidence of diminishing marginal value in both laboratory and field experiments. Similarly, analyzing the demand behavior from the sale of a private good (strawberries) in an actual "real world" field setting and in a laboratory auction setting, Brookshire et al. (1987) report that the two settings yielded similar demand behavior.

In the context of matching markets, Kagel and Roth (2000) reproduce the basic facts of the British medical markets observed in Roth (1991), and McKinney et al. (2005) find similar behavioral patterns as documented in the field by Niederle and Roth (2003). In the area of social dilemmas, Fehr and Gachter (2000) provide evidence that adding a punishment option to a standard public goods game considerably increases cooperative play, a finding we trust will manifest itself in broader field applications. Likewise, Levine and Plott (1977) report that agenda influences are directionally consistent across laboratory experiments and a naturally-occurring setting that they manipulated to secure a more favorable outcome in their flying club. We could easily expand such a list—from voting games (see, e.g., the work of Charles Plott and Thomas Palfrey) to simple auction games (see, e.g., the work of John Kagel, Dan Levin, and colleagues) to various market games (see, e.g., Charlie Holt's website)—and are confident that we could scribe a tome on such successes that are consonant with our theoretical model.[40]

We also view laboratory experimentation as a useful first step in the area of policymaking. As Plott (1997, p. 606) notes, experimental economics is an "…inexpensive, and fast method for getting data on how various types of auctions might perform." Plott (2001) provides several seminal examples that link laboratory experiments with policy applications. Similarly, the lab can be used to rank mechanisms within broad areas, such as charitable

[40] Similarly, recent work in experimental economics has found interesting correlations across games/environments (see, e.g., the neat studies of Karlan, 2005 and Benz and Meier, 2006). As aforementioned, this work has its roots in the cross-situational consistency debated discussed in Mischel (1968) and Ross and Nisbett (1991).

fundraising. In this spirit, Rondeau and List (2005) explore the efficacy of different fundraising schemes and find that the lab and field yield identical rankings of the 4 mechanisms that they study.

V. 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 provide a framework for such a discussion, erring on the side of simplicity in an effort to present our ideas concretely. We argue that experiments may not always yield results that are robustly generalizable. The types of experiments we should be particularly wary of are those that purport to estimate "physical constants" such as particular parameters of individuals' preferences. Part of the reason is 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.

To the extent that lab and naturally-occurring environments systematically differ on any of these dimensions, the results obtained inside and outside the lab need not correspond. 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 stark 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.

We conclude that there remains an important role for traditional laboratory experiments in economics, though one that is more limited than some ardent experimentalists might subscribe. At a minimum, lab experiments can provide a crucial first understanding of

qualitative effects, suggest underlying mechanisms that might be at work when certain data patterns are observed, provide insights into what can happen, and evoke empirical puzzles. In this light, experimenters are like aerodynamicists who use wind tunnels to test models of proposed aircraft, helicopters, cars, and trains. The wind tunnel provides the engineer with valuable data on scale models much like the lab provides economists with important insights on an economic phenomenon.[41] In the best cases – well designed experiments asking questions most suited to laboratory analysis – lab experiments can of course go beyond this role.

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 simple one 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.[42]

Second, by focusing on qualitative rather than quantitative insights much can be learned. Further, by adopting experimental designs that recognize the potential weaknesses of the lab, the usefulness of lab studies can be enhanced. For instance, even in games that pit wealth and morality, by "nesting" treatments and performing a difference-in-difference estimation that

[41] Some argue that the wind tunnel is the most lasting contribution of the Wright brothers to the science of aerodynamics. Interestingly, it is estimated that it took the Wright Brothers less than 20 hours of wind tunnel testing to produce their successful flyer.

[42] As we noted earlier, any empirical estimate requires an appropriate theory for proper inference—whether the data are obtained in the lab, from coin collector shows, or from government surveys. Knight (1921) recognized this problem when he noted that "The existence of a problem in knowledge depends on the future being different from the past, while the possibility of a solution of the problem depends on the future being like the past." We see this as a useful metaphor when thinking about the transference of results from the lab to naturally-occurring markets. In this sense, even in cases where lab results are believed to have little generalizability, some number is better than no number provided the proper theoretical model is used to make inference.

effectively "nets" out laboratory effects one can learn more about deep structural parameters than running a simple, more traditional, design.

Finally, we believe that the sharp dichotomy sometimes drawn between lab experiments and data generated in natural settings is a false one. The same concerns arise in both settings regarding the interpretation of estimates and their generalizability outside of the immediate application, circumstances, and treated population. Each approach has a different set of strengths and weaknesses, and thus a combination of the two is likely to provide more insight than either in isolation. Field experiments, which incorporate the strengths of the two approaches, can serve as a bridge connecting these two empirical approaches.

References

Akerlof, George A. "Labor Contracts as Partial Gift Exchange." *Quarterly Journal of Economics,* November 1982, *97*(4), pp. 543-569.

