

The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions

John A. List

University of Chicago and National Bureau of Economic Research

The role of the market in mitigating and mediating various forms of behavior is perhaps the central issue facing behavioral economics today. This study designs a field experiment that is explicitly linked to a controlled laboratory experiment to examine whether, and to what extent, social preferences influence outcomes in actual market transactions. While agents drawn from a well-functioning marketplace behave in accord with social preference models in tightly controlled laboratory experiments, when they are observed in their naturally occurring settings, their behavior approaches what is predicted by self-interest theory. In the limit, much of the observed behavior in the marketplace that is consistent with social preferences is due to reputational concerns: suppliers who expect to have future interactions with buyers provide higher product quality only when the buyer can verify quality via a third-party certifier. The data also speak to theories of how reputation effects enhance market performance. In particular, reputation and the monitoring of quality are found to be *complements*, and findings suggest that the private market can solve the lemons problem through third-party verification.

Orley Ashenfelter, Raymond Battalio, Roland Benabou, Daniel Benjamin, Gary Charness, Armin Falk, Edward Glaeser, Uri Gneezy, Glenn Harrison, Daniel Kahneman, Liesl Koch, David Laibson, Steven Levitt, Derek Neal, Matthew Rabin, and Al Roth provided remarks on an earlier version of this study that considerably improved the paper. Seminar participants at the University of Chicago (Graduate School of Business and the Economics Department), University of California at Berkeley, Cornell University, Harvard University, Princeton University, Texas A&M, University of Texas at Austin, and the University of Wisconsin provided comments that helped to shape the paper. Thanks to Michael Price for research assistance.

[*Journal of Political Economy*, 2006, vol. 114, no. 1]
© 2006 by The University of Chicago. All rights reserved. 0022-3808/2006/11401-0001\$10.00

I. Introduction

More than two decades ago, Stigler (1981, 176) wrote that when “self-interest and ethical values with wide verbal allegiance are in conflict, much of the time, most of the time in fact, self-interest theory . . . will win.” While this is the conventional wisdom among economists, an influential collection of laboratory experiments on “gift exchange” has called into question the validity of Stigler’s position (see, e.g., Camerer and Weigelt 1988; Fehr, Kirchsteiger, and Riedl 1993; Berg, Dickhaut, and McCabe 1995). This literature is complemented by an entire body of theoretical research exploring the economic consequences of “social preferences,” wherein agents have preferences that are measured over their own and others’ material payoffs (for models of reciprocity, see Rabin [1993], Charness and Rabin [2002], Dufwenberg and Kirchsteiger [2004], and Falk and Fischbacher [forthcoming]; for models of inequity aversion, see Fehr and Schmidt [1999] and Bolton and Ockenfels [2000]; on altruism, see Andreoni and Miller [2002]).¹ A second generation of experimental studies has subsequently emerged examining the nature of social preferences and underscoring the robustness of the gift exchange results (e.g., Charness 1996; Fehr, Gächter, and Kirchsteiger 1997; Fehr and Falk 1999; Charness and Rabin 2002; Gächter and Falk 2002; Hannan, Kagel, and Moser 2002; Brown, Falk, and Fehr 2004; Fehr and List 2004).²

The gift exchange results, which are consistent with the notion that people behave in a reciprocal manner even when the behavior is costly and yields neither present nor future material rewards, have attracted much attention, since many have argued that they are relevant beyond the context inherent in the laboratory. For example, some view the experimental results as providing key support for the labor market predictions in Akerlof (1982) and Akerlof and Yellen (1988, 1990), whereby higher than market-clearing wages and involuntary unemployment are potential outcomes of fairness considerations in the workplace.³ Indeed, Fehr et al. (1993, 437) note that their results “provide . . . experimental

¹ In this study, I explore social preferences under this broad definition and am not interested in pinpointing whether the behavior consistent with social preferences is due to altruism, reciprocity, fairness, inequality aversion, or another motive. Yet within the gift exchange literature, reciprocity motives have been highlighted; thus I shall continue this spirit in the discussion below. For a parsing of trust and reciprocity in a laboratory experiment, see Cox (2004).

² Fehr and Gächter (2000) provide an overview. The interested reader should also see the related literature on “lemons” markets (e.g., Miller and Plott 1985; Holt and Sherman 1990).

³ This conjecture is typically termed the “fair wage–effort” hypothesis. Alternatively, note that the “efficiency wage theory” surmises that wages above market-clearing levels occur because these wage profiles induce workers to be motivated in an effort to avoid being fired, which economizes on firm-level monitoring (see, e.g., Katz 1986).

support for the fair wage–effort theory of involuntary unemployment.” Of course, social preferences might be important in many other strategic situations as well (for overviews, see, e.g., Sobel [2002] and Camerer [2003]), and therefore such results have broad implications for economists and noneconomists alike. Despite these advances and the topic’s importance, it is fair to say that little is known about whether, and to what extent, social preferences influence economic outcomes in naturally occurring markets.⁴

The major goals of this study are to explore the nature of such preferences among real-market players in both the laboratory and naturally occurring environments. In doing so, the study provides a framework with which to disentangle social preferences and reputation effects. Measuring and disentangling social preferences and reputation effects are important in both a positive and a normative sense, since optimal contracting and proposed government intervention in principal-agent settings, appropriate design of collective choice mechanisms, and theory testing all depend critically on proper measurement of these effects. Equally important, the experimental design permits an examination of whether individual behavior in laboratory experiments provides a reliable indicator of behavior in the field—an issue fundamental to experimental economics.

To complete these tasks, I use several distinct experimental treatments to create a bridge between the laboratory and the field. A major attraction of this approach is that if behavioral differences are observed, I can pinpoint the important factors driving the disparities. I begin with a gift exchange laboratory treatment (which is in effect a sequential prisoner’s dilemma game) that closely follows the received literature (e.g., Fehr et al. 1993). Rather than using student subjects, however, I make use of subjects drawn from a well-functioning marketplace—the sports card market. In this setting, I place the experimental participants in their typical roles: consumers are placed in the role of buyers and dealers are placed in the role of sellers. This treatment is potentially important in that a fundamental feature of markets is that they sort agents into roles, whereas laboratory experiments randomly assign positions to agents. Experimental results are nevertheless consistent with those in the literature that uses students randomly allocated to roles: gift exchange is observed, and such behavior has an important influence

⁴ There is some survey evidence reported from interviews with managers that social preference considerations are important in the workplace (Blinder and Choi 1990; Bewley 1995). Furthermore, in a novel paper exploring the role of fairness in the marketplace, Kahneman, Knetsch, and Thaler (1986) report results from telephone surveys of residents of two Canadian metropolitan areas (Toronto and Vancouver). They use a “dual entitlement” theory to explain their data: previous transactions establish a reference level of consumer and producer surplus, and fairness considerations arise from outcomes relative to these “entitlements.”

on economic outcomes. This finding provides a validity check of the extant laboratory results on gift exchange, since it suggests that the major results can be replicated with real economic players from a much different population.

I proceed to explore several further treatments in the laboratory by varying the “context” in the experimental instructions; previous experimental studies typically use “context-free” instructions, such as neutral wording and the avoidance of words that might provide familiar contextual cues. Of course, this traditional approach potentially attenuates important elements of the exchange process and therefore may suppress important psychological effects. A final laboratory treatment in this spirit moves a step toward the naturally occurring marketplace by creating an experimental lab market in which buyers and sellers play a sequential prisoner’s dilemma game by exchanging cash for goods of uncertain quality in face-to-face transactions. If one ignores the artificiality invoked by the laboratory experimental setting, this particular treatment provides an environment that mirrors the actual decision-making process in the marketplace from which these subjects are drawn. As a whole, these design changes yield some behavioral differences, but gift exchange in these settings remains alive and well, both statistically and economically.

When one moves from the lab to the field, an important consideration is to remain parallel to the important lab features while ensuring that the transaction is a natural one in the field. The field experimental treatments mirror the laboratory gift exchange treatments and resemble many types of markets for goods or services: after receiving a price offer, sellers determine the good’s quality, which cannot be perfectly measured by buyers. In the first field treatment, subjects approach dealers (who are unaware that they are taking part in an experiment) and offer either \$20 or \$65 for a sports card of certain quality. Since quality is difficult for untrained consumers to detect in this market and the approached dealers have a sufficient stock of cards on hand to provide the requested quality levels, if social preferences play a role in this case, the card’s grade and the price offer should be positively correlated. Once the buying agents had purchased each of the cards from the dealers, I had every card professionally graded. I do find a positive correlation between the prices and grades received, but only among dealers who are “locals”; among dealers who are likely to have little future interaction with the buying agents (“nonlocals”), no such relationship emerges.

This result is interesting, but the data do not allow an unequivocal insight into the underlying mechanism at work. For instance, such data patterns might be due to several factors, including two competing alternatives critical to the issue at hand: selection effects—local dealers have social preferences and nonlocal dealers do not—or reputation

effects—local dealers are concerned with their reputations whereas non-local dealers are not. A final set of three treatments in the marketplace provide insights into what is driving these behavioral differences by examining outcomes in an identical experiment for collector tickets and ticket stubs. Tickets and ticket stubs provide a unique test because *no* third-party verification service existed to grade tickets until June 2003, though the major grading company announced in April 2003 that it would soon begin grading tickets. By comparing temporal outcomes, not only am I permitted a unique opportunity to examine the nature of market exchanges with and without third-party enforcement, but I am also able to explore the role of social preferences in such settings.⁵

In stark contrast to the results obtained from the sports card data, the empirical results in the time period in which no grading service was available and the public was unaware that a service was imminent (pre-April 2003) provide little evidence consistent with social preferences: ticket quality is not correlated with price for either dealer type, and local and nonlocal dealers provide similar quality levels. One could reason that dealers had little idea how to grade tickets since they had never been professionally graded (though many dealers made quality claims), and therefore the inability of this treatment to reject the homogeneity null is consistent with informational problems.

This potential drawback is rectified in an experimental treatment conducted after the announcement of grading (April 2003) but before the grading company released its grading criteria (June 2003). Purchasing identical tickets and using analogous protocol, I find that during this time period quality and price are correlated for tickets sold by local dealers, but no correlation is present in ticket sales among nonlocal dealers. Completing the experimental design is an identical treatment conducted after grading services commenced (post-June 2003). Insights gained from this treatment are quite similar to those obtained from the treatment conducted between April and June: gift exchange is evident among local dealers but not among nonlocals. This result stands to reason because Professional Sports Authenticators' (PSA's) ticket grading criteria are similar to its scheme for grading sports cards.

