ELSEVIER
JAI

ADVANCES IN APPLIED MICROECONOMICS
VOLUME 13

EXPERIMENTAL AND BEHAVIORAL ECONOMICS

JOHN MORGAN
Editor

EXPERIMENTAL AND BEHAVIORAL ECONOMICS

ADVANCES IN APPLIED MICROECONOMICS

Series Editor: Michael R. Baye

ADVANCES IN APPLIED MICROECONOMICS   VOLUME 13

ADVANCES IN APPLIED MICROECONOMICS

EDITED BY

JOHN MORGAN
University of California, Berkeley, USA

2005

ELSEVIER
JAI

Amsterdam – Boston – Heidelberg – London – New York – Oxford
Paris – San Diego – San Francisco – Singapore – Sydney – Tokyo

ELSEVIER B.V.
Radarweg 29
P.O. Box 211
1000 AE Amsterdam
The Netherlands

ELSEVIER Inc.
525 B Street, Suite 1900
San Diego
CA 92101-4495
USA

ELSEVIER Ltd
The Boulevard, Langford
Lane, Kidlington
Oxford OX5 1GB
UK

ELSEVIER Ltd
84 Theobalds Road
London
WC1X 8RR
UK

First edition 2005

British Library Cataloguing in Publication Data
A catalogue record is available from the British Library.

ISBN: 0-7623-1194-0
ISSN: 0278-0984 (Series)

∞ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).
Printed in The Netherlands.

CONTENTS

This page intentionally left blank

LIST OF CONTRIBUTORS

Lisa R. Anderson	Department of Economics, College of William and Mary, Williamsburg, VA, USA
Nancy Buchan	School of Business, University of Wisconsin, Madison, WI, USA
Anthony Burton	Department of Health, Richmond House, London, UK
Kay-Yut Chen	Decision Technology Department, Hewlett Packard Laboratories, Palo Alto, CA, USA
Syngjoo Choi	Department of Economics, New York University, New York, NY, USA
Rachel Croson	Wharton School of Business, University of Pennsylvania, Philadelphia, PA, USA
Richard Engelbrecht-Wiggans	College of Business, University of Illinois, Champaign, IL, USA
Douglas Gale	Department of Economics, New York University, New York, NY, USA
Teck H. Ho	Haas School of Business, University of California at Berkeley, Berkeley, CA, USA
Eric Johnson	Columbia Business School, Columbia University, New York, NY, USA
Shachar Kariv	Department of Economics, University of California at Berkeley, Berkeley, CA, USA
Elena Katok	Smeal College of Business, Penn State University, University Park, PA, USA
Graham Loomes	School of Economics and Social Studies, University of East Anglia, Norwich, UK

Jennifer M. Mellor	Department of Economics, College of William and Mary, Williamsburg, VA, USA
Jeffrey Milyo	Department of Economics and Truman School of Public Affairs, University of Missouri, Columbia, MO, USA
Maurice E. Schweitzer	Wharton School, University of Pennsylvania, Philadelphia, PA, USA
Martin Sefton	School of Economics, University of Nottingham, Nottingham, UK
George Wu	Graduate School of Business, University of Chicago, Chicago, IL, USA

PREFACE

As the demand to substantiate predictions from economic theory with causal empirical evidence increases, economists have begun relying on controlled laboratory experiments. As this field has blossomed, it has provided evidence confirming some of the key predictions of economic theory, exposing some of the weaker theoretical predictions, and highlighting the importance of non-pecuniary incentives such as trust and reciprocity in economic decision-making. This has resulted in a symbiotic relationship where experimental evidence is not only used to support theoretical conclusions but has pointed economists into bold and exciting new areas of investigation. In this volume I am pleased to present some of the most recent stimulating work in this field.

The first three chapters provide a fresh look at some of the classical issues in experimental economics. These papers provide novel insights into psychology in ultimatum games, the impact of social interaction on learning, and communication in coordination games. The next two chapters look at how experiments can illuminate our understanding of what determines trust. These papers examine how the monitoring within an organization influences trust as well as examining how individual political ideologies are related to an individual's level of trust. The final two chapters show how experiments can be fruitfully applied to vertical relationships and auction design, two of the most important areas in contemporary contract theory.

I am especially proud of the diversity and excellence of all of the contributions to this volume. These authors come from a variety of disciplinary backgrounds from some of the world's leading academic institutions. I feel that this volume is an exceptional example of the returns to cross-disciplinary interaction and I am very proud of what these authors have accomplished.

I am also very grateful to Elsevier Press and especially Valerie Teng for providing a visible and important forum for disseminating these new and important findings.

This page intentionally left blank

GAIN AND LOSS ULTIMATUMS

Nancy Buchan, Rachel Croson, Eric Johnson and George Wu

ABSTRACT

This chapter investigates the difference between ultimatum games over gains and over losses. Although previous research in decision making has found that individuals treat losses and gains differently, losses have not previously been investigated in strategic situations. In the field, however, the problem of negotiating over losses is as unavoidable and problematic as the problem of negotiating over gains. In addition, data on how we bargain over losses can shed some theoretical light on fairness preferences. Two experiments use within-subject designs, the first in the U.S. and the second in the U.S., China and Japan. We find that offers and demands are higher in losses than in gains, and that these results hold across the three countries. We adapt Bolton's (1991) model of fairness to explain the results. Specifically, we extend prospect theory's loss aversion to unfairness, suggesting that unfairness in losses looms larger than unfairness in gains.

1. INTRODUCTION

In the past decade, the ultimatum game has been the source of great empirical and theoretical interest. One attraction is the game's simplicity: it

Experimental and Behavioral Economics
Advances in Applied Microeconomics, Volume 13, 1–23

ISSN: 0278-0984/doi:10.1016/S0278-0984(05)13001-6

is remarkably easy for subjects to grasp and almost as simple for experimenters to conduct. More substantively, the ultimatum game is a building block for understanding more complex forms of bargaining behavior (Stahl, 1972; Rubinstein, 1982) as well as posted price markets (Thaler, 1988). Finally, the game sheds light on fairness and equity, two factors which have become increasingly important components in models of market and organizational transactions (Solow, 1979; Akerlof, 1982; Frank, 1985; Mellers & Baron, 1993).

In the standard ultimatum game, player 1 (the proposer) makes an offer to player 2 (the responder). The offer consists of division of a sum of money (the pie, π) between the two players. Usually this offer takes the form of "player 2 can have x, player 1 will get $\pi - x$." The responder can either accept or reject the offer made. If she accepts, the pie is divided as proposed and the game ends. If she rejects, neither player receives any money, and the game also ends. The unique subgame-perfect equilibrium of the ultimatum game (assuming pure self-interest) is for proposers to make the smallest possible offer, say ε. In turn, responders will accept ε, since ε is better than nothing.[1]

The ultimatum game has been studied through a variety of experimental methods (see summaries in Thaler, 1988; Güth & Tietz, 1990; Camerer & Thaler, 1995; Roth, 1995; Camerer, 2003). In every study, actual behavior deviates substantially from the equilibrium prediction: the average proposer makes offers of 40–50% of the pie. These non-zero offers are made for good reason, since responders reject small but positive offers, again contrary to the predictions of game theory.

This chapter investigates the difference between ultimatum games over gains and over losses. Previous research suggests that subjects in individual decision-making studies treat losses and gains differently (Kahneman & Tversky, 1979; Thaler & Johnson, 1990; Tversky & Kahneman, 1986, 1991, 1992). However, losses are typically not used in experiments and have not been investigated in strategic situations.

There are two reasons to extend the ultimatum game to losses. First, the problem of negotiating over losses is as unavoidable and as problematic as the problem of negotiating over gains. Consider, for example, insurance companies negotiating compensation for damages, joint venture partners bargaining over the division of capital expenditures, business colleagues dividing tasks or committee assignments, or debt holders negotiating over liabilities.

A second reason to consider losses is to shed some theoretical light on fairness. Ultimatum results have fueled alternative theories which "generalize" game theory by formally incorporating desires for fairness into the utility function of agents (e.g. Bolton, 1991; Rabin, 1993; Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000). Examining variants of the ultimatum game is important for theorizing, because the new results can suggest avenues for extending or correcting these generalized theories.

Although negotiating over bads is common in practice and important to theory, relevant literature is sparse. A number of studies compare *third party allocations* over goods and bads (e.g. Griffith & Sell, 1988; Kayser & Lamm, 1980; Lamm, Kayser, & Schanz, 1983), in contrast we examine *interested party bargaining* over bads.

This paper reports the results of two experiments in ultimatum games over losses and over gains; the first concerns data from a strictly American subject pool and only one decision of one pair is chosen for payment (thus the design is somewhere between an experiment and a survey), while the second is drawn from subjects in three countries in a more traditional laboratory setting where all participants are paid for all their decisions. While the experiments were run using different subject populations and procedures, the results are remarkably consistent. Ultimatum bargaining over losses elicits higher demands on the part of responders and higher offers on the part of proposers than does ultimatum bargaining over gains.

We explain the results by extending the intuition of loss aversion to unfairness. If responders are more unfairness averse for losses than gains, responders will be more demanding in the loss game than in the gains game, and this increase in demands makes proposers self-interestedly more generous. We describe the two experiments and their results in Sections 2 and 3. In Section 4, we provide a theoretical explanation of our results based on prospect theory. We conclude in Section 5.

2. EXPERIMENT 1

2.1. Experimental Design

Participants in this experiment were 74 MBA students who were enrolled in one section of a required first-year course. Each subject received a packet containing a set of four questions, counterbalanced for order.[2] The instructor asked the participants to read through the packet and to record their responses in the correct spaces. Subjects were told that one pair of them would be chosen at random in the next class to play the gains game for real, in accordance with the answers they recorded that day. Each participant then recorded their offers and demands for the two games. All participants played both roles of both games yielding four responses per subject, therefore "before" we will use within-subject comparisons.

The questions concerned a \$100 gains ultimatum game and a \$100 loss ultimatum game. Both games were run using the *strategy method* (Selten, 1967). In the gains game, proposers stated how much they wished to offer responders, as described above. Responders recorded their minimum acceptable demand; the amount they would need to receive in order to accept an offer. If the proposer's offer was at least as large as the responder's demand, then the offer was accepted and the responder received what the proposer offered (*not* her own demand), while the proposer received the residual. If the responder's demand was greater than the proposer's offer, then the offer was rejected and both parties received \$0.