Akerlof, George A. and Rachel E. Kranton. "Economics and Identity." *Quarterly Journal of Economics*, August 2000, *115*(3), pp. 715-753.

Akerlof, George A and Rachel E. Kranton. "Identity and the Economics of Organizations." *Journal of Economic Perspectives*, Winter 2005, *19*(1), pp. 9-32.

Alevy, Jonathan E., Michael S. Haigh, and John A. List. "Information Cascades: Evidence from a Field Experiment with Financial Market Professionals." *Journal of Finance*, 2006 forthcoming.

Allen, Vernon. "Situational Factors in Conformity," in Leonard Berkowitz, ed., *Advances in Experimental and Social Psychology*, 1965,Vol. 2. New York: Academic Press.

Andersen, Steffen, Glenn W. Harrison, Morten Igel Lau, and E. Elisabet Rutström. "Preference Heterogeneity in Experiments: Comparing the Lab and Field." 2004, Working Paper 04-21, Department of Economics, College of Business Administration, University of Central Florida.

Andreoni, James. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy*, December 1989, *97*, pp. 1447-58.

Andreoni, James. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *Economic Journal*, June 1990, *100*, pp. 464-477.

Andreoni, James. "Cooperation in Public Goods Experiments: Kindness or Confusion?" *American Economic Review*, 1995, 85(4): 891-904.

Andreoni, James and Rachel Croson. "Partners versus Strangers: Random Rematching in Public Goods Experiments." *Handbook of Experimental Economic Results, ed. C. Plott and V. Smith, 2003*

Andreoni, James and John Miller. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. " *Econometrica*, March 2002, *70*(2), pp. 737-753.

Ashraf, Nava, Colin F. Camerer, and George Loewenstein. "Adam Smith, Behavioral Economist." *Journal of Economic Perspectives*, Summer 2005, *19*, no. 3, pp. 131-145.

Ashraf Nava, Dean Karlan, and Wesley Yin. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." *Quarterly Journal of Economics*, May 2006, 121(2), pp. 635-672.

Bandiera, Oriana, Iwan Rasul, and Imran Barankay. “Social Preferences and the Response to Incentives: Evidence from Personnel Data.” *Quarterly Journal of Economics*, 2005 120(3): pp. 917-962.

Bardsley, Nicholas. “Altruism or Artifact? A Note on Dictator Game Giving.” University of Nottingham working paper, 2005a.

Bardsley, Nicholas. “Experimental Economics and the Artificiality of Alteration.” *Journal of Economic Methodology*, 2005b, 239-251.

Bazerman, M.H. and W.F. Samuelson. "I Won the Auction But Don't Want the Prize." *Journal of Conflict Resolution*, 1983, *27,* pp. 618-634.

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

Benz, Matthias and Stephan Meier. "Do People Behave in Experiments as in Real Life? – Evidence from Donations." Institute for Empirical Research in Economics - Working Papers iewwp248, 2006

Berg, Joyce, John W. Dickhaut, and Kevin A. McCabe. "Trust, Reciprocity, and Social History." *Games and Economic Behavior*, July 1995, *10*(1), pp. 122-142.

Bernheim, B.D. “A Theory of Conformity.” *Journal of Political Economy*, 1994, *102*(5), pp. 841-877.

Binmore, Ken. *Game Theory and the Social Contract, Volume 1: Playing Fair.* Cambridge and London: MIT Press, 1994.

Binzel, Richard P. “Pluto,” *Scientific American*, 1990, *262*, pp. 50-56.

Bohm, Peter. “Pitfalls in Experimental Economics.” *Experimental Economics: Financial Markets, Auctions, and Decision Making*, edited by Fredrik Andersson and Hakan Holm, Kluwer Academic Publishers, 2002, pp 117-126.

Bohnet, Iris and Robert D. Cooter. "Expressive Law: Framing or Equilibrium Selection?" KSG Working Paper No. RWP03-046; and UC Berkeley Public Law Research Paper No. 138, 2005.

Bohnet, Iris, and Bruno S. Frey. “Social Distance and Other-Regarding Behavior in Dictator Games.” *American Economic Review,* 1999a, 89, 335-339.

_____. “The Sound of Silence in Prisoner’s Dilemma and Dictator Games.” *Journal of Economic Behavior and Organization,*199b, 38, 43-57

Bolton, Gary E., and Axel Ockenfels. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, March 2000, *90*(1), pp. 166-193.

Bolton, Gary E., Rami Zwick, and Elena Katok. “Dictator Game Giving: Rules of Fairness Versus Acts of Kindness.” *International Journal of Game Theory, 1998*, 27, 269-299.

Botelho, Anabela, Glenn W. Harrison, Ligia M. Costa Pinto, and Elisabet Rutstrom. “Social Norms and Social Choice.” Working Paper, Department of Economics, College of Business Administration, University of Central Florida, 2005.