In summary, several insights follow. First, even though the data collected from one-shot laboratory experiments suggest that social preferences are quite important among these agents, parallel treatments in

⁵ Brown et al. (2004, 751) summarize the attractiveness of such treatments in motivating their laboratory experiments by noting that "the ideal data set for studying the effects of the absence of third party enforceability on market interactions . . . is based on a truly exogenous *ceteris paribus* variation in the degree of third party enforceability. . . . The problem is, however, that it seems almost impossible to find or generate field data that approximates this ideal data set." This is exactly what these three treatments offer, and to the best of my knowledge such exogeneity has not heretofore been achieved in the literature.

the field suggest that such effects have minimal influence in naturally occurring transactions. In this sense, dealer behavior in the marketplace approaches what is predicted by self-interest theory. From a methodological viewpoint, it is important to note that several changes to the laboratory environment had little influence on behavior, whereas moving from the lab to the field had striking effects. Second, empirical results provide insights into how reputation effects and professional certification influence market performance (see, e.g., Akerlof 1970; Klein and Leffler 1981). For example, I find that (i) reputation effects enhance the quality of goods, and (ii) reputation and the monitoring of quality are *complements*. In this spirit, the data suggest that the private market can solve the lemons problem through third-party verification.

The remainder of this study is organized as follows. Section II describes the experimental design and summarizes the institutional details of the market. Section III provides a summary of the empirical findings, highlighting differences in results across the various treatments and describing the effects of reputation and social preferences on market outcomes across both local and nonlocal dealers. Section IV concludes with a more general discussion of the empirical results.

II. Experimental Design and Hypotheses

The experimental investigation begins with an examination of behavior in standard laboratory gift exchange games. Treatment Lab-R (R denotes laboratory replication; see table 1 for a summary of the experimental design) makes use of the typical gift exchange experimental design.⁶ One session was run in this treatment. In this session, each participant's experience typically followed four steps: (1) consideration of the invitation to participate in an experiment, (2) a session to learn the experimental rules, (3) actual participation, and (4) conclusion of the experiment and exit interview. In step 1, the monitor approached dealers on the floor of a sports card show and inquired about their interest in participating in an economics experiment that would take about an hour. If the dealer agreed, the monitor summarized the meeting time and place. A similar approach was used to recruit consumers (non-dealers).

Subjects met in a large room adjacent to the floor of the sports card show: dealers entered on one side of the room and nondealers on the other side, and a divider was in place to ensure that identities were not revealed. The session consisted of five periods, with five dealers acting as sellers and five nondealers acting as buyers. Each participant received

⁶ Appendix A in List (2005) contains a copy of the experimental instructions, which are closely related to those in Fehr et al. (1993, 1997) and Gächter and Falk (2002).

TABLE 1
EXPERIMENTAL DESIGN

	(1)	(2)	(3)
Treatment Lab	Treatment Lab-R: Replicate lab studies $N = 25$	Treatment Lab-RF: Extend to field values $N = 25$	Treatment Lab-RF1: Extend to one-shot environment $N = 27$
Treatment Lab- Context	Treatment Lab-Con- text: Adds market context $N = 32$	Treatment Lab-Mar- ket(\$20): Adds market interaction $N = 30$	Treatment Lab-Mar- ket(\$65): Adds market interaction $N = 30$
Treatment Floor (Cards)	Treatment Floor- \$20: Naturally occurring sports card market $N = 50$	Treatment Floor- \$65: Naturally occurring sports card market $N = 50$	
Treatment Floor (Tickets)	Treatment Floor- NoGrading: Naturally occurring ticket market be- fore grading was available $N = 60$	Treatment Floor- AnnounceGrading: Naturally occurring ticket market af- ter grading announcement $N = 54$	Treatment Floor- Grading: Naturally occurring ticket market when grading ser- vice was available $N = 36$

NOTE.—Each cell represents one (or two, in the case of Treatment Floor [Tickets]) unique treatment. For example, Treatment Lab-R in row 1, col. 1, denotes that 25 dealer and 25 nondealer observations were gathered to replicate the laboratory gift exchange studies in the literature.

a copy of the instructions, and to ensure common information, the monitor read the instructions aloud as the subjects followed along. The instructions noted that in each of the five periods each buyer would be paired with a different seller. In every period, the buyer determines an integer value (denoted p for price) to send to the seller and requests a specific quality of the good (denoted q , for quality request). Only the seller who is paired with the buyer is aware of these two choices. After the buyer makes these private decisions on the decision sheet, the monitor collects the sheets and walks them to the seller partners. Sellers then choose a quality level (denoted q for quality chosen), with an associated cost of quality that is increasing monotonically with product quality (denoted $c(q)$; see List [2005, app. A] for the cost of product quality parameters, which closely follow the literature). The product quality choice is revealed only to the buyer partner, and as in the literature, all choices are revealed to the monitor.

Individual p and q choices combine to determine monetary payoffs for the pair according to the following payoff functions:

$$\text{seller payoff: } \Pi_s = p - c(q),$$

$$\text{buyer payoff: } \Pi_b = (v - p)q, v = \$80, p \in [\$5, \$80], q \in [0.1, 1]. \quad (1)$$

All payoff information was common information, and before the experiment began, several hypothetical exercises were completed to ensure that everyone understood the instructions and payoff functions. Subjects were also aware that one of the five periods would be selected randomly and that that particular period would determine payoffs. After the fifth period, subjects were paid in private after they completed a survey (see List 2005, app. B).

These parameter values yield a standard prediction under the assumption of common knowledge, self-interest theory, and appropriate backward induction. Since product quality is costly, sellers will choose the minimum level ($q_{\min} = 0.1$). A buyer's best response is to choose p_{\min} , which is $p = \$5$. Thus the subgame perfect equilibrium outcome is $q^* = 0.1$ and $p^* = \$5$, with associated profits of $\Pi_s = \$5$ and $\Pi_b = \$7.50$, much less than more efficient profit levels (i.e., $p = \$30$ and $q = 0.5$ yields $\Pi_s = \$24$ and $\Pi_b = \$25$).

Previous experimental efforts have found that typically $q > q^*$ and $p > p^*$ and that $\partial q / \partial p > 0$ in a reduced-form regression model, leading authors to conclude that reciprocity is important in economic interactions. The reciprocity inference is generally traced to Rabin's (1993) model of reciprocity (Fehr et al. 1997, 839), which describes a person with positive reciprocal motives as someone who responds to acts that are perceived as kind in a kind manner, even though there is no future pecuniary gain tied to this action. For the purposes herein, the literature has taken the qualitative implications of Rabin's model as meaning that the probability of nonshirking is increasing in the level of the perceived generosity of the offer. How generous an action is perceived is a difficult question to answer, however, and is surely quite heterogeneous across agents, inducing the literature to operationalize reciprocity as meaning that $\partial q / \partial p > 0$ in a reduced-form regression model (see, e.g., Fehr et al. 1993; Gächter and Falk 2002).

In column 2 of table 1, Treatment Lab-RF (RF denotes replication with field values) simply manipulates the environment in Treatment Lab-R by setting⁷

$$\text{seller payoff: } \Pi_s = p - c(q),$$

$$\text{buyer payoff: } \Pi_b = v(q) - p, \quad p \in [\$5, \$80], \quad q \in [1, 5]. \quad (2)$$

For $c(q)$ values, I use $c(q) = \$4, \$5, \$8, \15 , and $\$50$ for $q = 1, 2, 3, 4$, and 5 ; for $v(q)$ values, I use $\$6, \$8, \$15, \30 , and $\$80$ for $q = 1, 2, 3$,

⁷ The payoff function for the buyer is now similar to the S13–S16 treatment in Fehr et al. (1997). In this case, now the price represents a pure lump-sum transfer, which differs from the earlier joint profit equation, which was characterized by price increases leading to an increase in the sum of payoffs when $q < 1$.

4, and 5 (PSA 6, 7, 8, 9, and 10).⁸ While these chosen values are admittedly only a rough estimate of the gains to trade available in this market, use of these parameters provides the necessary tension between the dominant strategy and the joint-profit maximization actions. Under this design, the Nash purely selfish prediction is $p^* = \$5$; and for sellers to send minimal card quality, $q^* = 1$. These actions result in $\Pi_s = \$1$ and $\Pi_b = \$1$. The efficient quality level is $q = 5$, which ensures a joint surplus of \$30. Note that there could be losses of up to \$74 (the buyer sends \$80 and receives the lowest-quality Frank Thomas card); as in the other induced value laboratory treatments herein—Treatments Lab and Lab-Context—after these treatments were carried out, I had subjects participate in other unrelated experiments that did not involve interaction to ensure that they would leave with positive cash balances.

Treatment Lab-RF1 (RF1 denotes replication with field values in a purely one-shot setting) is identical to Treatment Lab-RF in every manner except that it is not executed over five periods with five different partners; rather it is a one-shot game. Since, in the above treatments, by design subjects should have construed the setting as one-shot, Treatment Lab-RF and Treatment Lab-RF1 should yield similar data patterns if (i) subjects interpret Treatment Lab-RF as several one-shot games and (ii) experience does not unduly influence play. In total, Treatment Lab yields 77 data points for buyers and 77 data points for sellers in the gift exchange game.

In row 2 of table 1, Treatment Lab-Context adds context to Treatment Lab-RF1. In this case, rather than buyers and sellers transacting with abstract commodities, Treatment Lab-Context adds context that resembles the subjects' naturally occurring environment. For example, buyers make an offer to a seller to buy one 1990 Leaf Frank Thomas card, and the buyer requests a certain PSA grade. As in Treatment Lab-RF1, sellers have five PSA grades available (PSA 6, 7, 8, 9, and 10) and subsequently choose the quality of the Thomas baseball card to give the buyer if they accept the buyer's offer.⁹ Treatment Lab-Context includes 32 buyers and 32 sellers.

Completing the laboratory treatments is Treatment Lab-Market, more specifically, Treatment Lab-Market(\$20) and Treatment Lab-Market(\$65). Treatment Lab-Market is the laboratory market parallel to Treatment Lab-Context: buying agents approach dealers in the experimental market to purchase 1990 Leaf Thomas baseball cards in face-to-face transactions. Each participant's experience in Treatment Lab-Market

⁸ Please see app. C in List (2005) for a discussion of how these values were obtained.

⁹ PSA grades 6–10 were chosen because little trading of Thomas cards below PSA 6 is carried out in the actual market. Note that in this treatment I am not actually having agents transact with real commodities; rather subjects are told to act as though they are using graded Thomas cards.

followed four steps: (1) consideration of the invitation to participate in an experiment, (2) a session to learn the market rules, (3) actual market participation, and (4) conclusion of the experiment and exit interview.

In step 1, potential subjects approached the monitor's dealer table on the floor of the sports card show and inquired about purchasing late 1980s/early 1990s baseball cards displayed on the table. If the subject was a white male roughly 25 years of age, the monitor asked if he was interested in participating in an experiment that would last about 30 minutes.¹⁰ If the agent agreed to participate, the administrator explained that at a prespecified time the subject should enter an adjacent room to take part in the experiment. Directions to the room were provided, and the subject was informed that he would receive \$20 to participate in the experiment. To gather the dealer subject pool, I visited numerous dealers' tables and examined whether the dealer had a fair number (more than five) of Thomas ungraded 1990 Leaf cards for sale that were of sufficiently heterogeneous quality. If the dealer had a sufficient number, he was asked if he would like to participate in a market experiment in which he could *potentially* sell some of the Thomas cards. Directions to the room and the appropriate times to enter the room were then provided to those dealers who agreed to participate. No show-up fee was given to dealers.

