Viewpoint: On the generalizability of lab behaviour to the field

Steven D. Levitt and John A. List *Department of Economics, University of Chicago*

Abstract.   We can think of no question more fundamental to experimental economics than understanding whether, and under what circumstances, laboratory results generalize to naturally occurring environments. In this paper, we extend Levitt and List (2006) to the class of games in which financial payoffs and 'doing the right thing' are not necessarily in conflict. We argue that behaviour is crucially linked to not only the preferences of people, but also the properties of the situation. By doing so, we are able to provide a road map of the psychological and economic properties of people and situations that might interfere with generalizability of laboratory result from a broad class of games. JEL classification: C9

A propos de la possibilité de généraliser les comportements de laboratoire à ce qui se passe sur le terrain.   Il n'y a pas de question plus fondamentale en économie expérimentale que de savoir si et en quelles circonstances les résultats de laboratoire peuvent être généralisés à ce qui se passe sur le terrain. Dans ce texte, on développe les résultats de Levitt et List (2006) pour les appliquer à une classe de jeux dans lesquels les résultats financiers et ≪faire la bonne chose≫ ne sont pas nécessairement des choix conflictuels. Le comportement n'est pas une simple question de préférences des gens mais aussi un écho des propriétés de

l'environnement. Il devient alors possible de proposer une image des propriétés psychologiques et économiques des gens et des situations qui pourraient empêcher qu'on généralise les résultats de laboratoire à une classe plus vaste de jeux.

Jane is a college sophomore with an undeclared major. On Tuesday mornings she usually scuttles through the quadrangle to her Biology 101 course, but today she has opted for a different route – one that takes her through the Social Sciences Building. Her hope is to find temporary employment to help to fund planned activities for homecoming in three weeks. Perhaps a wayward student needs tutoring or a professor needs data entry. As she scours the announcement board, one brightly coloured posting grabs her attention. It asks for student volunteers to participate in an economics experiment on decision making. No more than an hour of 'work,' and more important, subjects receive $5 merely for showing up and can earn considerably more money in the experiment. Jane jots down the particulars, as her need for cash, personal curiosity, and hope in some small way to contribute to the advancement of science draw her interest.

Three days later Jane finds herself entering an environment that is entirely foreign to her. Not that this particular room is different; it isn't, as the experimental laboratory is a room that shares similarities with many rooms on campus: chairs, tables, proper lighting, and a blackboard. Rather, the environment is different in that upon entering the room she is seated at a computer terminal where she cannot view any other participants, is told not to talk, and is asked to complete an 'informed consent' form that ensures the experimenter that she is here on her own volition and that she understands her decisions are being monitored, recorded, and subsequently scrutinized for scientific purposes. Jane realizes that she is entering a relationship that has no useful parallels in her everyday life.

Upon completing the introductory details, Jane is told that her task is to choose a number between 1 and 10. This number, coupled with the number that Jane's partner chooses, will determine how much money Jane earns. While the instructions go by quickly and Jane suspects that she is missing something, she figures that, since she is making 25 choices, she will eventually learn exactly how the game works. Besides, she wonders exactly how much money she can lose by choosing #4 rather than #7. The experimenter even reassures the group, noting that 'while this decision might seem daunting now, you will all get the hang of it after a few periods.' He is correct; after several choices, Jane begins to become comfortable with the 10×10 payoff matrix that is provided. She does wonder, however, exactly what the experimenter is attempting to learn from this experiment, and what the professor doing the research thinks are the 'correct' choices.

Jane enjoys the experience, and in the end she takes home $13 for an hour of work and also feels satisfied that she has done her part for science. In fact, the experience was most gratifying, inducing Jane to participate in future economics experiments that seem scientifically interesting and promise to pay well. In each

subsequent experiment the environment feels less foreign, but Jane nevertheless has difficulty providing a mental abstraction of an exact replicate in her extra-lab world.

Four years pass and Jane is confidently proceeding through the second year of her PhD program in economics. She is at the idea stage and voraciously consumes everything written on collusion in markets. A new empirical study has just been published that purports to show evidence that, consistent with the theoretical work beginning with Stigler (1964), firms have difficulty maintaining strict compliance with the collusive agreement in a repeated game with secret price cuts and demand uncertainty. The authors conclude the abstract by arguing that, since such conditions realistically define many arrangements, the inherent problems they found in maintaining collusive agreements precludes conspiracies from having considerable influence on prices in markets. Jane finds this fascinating, as she learned in her classes that such empirical evidence is difficult to obtain, given the multiple moving parts associated with any economy. Upon downloading the article, to her delight Jane realizes that the authors are two of her undergraduate economics professors. The article informs her that the evidence is from a laboratory experiment. Upon completing the study and carefully reading the experimental instructions the work finally hits home – these data are from Jane's very first experiment. Given the make-up of the experimental environment, Jane wonders about the applicability of her behaviour to models of firm-level collusion and the subsequent policy implications drawn.

Some economists seemingly share Jane's reservations. Given that the experimental approach has exerted an increasingly important influence on the profession of economics, questions of generalizability merit closer scrutiny.[1] In order for the laboratory to achieve its full potential as an invaluable empirical tool in economics, we need to understand when, and under what circumstances, we can reliably generalize lab results to naturally occurring markets that are of interest to economists. In those cases where generalizability to the field is not expected, a model of the nature and extent of behavioural differences across situations is necessary to calibrate behavioural departures.[2]

---

1 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, 40). In addition, 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 sceptical dissent of some of these views, see Bardsley 2005). Another example includes using the lab for methodological purposes – that is, to inform field designs by abstracting from naturally occurring confounds.