For the loss game, the identical game is constructed by subtracting \$100 from all payoffs. In the transformed game, the proposer's offer indicated how much the responder should pay of a \$100 loss. The responder indicated the maximum she would be willing to pay. If the responder was willing to pay at least as much as the proposer suggested that she pay, the costs were divided according to the proposer's proposal. If the proposal was higher than the responder's willingness-to-pay, then there was no deal and *both* parties paid \$100. All participants played both roles of both games.

Figure 1 depicts the two games in the strategy-method format. The equilibrium of the transformed game is simply a transformed equilibrium of the original game: the proposer offers that the responder pays \$99.99 and the responder accepts, leaving the responder with 1¢ of surplus and the proposer with \$99.99 worth of surplus. Note that accepting this offer and paying \$99.99 is better for the responder than rejecting the proposer's offer and paying \$100.

An example may be illuminating here. Assume the proposer offers that the responder pays \$70 out of the \$100 loss and that the responder indicates that she would be willing to pay at most \$80 out of the \$100 loss. Since the responder is willing to pay at least as much as the proposer suggests, the responder accepts the proposer's offer of \$70 by paying it while the proposer pays \$30. However, if instead the responder had indicated she would be willing to pay at most \$50 of the \$100 loss, then she would reject the proposer's offer of her paying \$70. Then both players pay a full \$100.

Note that in both treatments there is \$100 of surplus to be divided between the two parties. In the gains treatment, it is \$100 of gains. In the loss

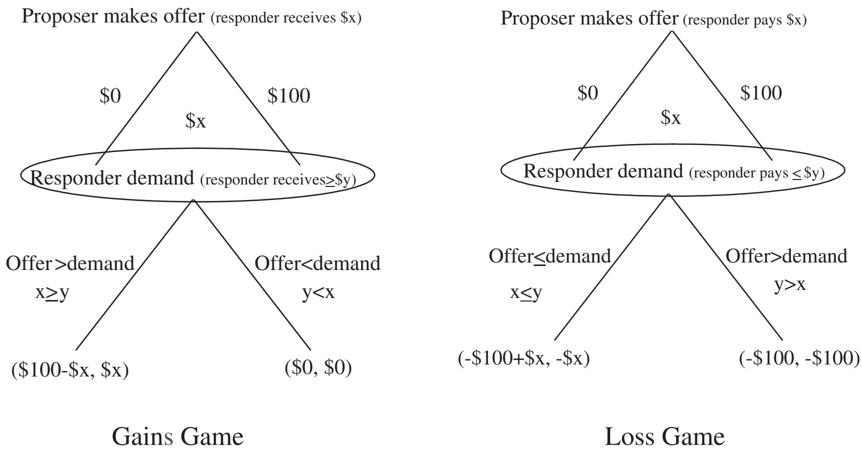


Fig. 1. Ultimatum Games in Gains and Losses.

treatment, it is \$100 of foregone losses. If the offer is rejected, both parties pay \$100 for a total payment of \$200. If the offer is accepted, both parties together pay only \$100. Thus, there is \$100 of surplus to be gained by coming to an agreement in both games. More generally, any offer in the loss game can be transformed into a gains-comparable offer by subtracting the loss offer from \$100. For purposes of comparison, we transform demands and offers from both games into the percentage of the surplus offered to the responder. In the equilibrium 0.01% of the surplus is being offered to the responder (1¢ in the gains game and \$99.99 in the losses game).

2.2. Experimental Results

To compare these treatments we transform offers in the loss game to their equivalent in the gains game, as described above.[3] We find, on average, that offers and demands are significantly higher when individuals bargain over losses rather than gains (see Table 1). Average demands in the loss treatment are for 37.3% of the pie, compared with 29.7% in the gains treatment, yielding a 7.6% difference (\$7.65 out of a \$100 pie). The difference between loss offers and gains offers is smaller (\$2.73). Figure 1 shows the distributions of offers and demands for gains and losses.

Table 1. Offers and Demands in $100 Ultimatum.

	Gain $100 (%)	Loss $100 (%)	Difference (%)
Offers			
Average	39.4	42.2	2.7
Standard Deviation	16.7	12.5	15.6
Demands			
Average	29.7	37.3	7.6
Standard Deviation	19.2	20.1	23.7

As in previous ultimatum experiments, the demands are lower than the offers, suggesting either risk-aversion or other-regarding preferences on the part of the proposers. Similarly, with a few exceptions, offers and demands are bounded above at 50%.

Since we used a within-subject design, we can measure the differences in demands and offers between the two treatments. We found that 21 participants offered more under losses than gains, compared with 9 who offered less (with 44 ties). The difference is statistically significantly (Wilcoxon test, $z = 2.08$, $p < 0.05$). Similarly, 29 participants demanded more for losses than gains, while 10 demanded less (with 35 ties) (Wilcoxon test, $z = 3.01, p < 0.01$). As can be seen from Fig. 2, the 50/50 division serves as a strong anchor for this task. This tendency of individuals to split equally in all situations increases the difficulty of finding significant differences between the treatments. Nonetheless, these differences are still present.

These results indicated that dividing losses is treated differently than dividing gains. We believe that losses loom larger for responders, making them more likely to reject unfair offers. In turn, an increase in anticipated demands induces proposers to offer more in order to have their offers accepted. We explore this explanation in greater detail in Section 4.[4]

Note that while this experiment involved incentive payments, only one decision of one pair out of 37 pairs were chosen for payment. Thus, this experiment has the flavor of a survey rather than a full-blown experiment. Furthermore, participants completed the task in their classrooms, which further limits the generalizability of the results. In experiment 2, discussed below, we address these issues by paying all participants based on all their decisions, and by running the experiment in a more traditional laboratory setting.

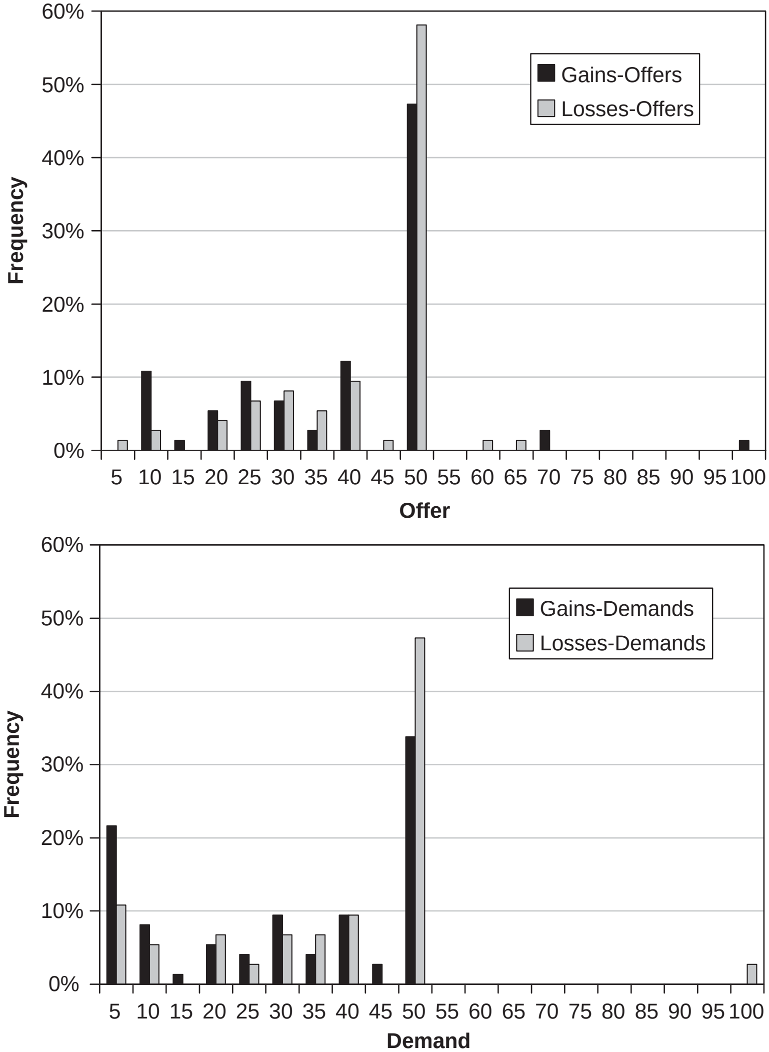


Fig. 2. Distribution of Offers and Demands in $100 Ultimatum.

3. EXPERIMENT 2

3.1. Motivation

The results of experiment 1 demonstrated that people – specifically Americans – behave differently when dividing losses than when dividing gains. But does this result generalize to other populations across the world? The work of Roth, Prasnikar, Okuna-Fujiwara, and Zamir (1991) and later Henrich et al. (2001) suggests that there is reason to question its generalizability. For example, the four-country study conducted by Roth et al. (1991) shows disparity in gain ultimatum games in Jerusalem, Ljubljana (in Slovenia), Pittsburgh, and Tokyo suggests that different culturally influenced beliefs about fairness may have impacted bargaining behavior. The (gain) ultimatum game results of Henrich et al. (2001) among 15 small scale societies are even more startling revealing offers that range from 26 to 58% among their sample populations. If people across the world exhibit differences in offers and demands when bargaining for gains – possibly due to different culturally influenced fairness norms – will there also be differences across countries when bargaining over losses? Furthermore, and perhaps more central to the current research, will people across the world exhibit the same pattern when facing a negotiation involving losses than when facing one involving gains?

In this second experiment, we examine whether (a) the disparity in behavior (and potentially fairness norms) seen in gain ultimatum games extends to behavior in loss ultimatum games, and (b) whether the pattern of higher offers and demands when people bargain over losses than over gains is a robust finding across countries, or is limited to a characterization of behavior of these particular subjects in the United States or to the methodology we used.