Upon subjects' arrival to the experimental market, in step 2 a monitor thoroughly explained the market rules to them privately (buyers in one room and sellers in another). Consumers were informed that they would be "buyers" of 1990 Leaf Thomas baseball cards in the experiment. The agents were told that they (typically in groups of five) would enter the market and approach a prespecified dealer, who had his Thomas cards displayed on his table in the experimental market. Importantly, in the spirit of the literature that suggests that contracted negotiations can crowd out reciprocity (see, e.g., Fehr and List 2004), I was careful to instruct buying agents to avoid haggling, while keeping the transaction as natural as possible.¹¹ In practice, negotiations are typically quite short or do not occur at all in this market (see List 2004a, table 2); thus, besides realism, this approach gives social preferences their best shot,

¹⁰ Given the results in List (2004a), I wished to avoid any confounds associated with statistical discrimination in this marketplace; hence I opted to use "majority" subjects as my buying agents in all treatments. This design choice may well give social preferences their best chance since the data in my earlier paper suggest that these types of buying agents receive the best offers from dealers. Note, however, that any agent who desired to participate in an experiment was able to do so since the minority agents were asked to participate in an unrelated pilot experiment.

¹¹ Macaulay (1963, 56) reports that "detailed negotiated contracts can get in the way of creating good exchange relationships between business units." Sitkin and Roth (1993, 376) assert that "legalistic remedies can erode the interpersonal foundations of a relationship they are intended to bolster because they replace reliance on an individual's good will with objective, formal requirements."

since buying agents are signaling a fair amount of trust in the dealer when purchasing nongraded sports cards without much detailed negotiations. To ensure that buying agents did not aggressively bargain, their payoffs were not tied to quality or price; rather, they were paid \$20 for approaching two dealers. And, to maintain consistency with Treatment Lab-Context and afford the dealers reasonable price offers, the buying agent offered \$20 (or \$65) and requested a 1990 Leaf Thomas card that would merit a PSA 9 (10) if graded.

These parameter values were guided by the empirical results in Treatment Lab-Context (discussed below), current market values of sports cards, and what would be naturally demanded in this environment. First, since the average buying agent sent \$20 to dealers in Treatment Lab-Context and requested a PSA 9 Thomas card, Treatment Lab-Market(\$20) is the naturally occurring analogue. Treatment Lab-Market(\$65) used the same dealers who were visited in Treatment Lab-Market(\$20) and was identical in every sense except that in this case buying agents offered \$65 for the Thomas card and requested a PSA 10. I chose \$65 because it is roughly 33 percent greater than $c(10) = \$50$, matching the relationship of $c(9) = \$15$ and the \$20 value chosen in Treatment Lab-Market(\$20). Second, use of lower-quality card levels would have been unnatural since the bulk of demanded volume in the market is for higher-end card types, such as PSA 9 and 10.

In step 3, the buying subjects each approached one dealer in round 1. Each interaction lasted less than three minutes and resulted in the purchase of a Thomas Leaf sports card. Upon completing the transaction, the buyer departed the experimental market and physically gave the monitor the Thomas card in an adjacent room. After all transactions in round 1 were completed, buyers received instruction on which dealer to approach in round 2. Dealers were not allowed to communicate during this time period. The buying agents then reentered the experimental market and approached a different dealer for the final buying period. Every dealer was approached twice: once with an offer of \$20 and a request for a PSA 9 card, and once with an offer of \$65 and a request for a PSA 10 card. The ordering of the offers was random.

Step 4 concluded the experiment: after subjects completed a confidential survey (see List 2005, app. B), they departed. In total, I observed the behavior of 30 dealers who were each approached by two different buying agents offering either \$20 or \$65; thus I have a sample size of 60 in Treatment Lab-Market.

Following the received gift exchange literature, if social preferences play a role in this case, then the card's grade and the offer price should be positively correlated: $\partial q/\partial p > 0$ in a reduced-form regression model. Once the buying agents had purchased each of the cards in these treat-

ments, the last step was to have the cards professionally graded. This was completed by having every card graded by a PSA representative.

Treatment Floor moves the exploration out of the laboratory and into the marketplace in which these agents actually consummate business: the floor of the sports card show.¹² Treatments Floor-\$20 and Floor-\$65 represent the naturally occurring analogues to Treatment Lab-Market and are identical whenever possible. Again, the buying agent's experience typically followed four steps. In step 1, white males roughly 25 years old who were interested in late 1980s/early 1990s baseball cards were asked to participate in an experiment. If the agent agreed to participate, in step 2 a monitor thoroughly explained the experimental rules. The agent was informed that he would be a "buyer" of 1990 Leaf Thomas baseball cards in the experiment.

The agent was told that he would approach 10 different dealers on the floor of a sports card show to purchase the Thomas card. I was able to preselect the dealers to be approached before the show by visiting their dealer tables and examining whether they had more than five Thomas ungraded 1990 Leaf cards for sale that were of sufficiently heterogeneous quality. It is common practice for dealers to mill around the show looking at others' goods, and I was merely behaving in accordance with this norm when visiting dealer tables. As in Treatment Lab-Market, I was careful to instruct buying agents to avoid haggling, while keeping the transactions as natural as possible. And the buying agent offered \$20 (or \$65) and requested a 1990 Leaf Thomas card that would merit a PSA 9 (10) if graded. In this spirit, much as in Treatment Lab-Market, buying agents represented sophisticated buyers in that they expressed interest in a PSA-graded card.

In step 3, the subject approached dealers one at a time. As in Treatment Lab-Market, each interaction lasted less than three minutes and resulted in the purchase of a Leaf Thomas sports card. It should be noted that throughout the experiment the sports card dealers were not aware that an experiment was occurring. This ensured that the process was as natural as possible for the dealers, whose behavior was of primary interest in this field experiment. Step 4 concluded the experiment: after subjects completed a confidential survey, they were paid \$20 in private.

A few noteworthy design issues should be mentioned before I proceed.

¹² As I have noted elsewhere (e.g., List 2004*b*, 2004*c*), with the rise in popularity of collector sports cards and memorabilia over the past two decades, markets that organize buyers and sellers have naturally arisen. Temporal assignment of the physical marketplace is typically done by a professional association or local sports card dealer, who rents a large space, such as a gymnasium or hotel conference center, and allocates 6-foot tables to dealers for a nominal fee. When the market opens, consumers mill around the marketplace, haggling and bargaining with dealers, who have their merchandise prominently displayed on their tables. The duration of a typical sports card show is a weekend, and subjects enter the market ready to buy, sell, and trade.

First, each dealer was approached twice: once in Treatment Floor-\$20 and once in Treatment Floor-\$65. Visits were spaced so as to attenuate any suspicion: one example is that dealer i was approached by agent n on Friday night and by agent m on Sunday morning. And, the ordering of the visits was random: some dealers were approached in the \$20 treatment first, others were approached in the \$20 treatment second. I observed no ordering effect, so I suppress further discussion of this issue.

Second, in contrast to audit studies that test for market discrimination, I am directing the agent to buy the good. In this sense, these are not transactors who obliquely discontinue bargaining if the dealer accepts an offer; these are actual transactions. And, since transactions are typically in cash at sports card shows, I provided the necessary funds to purchase the cards. Third, note that great care was taken to ensure that the data were gathered from interactions that would naturally occur in the marketplace. Subjects were entering the market to buy goods that were very similar to the good that I had them buying. Fourth, Treatments Lab-Market and Floor were carried out at several different sports card shows in the same region in the United States, from October 2002 to July 2004.

In total, I observed the behavior of 50 dealers who were each visited by two different agents (one in Treatment Floor-\$20 and one in Treatment Floor-\$65); thus I have a sample size of 100 in Treatment Floor. As in Treatment Lab-Market, the last step of the experiment was to have the cards professionally graded. In addition, I should note that in every case I was able to obtain important subject-specific information from the dealers, either via a survey they completed during an experiment in which they later participated or through a survey (see List 2005, app. B) they filled out in exchange for a payment of \$1.

To explore a level deeper into the underlying structure that organizes behavior in this market and control for potential selection effects, I completed three final treatments making use of natural exogeneity that the market offered during the sample period: while a third party (PSA) has graded sports cards since 1987, no service existed prior to June 2003 to grade sporting event tickets and ticket stubs. PSA announced its grading intentions in April 2003, but it did not provide grading criteria until June 2003. As noted earlier, Brown et al. (2004) highlight the attractiveness of such natural variation by arguing that such exogeneity is impossible to find in field data. I believe that these three field experimental treatments offer this useful characteristic.

Treatment Floor-NoGrading (denotes no grading available) is identical to Treatment Floor in that buyers approached dealers on the floor of a sports card show (from October 2002 to March 2003) with a low and a high price (either \$10 or \$30) to purchase an unused ticket or

ticket stub that would receive a PSA grade of 9 or 10 if tickets were graded like sports cards. Given the thinness of the ticket market, it was necessary to use five different ticket types in the purchasing tasks (Cal Ripken's last game at Camden Yards, his final game of "the Streak," his "consecutive world-record-breaking" game, and two World Series games). I was careful to choose tickets that were in the same price range to increase the likelihood of having the luxury of pooling the data. In total, I observed the behavior of 30 dealers in this treatment and therefore gathered 60 data points since each dealer was approached twice.

Treatment Floor-AnnounceGrading (denotes after announcement of grading) was completed at sports card shows after PSA announced that it would begin grading ticket stubs (April 2003) but before it released its grading scheme (June 2003). In this treatment, I purchased the same tickets and used the same protocol as in Treatment Floor-NoGrading. As outlined in row 4 of column 2 of table 1, I observed 54 dealer decisions in this treatment.

Completing the experimental design is Treatment Floor-Grading (denotes grading available), which is identical to Treatments Floor-NoGrading and Floor-AnnounceGrading but was completed after June 2003, a time period in which grading services of tickets and ticket stubs existed. I observed 36 total dealer decisions in this final treatment. Accordingly, I purchased 150 tickets in these three treatments; and as in Treatment Floor, I subsequently had every ticket graded by a PSA representative. In the Appendix, I provide the necessary institutional details about the sports card and sports ticket grading industry to motivate the experimental design.

Identification and Hypotheses

The sports card marketplace includes both local and nonlocal sellers (dealers). Accordingly, by parsing dealers into types, my experimental design permits two distinct identification strategies. First, with the maintained assumption that nonlocal dealers have no reputational concerns, any reciprocal behavior observed among the subjects in the floor treatments can be attributable to social preferences. Alternatively, behavior of local dealers might include reputational as well as social preference effects. Second, even if I observe disparate data patterns across local and nonlocal dealers, it might be the case that selection effects are at work. Indeed, one of the key points of the theoretical work on social preferences is that there is a distribution of types in the population—some purely selfish types and others with social preferences. In this spirit, it is important to consider a second means of identification: behavior *within* dealer type—across the laboratory and field settings, as well as within the field treatments.