Brookshire, David and Don Coursey. “Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures.” *American Economic Review,* September 1987, 77(4), pp. 554-566.

Brookshire, David S., Don L. Coursey, and William D. Schulze (1987) “The External Validity of Experimental Economics Techniques: Analysis of Demand Behavior,” *Economic Inquiry*, 25:2 (April), pp. 239-250.

Brown, Martin, Armin Falk, and Ernst Fehr. "Relational Contracts and the Nature of Market Interactions." *Econometrica*, 2004, *72*(3), pp. 747-780.

Brunswik, Egon. “Representative Design and Probabilistic Theory in a Functional Psychology.” *Psychological Review*, 1955, 62(3), pp. 193-217.

Brunswik, Egon. *Perception and the Representative Design of Psychological Experiments* (2nd ed.). Berkeley, CA: University of California Press, 1956.

Burnham, Terence C. “A Theoretical and Experimental Investigation of Anonymity and Gift Giving.” *Journal of Economic Behavior and Organization,* 2003, 50, 133-144.

Burnham, T., K. McCabe, V. Smith. “Friend-or-Foe Intentionality Priming in an Extensive Form Trust Game.” *Journal of Economic Behavior and Organization,* 2000, 1244, pp. 1–17.

Burns, Penny. "Experience and Decision Making: A Comparison of Students and Businessmen in a Simulated Progressive Auction.” *Research in Experimental Economics*, (JAI), ed. Vernon L. Smith, 1985, pp. 139-153.

Camerer, Colin F. “Do Biases in Probability Judgment Matter in Markets? Experimental Evidence.” *American Economic Review,* 1987, 77(5): 981-97.

Camerer, Colin F. “The Rationality of Prices and Volume in Experimental Markets.” *Organizational Behavior and Human Decision Processes,* 1992, 51, 237-72.

Camerer, Colin F. "Behavioral Game Theory: Experiments in Strategic Interaction. " *Roundtable Series in Behavioral Economics,* 2003, *Roundtable Series in Behavioral Economics*. Princeton: Princeton University Press, pp. xv, 550.

Camerer, Colin F. and Robin M. Hogarth. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework." *Journal of Risk and Uncertainty*, 1999, 19, pp. 7-42.

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

Campbell, D. T., and J.C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin, 1963.

Carpenter, Jeffrey, Eric Verhoogen, and Stephen Burks. "The Effect of Stakes in Distribution Experiments." *Economics Letters*, March 2005a, *86*(3), pp. 393-98.

_____. "Comparing Students to Workers: The Effects of Stakes, Social Framing, and Demographics on Bargaining Outcomes," in *Field Experiments in Economics*, (JAI), eds. Jeffrey Carpenter, Glenn Harrison, John A. List, 2005b, pp. 261-290.

Casari, Marco, John C. Ham, and John H. Kagel. "Selection Bias, Demographic Effects and Ability Effects in Common Value Auction Experiments." Ohio State University working paper, 2005.

Cason, Timothy N., and Vai-Lam Mui. "A Laboratory Study of Group Polarization in the Team Dictator Game." *Economic Journal*, 1997, 107, 1465-1483.

Charness, Gary, "Attribution and Reciprocity in an Experimental Labor Market," mimeo, 1996.

Charness, Gary and Dan Levin. "The Origin of the Winner's Curse: A Laboratory Study." UCLA Department of Economics, 2005.

Charness, Gary and Matthew Rabin. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, August 2002,117(3), August 2002, 817-869.

Chen, Paul. "Reciprocity at the Workplace: Do Fair Wages Lead to Higher Effort, Productivity, and Profitability? Working paper, 2005.

Chen, Yan, Peter Katuscak and Emre Ozdenoren. "Why Can't a Woman Bid More Like a Man?" Working paper, 2005.

Clark, Kenneth, and Martin Sefton. "The Sequential Prisoner's Dilemma: Evidence on Reciprocation." *Economic Journal*, January 2001, 111, 61-68.

Collard, David. *Altruism and Economy: A Study in Nonselfish Economics*. New York, NY: Oxford University Press, 1978.

Cookson, R. "Framing Effects in Public Goods Experiments." *Experimental Economics*, 2000, 3, 55-79.

Cooper, David J., John H. Kagel, W. Lo, Q.L. Gu. "Gaming Against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers." *American Economic Review*, September 1999, *89*(4), pp. 781-804.

Coursey, Don, Hovis, John, and Schulze, William. "The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value." *Quarterly Journal of Economics*, 1987, 102(3), pp. 679-90.

Cox, James C., Sam Dinkin, and James T. Sawrthout. "Endogenous Entry and Exit in Common Value Auctions." *Experimental Economics*, October 2001, *4*(2), pp. 163-81.