2 We define 'field' as an environment or situation (i.e., market) in the 'extra-lab' world. This is to be distinguished from 'field experiment' in that this term is used to denote a controlled

Jane's experience highlights several relevant factors that might vary across situations. First, the patterns of self-selection of participants into lab experiments might differ from those in naturally occurring environments, where individuals suffering from decision-making defects tend either to adapt, to 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, the nature and extent of scrutiny associated with the lab potentially influences behaviour. For example, there is evidence that subjects in laboratory environments might alter their actions to conform to the behaviour that they believe the experimenter desires (see, e.g., Orne 1959a, 1959b, 1962).

Third, the particulars of the lab context also potentially exert a powerful influence over individual behaviour. For example, when the game played in the lab rewards rules of thumb used by experts in natural environments, these experts typically outperform novices. But, slight alterations to the laboratory action set or optimal strategies that are not well aligned with real-world rules of thumb lead the experts to perform worse than novices. Fourth, the stakes associated with laboratory experiments might not mirror those in play in the field environment of interest. Finally, in some 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 to allow this experience to arise. Experiments typically last minutes or hours, whereas many real life decisions (e.g., effort to exert in labour market settings) are made over much longer time frames. The lab experience also suppresses learning from peers and mentors, a powerful and widely utilized strategy in naturally occurring environments.[3]

experimental exercise that has some 'field' aspect. Harrison and List (2004) propose six factors that can be used to determine the field context of an experiment: the nature of the subject pool, the nature of the information that the subjects bring to the task, the nature of the commodity, the nature of the task or trading rules applied, the nature of the stakes, and the environment in which the subjects operate. Using these factors, they discuss a broad classification scheme that helps to organize one's thoughts about the factors that might be important when moving from the lab to the field. We should stress that the relationship between the lab and the field has long been an object of study, with many different types of phrases used to characterize relevant concepts, with 'parallelism,' 'external validity,' and 'ecological validity' being the most popular. Parallelism, which we trace 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) introduce 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?' (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 behaviour in the world (Schmuckler 2001, 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).

3 Other arguments against generalizability exist. See, for example, Kagel et al. (1979), Cross (1980), Starmer (1999a, b), Bohm (2002), and Hertwig and Ortmann (2001). Harrison and List (2004) present arguments comparing data from lab and field experiments, some of which overlap with the arguments herein. The interested reader should 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.

In earlier work (Levitt and List 2006) we developed a model that explicitly allowed behaviour to be determined by both the properties of people and situations. We argued that individual choices were influenced not just by financial concerns, but by a desire to 'do the right thing.' The crux of the argument revolves around the notion that putting agents on an artificial margin limits the ability of the experimenter to make direct inference about markets of interest. We briefly review this earlier work in section 2. Using that model as a guide, Levitt and List (2006) examine the existing empirical literature on games that measure social preferences (e.g., Camerer and Weigelt 1988; Fehr et al. 1993). We argue that in settings where financial returns and 'doing the right thing' are at odds, the special features of the lab might preclude measurements of prosocial preferences to be directly applicable to other populations of people and situations.

The primary focus of this viewpoint article is to extend the analysis to the broad class of games in which 'doing the right thing' is not a primary concern (e.g., 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). We tackle this question in section 3. In situations where morality and wealth are not competing objectives, we expect that lab findings will be more readily generalizable than is the case for lab experiments that measure social preferences, but a number of obstacles to generalizability nonetheless are worthy of discussion.

Although early portions of the study focus on threats to the generalizability of laboratory experiments, section 4 highlights the many useful roles that lab experiments can serve. In particular, we argue that lab experiments can provide a crucial first understanding of qualitative effects and suggest underlying mechanisms that might be at work when certain data patterns are observed. In this spirit, even for those experiments that are affected by our concerns, it is likely that the qualitative findings of the lab are generalizable, even when the quantitative magnitudes are not. Section 5 concludes.

1. Reviewing the earlier arguments in Levitt and List (2006)

Our earlier work began with a simple theoretical model to make precise the arguments regarding potential factors that influence individual decision making. That model explicitly recognizes that any decision is determined by people's preferences and several situation-specific elements that people might find important. The model builds on the arguments of a long line of economists, including Smith (1759) and Becker (1974), who emphasize that decisions can have an impact on individual utility that goes beyond changes in wealth.[4]

---

4 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

In that model, a utility-maximizing individual is faced with a choice regarding a single action. We focused on the case in which utility is additively separable in the moral and wealth arguments; thus the utility function for when individual i takes action a is given by

$$U_i(a, v, n, s) = M_i(a, v, n, s) + W_i(a, v), \tag{1}$$

where the choice of action, a, 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. 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 that an individual views as immoral, anti-social, at odds with his or her own identity (Akerlof and Kranton 2000, 2005) may impose important costs on the decision maker (see also Gazzinga 2005). 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, perhaps owing to personal dissatisfaction or because the agent desires to be perceived by others as fair (see, e.g., Andreoni and Bernheim 2006). This moral payoff might vary across people, religions, or societies.[5]

In practice, many factors influence the moral costs associated with an action, but for modelling purposes, we focus on just three aspects of the moral determinant: the financial externality that an action imposes on others, v, the set of social norms or legal rules that govern behaviour in a particular society, n, and the effect of scrutiny, s.

makes' (Grampp 1948, 317, as cited in Ashraf, Camerer, and Loewenstein 2005). For further discussion on this issue see Meardon and Ortmann (1996), Palacios-Huerta (2003), and Ashraf et al. (2005). Becker (1974) introduced altruism as an explanation for why individuals give (or cooperate) when it is money maximizing not to give (or cooperate). Scores of economists in between have written about behaviour 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 modelling approach to the pure altruism framework is to consider moral or group-interested behaviour (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, Oberholzer-Gee, and Eichenberger (1996), Karni (1998), and Kotchen (2005), for example. Relatedly, in Bernheim's (1994) conformity model agents value status, and behavioural 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.