Under this design, a clean test of the predictions from a purely self-interested model and a model with social preferences is permitted. Table 2 provides a summary of what can be examined across the various experimental treatments. In a general sense, table 2 highlights that laboratory experiments that estimate propensities might not provide the necessary environment to measure them accurately. Consider column 1 of table 2, which summarizes predictions in the lab treatments. In this case, I delineate between two potential situations: one that has “experimenter” effects and one that does not have experimenter effects. For example, in the lab treatments, reputational concerns (i.e., sellers are fully aware that the experimenter can document delivered quality), experimenter demand effects, Hawthorne effects, or simply the fact that the task is undertaken in an artificial setting can each potentially confound any preferred interpretation.¹³ Though the definition is much too narrow, for minimalism I denote such *potential* laboratory effects simply as “reputational” concerns in table 2.

These effects are important to consider when delineating between predictions of the self-interest and social preference models. If such effects are present in the lab, then both models predict that there will be gift exchange, or a positive relationship between price and quality. Of course, in the field treatments (predictions contained in cols. 2 and 3 of table 2), by design, such effects are varied, permitting a clean test of the two models. As I progress through the summary of the results below, I shall highlight whether evidence is consistent with each of the theories.

III. Experimental Results

Table 3 provides a summary of the raw data. The table can be read as follows: Treatment Lab-R in row 1 denotes that the average price in this treatment was \$28.40, average quality was 3.5, and average requested quality was 6.1. Note that in table 3, for comparability reasons, I have scaled Treatment Lab-R data to range from 1 to 10, and PSA 6, 7, 8, 9, and 10 are denoted as quality levels 1, 2, 3, 4, and 5.¹⁴ A first result

¹³ In social psychology, several studies due to Martin Orne, Robert Rosenthal, and others discuss the important effects of the experimenter-subject relationship (see, e.g., Orne [1962] and Rosenthal’s [2002] summary). While efforts to expunge such effects have been explored in the experimental literature using double-blind (e.g., Hoffman, McCabe, and Smith 1996), randomized response (e.g., List et al. 2004), and related techniques, such approaches may attenuate a number of laboratory phenomena but seem incapable of completely eliminating them, and might even introduce other biases.

¹⁴ Average individual payoffs (ranges) are as follows: Treatment Lab-R: buyers, \$14.90 (\$6.50 to \$24), and sellers, \$18.60 (\$5 to \$34); Lab-RF: buyers, \$2.40 (−\$59 to \$25), and sellers, \$8.00 (\$1 to \$61); Lab-RF1: buyers, \$0.22 (−\$25 to \$25), and sellers, \$9.81 (\$1 to \$35); Lab-Context: buyers, −\$0.09 (−\$67 to \$25), and sellers, \$8.44 (\$1 to \$70).

TABLE 2
PREDICTIONS: SELF-INTERESTED MODEL VS. SOCIAL PREFERENCE MODEL

LAB TREATMENTS	FLOOR TREATMENTS		
	Sports Cards	Tickets	
		Floor-NoGrading	Floor-Announce Grading and Floor-Grading
Self-Interested Model			
Without experimenter effects: no relationship between price and quality	Local dealers: positive relationship between price and quality due to, e.g., reputation effects	Local dealers: no relationship between price and quality	Local dealers: positive relationship between price and quality due to, e.g., reputation effects
With experimenter effects: positive relationship between price and quality due to, e.g., reputation effects	Nonlocal dealers: no relationship between price and quality	Nonlocal dealers: no relationship between price and quality	Nonlocal dealers: no relationship between price and quality
Social Preference Model			
Without experimenter effects: positive relationship between price and quality due to social preferences	Local dealers: positive relationship between price and quality due to, e.g., reputation effects and social preferences	Local dealers: positive relationship between price and quality due to social preferences	Local dealers: positive relationship between price and quality due to, e.g., reputation effects and social preferences
With experimenter effects: positive relationship between price and quality due to, e.g., reputation effects and social preferences	Nonlocal dealers: positive relationship between price and quality due to social preferences	Nonlocal dealers: positive relationship between price and quality due to social preferences	Nonlocal dealers: positive relationship between price and quality due to social preferences

NOTE.—Each column represents predictions of the self-interested model vs. the social preference model across the three major experimental types. In a split of the dealer types, a dealer is labeled as a nonlocal if he or she is unlikely to be concerned with reputation effects; e.g., if he or she rarely attends sports card shows in the area (fewer than three times in a typical year), does not plan to attend more frequently than this in the future, does not own a sports card shop, and does not have an Internet sports card business. All other dealers are labeled as locals.

TABLE 3
RESULTS SUMMARY

	\bar{p} (1)	\bar{q} (2)	\bar{q}_r (3)
Treatment Lab			
Treatment Lab-R	28.4 (16.1)	3.5 (2.0)	6.1 (2.1)
Treatment Lab-RF	22.6 (20.7)	2.3 (1.4)	4.1 (.9)
Treatment Lab-RF1	24.8 (22.1)	2.5 (1.7)	4.0 (1.3)
Treatment Lab-Context			
Treatment Lab-Context	19.5 (19.6)	2.3 (1.5)	4.2 (1.1)
Treatment Lab-Market(\$20)	\$20	3.1 (.9)	4
Treatment Lab-Market(\$65)	\$65	4.1 (.6)	5
Treatment Floor (Cards)			
Treatment Floor-\$20	\$20	2.1 (.9)	4
Treatment Floor-\$65	\$65	3.2 (1.0)	5
Treatment Floor (Tickets): Treatment I			
Treatment Floor-NoGrading	\$10	2.7 (.6)	4
Treatment Floor-AnnounceGrading	\$10	2.9 (.6)	4
Treatment Floor-Grading	\$10	3.1 (.8)	4
Treatment Floor (Tickets): Treatment II			
Treatment Floor-NoGrading	\$30	2.7 (.7)	5
Treatment Floor-AnnounceGrading	\$30	3.4 (.8)	5
Treatment Floor-Grading	\$30	3.6 (1.1)	5

NOTE.—Summary statistics from one (or two in the case of Treatment Floor [Tickets]) unique treatment. \bar{p} is the average price, \bar{q} is the average quality, and \bar{q}_r is the average requested quality. Treatment I-R data are scaled to range from 1 to 10, and PSA 6, 7, 8, 9, and 10 are denoted as quality levels 1, 2, 3, 4, and 5 in the table. Standard deviations are in parentheses.

relates to the comparison between the behavior of this subject pool and that of students. As Fehr and List (2004) note, a typical criticism levied against experimental results concerns the fact that most economics experiments are conducted with students. This may be problematic for several reasons. For example, owing to selection effects, those who do not behave like students may have selected into roles and be overrepresented in certain parts of the economy (e.g., sellers in the marketplace). The first result addresses this issue.

RESULT 1. Behavior of sports card enthusiasts in laboratory games is in line with the gift exchange literature using student subjects, and the results extend well to one-shot environments.

Evidence for result 1 is contained in the raw statistics in row 1 of table 3, which are consistent with the raw data gathered in laboratory experiments with student subjects (see, e.g., Fehr et al. 1993; Charness 1996). Overall, a graphical depiction of the trajectory of the data clearly shows that product quality and prices are positively related (raw data figures that complement table 3 appear in the working paper of this study [List 2005]). In addition, when I examine the temporal aspect of the data, there is little variation over time, consistent with previous studies on gift exchange (for an exception, see Charness, Frechette, and Kagel [2004]).

To provide the necessary statistical link to the literature, I follow previous work and estimate Tobit random-effects regression models. The dependent variable in the regressions is the quality of the good, which is regressed on the price transfer and controls for dealer-specific effects:

$$q_{it} = \beta p_{it} + \omega_{ir} \quad (3)$$

In equation (3), q_{it} represents the product quality that dealer i sent to the buyer in period t , p_{it} denotes the buying agent's offer price to dealer i in period t , and ω_{ir} includes a white-noise error term with mean zero and a constant in the Tobit model. This specification is augmented by inclusion of dealer random effects in the Tobit random-effects regression model.¹⁵

Regression results presented in columns 1–3 of table 4 provide evidence that dealers reward buyers for paying higher prices. In each of the three treatments the marginal price effect is positive and statistically significant at the $p < .10$ level with a two-sided alternative. This result is consistent with the received gift exchange literature. When applicable, I also present an estimate of θ in table 4, where θ is equal to $\partial v(q)/\partial P$

¹⁵ Including a time trend does not change the qualitative results. And when the models converged, controlling for buyer-specific effects does not change the qualitative empirical results presented below. In addition, I have explored empirically modeling the relationship between buyer and seller rents: $v - p = f(p - c)$ in a regression framework. Insights similar to those presented below are obtained, so I suppress further discussion.

TABLE 4
MARGINAL EFFECTS ESTIMATES FOR THE SELLERS' PROVIDED QUALITY

VARIABLE	TREATMENT TYPE									
	Lab-R (1)	Lab-RF (2)	Lab-RF1 (3)	Lab- Context (4)	Lab- Market (5)	Floor (Cards) (6)	Floor- NoGrading (7)	Floor- Announce Grading (8)	Floor- Grading (9)	Floor- Pooled (10)
Price	.05 (1.8)	.05 (3.3)	.07 (4.3)	.05 (4.3)	.02 (4.4)	.02 (6.6)	-.001 (.01)	.02 (2.1)	.02 (1.1)	.02 (2.6)
Constant	.6 (.7)	-.4 (.7)	-.9 (3.3)	-.8 (2.9)	1.6 (6.2)	.6 (3.1)	1.7 (8.0)	1.7 (5.8)	1.8 (3.3)	1.7 (7.3)
θ	. . .	\$.72 (3.6)	\$.1.1 (6.9)	\$.65 (4.7)	\$.45 (2.1)	\$.21 (5.0)	\$.01 (.3)	\$.17 (1.1)	\$.23 (1.1)	\$.19 (2.3)
Dealer random effects	yes	yes	no	no	yes	yes	yes	yes	yes	yes
Observations	25	25	27	32	60	100	60	54	36	90

NOTE.—The dependent variable is the sellers' product quality given to the buyer. Floor-Pooled pools Floor-AnnounceGrading and Floor-Grading data. θ is the estimate of the monetary gift exchange, computed as $\partial v(q)/\partial P$. *t*-ratios (in absolute value) are beneath marginal effect estimates.

and provides a natural benchmark of gift exchange expressed in monetary units. In the case of Treatments Lab-RF and Lab-RF1, both estimates of θ are significantly different from zero, suggesting that gift exchange occurs at the margin. In terms of economic significance, a θ estimate of 1.1 in Treatment Lab-RF1 suggests that a \$1 increase in p leads to a \$1.10 increase in the reciprocated gift, $v(q)$.

While these results provide a robustness check of the data gathered in the laboratory with student subjects and represent good news in that the major laboratory results seem to spill over to different pools of subjects who are commonly engaged in similar exercises in their everyday lives, one can push the comparability notion a bit harder by adding field context to the laboratory environment. This approach is inherent in Treatment Lab-Context, which yields the following result.

RESULT 2. Adding natural context influences behavior, but gift exchange remains alive and well.