Crack, Timothy. *Heard on the Street: Quantitative Questions for Wall Street Interviews*. Bloomington, IN, Timothy Crack Publishing, (2000).

Croson, Rachel and Uri Gneezy, "Gender Differences in Preferences: A Review of Economics Experiments." Working paper, 2005.

Cross, J. *Some Comments on the Papers by Kagel and Battalio and by Smith* in J. Kmenta and J. Ramsey (eds.) Evaluation of Econometric Models, New York: New York University Press, 1980.

Cummings, Ronald G. and Laura O. Taylor. "Does Realism Matter in Contingent Valuation Surveys?" *Land Economics,* 1998, 74, pp. 203-15.

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

DellaVigna, Stefano, and Ulrike Malmendier. "Paying Not to Go to the Gym." forthcoming, *American Economic Review*., 2005.

DeLong, J.B., A. Shleifer, L.H. Summers, and R. J. Waldman. "Noise Trader Risk in Financial Markets." *Journal of Political Economy* 1990a, 98(4), pp. 703-38.

_____. "Positive Feedback Investment Strategies and Destabilizing Rational Speculation." *Journal of Finance* 1990b, 45(2), pp. 379-95.

_____. "The Survival of Noise Traders in Financial Markets." *Journal of Business,* 1991, 64, pp. 1-19.

DeMarchi, G.W. and J.E. Tong. "Psychophysiology." 1972, 9, pp. 31

Doty, Richard L. and Colin Silverthorne. "Influence of Menstrual Cycle on Volunteering Behavior." *Nature*, 1975, 254, pp. 139-140.

Duffy, John and Jack Ochs. "Cooperative Behavior and the Frequency of Social Interaction." Working paper, 2005.

Dyer, Douglas and John H. Kagel. “Bidding in Common Value Auctions: How the Commercial Construction Industry Corrects for the Winner’s Curse.” *Management Science*, 1996, 42, pp. 1463-76.

Eckel, Catherine and Phillip Grossman. "Altruism in Anonymous Dictator Games." *Games and Economic Behavior*, 1996, 16, pp. 181-191.

Eckel, Catherine C. and Philip J. Grossman. “Volunteers and Pseudo-Volunteers: The Effect of Recruitment Method in Dictator Experiments.” *Experimental Economics*, 2000, 3, pp. 107-120.

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

Engelmann Dirk and Urs Fischbacher. “Indirect Reciprocity and Strategic Reputation Building in an Experimental Helping Game.” 2002, Working Paper No. 132, University of Zurich.

Fehr, Ernst and Armin Falk. “Wage Rigidity in a Competitive Incomplete Contract Market.” *Journal of Political Economy*, 1999, 107, pp. 106-134.

Fehr, Ernst and Urs Fischbacher. “The Nature of Human Altruism.” *Nature*, 2003, 425.

Fehr, Ernst and Simon Gächter. “Fairness and Retaliation: The Economics of Reciprocity.” *Journal of Economic Perspectives*, 2000, 14(3), pp. 159-181.

_____. “Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 2000, 90(4), pp. 980-994.

Fehr, Ernst; Simon Gachter, and Georg Kirchsteiger. "Reciprocity as a Contact Enforcement Device: Experimental Evidence." *Econometrica*, 1997, 65(4), pp. 833-60.

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

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

Fehr, Ernst and Klaus Schmidt. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 1999, 114, pp. 817-868.

Fehr, Ernst, Urs Fischbacher, and E. Tougareva. “Do High Stakes and Competition Undermine Fairness - Evidence from Russia.” Institute for Empirical Research in Economics, University of Zürich, Working Paper No. 120, 2002.

Fehr, Ernst and Jean R. Tyran. “Does Money Illusion Matter?” *American Economic Review* 2001, 91(5), pp. 1239-62.

_____. "Individual Irrationality and Aggregate Outcomes." *Journal of Economic Perspectives,* forthcoming, 2006.

Ferraro, Paul J. and Christian A. Vossler. "The Dynamics of Other-Regarding Behavior and Confusion: What's Really Going on in Voluntary Contributions Mechanism Experiments?" Working paper, 2006.

Feynman, R.P., R.B. Leighton and M. Sands. *The Feynman Lectures on Physics*. Reading, MA: Addison-Wesley Publishing Company, 1963.

Frey, Bruno S., Felix Oberholzer-Gee, and Reiner Eichenberger. "The Old Lady Visits Your Backyard: A Tale of Morals and Markets." *The Journal of Political Economy*, 1996, 104(6), pp. 1297-1313.

Frohlich, Norman, Joe Oppenheimer, and J. Bernard Moore. "Some Doubts About Measuring Self-Interest Using Dictator Experiments: The Costs of Anonymity." *Journal of Economic Behavior and Organization*, 2001, 46, pp. 271-290.