5 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 utility function we describe has potential relevance for a wide variety of behaviours. 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 one is with clients, on a first date, or when another diner is closely observing your actions.[6]

A number of predictions emerge from the model. 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, they will generally make different choices when faced with the same decision problem. Fourth, as the stakes of the game rise, wealth concerns will increase in importance relative to issues of fairness.

The bulk of our earlier work was devoted to examining the empirical evidence to develop an understanding of the import of these considerations for the generalizability of laboratory findings on social preferences. We concluded that the high degree of scrutiny (and the special nature of that scrutiny) might influence the measurement of social preferences in the lab, as might the typically low stakes, and the particular selection mechanisms that potentially lead to systematic differences between lab volunteers and market participants.

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

In our previous study, 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, many of these concerns become inconsequential. For instance, the large lot of experiments measuring psychological biases such as the endowment effect (see, e.g., Kahneman, Knetsch, and Thaler 1990), or the series of studies exploring bidding in auctions where the demand and cost functions are private (see, e.g., Kagel and Levin 1999). Further, the early market experiments of Chamberlain (1948) and Smith (1962) and the vast literature on market institutions that followed (e.g., List 2004a) fall under this class of experiments.

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

---

6 Conlin, Lynn, and O'Donoghue (2003) present results from an extensive data set on tipping that confirms many of these intuitions.

to these two considerations, three additional issues are relevant for this class of experiments: (1) intricately related to selection, the amount of experience one has with a task is likely to be much greater in naturally occurring environments than in the lab, (2) 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 (3) there may be group differences (e.g., male vs female) in how the lab affects behaviour and in self-selection. These five issues – which encompass the economic and psychological properties of situations and agents that we focus on in this section – are addressed in turn, with much of the discussion centred on individual bidding behaviour in auctions, since this area represents a literature that is vast and therefore presents some empirical evidence on the issues at hand.

3.1. 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 31 published laboratory experiments. Camerer and Hogarth (1999) extend Smith and Walker's survey by examining 74 experimental papers and find evidence in favour 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 behaviour than did lower-priced goods. Likewise, in 'ratchet effect' games, Cooper et al. (1999) report that higher stakes leads to faster development of strategic play.

With this empirical evidence in hand, the next step is to discuss how the lab might differ from the field across these important dimensions. While certainly there are choices made over small stakes in the field, 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 that 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. 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, List, and Price 2005 for details):

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

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 or dimes (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.

3.2. Non-random selection of participants

If participants in laboratory studies differ in systematic ways from the actors engaged in the targeted real-world settings, attempts to directly generalize lab results might be frustrated. One approach to investigating biases from subject pool differences is to examine whether professionals, or other representative agents, and students behave similarly in laboratory experiments (Harrison and List denote lab exercises with professionals as 'artefactual field experiments'). Yet it is important to understand that subjects enter the lab with rules of behaviour 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.[7] Thus, one should take care in designing such exercises and understand the field environment particulars of the professional subjects.

A related issue concerns the possibility that only certain types of participants – students or professionals – are willing to take part in the experiment. In this case, if the selection rules into the laboratory differ across subject pools, then valid inference might be frustrated. 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 decision making occurs in the field.

---

7 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, 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 are required in their laboratory game. In this sense, age by itself can be an important confound.

The selection issue is perhaps best illustrated by Crack (2000, 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 behaviour 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 overestimated the value of the object, and they correct their bids for that fact.[8] Cox, Dinkin, and Sawrthout (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 possible WC element. The worst offenders chose to stay out, largely mitigating the WC problem. We should note that Kagel and Levin (2002) argue that when looked at carefully, the Cox et al. (2001) data do not show the effects discussed, and they even argue that the selection effect is opposite to what is claimed.

Nevertheless, selection effects have been shown elsewhere as well. For example, continuing with the exploration of selection effects strictly internal to the interpretation of results from common value auctions, Casari, Ham, and Kagel (2005, 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.'

Such selection effects potentially permeate the literature outside of auctions 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. Similar selection effects are found in Rutström (1998), who explores the role of recruitment fees in securing experimental participants and finds some evidence of group differences (male vs female, white

---

8 A first illustration of the phenomenon is due to Bazerman and Samuelson (1983), who used a classroom jar experiment whereby an auction was conducted for a jar of coins worth $8.00, with prizes for the best guess. As discussed in the survey of Kagel and Levin (2002), subsequent laboratory experiments have refined the analysis and have shown that the WC is 'alive and well' across student subject pools including super-experienced' bidders (Kagel and Levin 1999) and is a robust phenomenon in many common value auction forms (see, e.g., Kagel, Battalio, and Walker 1995). Charness and Levin (2005, 4-5) note, 'It seems that presently only a few economists doubt the existence of WC-type of behavior in the lab and more and more are willing to accept that such behavior may indeed exist outside the lab in real markets.'

vs non-white) in bidding in auctions in some of the treatments and no differences in other treatments.