Evidence for this result can be found in the summary of the Treatment Lab-Context data contained in table 3. Treatment Lab-Context data reveal that average prices and quality levels are only slightly lower than what was observed in Treatment Lab-RF1 (the comparable context-free treatment). Slight behavioral differences are also revealed in comparisons of scatter plots of these data (List 2005), which show (i) that the positive relationship remains in the contextual data but that there is a slightly greater mass at the subgame perfect equilibrium prediction: 13 of 32 (41 percent) observations in Treatment Lab-Context versus nine of 27 (33 percent) observations in Treatment Lab-RF1; and (ii) that there is a greater number of price (quality) realizations at \$25 (three) and below in Treatment Lab-Context.

For the data from Treatment Lab-Market, table 3 shows that the positive relationship between price and product quality is evident in the aggregate data: whereas the average quality was 3.1 (PSA 8.1) in Treatment Lab-Market(\$20), it was 4.1 (PSA 9.1) in Treatment Lab-Market(\$65). Figure 1 provides a visual view of these quality differences. Comparing the proportion of sellers who provided various quality levels across the \$20 and \$65 treatment yields a discernible rightward shift in the distribution of Treatment Lab-Market(\$65) data.¹⁶ In terms of the average monetary value of the return gift ($v(q)$), sellers provided \$19.73 in Treatment Lab-Market(\$20) and \$41.33 in Treatment Lab-Market(\$65).

To compare gift exchange on the margin across these two treatments, I return to equation (3) and estimate a Tobit model. For Treatment

¹⁶ In some instances dealers made quality claims, and these included statements that they could not provide the requested quality. As in Treatments Floor (Cards) and Floor (Tickets), I still had my buyers purchase the good in Treatment Lab-Market and provide this information to me in the survey. I consider mendacious claims below.

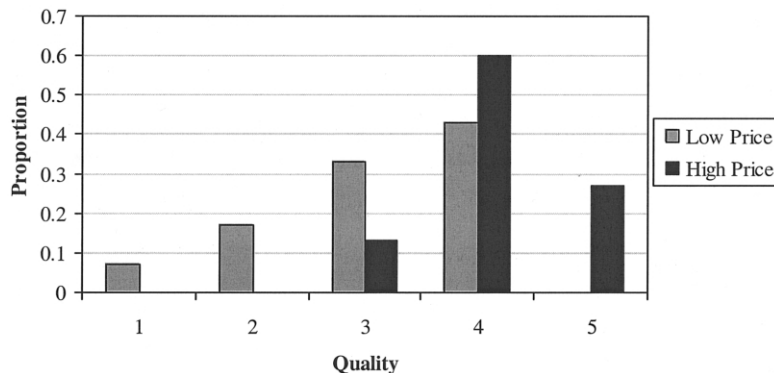


FIG. 1.—Treatment Lab-Market

Lab-Context data, the marginal price effect is positive and statistically significant at conventional levels (see col. 4 of table 4). It is interesting to note that the marginal effect estimate (0.05) is slightly lower than the marginal effect estimate in Treatment Lab-RF1 (0.07), and θ is considerably lower: \$0.65 versus \$1.10. When Treatment Lab-RF1 and Treatment Lab-Context data are pooled and equation (3) is estimated, however, a likelihood ratio test suggests that the homogeneity null should not be rejected, suggesting that behavioral differences do not exist across Treatments Lab-RF1 and Lab-Context.

Considering Treatment Lab-Market data, I provide marginal effects estimates from a Tobit random effects model in column 5 of table 4. The marginal effect estimate of 0.02 is positive and significant at conventional levels; this estimate suggests that card quality increases by roughly one grade when the buyer offers \$65 rather than \$20; in this case, $\theta = \$0.45$.¹⁷ Accordingly, the overall pattern of results suggests that gift exchange is alive and well, even when market context is utilized in the experimental design.

Results 1 and 2 provide a nice validity check of the extant gift exchange literature. Yet, as table 2 highlights, this evidence does not unequivocally show that these subjects exhibit social preferences. It might be the case that these dealers are purely selfish and, owing to effects associated with the lab environment, they behave in a manner consistent with gift exchange. A necessary next step is to explore behavior in

¹⁷ In addition to the Tobit random-effects estimation strategy, which is heavily utilized in the literature, since there is a natural ordering in the data and there are only five cells (i.e., PSA 6–10), I supplement these results by using a panel data ordered probit model, as described in app. D in List (2005). Empirical estimates from the panel data ordered probit model are suppressed because they always coincide with insights gained from eq. (3).

naturally occurring environments in which the controls of the experiment are relaxed appropriately. In such a setting, experimenter demand effects, Hawthorne effects, and the like are absent since, unbeknownst to them, experimental subjects (sellers) are randomly chosen from the dealers who have certain goods. A first insight from the field treatments is as follows.

RESULT 3. When third-party verification is available, behavior in naturally occurring transactions is consonant with both gift exchange and a concern for reputation.

Tables 3 and 4 as well as figure 2 provide evidence for result 3. Row 3 in table 3 shows that the positive relationship between price and product quality is evident in the aggregate data: whereas the average quality was 2.1 (PSA 7.1) in Treatment Floor-\$20, it was 3.2 (PSA 8.2) in Treatment Floor-\$65. In terms of the average monetary value of the return gift ($v(q)$), however, sellers provided much less than they provided in Treatment Lab-Market: roughly \$8 in Treatment Floor-\$20 and \$20 in Treatment Floor-\$65 (vs. \$19.73 and \$41.33 in Treatment Lab-Market). This difference is highlighted via a comparison of figures 1 and 2, which reveals the significant leftward shift in the quality distributions across both the \$20 and \$65 price offers when one moves from the lab to the field. Concerning the ticket stub data, I find that Treatments Floor-AnnounceGrading and Floor-Grading displayed in row 4 of table 3 support the positive relationship found in the sports card data.

Regression results in table 4 yield similar insights: estimates in column 6 of table 4 provide evidence that product quality and price are positively correlated in Treatment Floor (Cards), since the marginal effect estimate of 0.02 is positive and significant at conventional levels. This estimate, which is quite similar to the marginal effect in Treatment Lab-Market, suggests that card quality increases by roughly one grade when the buyer offers \$65 rather than \$20. In this case, however, since the quality changes from PSA 7 to PSA 8 (rather than PSA 8 to PSA 9 in Treatment Lab-Market), $\theta = \$0.21$, considerably lower than the θ estimate of \$0.45 in Treatment Lab-Market. A similar result is found in the Treatment Floor-AnnounceGrading and Treatment Floor-Grading data presented in columns 8 and 9 of table 4, although the marginal price effect is not statistically significant in the Treatment Floor-Grading data at conventional levels. When the Treatment Floor-AnnounceGrading and Treatment Floor-Grading data are pooled (a likelihood ratio test indicates that pooling is appropriate: $\chi^2 = 5.8$), however, the marginal price effect, contained in column 10 of table 4, is statistically significant. Interestingly, across all three specifications the marginal price effect estimate is 0.02, and θ is approximately \$0.20.¹⁸

¹⁸ In computations of θ in the ticket specifications, $v(q)$ is equivalent to one-half the value of $v(q)$ in the sports card data.

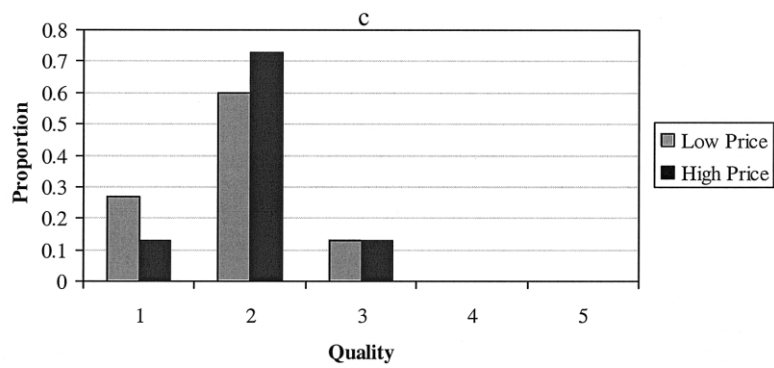
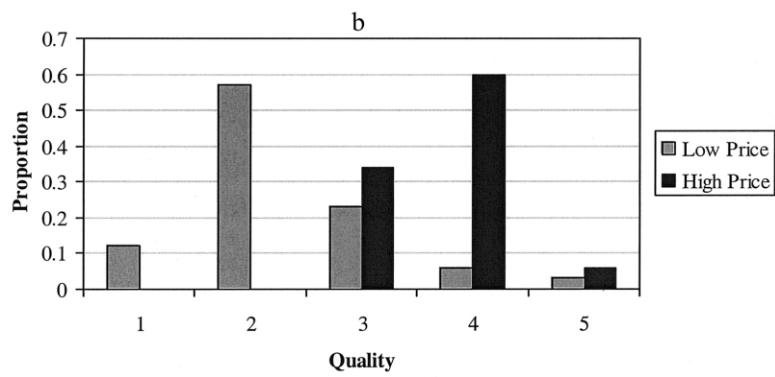
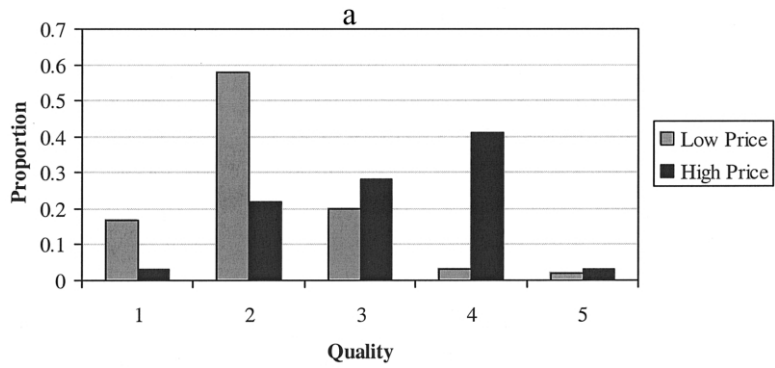


FIG. 2.—Treatment Floor: *a*, sports cards; *b*, local dealers; *c*, nonlocal dealers

As table 2 shows, since this data pattern is observationally equivalent to predictions from a model based purely on reputational effects (e.g., Klein and Leffler 1981), again, these insights are not unequivocal evidence in favor of gift exchange. One can explore a level deeper by recognizing that some of the dealers in the sample may have had an economic reason to uphold their reputations, whereas others may not have had similar incentives. A next result follows.

RESULT 4. When third-party verification is possible, local dealer behavior in naturally occurring transactions is consonant with both gift exchange and a concern for reputation, whereas nonlocal dealers' behavior is in line with self-interest theory.

Table 5 and figures 2, 3, and 4 provide evidence for this result. In a split of the dealer types, a dealer is labeled as a "nonlocal" if he or she is unlikely to be concerned with reputation effects, for example, if he or she rarely attends sports card shows in the area (fewer than three times in a typical year), does not plan to attend more frequently than this in the future, does not own a sports card shop, and does not have an Internet sports card business. All other dealers are labeled as "locals"; in practice, these are primarily dealers who frequent the area often. This information was obtained from a survey (see List 2005, app. B). Note that besides this difference, across all other observables, such as years of experience and age, dealers are similar. I return to the issue of selection effects below, however.¹⁹

The raw data displayed in figures 2, 3, and 4 provide initial support for result 4. In transactions with local dealers, higher price offers yield superior quality in Treatments Floor, Floor-AnnounceGrading, and Floor-Grading, as illustrated in figures 2*b* and 3. Alternatively, while delivered quality is positively related to price across these three treatments among nonlocal dealers (see figs. 2*c* and 4), the differences are minute.