3.2. Experimental Design

One hundred and twenty-six subjects participated in this experiment: 40 students from Nankai University (Tianjin, China), 40 students from Osaka University (Japan), and 46 students from the University of Pennsylvania (United States). Subjects were primarily sophomore or junior students in economics or business classes who were paid their actual monetary earnings from the experiment. Upon arrival at the experiment subjects were randomly assigned to one of two rooms, the proposer room or the responder room, and were each given a \$10 (or purchasing power equivalent in China and Japan) participation fee. Subjects were then given an experimental packet that contained instructions for the gain and loss ultimatum game, a short quiz concerning the procedure to make certain subjects understood the game and the potential outcomes, and an offer or demand form. The order in which the gain or loss treatments were administered was counterbalanced.[5] To avoid potential effects of easily anchoring on a 50% equal split of the experimental pie as we saw in experiment 1, the gain condition involved a \$10.26 (or foreign equivalent) stake, and the loss condition involved a \$9.72 (or foreign equivalent) debt.

Subjects were told to read the instructions, complete the quiz, and then fill in their offer or demand form. Subjects were notified that all decisions they made would be confidential and that their identities would remain anonymous to their counterparts in the other room.

We again used the strategy method in this experiment. Assume, for example that Round 1 was the gain round, in which subjects would decide how to divide a pool of money which they were "owed by" the experimenter. Each proposer would record on an offer form the amount of the gain he proposed to share with the responder. Simultaneously, the responder recorded on his form the minimum amount he was willing to accept from the proposer. The proposer's offer form was collected and given to the responder. The responder then compared the offer with the demand he previously recorded. If the proposer's offer equaled or exceeded the responder's demand, the responder would indicate "acceptance" on the offer form, if the offer was less than the demand, the responder circled "reject". The offer form with the response was returned to the proposer and the round was finished. If the bargaining ended in acceptance, the players would be paid by the experimenter. If the bargaining ended in rejection, neither player received anything. After Round 1 was completed, subjects were paid their earnings.

Subjects would then bargain with a different anonymous partner for Round 2, in this case, the loss round. In this round, subjects were to decide how to divide a loss of money which they "owed to" the experimenter. Each proposer would offer that the responder pay a certain amount of the loss, while simultaneously, the responder was stating the maximum amount of the loss he was willing to pay. Offers and demands were recorded and communicated in the same manner as in the gain round. If the amount the proposer offered that the responder pay was less than or equal to the amount the responder was willing to pay, the round ended in acceptance. If the amount the proposer offered that the responder pay was greater than the maximum the responder was willing to pay, the round ended in rejection. In the event of acceptance, each player paid the decided-upon amount. In the event of rejection, each player paid the full amount of the loss. After Round 2 was completed subjects were paid their earnings for the second round, and dismissed.

3.2.1. Cross-country Controls

The international character of this research warranted that we control for country or culture-specific variables that could influence our results. Specifically, we addressed the following issues as suggested by Roth et al. (1991).

1. *Controlling for subject pool equivalency.* We controlled for equivalency in educational background and knowledge of economics among the subject populations in two ways: Firstly, the universities chosen for the experiment were all well-known universities in their countries. Secondly, subjects were all sophomore or junior economics or business undergraduate students and were paid for their earnings in the experiment.
2. *Controlling for currency effects.* We controlled for purchasing power parity by choosing denominations such that monetary incentives relative to subject income and living standards were approximately equal across countries (as in Kachelmeier & Shehata, 1992). Amounts used were Japan (2,000 yen participation fee, 2,052 yen in the gain game, and 1,944 yen in the loss game), China (10 yuan participation fee, 11 yuan in the gain game, and 9 yuan in the loss game), United States ($10 participation fee, $10.26 in the gain game, and $9.72 in the loss game). These amounts were based on information from the U.S. Bureau of Labor Statistics (*Monthly Labor Review,* 1998), and on the recommendations of three independent experts on each economy.
3. *Controlling for Language Effects.* To control for any nuances in language which may impact results across countries, instructions for the experiments in China and Japan were translated into the native language and back-translated into English using separate external translators.
4. *Controlling for Experimenter Effects.* Various measures were taken to control for differences among experimenters in different countries. Firstly, in each country, the lead experimenter was an advanced student in business, and a native of that country. Secondly, an extremely thorough experimental protocol was designed based upon the procedure used in the United States and used in all three countries. The protocol included information such as the positioning of the experimenter in the room, and

the method to be used in answering subject questions. Thirdly, the experimenter from the United States met with the lead experimenter in each country prior to each experiment to brief them on the protocol and to run through a practice (no subjects) session with them. Finally, the American experimenter was present in the data recording room while each experiment was being conducted.

3.3. Experimental Results

Once again, the percentages are transformed to reflect the surplus for the responder. Figure 3 presents a distribution of the offers and demands. Table 2 presents average offers and demands overall and Table 3 provides the breakdown by country.

Firstly, we find that on average, the amount demanded and the amount offered are again higher under losses than under gains with our new experimental conditions. As with experiment 1 and most ultimatum games, offers are higher than demands. Surprisingly, the average offer in the loss condition is substantially above 50%, the typical benchmark for fair offers, and an extremely rare occurrence for standard ultimatum games.

Once again we code the within-subjects differences between the amount demanded and offered under the two conditions. We found 30 participants who offered more under losses than gains, compared with 21 who offered less (and 12 ties). The difference is marginally significant (Wilcoxon test, $z = 1.81, p < 0.06$). We found 35 participants who demanded more for losses than gains, compared with 15 who demanded less (and 13 ties); this difference is significant (Wilcoxon, $z = 2.38, p < 0.05$). In sum, as in experiment 1, individuals demand and offer more when bargaining over losses than when bargaining over gains.

Secondly, we find consistent variance across countries, as shown in Table 3. Average offers and demands significantly differed at the $p < 0.05$ confidence level across countries as shown in two separate analyses of variance. Chinese proposers were the most generous with average offers of 56% compared with 52% for Japanese proposers and 51% for American proposers. American responders made significantly higher demands than responders from other countries: American demands = 47%, Chinese demands = 38%, and Japanese demands = 31%. These results are consistent with the work of Roth et al. (1991) and Henrich et al. (2001) in demonstrating variability in bargaining behavior across countries, and extends this demonstration of variability to games involving losses as well as gains.

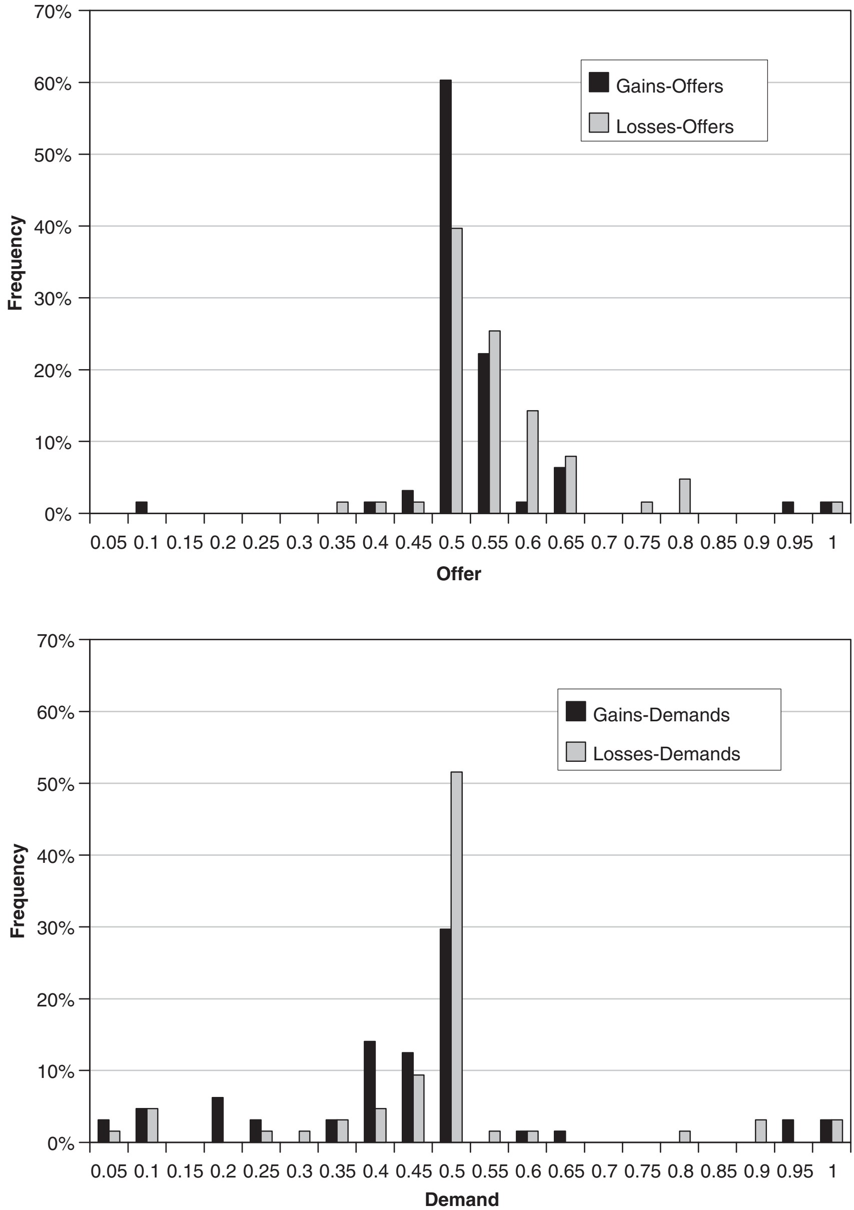


Fig. 3. Distribution of Offers and Demands in Experiment 2.

Table 2. Offers and Demands in $10 Ultimatum.

	Gain (%)	Loss (%)	Difference (%)
Offers			
Average	51.3	54.4	3.1
Standard Deviation	10.5	9.7	11.3
Demands			
Average	36.8	42.0	5.2
Standard Deviation	24.6	22.9	25.4

Table 3. Average Offers and Demands in $10 Ultimatum by Country.

	Gain (%)	Loss (%)	Difference (%)
Offers			
China	55.0	57.2	2.2
Japan	50.1	54.1	4.0
United States	49.1	52.1	3.0
Demands			
China	37.2	39.1	1.9
Japan	28.1	35.5	7.4
United States	44.4	50.4	6.0

Interestingly there was no interaction here. That is, the influence of the gain and loss frames did not vary across countries; in all three countries offers and demands were higher in the loss condition than under the gain condition. Thus, although average offers and demands did differ across countries, these results suggest that regardless of country, the bargaining frame exerts a robust effect on how people react to potential losses in bargaining as opposed to potential gains.