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. Andersen et al. (2005) illustrate this point by considering preferences over risk and time. They find that there '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' (1).

3.3. 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 many cases, work towards eliminating decision-making defects.[9] Of course, this intuition is critical not of lab experiments per se, but rather of their typical implementation. For example, as discussed earlier, it is not difficult to gather a pool of sophisticated and well-practised agents as experimental participants. Harrison and List (2005), for instance, examine the behaviour 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, Dinkin, and Sawrthout (2001), is consonant with the notion that agents with decision-making defects tend either to adapt, disappear from the market, or never enter the market at all.[10]

---

9   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 decision making, 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, Haigh, and List 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 non-existent or too weak to signal to the individual that his behaviour 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.

10 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

Similarly, there is lab and field experimental evidence that is consistent with the hypothesis that the endowment effect (and the willingness to pay – WTP – and willingness to accept (WTA) disparity) shrinks with market experience (e.g., Knez, Smith, and Williams 1985; Coursey, Hovis, and Schulze 1987; Brookshire and Coursey 1987; Myagkov and Plott 1997; List 2003, 2004b).[11] 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 behaviour 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 behaviour across investors and owner-occupants is different: owner-occupants exhibit about *twice* the degree of loss aversion compared with investors. Perhaps providing a cleaner result for the purposes herein, Shapira and Venezia (2005) analyse 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 behaviour than their less successful rivals.

Such results have led even noted sceptics of neoclassical theory to conclude that market experience can eliminate disparities between WTA and WTP (see, e.g., Loomes, Starmer, and Sugden 2005). One interesting data pattern found in the field experiments is that behaviour among the least experienced agents is similar to individual behaviour 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.[12]

---

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. This also highlights important effects that are outside the theory.

11 Although we do not formally discuss contextual effects in this section, they certainly can be quite important. In the WTA/WTP literature, for example, Plott and Zeiler (2005) reveal the strength of context in experimental design. Perfectly consistent with our model, their results suggest that experimenters can eliminate or create the value disparity by modifying experimental procedures.

12 There are two standard counter-arguments to the points we make above. First, 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, 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.

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 less experienced bidders. In sum, experience influences behaviour, leading us to conclude that if real market experience is important for the case at hand, then student behaviour might not provide an adequate ideal for extrapolation purposes.

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, for that matter how it influences equilibrium outcomes in any naturally occurring settings. 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.

3.4. Short-run effects versus long-run effects

In most cases laboratory experiments are designed to last no more than a few hours. Yet inference is often 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 labour market predictions of Akerlof (1982) on gift exchange.

One result from the psychology literature is that there are important behavioural differences between short-run (*hot*) and long-run (*cold*) decision making. In the hot phase, visceral factors and emotions might prove quite important, whereas in the cold phase immediate reactions are more carefully suppressed. In this sense, the hot/cold settings can lead to much different behaviours (see, e.g., Loewenstein and Schkade 1999). Loewenstein (2005) reviews some of the empirical evidence on behavioural differences across cold and hot states.[13]

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 labour market. Gneezy and List (2006) find that worker effort in the

---

13 Likewise, whether real effort decisions are important remains an open empirical question.

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 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.

3.5. Endogenous institutions

In addition to these factors, there are also other forces at work in naturally oc-curring markets that are absent in the lab. As Glaeser (2004) notes, it may be in the best interests of sophisticated agents to design institutions in such a way as to exploit the anomalous tendencies of others with whom they interact. Della Vigna and Malmendier (2006) provide evidence that health clubs structure fees in a man-ner that capitalizes on the overly optimistic expectations of potential customers. Levitt (2004) similarly shows that bookmakers set lines that take advantage of the inherent biases of bettors.

The impact of endogenously generated institutions on the amount of anoma-lous behavior is ambiguous. 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 that 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 (of course, one must be cognizant of the fact that if feedback mechanisms are weak in the field such effects will generally not be observed). Relatedly, in important cases third-parties might 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).

Yet it is possible that anomalous behaviour is even greater in the field than in the lab, owing to endogenously generated institutions. First, since the latter environment is chosen by the experimenter, whose interests and goals are not nec-essarily similar to those of entrepreneurs in the field who endogenously generate profit opportunities, the lab might yield less anomalous behaviour. 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 behaviours in markets by creating extractive institutions, whereas such pressures are usually absent in 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 behaviour of the naïve types, causing the aggregate market outcome to look more 'behavioural' than the lab outcomes. The intuition is that in this case the naïve types are forgoing relatively little utility and therefore have less reason to adapt their behaviour or drop out of the market. Fehr and

Tyran (2006) present a nice discussion of factors affecting the aggregation of individual decisions to the market level.

3.6. 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' behaviour relative to the wealth-maximizing behaviour. 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; 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, Katuscak, and Ozdenoren (2005) provide some insights in their study of bidding behaviour 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.

4. 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 great value of lab experiments. Before doing so, it is worthwhile to again note that we have taken as given that the issue of generalizability is important. As mentioned earlier, this does not suggest that we think

generalizability is the trump card when considering the value of every experiment. In some cases, such as when testing a general theory, generalizability might not be of first-rate importance. In fact, as a first test of theory an experimenter might wish to create an artificial environment for its own purpose: to create a 'clean' test of said theory.

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 behaviour.

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, List, and McConnell (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, analysing the demand behaviour from the sale of a private good (strawberries) in an actual real-world field setting and in a laboratory auction setting, Brookshire, Coursey, and Schulze (1987) report that the two settings yielded similar demand behaviour.