Table 5 provides regression results to support result 4. Columns 1 and 2 split the Treatment Floor data into two subsamples: Floor(Cards)_L (local dealers over sports cards) and Floor(Cards)_N (nonlocal dealers over sports cards). In the former subsample, the marginal price effect is positive and significant at conventional levels. In terms of economic significance, the coefficient estimate in column 1 of 0.03 results in an estimated marginal effect of roughly 1.5 grades; that is, in the \$65 treatment, local dealers provided a quality that was 1.5 grades above the quality level they provided in the \$20 treatment. Measured at the sample means, this 1.5 quality increment yields the buyer a PSA-rated 8.6 card

¹⁹ When I categorize dealer data from Treatment Lab and Treatment Lab-Market in a similar manner, I find that there is no difference in behavior across local and nonlocal dealers.

TABLE 5
MARGINAL EFFECTS ESTIMATES FOR THE SELLERS' PROVIDED QUALITY SPLIT BY DEALER TYPE

VARIABLE	TREATMENT TYPE								
	Floor (Cards) _L (1)	Floor (Cards) _N (2)	Floor-No Grading _L (3)	Floor-No Grading _N (4)	Floor- Announce Grading _L (5)	Floor- Announce Grading _N (6)	Floor- Grade _L (7)	Floor- Pool _N (8)	Floor- Pool _L (9)
Price	.03 (8.6)	.004 (.7)	.002 (.2)	-.005 (.5)	.04 (2.1)	.003 (.3)	.04 (2.7)	.003 (.1)	.04 (4.8)
Constant	.6 (4.1)	.6 (4.6)	1.6 (5.0)	1.8 (5.2)	1.7 (5.2)	1.5 (4.6)	1.8 (5.0)	1.8 (1.7)	1.8 (10.0)
θ	\$.31 (5.2)	\$.01 (.5)	\$.02 (.4)	-\$0.006 (.5)	\$.32 (1.4)	\$.02 (.6)	\$.42 (1.5)	\$.03 (.1)	\$.35 (2.1)
Dealer random effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	70	30	36	24	30	24	20	16	50

NOTE.—The dependent variable is the sellers' product quality given to the buyer. Floor-Pool_L pools Floor-AnnounceGrading_L and Floor-Grading_L data. θ is computed as $\partial v(q)/\partial P$. t -ratios (in absolute value) are beneath marginal effect estimates. Subscripts N and L after treatment type denote regressions with local and nonlocal dealer data only.

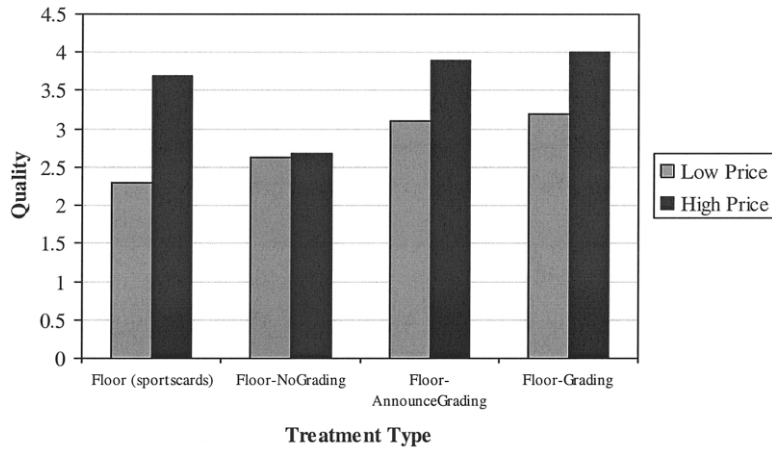


FIG. 3.—Price/quality relationship for local dealers

rather than a PSA-rated 7.1 card. With the $v(q)$ values discussed earlier, this quality increase maps into an increase in market value of roughly \$20, much less than the extra \$45 spent to obtain the card. A θ estimate of \$0.31 complements this finding.

Alternatively, for nonlocal dealers, gift exchange is not evident in Treatment Floor (see col. 2 of table 5), since the marginal price effect is not statistically significant at conventional levels. Regression results for Treatments Floor-AnnounceGrading and Floor-Grading provide fur-

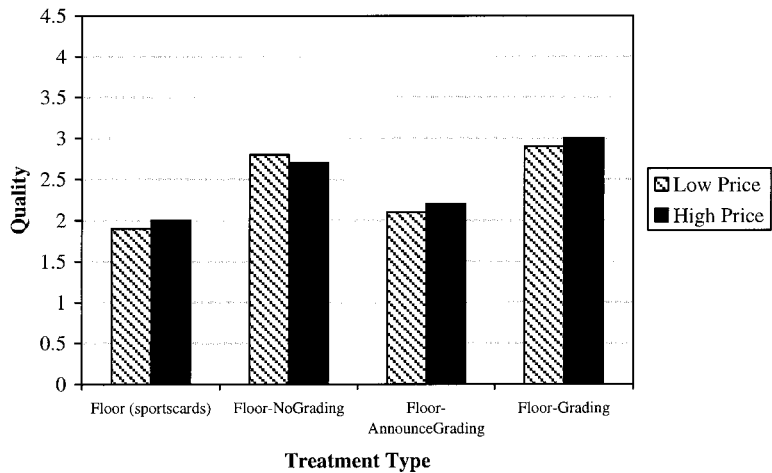


FIG. 4.—Price/quality relationship for nonlocal dealers

ther support for result 4: in both cases the marginal price effect in the local dealer data is positive and significant at conventional levels (cols. 5 and 7 of table 5), whereas there is no such effect found in the nonlocal dealer data (cols. 6 and 8 of table 5). For both the Treatment Floor-AnnounceGrading and Floor-Grading local dealer data, the marginal effect estimate is 0.04, and $\theta = \$0.32$ and $\$0.42$, though neither θ estimate is statistically significant at conventional levels. When these data are pooled (likelihood ratio test: $\chi^2 = 1.4$), $\theta = \$0.35$ and is significant at the $p < .05$ level (col. 9 of table 5). Treating nonlocal dealer data similarly by pooling data from Treatments Floor-AnnounceGrading_N and Floor-Grading_N provides little new information: gift exchange is not evident among nonlocal dealers. Accordingly, as table 2 suggests, the nonlocal dealer data are consonant with the self-interest model.

A natural question that arises concerns whether the local dealer behavior is driven primarily by reputation effects or social preferences: given the identification problem, from the above results alone one cannot determine the extent to which reputation effects and social preferences are influencing market outcomes. One nice characteristic of the current experimental design is that I can examine behavior in markets that are void of third-party verification to explore this issue. In such cases, in economic terms the situation faced by the local and nonlocal dealers is identical. Treatment Floor-NoGrading provides a first result.

RESULT 5. When third-party verification is not available, supply-side behavior in naturally occurring transactions is consonant with purely selfish money-maximizing theory, suggesting that reputational considerations, rather than social preferences, are driving the earlier results.

Evidence for this result can be seen in tables 3–5 as well as figures 3 and 4. Table 3 shows that there is very little quality difference between the \$10 and \$30 offers in Treatment Floor-NoGrading. This result is highlighted in figures 3 and 4, where both local and nonlocal dealers do not provide different quality levels across offers of \$10 and \$30 in Treatment Floor-NoGrading. Empirical results displayed in tables 4 and 5 support the raw data patterns, since the marginal price effect is insignificant in the aggregate data (col. 7 of table 4) and in both specifications that split the data by dealer type (cols. 3 and 4 of table 5).

This finding, which according to the predictions outlined in table 2 is in line with the self-interest model, leads to the tentative conclusion that reputation effects rather than social preferences are responsible for driving a large part of the price/quality tendencies observed in the naturally occurring data. While there is some evidence in favor of social preferences in this market, as price and quality are directionally related (positively) in various places in the nonlocal dealer data and in the local dealer Treatment Floor-NoGrading data, it seems to be of second-order

importance in real market transactions.²⁰ This insight can be viewed in List (2005, app. E), which provides several supplementary figures that summarize the raw data across local and nonlocal dealers in the ticket stub treatments.

Mendacious Claims

Empirical estimates presented above provide measures of gift exchange in the spirit of the extant literature and highlight a framework that can measure social preference effects and reputation effects. Yet it is important to recognize and examine the degree of mendacious claims in the marketplace. If dealers do not have the necessary inventory to fulfill the quality request (e.g., as a result of my misjudgment of quality during my perusal of sales during the show) but provide quality disclaimers, then it is important to explore this aspect of behavior. In this spirit, an important complement to the above results is a thorough analysis of the statistical association between quality claimed and quality delivered. A first result follows.

RESULT 6. When third-party verification is possible, local dealers provide fewer claims of quality than nonlocal dealers in the field and, conditional on claiming quality, shirk less frequently.

Table 6 summarizes dealer behavior across Treatments Floor (Cards) and Floor (Ticket Stubs). Evidence for the first part of result 6 can be obtained by computing the percentage of local and nonlocal dealers who claim quality in Treatments Floor, Floor-AnnounceGrading, and Floor-Grading. The second part of result 6 follows from a comparison of the quality claimed and the quality actually delivered. Before I discuss the evidence for result 6, it is important to point out that in some cases dealers provide quality ranges; for example, “this card would grade at PSA 8 or 9.” In these cases I use the midpoint of the range (e.g., 8.5). A few other dealers were agnostic about the grading system. I label these types as not claiming quality (similar results are obtained if I simply delete these observations). And, in some instances the dealer stated “this one is top quality” or “this is a gem” when describing the good. I label these dealers as not claiming quality, but note that if I take the literal word of the dealer and pair these statements with the appropriate PSA grade, the fundamental results do not change.

Upon pooling data from the Treatments Floor, Floor-Announce Grading, and Floor-Grading in table 6, I find that 94 of 190 (49 percent)

²⁰ I also gathered information on length of buyer/seller relationships. While speculative, an upper-bound estimate of social preferences within long-term relationships (where a long-term relationship is defined as one wherein the buyer and dealer have had five or more interactions in the previous 12 months or have had two or more interactions annually over the past three plus years) suggests that they influence the price and quality correlation.