Gachter, Simon and Armin Falk. "Reputation and Reciprocity: Consequences for the Labour Relation." *Scandinavian Journal of Economics*, 2002, 104(1), pp. 1-26.

Gintis, Herbert. "The Contribution of Game Theory to Experimental Design in the Behavioral Sciences." *Behavioral and Brain Sciences*, 2001, 24, pp. 411-412.

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

Gneezy, Uri and John A. List. "Putting Behavioral Economics to Work: Field Evidence of Gift Exchange," *Econometrica,* forthcoming, 2006.

Gneezy, Uri, Ernan Haruvy, and H. Yafe. "The Inefficiency of Splitting the Bill: A Lesson in Institution Design." *The Economic Journal,* April 2004, 114, pp. 265-280.

Grampp, William. "Adam Smith and the Economic Man." *The Journal of Political Economy*, August 1948, 56(4), pp. 315-336.

Haltiwanger, C.J. and M. Waldman. "Rational Expectations and the Limits of Rationality: An Analysis of Heterogeneity." *American Economic Review*, 75(3), pp. 326-40.

_____. "Limited Rationality and Strategic Complements: The Implications for Macroeconomics." *Quarterly Journal of Economics*, 1989, 104(3), pp. 463-83.

Hanemann, W. Michael. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review*, June 1991, 81(3), pp. 635-47.

Haley, Kevin J. and Daniel M.T. Fessler. “Nobody’s Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game.” *Evolution of Human Behavior*, 2005, 26, pp. 245-256.

Haney, C., W.C. Banks, and P.G. Zimbardo. “Interpersonal Dynamics in a Simulated Prison.” *International Journal of Criminology and Penology*, 1973, 1, pp. 69–97.

Hannan, R. Lynn, John H. Kagel, and Donald V. Moser. “Partial Gift Exchange in an Experimental Labor Market: Impact of Subject Population Differences, Productivity Differences, and Effort Requests on Behavior.” *Journal of Labor Economics,* 2004, 20(4), pp. 923-951.

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

Harbaugh, William T. Kate Krause, and Steven Liday. “Bargaining by Children.” University of Oregon working paper, 2003b.

Harris, S.J. and K. Munger. “Effects of an Observer on Hand Washing in Public Restrooms.” Perceptual and Motor Skills, 1989, 69, pp. 733-735.

Harrison Glenn W. “Theory and Misbehavior of First-Price Auctions.” *American Economic Review*, September 1989, 79(4), pp. 749-762.

Harrison, Glenn W., Morten Igel Lau, and Elisabet E. Rutström. "Risk Attitudes, Randomization to Treatment, and Self-Selection Into Experiments." Working Paper 05-01, Department of Economics, College of Business Administration, University of Central Florida, 2005.

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

Harrison, Glenn W. and John A. List. “Naturally Occurring Markets and Exogenous Laboratory Experiments: A Case Study of Winner's Curse.” UCF working paper, 2005.

Harrison, Glenn and Elisabet Rutstrom. “Doing It Both Ways -- Experimental Practice and Heuristic Context.” *Behavioral and Brain Sciences*, 2001, 24(3), pp. 413-414.

Heiner, R. “The Origin of Predictable Behavior.” *American Economic Review*, 1983, 73(4), pp. 560-595.

Henrich, J. et al. “In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies.” *American Economic Review*, May 2001, 91(2), pp. 73-78.

Henrich, J., R. Boyd, S. Bowles, H. Gintis, E. Fehr, C. Camerer, R. McElreath, M. Gurven, K. Hill, A. Barr, J. Ensminger, D. Tracer, F. Marlow, J. Patton, M. Alvard, F. Gil-White, and N.

Smith. "'Economic Man' in Cross-Cultural Perspective: Ethnography and Experiments from 15 Small-Scale Societies." *Behavioral and Brain Sciences*, forthcoming, 2004.

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

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. "Preferences, Property Rights, and Anonymity in Bargaining Games." *Games and Economic Behavior*, 1994, 7(3), pp. 346-380.

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

Holt, Charles A. *Markets, Games, and Strategic Behavior: Recipes for Interactive Learning*. 2005, Addison-Wesley, forthcoming.

Horowitz, John, John A. List, and Kenneth McConnell. "Diminishing Marginal Value: Evidence from Field Experiments,' *Oxford Economic Papers*, forthcoming, 2006.

Houser, Daniel and Robert Kurzban. "Revisiting Kindness and Confusion in Public Goods Experiments." *American Economic Review*, 2002, 92(4), pp. 1062-1069.

Jones, Stephen. *The Economics of Conformism*. Oxford & New York: Basil Blackwell, 1984.

Kagel, John H., Raymond C. Battalio, and James M. Walker. "Volunteer Artifacts in Experiments in Economics; Specification of the Problem and Some Initial Data from a Small-Scale Field Experiment." *Research in Experimental Economics*, (JAI), ed. Vernon L. Smith, pp. 169-197 (1979).