4. PROSPECT THEORY AS AN EXPLANATION

In this section, we propose a model that explains how responders can behave differently between losses and gains. Our model follows Bolton (1991), who explained rejections of positive ultimatum offers by positing a two-argument utility function which includes both money and fairness as arguments. We depart, however, in introducing a reference point, thus permitting responder behavior to be sign-dependent. Our model extends the intuition of losses looming larger than gains to the concept of unfairness and uses the prospect-theory value function as the responder's utility function.

Let π be the size of the pie or the amount of surplus to be divided between the proposer and the responder. Next let d be the disagreement point, or the amount each party will receive or pay in the event of a disagreement. For the loss condition, $d = -\pi$, and for the gain condition, $d = 0$. Finally, we denote the offer to the responder x, $0 \leqslant x \leqslant \pi$, defined as the improvement over the disagreement point d.

The responder has an additive utility function, $U(x, \pi, d) = v(x - d) - f(x, \pi, d)$, which is used to determine whether an offer is accepted or rejected. The first argument, which we call monetary utility for short, $v(x-d)$, captures the utility of money relative to a disagreement point d. The second argument, which we call unfairness utility, $f(x, \pi, d)$, measures the disutility of receiving x from π when the disagreement point is d.

We first assume that $U(\cdot)$ satisfies the following monotonicity conditions:

Assumption 1. $f_x \leqslant 0$ and $f_\pi \geqslant 0$.

Assumption (1) indicates that unfairness utility decreases with the offer x and increases with the size of the pie π.

Next, we assume that the utility of money is described by the prospect theory value function.

Assumption 2. The utility of money is governed by the prospect theory value function. Thus, $v(0) = 0$, $v(x)$ is concave for gains ($v''(x) < 0, x > 0$), convex for losses ($v''(x) > 0, x < 0$), and steeper for losses than gains ($v'(-x) > v'(x), x > 0$).

A responder is willing to take x from a pie π provided that

$$v(x + d) - v(d) \geqslant f(x, \pi, d) \tag{4.1}$$

We first consider gains, where $d = 0$. Let $\hat{x}_g$ be the proposal that the responder is indifferent between accepting and rejecting:

$$v(\hat{x}_g) - v(0) = f(\hat{x}_g, \pi, 0) \tag{4.2}$$

Thus, the responder will accept any offer $x > \hat{x}_g$. Losses ($d = -\pi$) are treated analogously with $\hat{x}_l$ denoting the proposal that leaves the responder indifferent between accepting and rejecting:

$$v(\hat{x}_l - \pi) - v(-\pi) = f(\hat{x}_l, \pi, -\pi) \tag{4.3}$$

In our empirical study, we observed that responders demanded more for losses than for gains, i.e. $\hat{x}_l > \hat{x}_g$. We next examine what restrictions on $v(\cdot)$

and $f(\cdot)$ imply $\hat{x}_l > \hat{x}_g$. Sufficient conditions for $\hat{x}_l \geqslant \hat{x}_g$ are

$$v(\hat{x}_g - \pi) - v(-\pi) \leqslant v(\hat{x}_g) - v(0) \tag{4.4}$$

and

$$f(\hat{x}_g, \pi, -\pi) \leqslant f(\hat{x}_g, \pi, 0) \tag{4.5}$$

Although one of the conditions must hold with strict inequality, the other condition may be violated. We concentrate on the second condition, but briefly discuss the first. We assume prospect theory's S-shaped value function. The condition in (4.4) requires that the value function be sufficiently S-shaped to overcome the loss aversion. Estimated parameters offer mixed results as to the direction of the inequality in (4.4). The parameters in Tversky and Kahneman (1992) predict that $\hat{x}_l < \hat{x}_g$, while estimates found in Camerer and Ho (1994), Wu and Gonzalez (1996), and Gonzalez and Wu (1999) are consistent with the desired result, $\hat{x}_l > \hat{x}_g$. Thus, we concentrate on the second condition (4.5).

We consider a convenient parametric form of the unfairness function to examine what restriction is needed for (4.5).

Assumption 3. Unfairness utility is captured by the following function:

$$f(x, \pi, d) = \begin{cases} \beta_d\left(1 - \frac{x}{\pi/2}\right), & x < \pi/2 \\ 0, & x \geqslant \pi/2 \end{cases}$$

We make the simplifying assumption that offers are most unfair when $x = 0$, and that unfairness decreases linearly in x until the responder receives $\pi/2$. Note that $f(\cdot)$ is reference-dependent: the intercept is indexed by the disagreement point d.

To isolate the role of the unfairness function, we assume that the value function plays no role in producing different demands for losses and gains, i.e., $v(\hat{x}_g - \pi) - v(-\pi) = v(\hat{x}_g) - v(0)$. If we fit the mean data for losses and gains from Section 2, we have $\hat{x}_l = 37.3$ and $\hat{x}_g = 29.7$. Letting $f(29.7, 100, 0) = f(37.3, 100, -100)$, we get that $\beta_{-\pi}(1 - 37.3/50) = \beta_0(1 - 29.7/50)$, or $\beta_{-\pi}/\beta_0 = 20.3/12.7 = 1.60$. The mean data from the international study provides similar coefficients, $\beta_{-\pi}/\beta_0 = 1.65$. Thus, the ratio of the intercepts for the loss treatment ($\beta_{-\pi}$) and the gain treatment (β_0) must be approximately 2-to-1, which is consistent with coefficients of loss aversion measured in endowment effect experiments (Kahneman, Knetsch, & Thaler, 1990) as well as risky choice studies (Tversky & Kahneman, 1992).

5. CONCLUSION

This chapter presents preliminary evidence that bargaining over losses may be different than bargaining over gains. We believe that many real-world negotiations happen over losses, and our experimental results suggest that features like fairness may take on a greater weight in the loss domain than in the gains domain. Thus, one contribution of this work is to suggest that new models (or possibly new parameterizations of existing models), may be needed to describe and predict behavior in the loss domain.

In particular, we find in two experiments that individuals both demand and offer more in ultimatum bargaining over losses than over gains. The first experiment is run in a classroom setting and only one decision of one pair is paid. The second experiment is run in a traditional laboratory setting in three different countries and all decisions of all participants are paid. The fact that the results from both experiments are consistent with each other is a reassuring robustness check on our claims of the asymmetries between bargaining over losses and gains.

Importantly, we demonstrate that although what is regarded as a fair offer or demand in both the gain and loss games may differ across countries, the pattern of both demanding and offering more when facing potential losses rather than potential gains is robust, regardless of the country of origin of the subjects. We extend the classic model of inequality-aversion from Bolton (1991) to include reference dependence.

Intuitively, we suggest that unfairness looms larger for splitting losses than dividing gains. When we calibrate our simple reference-dependent model to mean data, we find that unfairness for losses is roughly twice the magnitude as unfairness for gains. This ratio is close to parameters of loss aversion that have been observed in studies of risky and riskless choice.

Thus, our results suggest both that fairness matters, and that fairness matters *differently* when bargaining over losses as over gains. Evidence from other experiments has demonstrated that the profit-maximizing offer in the ultimatum game is far from zero. Evidence from this study suggests that this profit-maximizing offer is *context-dependent*, and, in particular, sensitive to whether the bargaining is happening over losses or over gains.

NOTES

1. While this is the unique subgame perfect equilibrium of the ultimatum game, any division of the pie is a Nash equilibrium.

2. Instructions are in the Appendix I and raw data are available from the authors.
3. We find no significant order effects in this data, thus the responses are pooled over the two orders.
4. As mentioned above, any division of the pie is a Nash equilibrium. One alternative explanation for our results that the framing of the negotiation (over losses or over gains) might select one of these Nash equilibria.
5. Instructions are contained in the Appendix II and raw data are available from the authors; as above, there were no order effects across countries.

REFERENCES

Akerlof, G. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, *87*, 543–569.

Bolton, G. (1991). A comparative model of bargaining: Theory and evidence. *American Economic Review*, *81*, 1096–1136.

Bolton, G., & Ockenfels, A. (2000). ERC – A theory of equity, reciprocity and competition. *American Economic Review*, *90*, 166–193.

Camerer, C. F. (2003). *Behavioral game theory: Experiments on strategic interaction*. Princeton: Princeton University Press.

Camerer, C. F., & Ho, T.-H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, *8*, 167–196.

Camerer, C. F., & Thaler, R. (1995). Anomalies: Ultimatums, dictators and manners. *Journal of Economic Perspectives*, *9*, 209–219.

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, *114*, 817–868.

Frank, R. (1985). *Choosing the right pond: Human behavior and the quest for status*. Oxford: Oxford University Press.

Gonzalez, R., & Wu, G. (1999). On the Shape of the Probability Weighting Function. *Cognitive Psychology*, *38*, 129–166.

Griffith, W. I., & Sell, J. (1988). The effects of competition on allocators' preferences for contributive and retributive justice rules. *European Journal of Social Psychology*, *18*, 443–455.

Güth, W., & Tietz, R. (1990). Ultimatum bargaining behavior: A survey and comparison of experimental results. *Journal of Economic Psychology*, *11*, 417–449.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of homo economicus: Behavioral experiments in 15 small scale societies. *American Economic Review*, *91*(2), 73–78.

Kachelmeier, S. J., & Shehata, M. (1992). Examining risk preferences under high monetary incentives: Experimental evidence from the People's Republic of China. *American Economic Review*, *82*, 1120–1141.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase Theorem. *Journal of Political Economy*, *98*, 1325–1348.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kayser, E., & Lamm, H. (1980). Input integration and input weighting in decisions on allocations of gains and losses. *European Journal of Social Psychology*, *10*, 1–15.