TABLE 6
SUMMARY OF RESULTS: PRODUCT QUALITY CLAIMS

	Claims (1)	Quality Claim (2)	Delivered Quality (3)	Delivered Promised Quality or Above (4)
Overall				
Treatment Lab-Market	10/60	3.95 (.4)	4.3 (.5)	10/10
Treatment Floor (Cards)	53/100	3.9 (.7)	2.7 (1.1)	15/53
Treatment Floor-NoGrading	36/60	3.8 (.6)	2.8 (.6)	8/36
Treatment Floor-Announce Grading	24/54	4.2 (.5)	2.9 (.9)	4/25
Treatment Floor-Grading	17/36	4.2 (.6)	3.1 (1.1)	4/17
Local Dealers				
Treatment Lab-Market	6/42	4.0 (.5)	4.3 (.5)	6/6
Treatment Floor (Cards)	27/70	3.9 (.7)	3.4 (1.1)	12/27
Treatment Floor-NoGrading	22/36	3.9 (.5)	2.8 (.6)	4/22
Treatment Floor-Announce Grading	7/30	4.1 (.3)	3.9 (.4)	4/7
Treatment Floor-Grading	4/20	4.3 (1.0)	3.8 (.5)	2/4
Nonlocal Dealers				
Treatment Lab-Market	4/18	3.9 (.3)	4.3 (.5)	4/4
Treatment Floor (Cards)	26/30	4.0 (.7)	2.0 (.6)	3/26
Treatment Floor-NoGrading	14/24	3.7 (.6)	2.8 (.6)	4/14
Treatment Floor-Announce Grading	17/24	4.3 (.6)	2.5 (.6)	0/18
Treatment Floor-Grading	13/16	4.2 (.4)	2.9 (1.2)	2/13

NOTE.—Standard deviations are in parentheses.

dealer observations involve product quality claims. In a split by dealer type, 38 of 120 (32 percent) local dealer observations involve product quality claims, whereas 56 of 70 (80 percent) nonlocal dealer observations involve product quality claims. Of those dealers who make quality claims, local dealers deliver the promised quality (or above) in 18 of 38 cases (47 percent), whereas nonlocals deliver the promised quality (or above) in only five of 57 (9 percent) cases.

To complement these insights, I estimate the bivariate probit model with sample selection due to van de Ven and van Praag (1981):

$$Y_1^* = \Phi(\beta_1'V) + \epsilon_1; \quad Y_1 = \begin{cases} 1 & \text{if } Y_1^* > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4a)$$

and

$$Y_2^* = \Omega(\beta_2'Z) + \epsilon_2; \quad Y_2 = \begin{cases} 1 & \text{if } Y_2^* > 0 \\ 0 & \text{otherwise;} \end{cases} \quad (4b)$$

$$\epsilon_1, \epsilon_2 \sim \text{bivariate normal}(0, 0, 1, 1, \rho).$$

Equation (4a) is the quality claim equation. The variable Y_1^* is unobserved, but I can observe its sign since $Y_1 = 1$ if the dealer provided a quality claim, and zero otherwise. Variables in V include a dichotomous variable indicating whether the seller is a local dealer and a control for the price offer/quality requested. Equation (4b) is the shirking equation and is observed only when $Y_1 = 1$; hence the selectivity model arises since a mendacious claim can occur only among dealers who provide a quality promise. In Z , I include a dichotomous variable for whether the seller is a local dealer. To account for data dependencies, I calculate the standard errors assuming that the error terms are independent across dealers but not within each dealer (i.e., clustered standard errors for dealers).

The individual quality claim and shirking equations could be estimated separately, but unless $\rho = 0$, such estimation results in sample selection bias. In estimation of the system, I use full-information maximum likelihood, where the log likelihood is given by

$$\begin{aligned} \sum_{y_1, y_2=1} \ln \phi_2(\beta_1'V, \beta_2'Z, \rho) + \sum_{y_1=1, y_2=0} \ln \phi_2(\beta_1'V, -\beta_2'Z, -\rho) \\ + \sum_{y_1=0} \ln \phi(-\beta_1'V), \end{aligned} \quad (5)$$

where ϕ_2 denotes the bivariate standard normal cumulative density function and ϕ denotes the univariate standard normal cumulative density function.

Empirical results in support of result 6 are presented in columns 1, 2, and 5–8 of table 7. The first part of result 6—when third-party verification is possible, in the field local dealers provide fewer quality claims—can be found in all three quality claim equations.²¹ For instance, empirical estimates in column 1 suggest that in the sports card treat-

²¹ The Floor-Grading model failed to converge; thus I present estimates from equations that are estimated separately.

TABLE 7
EMPIRICAL ESTIMATES FOR THE SELLERS' QUALITY CLAIMS AND SHIRKING RATES

VARIABLE	TREATMENT TYPE								
	Floor (Cards)		Floor (Tickets) NoGrading		Floor (Tickets) Announce Grading		Floor (Tickets) Grading		Lab-Market
	Quality Claim (1)	Shirk (2)	Quality Claim (3)	Shirk (4)	Quality Claim (5)	Shirk (6)	Quality Claim (7)	Shirk (8)	Quality Claim (9)
Local dealer	-1.4 (4.4)	-1.2 (1.8)	.06 (.2)	.24 (.6)	-1.3 (2.5)	-2.3 (4.8)	-2.2 (2.2)	-1.0 (1.4)	-.30 (.8)
Price	.01 (1.9)	. . .	-.03 (1.7)	. . .	-.01 (.3)04 (1.3)	. . .	-.01 (.7)
Constant	.6 (2.0)	1.1 (2.4)	.7 (1.7)	-.1 (.3)	.6 (1.5)	1.0 (2.8)	.3 (.5)	1.0 (2.4)	-.5 (.5)
Observations	100	53	60	36	54	24	36	17	60

NOTE.—The dependent variable in the quality claim specification equals one if the dealer claimed quality, zero otherwise. The dependent variable in the shirking specification equals one if the dealer shirked on the quality claim, zero otherwise. The Floor-Grading model did not converge; thus estimates are derived from equations that are estimated separately. The second-stage shirking equation for Lab-Market cannot be estimated because no dealers shirked in the Lab-Market treatment. *t*-ratios (in absolute value) are beneath coefficient estimates.

ment, local dealers provide fewer quality claims, and this estimate is statistically significant at the $p < .05$ level. As is clear from columns 5 and 7 of table 7, similar insights are obtained in the ticket treatments. Evidence in favor of the second part of result 6—conditional on claiming quality, when third-party verification is possible, local dealers shirk less often in the field—can also be found in all three shirking specifications. Whereas the coefficient estimates in Treatments Floor (Cards) and Floor-Grading specifications are significant only at the $p < .07$ ($p < .17$) level with a two-sided alternative (cols. 2 and 8), the coefficient estimate in the Floor Announce Grading model is statistically significant at the $p < .01$ level (col. 6).

As in the spirit of the inquiry into result 4, one can question whether the increased quality promises and deliveries from local dealers are due purely to reputational concerns or have an element of social preferences. Examining data collected in Treatment Lab-Market and Treatment Floor (Ticket Stubs) lends insights into this issue and leads to the final result.

RESULT 7. In the laboratory, or when third-party verification is not possible, local and nonlocal dealers make similar claims of quality, and conditional on claiming quality, shirk to the same extent.

As table 6 reveals, in Treatments Lab-Market and Floor-NoGrading, local and nonlocal dealers behave quite similarly. In Treatment Floor-NoGrading, local dealers make quality claims in 22 of 36 (61 percent) cases, whereas nonlocal dealers make quality claims in 14 of 24 (58

percent) cases. Likewise, conditional on claiming quality, local dealers in Treatment Floor-NoGrading shirk in 18 of 22 cases—that is, in 82 percent of transactions, local dealers provide lower quality than promised—whereas 71 percent (10 of 14) of nonlocal dealer transactions should be considered shirking. In Treatment Lab-Market, both dealer types make considerably fewer quality claims and never shirk.

Estimating the bivariate probit model with sample selection in equations (4a) and (4b), I find that the observed differences across local and nonlocal dealers are statistically insignificant at conventional levels for both Treatment Lab-Market and Treatment Floor-NoGrading data. For Treatment Floor-NoGrading, columns 3 and 4 of table 7 show that both local and nonlocal dealers make similar quality claims and, conditional on providing a quality claim, shirk to the same extent. For Treatment Lab-Market data, column 9 of table 7 provides similar evidence: local and nonlocal dealers provide a similar number of quality claims.²²

IV. Concluding Remarks

This study provides a framework for measuring social preferences and reputation effects using a series of laboratory and field experiments. In doing so, it showcases the desirability of building a bridge between the lab and the field. In a methodological sense, this bridge permits a test of whether laboratory behavior is a good indicator of behavior in the field. The finding that agents behave differently in tightly controlled laboratory experiments than in their naturally occurring environment poses an important challenge to laboratory studies that measure individual propensities. More generally, these results underscore the role that field experiments can play in empirical economics. For example, experimentalists typically take stock in results from a series of *laboratory* experiments. This study pushes this notion in a new direction by shedding light on the importance of including results from field experiments within this evidentiary system. In this light, the results show that field experiments can help to uncover the causes and underlying conditions necessary to produce data patterns observed in the lab.

While the data suggest that social preferences do not have a major impact in these particular markets, such results, of course, do not necessarily preclude social preferences from having import within other economic domains. Some scholars have argued that such preferences are evident in domains in which the pressures of the market are absent (e.g., the charitable fund-raising work of Falk [2004]). I view this class

²² As table 6 reveals, the shirking equation cannot be estimated for the Treatment Lab-Market data because no dealers—local or nonlocal—shirked in the laboratory experiment.

of studies as fruitful in that they represent good examples of domains in which such preferences might be significant.

The data are also sufficiently rich to speak to how reputation effects and professional certification influence market performance. For example, the data support the view that reputation effects enhance the quality of goods. This insight is consonant with results in the work of Akerlof (1970), who provides evidence on the operation of markets in developing countries that demonstrates a positive association between reputation effects and the quality of goods. Furthermore, empirical results suggest that third-party enforcement of contracts is important: the addition of professional quality certifiers enhances market performance and supports the conjecture that the private market can solve the lemons problem through third-party verification. This result might be viewed as a test of the Klein and Leffler (1981) model in that local sellers cheat less than nonlocal sellers when quality is measurable, but this is not the case when quality is not easily measurable. This finding indicates that reputations cannot work without information, suggesting that reputation and the monitoring of quality are *complements*.

Appendix

Further Experimental and Institutional Details

A. Determining $c(q)$ and $v(q)$ Values

The values of $c(q)$ were chosen to represent the dealer cost to replace a 1990 Leaf Frank Thomas card of various quality levels.²³ The values are taken from the standard price guide for baseball cards: *Beckett Baseball Cards Monthly*. For each single type of ungraded card, *Beckett* collects pricing information from about 110 card dealers throughout the country and publishes a “high” and “low” price reflecting current selling ranges for several quality variants. The high price represents the highest reported selling price and the low price represents the lowest price one could expect to find with extensive shopping. Assuming that dealers’ replacement costs are roughly equivalent to the reported “low” price, I use the “low” prices from *Beckett* for 1990 Leaf Thomas cards that would grade PSA 6, 7, 8, 9, and 10 to approximate $c(q)$ values.