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, Niederle, and Roth (2005) find similar behavioural patterns, as documented in the field by Niederle and Roth (2003). In the area of social dilemmas, experimenters have provided evidence that negative reciprocity is a more powerful force than positive reciprocity (see, e.g., Fehr and Gaechter 2000), a finding that we trust will manifest itself in broader field applications when the important lab conditions are fulfilled. 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 favourable 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.[14]

In this light, laboratory exercises facilitate an important dialog between theorists and experimentalists. Alvin Roth highlighted this virtue two decades ago at the symposium on experimental economics at the 5th World Congress of

---

14 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). This work has its roots in the cross-situational consistency debated discussed in Mischel (1968) and Ross and Nisbett (1991).

the Econometric Society, when he correctly noted that one of the important contributions of experimentalists is to 'speak to theorists' (see Roth 1987). Samuelson (2005) provides an excellent overview of the relationship between economic theory and experiments and how they work together to advance scientific knowledge (as do Roth 1988, and Crawford 1997). In this spirit, the interested reader should see Levin and Smith (1994), which is an example of 'theory to experiment,' and Kagel and Levin (1999), which represents 'experiment to theory.'

We also view laboratory experimentation as a useful first step in the area of policymaking. As Plott (1997, 606) notes, experimental economics is an 'inexpensive, and fast method for obtaining data on how various types of auctions might perform.' Plott (2001) provides several seminal examples – from designing spectrum auctions to designing procedures to allocate access to railroad tracks – that link laboratory experiments with policy applications. As Samuelson (2005) notes, other applications also abound, from designing multi-unit auctions to allocating airport take-off and landing slots. Similarly, the lab can be used to rank mechanisms within broad areas, such as charitable fundraising. In this spirit, Rondeau and List (2005) explore the efficacy of different fund-raising schemes and find that the lab and field yield identical rankings of the four mechanisms that they study.

Finally, although rarely discussed, one of the most important values of lab experiments is observability. Beyond measuring treatment effects, such observability permits a crisp view of the decision-making process at work. In many field settings, process is somewhere between difficult and impossible to study, but the lab permits a glimpse of not only what decisions are made, but how and why they are made. In this light, lab experiments can suggest underlying mechanisms that might be at work when certain data patterns are observed and provide insights into what can happen in other related settings.

5. Concluding remarks

This study summarizes a set of factors that we believe are important properties of the environment that should be considered when developing models of behavioral consistency across situations. These factors are salient features of both the economic and the psychological situations, as well as specifics of the actual agents. While there is a rich tradition in various fields studying and assessing people (see, e.g., Wiggins 1973), there is a clear lack of knowledge concerning the nature and extent of how situation specific factors influence behaviour. Our approach to assess the properties of the situation is to explore, both theoretically and empirically, how individual behaviour changes across judiciously chosen levels of these factors, as moderated by both the task and the agent type. Until this bridge is built between the lab and the field, any argument concerning behavioural consistency might be considered premature.

Our model predicts that for any given individual, behaviour will converge across situations as the economically and psychologically relevant factors converge. Yet unless considerable changes are made in the manner in which we conduct lab experiments, our model highlights that the relevant factors will rarely converge across the lab and many field settings. Jane's experience summarized in the introduction clearly makes this point. Thus, what is necessary are a model and a set of empirical estimates to inform us of when and where we should expect lab behaviour to be similar to a particular field environment and, alternatively, when we should expect large differences. In the latter case, at the least we should be armed with intuition as to why the empirical estimate is biased upwards or downwards.

Our goal in this viewpoint article was to summarize, in a provocative manner, some of the important factors at work when extrapolating results from laboratory experiments to the field. Even in the worst case, we conclude that there remains an important role for traditional laboratory experiments in economics. 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.[15] In the best cases carefully designed experiments asking questions most suited to laboratory analysis can, of course, go well beyond this role.

Clearly, *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, '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.

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.

---

15 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.

References

Akerlof, George A. (1982) 'Labor contracts as partial gift exchange,' *Quarterly Journal of Economics* 97, 543–69

Akerlof, George A., and Rachel E. Kranton (2005) 'Identity and the economics of organizations,' *Journal of Economic Perspectives* 19, 9–32

— (2000) 'Economics and identity,' *Quarterly Journal of Economics* 115, 715–53

Alevy, Jonathan E., Michael S. Haigh, and John A. List (2006) 'Information cascades: evidence from a field experiment with financial market professionals,' *Journal of Finance*, forthcoming

Andersen, Steffen, Glenn W. Harrison, Morten Igel Lau, and E. Elisabet Rustrröm (2004) 'Preference heterogeneity in experiments: comparing the lab and field,' Working Paper 04-21, Department of Economics, College of Business Administration, University of Central Florida

Andreoni, James, and B.D. Bernheim (2006) 'Social image and the 50-50 norm: a theoretical and experimental analysis of audience effects,' working paper, Stanford University.