Lamm, H., Kayser, E., & Schanz, V. (1983). An attributional analysis of interpersonal justice: Ability and effort as inputs in the allocation of gain and loss. *Journal of Social Psychology*, *19*, 269–281.
Mellers, B. A., & Baron, J. (Eds) (1993). *Psychological perspectives on justice*. Cambridge: Cambridge University Press.
Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, *83*, 1281–1302.
Roth, A. (1995). Bargaining experiments. In: J. Kagel & A. E. Roth (Eds), *Handbook of Experimental Economics* (pp. 253–348). Princeton: Princeton University Press.
Roth, A., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *American Economic Review*, *81*, 1068–1095.
Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, *50*, 97–109.
Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschraenkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In: H. Sauermann (Ed.), *Beiträge zurexperimentellen Wirtschatsforschung* (pp. 136–168). Tübingen: J.C.B. Mohr.
Solow, R. M. (1979). Another possible source of wage stickiness. *Journal of Macroeconomics*, *1*, 79–82.
Stahl, I. (1972). *Bargaining theory*. Stockholm: Economic Research Institute.
Thaler, R. H. (1988). Anomalies: The ultimatum game. *Journal of Economic Perspectives*, *2*, 195–205.
Thaler, R. H., & Johnson, E. J. (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science*, *36*, 643–660.
Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, *59*, S251–S278.
Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics*, *106*, 1039–1061.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.
Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, *42*, 1676–1690.

APPENDIX I. EXPERIMENT 1 INSTRUCTIONS

Dividing Losses

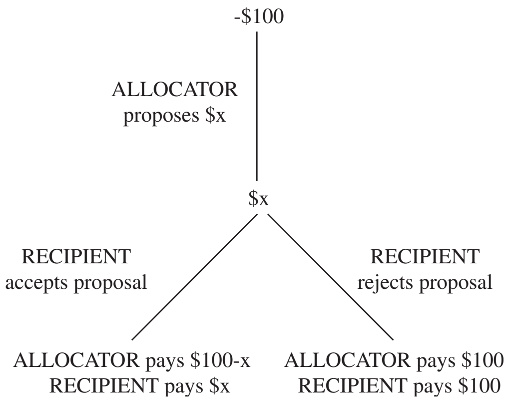
You have been designated ALLOCATOR [RECIPIENT]. In this case, you and another student have been selected to divide **losses** of $100, provided you agree how to share it.

1. The ALLOCATOR submits an ultimatum to the RECIPIENT:
   "I propose that you pay x and that I pay $100-x$."
2. The RECIPIENT can either

(i) agree to the ultimatum proposal, in which case the RECIPIENT pays x and the ALLOCATOR pays $100-x$;

or

(ii) reject the proposal, in which case **both players pay $100**.



As ALLOCATOR, I would propose that the RECIPIENT pays____
[As RECIPIENT, the maximum offer I would accept (the most I would pay) is ____]

Dividing Gains

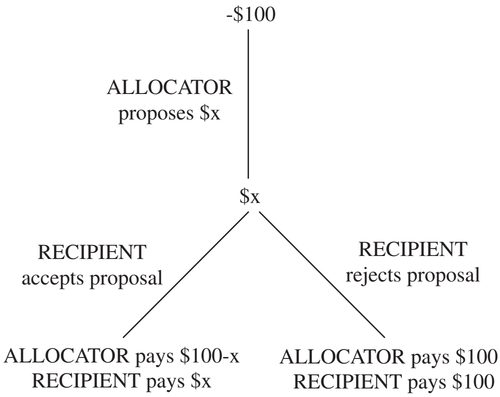
You have been designated ALLOCATOR [RECIPIENT]. In this case, you and another student have been selected to divide **losses** of $100, provided you agree how to share it.

1. The ALLOCATOR submits an ultimatum to the RECIPIENT: "I propose that you pay x and that I pay $100-x$."
2. The RECIPIENT can either

   (i) agree to the ultimatum proposal, in which case the RECIPIENT pays x and the ALLOCATOR pays $100-x$;

   or

   (ii) reject the proposal, in which case **both players pay $100**.



As ALLOCATOR, I would offer the RECIPIENT ____
[As RECIPIENT, the minimum offer I would accept is ____]

APPENDIX II. EXPERIMENT 2 INSTRUCTIONS

Player 1 Gains Game

PLAYER ID ____

You are Player 1. You and a randomly assigned Player 2 have an opportunity to earn some money.

The experimenter owes you pair $10.26. You will offer some amount of the money (less than or equal to $10.26) to Player 2. At the same time Player 2 will record the smallest offer he is willing to accept (less than or equal to $10.26). If his demand is less than or equal to your offer we say he *accepts* the offer. Then you get $10.26 minus the amount you offered, while Player 2 receives the amount you offered. If his demand is strictly greater than your offer we say is *refuses* the offer. Then you both get no money; the money returns to the experimenter.

To be sure you understand the procedure, fill in the blanks in the examples below and wait for someone to check your answers.

EXAMPLE: The experimenter owes your pair $10.26. You offer $Y to Player 2. ($Y ⩽ $10.26)
A: Player 2 is willing to accept $Z ⩽ $Y. Player 2 receives $____, you receive $ ____.

B: Player 2 is willing to accept $Z > $Y. Player 2 receives $ ____, you receive $ ____.

Any questions?

You will now offer some amount of the $10.26 to Player 2. Take as much or as little time as you like to decide. Once you have decided on your offer, write it in the appropriate place on the next page. We will compare your offer to the appropriate Player 2's recorded demand. If your pair has reached an agreement, we will pay each of you the agreed-upon amounts. If your pair has not reached an agreement, we will pay each of you nothing. At no time will Player 2 know your identity, nor will you know his.

Player 2 Gains Game

PLAYER ID ____

You are Player 2. You and a randomly assigned Player 1 have an opportunity to earn some money.

The experimenter owes you pair $10.26. Player 1 will offer some amount of the money (less than or equal to $10.26) to you. At the same time you will record the smallest offer you are willing to accept (less than or equal to $10.26). If your demand is less than or equal to the offer made we say you *accept* the offer. Then Player 1 gets $10.26 minus the amount he offered you, while you receive the amount offered to you. If your demand is strictly greater than the amount offered to you we say you *refuse* the offer. Then you both get no money; the money returns to the experimenter.

To be sure you understand the procedure, fill in the blanks in the examples below and wait for someone to check your answers.

EXAMPLE: The experimenter owes your pair $10.26. Player 1 offer $Y to you. ($Y ⩽ $10.26)
A: You are willing to accept $Z ⩽ $Y. You receive $____, Player 1 receives $ ____.
B: You are willing to accept $Z > $Y. You receive $ ____, Player 1 receives $ ____.

Any questions?

You will now record the smallest offer out of $10.26 as you are willing to accept. Take as much or as little time as you like to decide. Once you have decided on your demand, write it in the appropriate place on the next page. We will compare your demand to the appropriate Player 1's recorded offer.

If your pair has reached an agreement, we will pay each of you the agreed-upon amounts. If your pair has not reached an agreement, we will pay each of you nothing. At no time will Player 1 know your identity, nor will you know his.

Player 1 Losses Game

PLAYER ID ____

You are Player 1. You and a randomly assigned Player 2 have an opportunity to minimize your losses by agreeing on how to divide a loss.

Your pair owes the experimenter \$9.72. You will offer that Player 2 pay some of the loss (less than or equal to \$9.72). At the same time Player 2 will record the largest amount of the loss he is willing to pay (less than or equal to \$9.72). If he is willing to pay at least as much as you offer than he pay, we say he *accepts* the offer. Then you pay the \$9.72 minus the amount you offered, while Player 2 pays the amount you offered. If he is willing to pay strictly less than you offer that he pay, we say he *refuses* the offer. Then you each pay \$9.72.

To be sure you understand the procedure, fill in the blanks in the examples below and wait for someone to check your answers.

EXAMPLE: Your pair owes the experimenter \$9.72. You offer \$Y of the loss to Player 2. ($\$Y \leqslant \$9.72$)
A: Player 2 is willing to pay $\$Z \geqslant \$Y$. Player 2 pays \$ ____, you pay \$ ____.
B: Player 2 is willing to pay $\$Z < \$Y$. Player 2 pays \$ ____, you pay \$ ____.

Any questions?

You will now specify how much of the \$9.72 loss you offer for Player 2 to pay. Take as much or as little time as you like to decide. Once you have decided on your offer, write it in the appropriate place on the next page. We will compare your offer to the appropriate Player 2's recorded demand. If your pair has reached an agreement, we will collect the agreed-upon amounts from each of you. If your pair has not reached an agreement, we will collect \$9.72 from each of you. At no time will Player 2 know your identity, nor will you know his.

Player 2 Losses Game

PLAYER ID ____

You are Player 2. You and a randomly assigned Player 1 have an opportunity to minimize your losses by agreeing on how to divide a loss.

Your pair owes the experimenter \$9.72. You will offer that Player 2 pay some of the loss (less than or equal to \$9.72). At the same time you will record the largest amount of the loss you are willing to pay (less than or equal to \$9.72). If you are willing to pay at least as much as Player 1 has offered to have you pay, we say you *accept* the offer. Then Player 1 pays \$9.72 minus the amount he offered you, while you pay the amount he offered you. If you are willing to pay strictly less than Player 1 has offered to have you pay, we say you *refuse* the offer. Then you each pay \$9.72.

To be sure you understand the procedure, fill in the blanks in the examples below and wait for someone to check your answers.

EXAMPLE: Your pair owes the experimenter \$9.72. Player 1 offers \$Y of the loss to you. ($\$Y \leqslant \$9.72$)
A: You are willing to pay $\$Z \geqslant \$Y$. You pay \$ ____, Player 1 pays \$ ____.
B: You are willing to pay $\$Z < \$Y$. You pay \$ ____, Player 1 pays \$ ____.

Any questions?

You will now record the largest amount you are willing to pay out of \$9.72. Take as much or as little time as you like to decide. Once you have decided on your demand, write it in the appropriate place on the next page. We will compare your demand to the appropriate Player 1's recorded offer. If your pair has reached an agreement, we will collect the agreed-upon amounts from each of you. If your pair has not reached an agreement, we will collect \$9.72 from each of you. At no time will Player 2 know your identity, nor will you know his.