Kagel, J.H., D. Levin, and R. Harstad. "Comparative Static Effects of Number of Bidders and Public Information on Behavior in Second-Price Common-Value Auctions." *International Journal of Game Theory*, 1995, 24, pp. 297-319.

Kagel, John H. and Dan Levin. "Common Value Auctions with Insider Information." *Econometrica*, September 1999, 67(5), pp. 1219-38

Kagel, John H. and Dan Levin *Common Value Auctions and the Winner's Curse*. 2002, pp. xvi, 401, Princeton and Oxford: Princeton University Press

Kagel, John H. and A.E. Roth. "The Dynamics of Reorganization in Matching Markets: A Laboratory Experiment Motivated By a Natural Experiment." *Quarterly Journal of Economics*, February 2000, pp. 201-35.

Kahneman, Daniel, and Amos Tversky. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*, March 1979, 47, pp. 263-91.

Karlan, Dean. “Using Experimental Economics to Measure Social Capital and Predict Financial Decisions.” *American Economic Review*, December 2005, 95(5), pp.1688-1699.

Karni, Edi. “Impartiality: Definition and Representation.” *Econometrica*, 1998, *66*, pp. 1405–1415.

Knez, Peter, Vernon L. Smith, and Arlington Williams. “Individual Rationality, Market Rationality, and Value Estimation.” *American Economic Review*, May 1985, 75(2), pp. 397-402.

Knight, Frank H. *Risk, Uncertainty and Profit.* New York: Houghton Mifflin, 1921.

Kotchen, Matthew J. “Impure Public Goods and the Comparative Statistics of Environmentally Friendly Consumption.” *Journal of Environmental Economics and Management*, March 2005, 49(2), pp. 281-300.

Kremer, Michael. “Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons." *American Economic Review*, May 2003, 93(2), pp. 102-106.

Krueger, Alan B. and Alexandre Mas. “Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires.” *Journal of Political Economy*, 2004, 112(2), pp. 253-289.

Laffont, Jean-Jacques. "Macroeconomic Constraints, Economic Efficiency and Ethics: An Introduction to Kantian Economics." *Economica*, 1975, 42(168), pp. 430-437.

Lange, Andreas, John A. List, and Michael K. Price. “Auctions with Resale When Private Values Are Uncertain: Theory and Empirical Evidence.” NBER working paper #10639, 2005.

Laury, Susan K., James M. Walker, and Arlington W. Williams. “Anonymity and the Voluntary Provision of Public Goods.” *Journal of Economic Behavior and Organization*, 1995, 27, pp. 365-380.

Lazear, Edward P., Ulrike Malmendier, and Roberto A. Weber, “Sorting in Experiments,” working paper Stanford, 2004

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

Levine, Michael E. and Charles R. Plott. “Agenda Influence and its Implications.” *Virginia Law Review*, 1977, 63, pp. 561-604.

Levitt, Steven D. "Testing Theories of Discrimination: Evidence from the Weakest Link." *Journal of Law & Economics*, 2004, 47(2), pp. 431-452.

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

List, John A. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, 2003, 118(1), pp. 41-71.

List, John A., "Neoclassical Theory Versus Prospect Theory: Evidence from the Field." *Econometrica*, 2004, 72(2), pp. 615-625.

List, John A. "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions." *Journal of Political Economy*, 2005a, forthcoming.

List, John A. "Dictator Game Giving is an Experimental Artifact," working paper, University of Chicago, 2005b.

List, John A. and Todd L. Cherry. "Examining the Role of fairness in High Stakes Allocation Decisions." *Journal of Economic Behavior and Organization*, 2005, forthcoming.

List, John A. and David Lucking-Reiley. "Demand Reduction in a Multi-Unit Auction: Evidence from a Sportscard Field Experiment." *American Economic Review*, 2000, 90(4), pp. 961-972.

List, John A., Robert Berrens, Alok Bohara, and Joe Kerkvliet. "Examining the Role of Social Isolation on Stated Preferences." *American Economic Review*, 2004, 94(3), pp. 741-752.

Loewenstein, George and David Schkade. *Wouldn't It Be Nice? Predicting Future Feelings.* In D. Kahneman, E. Diener & N. Schwartz (Eds.), Well Being: The Foundations of Hedonic Psychology. Russell Sage Foundation, 1999.

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

Loewenstein, George. "Hot-cold Empathy Gaps and Medical Decision-Making." *Health Psychology*, 2005, forthcoming.

Loomes, Graham and Robert Sugden. "Regret Theory: An Alternate Theory of Rational Choice Under Uncertainty." *Economic Journal*, 1982, 92(368), pp. 805-24.

Loomes Graham, Chris Starmer, and Robert Sugden. "Preference Reversals and Disparities Between Willingness to Pay and Willingness to Accept in Repeated Markets" working paper, University of Nottingham, 2005

Margolis, Howard. *Selfishness, Altruism, and Rationality.* Chicago, IL: University of Chicago Press, 1982.