Ashraf, Nava, Colin F. Camerer, and George Loewenstein (2005) 'Adam Smith, behavioral economist,' *Journal of Economic Perspectives* 19, 131–145

Bardsley, Nicholas (2005) 'Experimental economics and the artificiality of alteration,' *Journal of Economic Methodology* 12, 239–51

Bazerman, M.H., and W.F. Samuelson (1983) 'I won the auction but don't want the prize,' *Journal of Conflict Resolution* 27, 618–34

Becker, Gary S. (1974) 'A theory of social interactions,' *Journal of Political Economy* 82, 1063–93

Benz, Matthias, and Stephan Meier (2006) 'Do people behave in experiments as in real life? Institute for Empirical Research in Economics – Working Papers iewwp248

Bernheim, B.D. (1994) 'A theory of conformity,' *Journal of Political Economy* 102, 841–77

Bohm, Peter (2002) 'Pitfalls in experimental economics,' in *Experimental Economics: Financial Markets, Auctions, and Decision Making*, ed. Fredrik Andersson and Hakan Holm (Norwell, MA: Kluwer Academic)

Brookshire, David, and Don Coursey (1987) 'Measuring the value of a public good: an empirical comparison of elicitation procedures,' *American Economic Review* 77, 554–66

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, 239–50

Brunswik, Egon (1955) 'Representative design and probabilistic theory in a functional psychology,' *Psychological Review* 62, 193–217

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

Burns, Penny (1987) 'Experience and decision making: a comparison of students and businessmen in a simulated progressive auction,' in *Research in Experimental Economics*, ed. Vernon L. Smith (Greenwich, CT: JAI Press)

Camerer, Colin F. (1987) 'Do biases in probability judgment matter in markets? experimental evidence,' *American Economic Review* 77(5): 981–97

— (1992) 'The rationality of prices and volume in experimental markets,' *Organizational Behavior and Human Decision Processes* 51, 237–72

Camerer, Colin F., and Robin M. Hogarth (1999) 'The effects of financial incentives in experiments: a review and capital-labor-production framework,' *Journal of Risk and Uncertainty* 19, 7–42

Camerer, Colin F., and Keith Weigelt (1988) 'Experimental tests of a sequential equilibrium reputation model,' *Econometrica* 56, 1–36

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

Casari, Marco, John C. Ham, and John H. Kagel (2005) 'Selection bias, demographic effects and ability effects in common value auction experiments,' working paper, Ohio State University

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

Charness, Gary, and Dan Levin (2005) 'The origin of the Winner's Curse: a laboratory study,' UCSB Department of Economics

Chen, Yan, Peter Katuscak, and Emre Ozdenoren (2005) 'Why can't a woman bid more like a man?' Working paper

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

Conlin, Michael, Michael Lynn, and Ted O'Donoghue (2003) 'The norm of restaurant tipping,' *Journal of Economic Behavior and Organization* 52, 297–321

Cooper, David J., John H. Kagel, W. Lo, and Q.L. Gu (1999) 'Gaming against managers in incentive systems: experimental results with Chinese students and Chinese managers,' *American Economic Review* 89, 781–804

Coursey, Don, John Hovis, and William Schulze (1987) 'The disparity between willingness to accept and willingness to pay measures of value,' *Quarterly Journal of Economics* 102, 679–90

Cox, James C., Sam Dinkin, and James T. Sawrthout (2001) 'Endogenous entry and exit in common value auctions,' *Experimental Economics* 4, 163–81

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

Crawford, Vincent P. (1997) 'Theory and experiment in the analysis of strategic interaction,' in *Advances in Economics and Econometrics: Theory and Applications*, Vol. 1, ed. Davis M. Kreps and Kenneth F. Wallis (Cambridge: Cambridge University Press)

Croson, Rachel, and Uri Gneezy (2005) 'Gender differences in preferences: a review of economics experiments,' working paper

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

DellaVigna, Stefano, and Ulrike Malmendier (2006) 'Paying not to go to the gym,' *American Economic Review* 96, 694–719

DeMarchi, G.W., and J.E. Tong (1972) 'Menstrual, diurnal, and activation effects on the resolution of temporally paired flashes,' *Psychophysiology* 9, 31

Doty, Richard L., and Colin Silverthorne (1975) 'Influence of menstrual cycle on volunteering behavior,' *Nature* 254, 139–40

Dyer, Douglas, and John H. Kagel (1996) 'Bidding in common value auctions: how the commercial construction industry corrects for the winner's curse,' *Management Science* 42, 1463–76

Fehr, Ernst, and Simon Gächter (2000) 'Fairness and retaliation: the economics of reciprocity,' *Journal of Economic Perspectives* 14, 159–81

Fehr, Ernst, and Jean R. Tyran (2001) 'Does money illusion matter?' *American Economic Review* 91, 1239–62

— (2006) 'Individual irrationality and aggregate outcomes,' *Journal of Economic Perspectives*, forthcoming

Fehr, Ernst, George Kirchsteiger, and Arno Riedl (1993) 'Does fairness prevent market clearing? An experimental investigation,' *Quarterly Journal of Economics* 108, 437–59

Frey, Bruno S., Felix Oberholzer-Gee, and Reiner Eichenberger (1996) 'The old lady Visits your backyard: a tale of morals and markets,' *Journal of Political Economy* 104, 1297–313

Gazzaniga, Michael S. (2005) *The Ethical Brain* (New York: Dana Press)

Genesove, David, and Christopher Mayer (2001) 'Loss aversion and seller behavior: evidence from the housing market,' *Quarterly Journal of Economics* 116, 1233–60

Glaeser, Edward (2004) 'Psychology and the market,' *American Economic Association Papers and Proceedings* 94, 408–13

Gneezy, Uri, and John A. List (2006) 'Putting behavioral economics to work: field evidence of gift exchange,' *Econometrica* 74, 1365–84

Grampp, William (1948) 'Adam Smith and the economic man,' *Journal of Political Economy* 56, 315–36

Haltiwanger, C.J., and M. Waldman (1985) 'Rational expectations and the limits of rationality: an analysis of heterogeneity,' *American Economic Review* 75, 326–40