This page intentionally left blank

BEHAVIORAL ASPECTS OF LEARNING IN SOCIAL NETWORKS: AN EXPERIMENTAL STUDY☆

Syngjoo Choi, Douglas Gale and Shachar Kariv

ABSTRACT

Networks are natural tools for understanding social and economic phenomena. For example, all markets are characterized by agents connected by complex, multilateral information networks, and the network structure influences economic outcomes. In an earlier study, we undertook an experimental investigation of learning in various three-person networks, each of which gives rise to its own learning patterns. In the laboratory, learning in networks is challenging and the difficulty of solving the decision problem is sometimes massive even in the case of three persons. We found that the theory can account surprisingly well for the behavior observed in the laboratory. The aim of the present paper is to investigate important and interesting questions about individual and group behavior, including comparisons across networks and information treatments. We find that in order to explain subjects' behavior, it is necessary to take into account the details of the network architecture as well as the information

☆This research was supported by the Center for Experimental Social Sciences (CESS) and the C. V. Starr Center for Applied Economics at New York University.

Experimental and Behavioral Economics
Advances in Applied Microeconomics, Volume 13, 25–61

ISSN: 0278-0984/doi:10.1016/S0278-0984(05)13002-8

structure. We also identify some "black spots" where the theory does least well in interpreting the data.

1. INTRODUCTION

In a modern economy, an individual knows only a small fraction of the information distributed throughout the economy as a whole. Consequently, he has a very strong incentive to try to benefit from the knowledge of others before making an important decision. Sometimes he learns from public sources, books, newspapers, the Internet, etc. At other times he needs information that is not available from these public sources and then he must try to find the information in his local environment. In social settings, where an individual can observe the choices made by other individuals, it is rational for him to assume that those individuals may have information that is not available to him and then to try to infer this information from the choices he observes. This process is called *social learning.* The literature on social learning contains numerous examples of social phenomena that can be explained in this way. In particular, it has been argued that the striking uniformity of social behavior is an implication of social learning.

Much of the social-learning literature has focused on examples of inefficient information aggregation. The seminal papers of Bikhchandani, Hirshleifer and Welch (1992) (BHW) and Banerjee (1992) show that social learning can easily give rise to herd behavior and informational cascades. Herds or cascades can be started by a small number of agents who choose the same action. Subsequently, other agents ignore their own information and join the herd. Once an agent decides to join the herd, his own information is suppressed. Since only a small amount of information is revealed by the agents who started the herd, the herd is likely to have chosen a suboptimal action. Smith and Sørensen (2000) extend the basic model to allow for richer information structures and to provide a more general and precise analysis of the convergence of actions and beliefs.[1]

The models of BHW and Banerjee (1992) are special in several respects. They assume that each agent makes a once-in-a-lifetime decision and the decisions are made sequentially. Furthermore, when each agent makes his decision, he observes the decisions of all the agents who have preceded him. An alternative model, described in Gale and Kariv (2003), assumes that agents are part of a social network and can only observe the actions of agents to whom they are connected through the network.[2] Information percolates through the network as an agent's action is first observed by his neighbors and then (indirectly) by his neighbors' neighbors and so on. In order to model the diffusion of information through the network, it is natural to assume that agents can revise their decisions as more information becomes available. More precisely, Gale and Kariv (2003) assume that all agents make simultaneous and repeated decisions.

Whereas herd behavior arises quickly in the sequential model of BHW and Banerjee (1992), in the social-network model learning may continue for some time as information diffuses through the network. Paradoxically, in spite of the agents' limited powers of observation, the informational efficiency of the network model may be greater than the sequential decision model.

Another difference between the network model and the sequential model is related to the complexity of decision-making. Because the history of actions is not common knowledge, the agents in the network model have to make inferences not just about their neighbors' private signals, but also about their neighbors' observations and inferences about *their* neighbors. The greater complexity of the learning process raises questions about the plausibility of a rational learning model. The absence of common knowledge of the history of actions requires agents to hold beliefs about beliefs about beliefs … about the actions and information of agents they cannot observe directly. This has led some authors, e.g. Bala and Goyal (1998), to suggest that models of bounded rationality are more appropriate for describing learning in networks.

Whether individuals can rationally process the information available in a network is ultimately an empirical question. To test the relevance of the theory, Choi, Gale, and Kariv (2005) (CGK) examined the behavior of subjects in a variety of three-person networks based on the model of Gale and Kariv (2003). The information structure for the experiments was adapted from BHW and the experiment utilized the procedures of Anderson and Holt (1997).[3] The family of three-person networks includes several nontrivial architectures, each of which gives rise to its own distinctive learning patterns. CGK studied three of these networks: the complete network, in which each agent can observe the other two agents, and two incomplete networks; the circle, in which each agent observes one other agent; and the star, in which the agent in the center of the star is connected to the two agents at the periphery.

In the experimental design, there are two decision-relevant events equally likely to occur *ex ante* and two corresponding signals. Signals are informative in the sense that there is a probability higher than 1/2 that a signal matches the label of the realized event. We allow subjects to be of two types: informed agents, who receive a private signal, and uninformed agents, who know the true prior probability distribution of the states but do not receive a private signal. Each experimental round consisted of six decision turns. At each decision turn, the subject is asked to predict which of the events has taken place, basing his forecast on a private signal and the history of his neighbors' past decisions.

CGK found that the theory can account for the behavior observed in the laboratory in most of the networks and informational treatments. In fact, the rationality of behavior is striking. The error rates calculated using deviations from the equilibrium strategies implied by the Gale–Kariv model are positive but moderate. To account for these errors, CGK adapted the model to allow for the effect of the "trembling hand" and estimated a recursive Quantal Response Equilibrium (QRE) model. They found that the QRE model appears to account for the large-scale features of the data with surprising accuracy. Of course, there may be other ways of accounting for the data, and one must consider whether the apparent success of the theory may be due to special features of the experimental design, such as the simplicity of the networks chosen or the fact that the optimal strategies are well approximated by simple heuristics.

The data generated by the CGK experiments can also be used to address a variety of important and interesting questions about individual and group behavior. In this paper, we use the same data set to investigate behavioral aspects of individual and group behavior, including comparisons across networks and information treatments. We also look more closely at the data in order to identify the "black spots" where the theory does least well in interpreting the data and ask whether additional "behavioral" explanations might be needed to account for the subjects' behavior.

Much of the theoretical and experimental literature on social learning has focused on the phenomenon of herd behavior, which is said to occur when every agent acts like others do, even though he would have chosen a different action on the basis of his own information alone. In this sense, individuals rationally "ignore" their own information and "follow the herd." Furthermore, since actions aggregate information poorly, despite the available information, herds need not adopt an optimal action. Therefore, the efficiency of information aggregation is one of the main concerns in the study of social learning.

We find that the experimental data exhibit a strong tendency toward herd behavior and a marked efficiency of information aggregation. The data also suggest that there are significant and interesting differences in individual and group behavior among the three networks and three information treatments. We argue that these differences can be explained by the symmetry or asymmetry of the network or the information treatment and the resultant differences in the amount of common knowledge.

We first provide information about the evolution of herd behavior. First, diversity of private information initially leads to diversity of actions, which then gives way to uniformity as the subjects learn by observing the actions of their neighbors (Result 1). Second, although convergence to a uniform action is quite rapid, frequently occurring within two to three turns, there are significant differences between the behavior of different networks (Result 2) and information treatments (Result 3). Finally, most herds tend to entail correct decisions (Result 4), which is consistent with the predictions of the parametric model underlying our experimental design.

Next, we use expected payoff calculations to measure the efficiency of the decisions made by our subjects in the laboratory. We compare the levels of efficiency across networks (Result 5) and information treatments (Result 6). We then provide information as to why the evolution of actual efficiency depends on the information treatment (Result 7). We also discuss the behavioral regularities at the individual level and how they are affected by the network (Result 8) and information treatment (Result 9). Comparing individual behavior indicates that there is indeed high variation in individual behavior across subjects.

Finally, we examine how well the theory approximates the actual behavior observed in the laboratory. We begin by computing the optimal strategies as predicted by the theoretical model and use these to compute the level of rationality (Result 10). At the first and second turns, the error rates are uniformly fairly low, although there are significant differences across information sets. In the sequel, we identify some "black spots" in which there are sharp drops in rationality and discuss the departures from the predictions of the theory that ought to be considered in future work.

The rest of the paper is organized as follows. We describe the theoretical model and the experimental design in Section 2. In Section 3, we introduce three summary measures of subject behavior that are used in the sequel. The results are contained in Section 4. We group our results under three headings, relating to group behavior, efficiency, and rationality. The important features of the QRE analysis are summarized in Section 5. Some concluding remarks and important topics for further research are contained in Section 6.

2. THEORY, PREDICTIONS AND DESIGN

In this section, we describe the theory on which the experimental design is based and the design itself. Gale and Kariv (2003) provide a more extensive description and analysis of the model and CGK provide a fuller description of the experimental design (the experimental instructions are available upon request).

2.1. The Model

We restrict attention to three-person networks. Each network has three locations, indexed by $i = A, B, C$, and, at each location i, a single (representative) agent i who maximizes his short-run payoff in each period. The network is a *directed graph* represented by a family of sets $\{N_i : i = A, B, C\}$, where N_i denotes i's neighbors, i.e. the set of agents $j \neq i$ who can be observed by agent i.

We study three networks: the *complete network*, in which each agent can observe the actions chosen by the other agents; the *star*, in which one agent, the center, can observe the actions of the other two peripheral agents, and the peripheral agents can only observe the center; and the *circle*, in which each agent can observe only one other agent and each agent is observed by one other agent. The networks are illustrated in Fig. 1, where an arrow pointing from agent i to agent j indicates that $j \in N_i$.

There are two equally likely *states of nature* represented by two urns, a red urn (R) and a white urn (W). The red urn contains two red balls (r) and one white ball (w); the white urn contains two white balls and one red ball. One of these urns is randomly selected by nature before the start of the game. This is the ball-and-urn social-learning experiments paradigm of Anderson and Holt (1997).

Once the urn is chosen, each agent receives a private signal with probability q. In this experiment, the signal consists of seeing the color of a ball that is randomly drawn from the urn (with replacement). Signals are informative in the sense that there is a probability $2/3$ that a signal matches the label of the realized event. With probability $1-q$ the agent does not receive a signal. An agent who receives a signal is called *informed*; otherwise he is called *uninformed*. The information structure is summarized in the diagram below.