Masclet, David et al. "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review*, 2003, 93(1), pp. 366-380.

McKinney, Nicholas, Muriel Niederle, and Alvin E. Roth, ""The collapse of a medical labor clearinghouse (and why such failures are rare)," *American Economic Review*, 2005, 95(3), pp. 878-889.

Meyer, Bruce, "Natural and Quasi-Experiments in Economics." *Journal of Business and Economic Statistics*, 13, pp. 151-161.

Milgram, Stanley. "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology*, 1963, 67, pp. 371-378.

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

Mook, D. G. "In Defense of External Invalidity." *American Psychologist,* 1983, *38*, pp. 379-387.

Moos, R.H. et al., Journal of Psychosomat. Res. 13, 37, 1969.

Mullainathan, Sendhil and Richard H. Thaler. "Behavioral Economics." *National Bureau of Economic Research,* Inc, 2000, NBER Working Papers: 7948.

Myagkov Mikhail and Charles Plott. "Exchange Economies and Loss Exposure: Experiments Exploring Prospect Theory and Competitive Equilibria in Market Environments." *American Economic Review*, 1997, 87(5), pp. 801-828.

Newton, Isaac. *Mathematical Principles of Natural Philosophy.* Los Angeles, CA: University of California Press, 1966.

Niederle, Muriel and Alvin E. Roth. "Unraveling Reduces Mobility in a Labor Market: Gastroenterology With and Without a Centralized Match," *Journal of Political Economy*, 2003, 6, pp. 1342-1352.

Orne, Martin T. *The Demand Characteristics of an Experimental Design and their Implications.* Paper read at American Psychological Association, Cincinnati, 1959a.

_____. "The Nature of Hypnosis: Artifact and Essence." *Journal of Abnormal and Social Psychology,* 1959b, *58,* pp. 277-299.

_____. "On the Social Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications," *American Psychologist*, *17*(10), 1962, 776-783.

Ortmann, Andreas and Ralph Hertwig. "Why Anomalies Cluster in Experimental Tests of One-shot and/or Finitely Repeated Games: Some Evidence from Psychology." working paper, 2000.

Ortmann, Andreas. “Charles R. Plott's "Collected Papers on the Experimental Foundations of Economic and Political Science.” *Journal of Economic Psychology*, 2003, 24, 2003, pp. 555-575.

Ortmann, Andreas. *Field Experiments in Economics: Some Methodological Caveats* in *Field Experiments in Economics*, (JAI), eds. Jeffrey Carpenter, Glenn Harrison, John A. List, (2005).

Parco, J. E., A. Rapoport, and W.E. Stein. “Effects of Financial Incentives on the Breakdown of Mutual Trust.” *Psychological Science*, 2002, 13, pp. 292-297.

Parlee, M.B. *Psychological Bulletin*, 1973, 80, pp. 545.

Pierce, A. H. “The Subconscious Again.” *Journal of Philosophy, Psychology, & Scientific Methods,* 1908, 5, pp. 264-271.

Plott, Charles R. “Industrial Organization Theory and Experimental Economics.” *Journal of Economic Literature,* 1982, 20, pp. 1485-1527.

_____. “Laboratory Experimental Testbeds: Application to the PCS Auction.” *Journal of Economics & Management Strategy*, 1997, 6(3), pp. 605-638.

_____. *Public Economics, Political Processes and Policy Applications. Collected Papers on the Experimental Foundations of Economics and Political Science*. Volume One. Cheltenham, UK: Edward Elgar Publishing (2001).

Pruitt Dean G. and Melvin J. Kimmel. “Twenty Years of Experimental Gaming: Critique, Synthesis, and Suggestions for the Future.” *Ann. Rev. Psychology*, 1977, 28, pp. 363-392.

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

_____. "Psychology and Economics." *Journal of Economic Literature*, 1998, 36(1), pp. 11-46.

Rapoport, A. Conflict *Resolution in the Light of Game Theory and Beyond.* In *The Structure of Conflict*, ed., P. Swingle, pp. 1-42, New York: Academic. 1970.

Rege, Mari and Kjetil Telle. "The Impact of Social Approval and Framing on Cooperation in Public Good Situations." *Journal of Public Economics*, 2004, 88(7-8), pp. 1625-1644.

Rondeau, Daniel and John A. List. “Exploring the Demand Side of Charitable Fundraising: Evidence from Field and Laboratory Experiments.” Working paper, 2005.

Rosenthal, R.L. “Covert Communication in the Psychological Experiment.” *Psychological Bulletin,* 1967, 67*,* pp. 356–367.