— (1989) 'Limited rationality and strategic complements: the implications for macroeconomics,' *Quarterly Journal of Economics* 104, 463–83

Harrison, Glenn W., and John A. List (2004) 'Field experiments,' *Journal of Economic Literature* 42, 1009–55

— (2005) 'Naturally occurring markets and exogenous laboratory experiments: a case study of Winner's Curse,' working paper, UCF

Harrison, Glenn W., Morten Igel Lau, and E. Elisabet Rutström (2005) '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

Hertwig, Ralph, and Andreas Ortmann (2001) 'Experimental practices in economics: a challenge for psychologists?' *Behavioral and Brain Sciences* 24, 383–451

Horowitz, John, John A. List, and Kenneth McConnell (2006) 'Diminishing marginal value: evidence from field experiments,' *Economica*, forthcoming

Jones, Stephen (1984) *The Economics of Conformism* (Oxford and New York: Basil Blackwell)

Kagel, J.H., D. Levin, and R. Harstad (1995) 'Comparative static effects of number of bidders and public information on behavior in second-price common-value auctions,' *International Journal of Game Theory* 24, 297–319

Kagel, John H., and Dan Levin (1999) 'Common value auctions with insider information,' *Econometrica* 67, 219–38

— (2002) *Common Value Auctions and the Winner's Curse* (Princeton: Princeton University Press)

Kagel, John H., and A.E. Roth (2000) 'The dynamics of reorganization in matching markets: a laboratory experiment motivated by a natural experiment,' *Quarterly Journal of Economics* 115, 201–35

Kagel, John H., Raymond C. Battalio, and James M. Walker (1979) 'Volunteer artifacts in experiments in economics; specification of the problem and some initial data from a small-scale field experiment,' in *Research in Experimental Economics*, ed. Vernon L. Smith (Greenwich, CT: JAI Press)

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler (1990) 'Experimental tests of the endowment effect and the Coase theorem,' *Journal of Political Economy* 98, 1325–48

Karlan, Dean (2005) 'Using experimental economics to measure social capital and predict financial decisions,' *American Economic Review* 95, 1688–99

Karni, Edi (1998) 'Impartiality: definition and representation,' *Econometrica* 66, 1405–15

Knez, Peter, Vernon L. Smith, and Arlington Williams (1985) 'Individual rationality, market rationality, and value estimation,' *American Economic Review* 75, 397–402

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

Kotchen, Matthew J. (2005) 'Impure public goods and the comparative statistics of environmentally friendly consumption,' *Journal of Environmental Economics and Management* 49, 281–300

Laffont, Jean-Jacques (1975) 'Macroeconomic constraints, economic efficiency and ethics: an introduction to Kantian economics,' *Economica* 42, 430–7

Lange, Andreas, John A. List, and Michael K. Price (2005) 'Auctions with resale when private values are uncertain: theory and empirical evidence,' NBER Working Paper No. 10639

Levine, Michael E., and Charles R. Plott (1977) 'Agenda influence and its implications,' *Virginia Law Review* 63, 561–604

Levitt, Steven D. (2004) 'Why are gambling markets organized so differently from financial markets?' *Economic Journal* 114, 223–46

Levitt, Steven D., and John A. List 'What do laboratory experiments measuring social preferences tell us about the real world?' *Journal of Economics Perspectives*, forthcoming

List, John A. (2003) 'Does market experience eliminate market anomalies?' *Quarterly Journal of Economics* 118, 41–71

— (2004a) 'Testing neoclassical competitive theory in multi-lateral decentralized markets,' *Journal of Political Economy* 112, 1131–56

— (2004b) 'Neoclassical theory versus prospect theory: evidence from the field,' *Econometrica* 72, 615–25

List, John A., and David Lucking-Reiley (2000) 'Demand reduction in a multi-unit auction: evidence from a sportscard field experiment,' *American Economic Review* 90, 961–72

Locke, Peter R., and Steven C. Mann (2000) 'Do professional traders exhibit loss realization aversion?' Working paper, Texas Christian University

Loewenstein, George (2005) 'Hot-cold empathy gaps and medical decision-making,' *Health Psychology* 24 (4 Suppl), S49–S56

Loewenstein, George, and David Schkade (1999) 'Wouldn't it be nice? Predicting future feelings,' in *Well Being: The Foundations of Hedonic Psychology*, ed. D. Kahneman, E. Diener, and N. Schwartz (New York: Russell Sage Foundation)

Loomes, Graham, Chris Starmer, and Robert Sugden (2005) 'Preference reversals and disparities between willingness to pay and willingness to accept in repeated markets,' working paper, University of Nottingham

Margolis, Howard (1982) *Selfishness, Altruism, and Rationality* (Chicago: University of Chicago Press)

McKinney, Nicholas, Muriel Niederle, and Alvin E. Roth (2005) 'The collapse of a medical labor clearinghouse (and why such failures are rare),' *American Economic Review* 95, 878–89

Meardon, Stephen J., and Andreas Ortmann (1996) 'Self-command in Adam Smith's *Theory of Moral Sentiments:* a game-theoretic reinterpretation,' *Rationality and Society* 8, 57–80

Mischel, Walter (1968) *Personality and Assessment* (New York: Wiley)

Mook, D.G. (1983) 'In defense of external invalidity,' *American Psychologist* 38, 379–87

Moos, R.H., et al. (1969) 'Fluctuations in symptoms and moods during the menstrual cycle,' *Journal of Psychosomatic Research* 13, 37–44