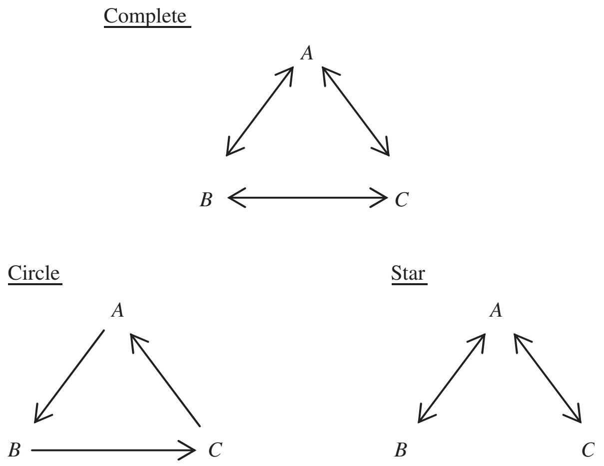


Fig. 1. The Complete, Circle and Star Networks with Three Agents.
Note: A line segment between any two types represents that they are connected and the arrowhead points to the participant whose action can be observed.

	State	
Signal	R	W
$\varnothing$	$1-q$	$1-q$
r	$2q/3$	$q/3$
w	$q/3$	$2q/3$

An uninformed agent has a uniform prior across the two states. An informed agent has a posterior probability that depends on his signal. For example, if he sees a red ball, he believes the true state is red with probability 2/3, and if he sees a white ball, he believes the true state is white with probability 2/3.

2.2. The Decision Problem

After the state of nature has been determined and some agents are informed, a simple guessing game is played. There are six stages or *decision turns* in the game. At each decision turn, each agent is asked to guess the true state of nature based on the information he has at that turn. The choice of agent i at date t is denoted by $x_{it} \in \{R, W\}$. An agent receives a positive payoff for guessing the correct state and zero for the wrong state. The payoff is received at the end of the game, after all the decisions are made, so there is no learning from payoffs.

At the first turn, the agent's information consists of his private signal, if he has one, and the structure of the game, which is common knowledge. An informed agent will maximize his expected payoff by choosing the state he thinks is more likely. An uninformed agent regards each state as equally likely and so is indifferent between them. We assume that whenever an agent has no signal, he chooses each action with equal probability; and when an agent is indifferent between following his own signal and following someone else's choice, he follows his own signal. One may assume different tie-breaking rules, but our experimental data support this specification.

At the end of the first turn, after all the agents have made their decisions, agents are allowed to observe the decisions of all the agents to whom they are connected by the network. At the second turn, the agents update their probability beliefs about the true state of nature based on the information obtained at the first turn, and are again asked to make their best guess of the true state.

After all the agents have made their decisions, they observe what their neighbors have chosen. This procedure is repeated until the agents have made six decisions.

Note that we restrict attention to equilibria in which myopic behavior is optimal, i.e. it is rational for agents in equilibrium to choose the actions that maximize their short-run payoffs at each date. A careful analysis shows that in our setting there is no incentive to sacrifice short-run payoffs in any period in order to influence the future play of the game, because full revelation obtains if agents switch actions whenever they are indifferent between continuing to choose the same action and switching in the next period.

2.3. The Complete Network

We illustrate the play of the game using the complete network, in which each agent can observe the other two, as an example. The complete network is defined by the following conditions:

$$N_A = \{B, C\}, \quad N_B = \{A, C\}, \quad N_C = \{A, B\}$$

There are three different information treatments, corresponding to different values of the probability q that an individual agent receives a signal. We refer to them as *full information* ($q = 1$), *high information* ($q = 2/3$), and *low information* ($q = 1/3$).

Full information: In the case of full information ($q = 1$), all three agents are informed. At the first turn, each agent chooses the state he thinks is more likely, i.e. R if he draws r and W if he draws w. Each agent's action reveals his private information, and since each agent can observe the actions of the other two, the private information becomes common knowledge at the end of the first round.

This means that at the beginning of the second turn, all the agents have the same information, they all choose the same action, and no further information is revealed. Both the actions and the beliefs of the agents will remain constant from the second turn onward. This is a simple example of herd behavior. Although the decisions from the second turn onward are based on all the information available, the herd will be incorrect with positive probability.

High information: The game changes in two ways when $q = 2/3$. First, an agent may be informed or uninformed, and second, the other agents do not know whether he is informed or not. Obviously, the informational value of observing another agent's action is smaller than in the full-information case.

Suppose, for example, that the pattern of signals is given by the diagram below.

Period	Agent/Signal		
	A	B	C
	r	w	$\varnothing$
1	R	W	W
2	W	W	W
3	W	W	W
…	…	…	…

At the first decision turn, agents A and B follow their signals, while C, being uninformed, randomizes and ends up choosing W. At the second turn, A sees that two others have chosen W, but he knows that they are informed with probability 2/3 and must take this into account in updating his belief of the true state. If B and C had observed exactly one w signal, that would make A indifferent, so it becomes a question of whether they observed two

w signals or no signals. When $q = 2/3$, two w signals is more likely than no signals, so A will switch. By similar reasoning, at the second turn B will not switch and C will be indifferent. Assuming that C does not switch when indifferent, we have reached an absorbing state at the second decision turn.

If C were to switch to R at the second decision turn, this would signal that he is uninformed. At that point, A should switch back to R at the third turn, thus revealing that he is informed. Then the fact that B continues to choose W at the fourth turn reveals that he is informed. At the fifth turn, everyone is indifferent, knowing that there is one w and one r signal, and they can continue to choose different actions in the remainder of the game. The lack of common knowledge (of private signals) postpones the development of herd behavior and allows information to be revealed over a longer period of time. Whether this results in better decision-making overall depends on the particular realization of the signals.

Low information: With low information, the probability that the other agents are uninformed increases. In that case, an informed agent will continue to follow his own information at the second date. Even if A observes the other two agents choosing W, the possibility that they are both uninformed is so high that he would rather ignore their actions and follow his own signal. This will reveal that A is informed at the second turn. At this point, C will imitate A, because of the possibility that B is uninformed, and switch to R. This reveals C to be uninformed.

At the third decision turn, B will be indifferent. If we assume that he follows his own signal when indifferent, he will be revealed to be informed, and from that point onward everyone is indifferent between the two states and may continue to choose different actions. This pattern is given by the diagram below. Comparing the high- and low-information examples, we can see that one effect of reducing q is to make it clearer who is informed and who is uninformed.

Period	Agent/Signal		
	A	B	C
	r	w	$\varnothing$
1	R	W	W
2	R	W	W
3	R	W	R
…	…	…	…

2.4. The Star

The first incomplete network we examine is the star, which is defined by the following conditions:

$$N_A = \{B, C\}, \quad N_B = \{A\}, \quad N_C = \{A\}$$

At each decision turn, agent A is informed about the entire history of actions taken, whereas B and C have imperfect information. The asymmetry of this network gives rise to effects not found in the other networks. The central agent A plays an important role because it is only through him that information can be transmitted between the other two.

Suppose that $q = 1/3$ and the signals are shown in the diagram below. At the first decision turn, A and B follow their signals and C randomizes and chooses R, say. At the second turn, A observes the complete history of actions but B and C only observe A's action. B will continue to choose W because he is informed, whereas A might be uninformed, and C will continue to choose R, because he is uninformed, whereas A might be informed. A will continue to choose R because, from his point of view, B and C cancel each other out.

At the third decision turn, A knows that B is informed and C may be informed or uninformed, so A continues to choose R. B now knows that A is either informed or observed that C chose R at the first turn. Eventually, if A and C continue to choose R, their information overwhelms B, and B will switch to R.

Period	Agent/Signal		
	A	B	C
	r	w	$\varnothing$
1	R	W	R
2	R	W	R
3	R	W	R
…	…	…	…

2.5. The Circle

The second incomplete network we examine is the circle, in which each agent observes exactly one of the others. The circle is defined by the following conditions:

$$N_A = \{B\}, \quad N_B = \{C\}, \quad N_C = \{A\}$$

The peculiarity of the circle is that, while the equilibrium strategies are very simple, the analysis is quite subtle because of the lack of common knowledge. Assuming that $q = 2/3$ and the signals are given in the diagram below, we can trace out one possible evolution of play.

Period	Agent/Signal		
	A	B	C
	r	w	$\varnothing$
1	R	W	W
2	R	W	R
3	R	W	R
4	W	W	R
…	…	…	…

At the first turn, the informed agents follow their signals and the uninformed agent randomizes (here we assume he chooses W). At the second turn, C switches to R because he is uninformed and observes A who might be informed. Imitating the behavior of the neighboring agent is always the optimal strategy for an uninformed agent. Conversely, it is always optimal for an informed agent to follow his signal, though this is far from obvious. Agent B, seeing C switch to R, will know that C is uninformed and has observed A choose R. If C continues to choose R, this will tell B that A has continued to choose R, which means that A is informed.

At this point, B is indifferent between R and W, and we will assume he continues to follow his signal. All that A knows is that B continues to choose W. He cannot infer what C is doing, so he simply has to rely on Bayes rule to take account of all the possibilities. A lengthy calculation would show that it is optimal for A to continue choosing R, but that eventually he will become indifferent. In the limit, everyone will be indifferent except for C.

2.6. Summary

In summary, the above examples have illustrated several features of the theory that can be tested in the laboratory. First, initially, diversity of private information leads to diversity of actions. But, as agents learn by observing the actions of their neighbors diversity is over time replaced by uniformity. Second, convergence to a uniform action is quite rapid, typically occurring within two to three periods. Thus, what happens in those first few periods is crucial for the determination of the social behavior. Third, significant differences can be identified in the behavior of different networks. In particular, in the complete network learning stops almost immediately, while in the incomplete networks learning can continue for a longer time. Finally, despite the fact that agents suppress their own information and follow a herd, a careful analysis shows that in all treatments, except with very small probability in the complete network under high information, herds always adopt an action that is optimal relative to the total information available to agents.

The theory clearly suggests that even in the three-person case the process of social learning in networks can be complicated, particularly in the incomplete networks. That is why we believe that insights obtained from an experiment may provide better understanding of social learning in networks.

2.7. Experimental Design

We studied three different network structures (the complete network, the star and the circle) and three different information treatments ($q = 1$, $2/3$, $1/3$). The network structure and the information treatment were held constant throughout a given experimental session. In each session, the network positions were labeled A, B, or C. A third of the subjects were designated type-A participants, one third type-B participants and one third type-C participants. The participant's type, A, B, or C, remained constant throughout the session.