Determining $v(q)$ values in equation (2) to approximate the gains from trade is more difficult since consumer demand curves are not readily observable. In this case I considered results from two approaches: (i) taking the “high” prices from *Beckett* for 1990 Leaf Thomas cards that would grade PSA 6, 7, 8, 9, and 10 and (ii) gathering statements of value for 1990 Leaf Thomas cards that would grade PSA 6, 7, 8, 9, and 10 via a contingent survey in the spirit of Cummings

²³ I chose this particular card for all treatments because of my experience in evaluating the attributes of the card over the past 15 years (as a dealer and a consumer), Thomas’s popularity, and the fact that this variant represents his “rookie card”—typically a player’s most sought-after card. These latter two factors help to explain the extensive interest in the card among broad classes of collectors.

and Taylor (1999).²⁴ The contingent valuation experiment, which was run on the floor of a sports card show, randomly allocated consumers into one of five treatments (PSA 6, 7, 8, 9, or 10). Thirty subjects were placed into each treatment, for a total of 150 subjects. Subjects were asked to state their true value for a 1990 Leaf Thomas card in a contingent valuation scenario. In addition, they were warned about hypothetical bias, which oftentimes arises in such situations, with a “cheap talk” script.²⁵ Previous efforts have found that a contingent survey that includes a cheap talk script has yielded consumer values that closely match actual values (e.g., List 2001). Most important for our purposes, mean values from the contingent survey are in the range of published *Beckett* “high” prices; thus I use these values and make $v(q) = \$6, \$8, \$15, \$30, \text{ and } \$80$ for $q = 1, 2, 3, 4, \text{ and } 5$ (PSA 6, 7, 8, 9, and 10).

B. Sports Card and Sports Ticket Grading

Each year, sports card companies design and print sets of sports cards depicting players and events from the previous season. Once the print run of a particular set has been completed, the supply of each distinct card in the set is fixed. The value of a particular card depends on its scarcity, the player depicted, and the physical condition of the card, that is, the condition of its edges, corners, and surface and centering of the printing. To track card condition, people often use a 10-point scale. For example, a card with flawless characteristics under microscopic inspection would rate a perfect 10, whereas defects, including minor wear on the corners, would decrease the card’s grade to a 7. The card’s overall grade is computed via the aggregation of the various characteristics.

Professional Sports Authenticators is the industry leader in grading services, and its parent company became publicly traded in 1999 (Collectors Universe, under NASDAQ ticker symbol CLCT). PSA has graded more than 7 million sports cards since its inception in 1987. Professional grading is voluntary and costs \$6–\$100 per card, depending on package size and requested turnaround time. Importantly, the fee is independent of the actual grade received. Graded cards are encased in plastic and sealed with a sonic procedure that makes it virtually impossible to open and reseal the case without evidence of tampering.

PSA adopted integer grades from 1 to 10, where a 10 is considered gem mint and commands a premium price. A PSA 9 card is considered mint and is the next most valuable card type. As witnessed by the $c(q)$ and $v(q)$ vectors used in treatments I and II, card values are convex in the grade received. Importantly, Jin, Kato, and List (2004) provide evidence suggesting that even under PSA’s coarse grading system, certification reveals important information to ordinary

²⁴ I could have gathered willingness to pay (or willingness to accept) values by auctioning off Thomas cards using an incentive-compatible auction institution (i.e., a Vickrey second-price auction), but market prices should influence bids, leaving me with a vector of bids that roughly estimate the perceived market price adjusted for transactions costs.

²⁵ The cheap talk script is similar to that in List (2001) and notes that “In most questions of this kind, folks seem to have a hard time doing this. They act *differently* in a hypothetical situation, where they don’t really have to pay money, than they do in a real situation, where they really have to pay money. We call this ‘hypothetical bias’. ‘Hypothetical bias’ is the difference that we continually see in the way people respond to hypothetical situations as compared to real situations. So, if I was in your shoes, and I was asked to make a choice, I would think about how I feel about spending *my* money this way. When I got ready to choose, I would ask myself: if this was a real situation, do I *really want* to spend my money this way?”

consumers. Yet they report that dealers gain no information from a card's PSA grade, suggesting that dealers are able to evaluate quality as well as PSA.

Sports tickets and ticket stubs have recently gained enough market acceptance to merit professional grading. Ticket supply, of course, depends on the stadium size of the event and the proportion of fans in attendance who preserved their ticket stubs (or in the case of unused tickets, the number of fans who left their tickets unused). Ticket grading is similar to sports card grading: an identical 10-point scale is used, and sharpness of corners, centering of printing, sharp focus, and original gloss are very important. Furthermore, staining, printing imperfections, and print quality of crucial game information are also important in determining ticket quality.

References

- Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Q.J.E.* 84 (August): 488–500.
- . 1982. "Labor Contracts as Partial Gift Exchange." *Q.J.E.* 97 (November): 543–69.
- Akerlof, George A., and Janet L. Yellen. 1988. "Fairness and Unemployment." *A.E.R. Papers and Proc.* 83 (May): 44–49.
- . 1990. "The Fair Wage–Effort Hypothesis and Unemployment." *Q.J.E.* 105 (May): 255–83.
- Andreoni, James, and John Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70 (March): 737–53.
- Berg, Joyce, John W. Dickhaut, and Kevin A. McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Econ. Behavior* 10 (July): 122–42.
- Bewley, Truman F. 1995. "A Depressed Labor Market as Explained by Participants." *A.E.R. Papers and Proc.* 85 (May): 250–54.
- Blinder, Alan S., and Don H. Choi. 1990. "A Shred of Evidence on Theories of Wage Stickiness." *Q.J.E.* 105 (November): 1003–15.
- Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *A.E.R.* 90 (March): 166–93.
- Brown, Martin, Armin Falk, and Ernst Fehr. 2004. "Relational Contracts and the Nature of Market Interactions." *Econometrica* 73 (May): 747–80.
- Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton Univ. Press.
- Camerer, Colin F., and Keith Weigelt. 1988. "Experimental Tests of a Sequential Equilibrium Reputation Model." *Econometrica* 56 (January): 1–36.
- Charness, Gary. 1996. "Attribution and Reciprocity in an Experimental Labor Market." Manuscript, Univ. California, Santa Barbara.
- Charness, Gary, Guillaume R. Frechette, and John H. Kagel. 2004. "How Robust Is Laboratory Gift Exchange?" *Experimental Econ.* 7 (June): 189–205.
- Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Q.J.E.* 117 (August): 817–69.
- Cox, James C. 2004. "How to Identify Trust and Reciprocity." *Games and Econ. Behavior* 46 (February): 260–81.
- Cummings, Ronald G., and Laura O. Taylor. 1999. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *A.E.R.* 89 (June): 649–65.
- Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A Theory of Sequential Reciprocity." *Games and Econ. Behavior* 47 (May): 269–98.

- Falk, Armin. 2004. "Charitable Giving as a Gift Exchange: Evidence from a Field Experiment." IZA Working Paper no. 1148, Inst. Study Labor, Bonn.
- Falk, Armin, and Urs Fischbacher. Forthcoming. "A Theory of Reciprocity." *Games and Econ. Behavior*.
- Fehr, Ernst, and Armin Falk. 1999. "Wage Rigidity in a Competitive Incomplete Contract Market." *J.P.E.* 107 (February): 106–34.
- Fehr, Ernst, and Simon Gächter. 2000. "Fairness and Retaliation: The Economics of Reciprocity." *J. Econ. Perspectives* 14 (Summer): 159–81.
- Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger. 1997. "Reciprocity as a Contract Enforcement Device: Experimental Evidence." *Econometrica* 65 (July): 833–60.
- Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Q.J.E.* 108 (May): 437–60.
- Fehr, Ernst, and John A. List. 2004. "The Hidden Costs and Returns of Incentives—Trust and Trustworthiness among CEOs." *J. European Econ. Assoc.* 2 (September): 743–71.
- Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Q.J.E.* 114 (August): 817–68.
- Gächter, Simon, and Armin Falk. 2002. "Reputation and Reciprocity: Consequences for the Labour Relation." *Scandinavian J. Econ.* 104 (1): 1–26.
- Hannan, R. Lynn, John H. Kagel, and Donald V. Moser. 2002. "Partial Gift Exchange in an Experimental Labor Market: Impact of Subject Population Differences, Productivity Differences, and Effort Requests on Behavior." *J. Labor Econ.* 20 (October): 923–51.
- Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith. 1996. "Social Distance and Other-Regarding Behavior in Dictator Games." *A.E.R.* 86 (June): 653–60.
- Holt, Charles A., and Roger Sherman. 1990. "Advertising and Product Quality in Posted-Offer Experiments." *Econ. Inquiry* 28 (January): 39–56.
- Jin, Ginger, Andrew Kato, and John A. List. 2004. "That's News to Me! Information Revelation in Professional Certification Markets." Working paper, Univ. Maryland.
- Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1986. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *A.E.R.* 76 (September): 728–41.
- Katz, Lawrence F. 1986. "Efficiency Wage Theories: A Partial Evaluation." In *NBER Macroeconomics Annual*, edited by Stanley Fischer. Cambridge, MA: MIT Press.
- Klein, Benjamin, and Keith B. Leffler. 1981. "The Role of Market Forces in Assuring Contractual Performance." *J.P.E.* 89 (August): 615–41.
- List, John A. 2001. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sports cards." *A.E.R.* 91 (December): 1498–1507.
- . 2004a. "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field." *Q.J.E.* 119 (February): 49–89.
- . 2004b. "Neoclassical Theory versus Prospect Theory: Evidence from the Marketplace." *Econometrica* 72 (March): 615–25.
- . 2004c. "Testing Neoclassical Competitive Theory in Multilateral Decentralized Markets." *J.P.E.* 112 (October): 1131–56.
- . 2005. "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions." Working Paper no. 11616 (September), NBER, Cambridge, MA.
- List, John A., Robert P. Berrens, Alok K. Bohara, and Joe Kerkvliet. 2004. "Ex-

- aming the Role of Social Isolation on Stated Preferences." *A.E.R.* 94 (June): 741–52.
- Macaulay, Stewart. 1963. "Non-contractual Relations in Business: A Preliminary Study." *American Sociological Rev.* 28 (February): 55–67.
- Miller, Ross M., and Charles R. Plott. 1985. "Product Quality Signaling in Experimental Markets." *Econometrica* 53 (July): 837–72.
- Orne, Martin T. 1962. "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications." *American Psychologist* 17 (11): 776–83.
- Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *A.E.R.* 83 (December): 1281–1302.
- Rosenthal, Robert. 2002. "Experimenter and Clinician Effects in Scientific Inquiry and Clinical Practice." *Prevention & Treatment* 5 (October 18). <http://journals.apa.org/prevention/>.
- Sitkin, Sim B., and Nancy L. Roth. 1993. "Explaining the Limited Effectiveness of Legalistic 'Remedies' for Trust/Distrust." *Organization Sci.* 4 (August): 367–94.
- Sobel, Joel. 2002. "Social Preferences and Reciprocity." Manuscript, Univ. California, San Diego.
- Stigler, George J. 1981. "Economics or Ethics?" In *The Tanner Lectures on Human Values*, vol. 2, edited by Sterling M. McMurrin. Cambridge: Cambridge Univ. Press.
- van de Ven, Wynand P. M. M., and Bernard M. S. van Praag. 1981. "The Demand for Deductibles in Private Health Insurance: A Probit Model with Sample Selection." *J. Econometrics* 17 (November): 229–52.