Myagkov, Mikhail, and Charles Plott (1997) 'Exchange economies and loss exposure: experiments exploring prospect theory and competitive equilibria in market environments,' *American Economic Review* 87, 801–28

Niederle, Muriel, and Alvin E. Roth (2003) 'Unraveling reduces mobility in a labor market: gastroenterology with and without a centralized match,' *Journal of Political Economy* 6, 1342–52

Orne, Martin T. (1959a) 'The demand characteristics of an experimental design and their implications,' paper read at the American Psychological Association, Cincinnati, Ohio

—(1959b) 'The nature of hypnosis: artifact and essence,' *Journal of Abnormal and Social Psychology 58*, 277–99

— (1962) 'On the social psychological experiment: with particular reference to demand characteristics and their implications,' *American Psychologist* 17, 776–83

Ortmann, Andreas (2003) 'Charles R. Plott's "Collected Papers on the Experimental Foundations of Economic and Political Science,"' *Journal of Economic Psychology* 24, 555–75

Palacios-Huerta, Ignacio (2003) 'Time-inconsistent preferences in Adam Smith and David Hume,' *History of Political Economy* 35, 241–68

Parlee, M.B. (1973) 'The premenstrual syndrome,' *Psychological Bulletin* 80, 454–65

Plott, Charles R. (1982) 'Industrial organization theory and experimental economics,' *Journal of Economic Literature* 20, 1485–527

— (1997) 'Laboratory experimental testbeds: application to the PCS auction,' *Journal of Economics & Management Strategy* 6, 605–38

Plott, Charles R., and Kathryn Zeiler (2004) 'The willingness to pay – willingness to accept gap, the endowment effect, subject misconceptions, and experimental procedures for eliciting valuations,' *American Economic Review* 95, 530–45

Pruitt, Dean G., and Melvin J. Kimmel (1977) 'Twenty years of experimental gaming: critique, synthesis, and suggestions for the future,' *Annual Revue of Psychology* 28, 363–92

Rapoport, A. (1970) 'Conflict resolution in the light of game theory and beyond,' in *The Structure of Conflict*, ed. P. Swingle (New York: Academic Press)

Rondeau, Daniel, and John A. List (2005) 'Exploring the demand side of charitable fundraising: evidence from field and laboratory experiments,' working paper

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

Roth, Alvin E. (1987) 'Laboratory experimentation in economics,' in *Advances in Economic Theory: Fifth World Congress of the Econometric Society*, ed. Truman Bewley (Cambridge: Cambridge University Press)

— (1988) 'Laboratory experimentation in economics: a methodological overview,' *Economic Journal* 98, 974–1031

— (1991) '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* 81, 415–40

Rutstrom, Elisabet (1998) 'Home-grown values and incentive compatible auction design,' *International Journal of Game Theory* 27, 427–41

Samuelson, Larry (2004) 'Economic theory and experimental economics,' *Journal of Economic Literature* 43, 65–107

Schmuckler, M.A. (2001) 'What is ecological validity? A dimensional analysis,' *Infancy* 2, 419–36

Schram, Arthur (2005) 'Artificiality: the tension between internal and external validity in economic experiments,' *Journal of Economic Methodology* 12, 225–37

Sen, Amartya K. (1977) 'Rational fools: a critique of the behavioral foundations of economic theory,' *Journal of Philosophy and Public Affairs* 6, 317–44

Shapira, Zur, and Itzhak Venezia (2000) 'Patterns of behavior of professionally managed and independent investors,' working paper, New York University

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

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

Smith, Vernon L. (1962) 'An experimental study of competitive market behavior,' *Journal of Political Economy* 70, 111–37

—(1982) 'Microeconomic systems as an experimental science,' *American Economic Review* 72, 923–55

Smith, Vernon L., and James M. Walker (1993) 'Rewards, experience and decision costs in first price auctions,' *Economic Inquiry* 31, 237–45

Starmer, Chris (1999a) 'Experimental economics: hard science or wasteful tinkering?' *Economic Journal* 109, F5–F15

—(1999b) 'Experiments in economics: should we trust the dismal scientists in white coats?' *Journal of Economic Methodology* 6, 1–30

Steele, C.M. (1997) 'A threat in the air: how stereotypes shape intellectual identity and performance,' *American Psychologist* 52, 613–29

Steele, C.M., and J. Aronson (1995) 'Stereotype vulnerability and the intellectual test performance of African Americans,' *Journal of Personality and Social Psychology* 69, 797–811

Stigler, George (1964) 'A theory of oligopoly,' *Journal of Political Economy* 72, 44–61

— (1981) 'Economics or ethics?' in *Tanner Lectures on Human Values*, ed. S. McMurrin (Cambridge: Cambridge University Press)

Sugden, Robert (1984) 'Reciprocity: the supply of public goods through voluntary contributions,' *Economic Journal* 94, 772–87

Sutherland, H., and Lance I. Stewart (1965) 'A critical analysis of the premenstrual syndrome,' *Lancet* 5, 1180–3

Thaler, Richard H. (1992) *The Winner's Curse: Paradoxes and Anomalies of Economic Life* (New York: Free Press)

Wiggins, J.S. (1973) *Personality and Prediction: Principles of Personality Assessment* (Reading, MA: Addison-Wesley)

Wilde, Louis (1981) 'On the use of laboratory experiments in economics,' in *The Philosophy of Economics*, ed. Joseph Pitt (Dordrecht: Reidel)