_____. *Artifact in Behavioral Research*, New York: Academic Press, 1969.

_____. *The Volunteer Subject*, New York: John Wiley and Sons, 1973.

Rosenthal, Robert W. "Market Games with Production and Public Commodities." 1976, Northwestern University, Center for Mathematical Studies in Economics and Management Science, Discussion Papers: 156

Rosenthal, Robert W. "A Note on Robustness of Equilibria with Respect to Commitment Opportunities." *Games and Economic Behavior*, May 1991, 3(2), pp. 237-43.

Ross, L. and A. Ward. *Naive Realism: Implications for Social Conflict and Misunderstanding* in T. Brown, E. Reed, and E. Turiel eds., *Values and Knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1996, pp. 103-135.

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

Roth, Alvin E. "A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the U.K." *American Economic Review*, 1991, 81, pp. 415-440.

Roth, Alvin E. *Bargaining Experiments* in Kagel, J.H., Alvin E. R., eds., *The Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press, 1995, pp. 253–342.

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir. "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study." *The American Economic Review*, 1991, 81(5), pp. 1068-1095.

Rutstrom, Elisabet. "Home-grown Values and Incentive Compatible Auction Design." *International Journal of Game Theory*, 1998, 27(3), pp. 427-441.

Rutstrom, Elisabet and Melonie B. Williams. "Entitlements and Fairness: An Experimental Study of Distributive Preferences." *Journal of Economic Behavior and Organization*, 2000, 43, pp. 75-80.

Schultz, D.P. "The Human Subject in Psychological Research," *Psychological Bulletin*, 1969, 72:214-228.

Seinen, I. and A.J.H.C. Schram. "Social Status and Group Norms: Indirect Reciprocity in a Repeated Helping Experiment." *European Economic Review*, 2006, 50, pp.581-602.

Shleifer, Andrei and R. Vishny. "The Limits of Arbitrage." *Journal of Finance*, 1997, 52, pp. 35-55.

Shleifer, Andrei. *Inefficient Markets – An Introduction to Behavioral Finance.* Oxford, Oxford University Press (2000).

Schmuckler, M. A. "What is Ecological Validity? A Dimensional Analysis." *Infancy*, 2001, 2, pp. 419-436.

Schram, Arthur. "Artificiality: The Tension Between Internal and External Validity in Economic Experiments." *Journal of Economic Methodology*, 2005, 12(2), pp. 225-237.

Sen, Amartya K. "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory." *Journal of Philosophy and Public Affairs*, 1977, 6, pp. 317-44.

Settle, Thomas B. "An Experiment in the History of Science." *Science*, 1961, 133, pp. 19-23.

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

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

Smith, Adam. *The Theory of Moral Sentiments.* eds. D.D. Raphael and A.L. Macfie. Indianapolis: Liberty Fund, 1982.

Smith, Vernon L. "Microeconomic Systems as an Experimental Science." *The American Economic Review*, 1982, 72(5), pp. 923-955.

Smith, Vernon L. and James M. Walker. "Rewards, Experience and Decision Costs in First Price Auctions." *Economic Inquiry*, April 1993, 31(2), pp. 237-45

Sobel, Joel. "Social Preferences and Reciprocity," mimeo, U.C. San Diego, 2002.

Soetevent, Adriaan R. "Anonymity in Giving in a Natural Context-A Field Experiment in Thirty Churches," *Journal of Public Economics*, forthcoming.

Starmer, Chris. "Experimental Economics: Hard Science or Wasteful Tinkering?" *Economic Journal*, 1999a, 109(453), pp. F5-15.

Starmer, Chris. "Experiments in Economics: Should we Trust the Dismal Scientists in White Coats?" *Journal of Economic Methodology*, 1999b, 6(1), pp. 1-30.

Steele, C.M. "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance." *American Psychologist*, 1997, 52, pp. 613 – 629.

Steele, C.M. and J. Aronson. "Stereotype Vulnerability and the Intellectual Test Performance of African Americans." *Journal of Personality and Social Psychology*, 1995, 69, pp. 797 – 811.

Stigler, George. "Economics or Ethics?" in S. McMurrin, ed. *Tanner Lectures on Human Values*. Cambridge: Cambridge University Press, 1981.

Sugden, Robert. "Reciprocity: The Supply of Public Goods Through Voluntary Contributions." *Economic Journal*, 1984, 94, pp. 772-787.

Sutherland, H. and Lancet I. Stewart, 1965, 1180.

Thaler, Richard H. *The Winner's Curse: Paradoxes and Anomalies of Economic Life.* 1992, pp. ix, 230, Princeton and Chichester, U.K: University Press

Tversky, Amos and Daniel Kahneman. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Quarterly Journal of Economics*, 1991, 106(4), pp. 1039-61

Wilde, Louis. "On the Use of Laboratory Experiments in Economics." In Joseph Pitt, ed., *The Philosophy of Economics*. Dordrecht: Reidel, 1981, pp. 137-143.