Each session consisted of 15 independent rounds, and each round consisted of six decision turns. The following process was repeated in all 15 rounds. Each round started with the computer randomly forming three-person networks by selecting one participant of type A, one of type B, and one of type C. The networks formed in each round depended solely upon chance and were independent of the networks formed in any of the other rounds. The computer also chose one of two equally probable urns, labeled R and W, for each network and each round. The urn remained constant throughout the round. The choice of urn was independent across networks and across rounds.

When the first round ended, the computer informed subjects which urn had actually been chosen. Then, the second round started by having the computer randomly form new groups of participants in networks and select an urn for each group. This process was repeated until all the 15 rounds were completed. Earnings at each round were determined as follows: at the end of the round, the computer randomly selected one of the six decision turns. Everyone whose choice in this decision turn matched the letter of the urn that was actually used earned $2. All others earned nothing. This procedure insured that at each decision turn, subjects would make their best guess as to what urn had been chosen.

The data were generated by experiments run during the summer and fall of 2003 at the Center for Experimental Social Science (CESS) at New York University. In total, we have observations from 156 subjects who had no previous experience in networks or social-learning experiments. Subjects were recruited from undergraduate classes at New York University, and each subject participated in only one of the nine experimental sessions. The diagram below summarizes the experimental design (the entries have the form a/b, where a is the number of subjects and b the number of observations per type and turn).

	Information		
Network	Full	High	Low
Complete	18/90	15/75	18/90
Star	18/90	18/90	18/90
Circle	18/90	18/90	15/75

3. THREE MEASURES

For a better understanding of the decision mechanism of our subjects we organize the data according to three measures: *stability, uniformity*, and *efficiency*. Next, we explain the three measures and their motivations.

Much of the theoretical and experimental literature on social learning has focused on the related phenomena of informational herd behavior and informational cascades. Herd behavior is said to occur when, after some point, all agents choose the same action. A herd may arise even if the agents would have chosen a different action on the basis of their own information alone. In this sense, agents rationally "ignore" their own information and "follow the herd." We characterize herd behavior by two related phenomena, stability and uniformity of actions.

Stability: At each turn t, stability is measured by the proportion of subjects who continue to choose the action they chose at turn $t-1$. For each network a stability variable is denoted by S_t and defined by

$$S_t = \frac{\#\{i : x_{it} = x_{it-1}\}}{n}$$

We report averages of S_t across different networks.

Uniformity: At each turn t, uniformity is measured by a score function that takes the value 1 if all subjects act alike and takes the value 0 otherwise. For each network a uniformity variable is denoted by U_t and defined by

$$U_t = \begin{cases} 1 & \text{if } x_{it} = x_{jt}, \forall i,j, \\ 0 & 0 \text{ otherwise} \end{cases}$$

We report averages of U_t across different networks.

Herd behavior arises in the laboratory when, from some decision turn onward, all subjects take the same action. Notice that uniformity of actions at some date t will persist and lead to herd behavior if and only if stability takes the value 1 at all subsequent stages or decision turns.

As the examples in the preceding section have illustrated, the theory predicts that convergence to a uniform action typically occurs within two to three periods. Furthermore, except with very small probability in the complete network under high information, herds always adopt the optimal action. As a benchmark for our empirical analysis of stability and uniformity, we first calculated the values of these measures predicted by the theory. The theoretical predictions are derived with the help of simulations, which are summarized in Table 1 and show, turn by turn, the average level of stability and uniformity and the percentage of herd behavior. The numbers in parentheses are the fractions of herds that choose the wrong action, defined relative to the information available.

Informational efficiency of markets is a natural question for economists, and the efficiency of information aggregation is one of the main concerns in the study of social learning. A central result of the literature is that herd behavior may result in most agents choosing the wrong action (where the right action is defined relative to the information available in the economy). This outcome is both informationally inefficient and Pareto inefficient. This failure of information aggregation is explained by two facts. First, an

Table 1. Theoretical Results: Uniformity, Stability and Herd Behavior, Turn by Turn, under the Different Information Structures and Networks (Average Level of Uniformity and Stability and the Percent of Rounds in which Subjects Followed a Herd from that Turn on).

		Decision turn						
Uniformity		1	2	3	4	5	6	Average
Full information	Complete	0.33	1.00	1.00	1.00	1.00	1.00	0.89
	Star	0.33	0.56	1.00	1.00	1.00	1.00	0.81
	Circle	0.33	0.33	0.33	0.33	0.33	0.33	0.33
High information	Complete	0.29	0.78	0.86	0.89	0.89	0.89	0.76
	Star	0.29	0.54	0.80	0.86	0.86	0.85	0.70
	Circle	0.29	0.47	0.58	0.58	0.58	0.58	0.51
Low information	Complete	0.26	0.49	0.77	0.78	0.77	0.76	0.64
	Star	0.26	0.54	0.68	0.88	0.86	0.86	0.68
	Circle	0.26	0.43	0.65	0.71	0.71	0.71	0.58
Stability		1	2	3	4	5	6	Average
Full information	Complete		0.78	1.00	1.00	1.00	1.00	0.96
	Star		0.93	0.85	1.00	1.00	1.00	0.96
	Circle		1.00	1.00	1.00	1.00	1.00	1.00
High information	Complete		0.68	0.89	0.97	0.97	0.97	0.90
	Star		0.78	0.85	0.95	0.98	0.98	0.91
	Circle		0.83	0.94	0.99	0.99	0.99	0.95
Low information	Complete		0.67	0.82	0.90	0.91	0.91	0.84
	Star		0.67	0.82	0.88	0.93	0.92	0.84
	Circle		0.67	0.78	0.93	0.91	0.90	0.83
Herds		1	2	3	4	5	6	Length
Full information	Complete	33.33	100.00	100.00	100.00	100.00	100.00	5.3
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
	Star	33.33	55.56	100.00	100.00	100.00	100.00	4.9
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
	Circle	33.33	33.33	33.33	33.33	33.33	33.33	2.0
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
High information	Complete	28.77	78.35	85.40	88.14	88.19	88.57	4.6
		(0.00)	(1.71)	(1.57)	(1.52)	(1.52)	(1.51)	
	Star	28.75	54.21	79.55	85.12	85.15	85.45	4.2
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
	Circle	28.70	46.60	57.72	57.77	57.86	58.43	3.1
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
Low information	Complete	25.80	48.68	71.69	74.18	74.37	76.21	3.7
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
	Star	25.99	54.33	66.77	82.95	83.23	85.58	4.0
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
	Circle	25.93	43.21	65.43	65.76	66.49	71.01	3.4
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	

agent's action is a coarse signal of his private information, and second, after some point, agents suppress their private information and join the herd, so that only a small fraction of the private information in the game may ever be revealed.

Like Anderson and Holt (1997), we use expected payoff calculations to measure the efficiency of the decisions made by our subjects in the laboratory. As a benchmark we use the payoff to a hypothetical agent who has access to all private signals in his network. Define the *efficient expected payoff* to be the expected earnings of an agent who makes his decision based on the entire vector of private signals; define the *private-information expected payoff* to be the expected earnings of an informed agent who makes his decision on the basis of his own private signal; and define the *random expected payoff* to be the expected earnings of an agent who randomizes uniformly between the two actions. Finally, for each turn, let the *actual expected payoff* be the expected earnings from the subject's actual decision in the laboratory.

The *sum* (over agents) of the efficient, private-information, random, and actual payoffs, for all rounds, will be denoted by π_e, π_p, π_r and π_a, respectively. We use these payoff calculations to assess the quality of aggregation and use of information within a network.

Efficiency: The efficiency of decisions is measured in two ways:

$$\text{actual efficiency} = \frac{\pi_a - \pi_r}{\pi_e - \pi_r}$$

which is calculated turn by turn, and

$$\text{private-information efficiency} = \frac{\pi_p - \pi_r}{\pi_e - \pi_r}$$

There are two normalizations in our measure of actual efficiency. First, since even uninformed random choices will be correct half the time, we subtract random efficiency π_r from actual efficiency π_a in order to get a more accurate measure of the benefit the subjects get from the information they use. Second, there is more information available in some treatments than in others, so we express the net actual efficiency $\pi_a - \pi_r$ as a fraction of the net efficiency $\pi_e - \pi_r$ that could be achieved if subjects pooled their information. Hence, efficient decisions have an efficiency of one and random decisions have an efficiency of zero. A similar rationale applies to the measure of private-information efficiency. The comparison of actual- and private-information efficiencies is useful in determining the extent to which subjects use the information revealed by their neighbors' actions, i.e. the extent to

Table 2. Theoretical Results: Actual- and Private-Information Efficiencies in all Networks and Information Structures.

Information	Network	Private-information	Actual						
			Turn 1	Turn 2	Turn 3	Turn 4	Turn 5	Turn 6	Average
Full	Complete	0.69	0.69	1.00	1.00	1.00	1.00	1.00	0.95
	Star	0.69	0.69	0.79	1.00	1.00	1.00	1.00	0.91
	Circle	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69
High	Complete	0.61	0.61	0.90	0.96	0.97	0.97	0.97	0.90
	Star	0.61	0.61	0.84	0.96	1.00	1.00	1.00	0.90
	Circle	0.61	0.61	0.81	0.88	0.88	0.88	0.88	0.82
Low	Complete	0.46	0.46	0.77	0.98	1.00	1.00	1.00	0.87
	Star	0.46	0.46	0.77	0.92	1.00	1.00	1.00	0.86
	Circle	0.46	0.46	0.77	0.98	0.98	0.98	0.98	0.86

which they did worse than choosing according to all the information and the extent to which they did better than choosing only according to their private information.

Note that the prediction of the theory is that, except in the circle network under full information in which agents can rationally choose different actions forever, complete learning occurs quite rapidly with the result that an efficient action, i.e., the action that would be chosen if all the signals were public information, will be chosen. Thus, the theory predicts a marked efficiency of information aggregation. Table 2 summarizes the theoretical predictions, which are derived with the help of simulations, in all networks and information structures.

4. RESULTS

4.1. Group Behavior

We characterize herd behavior in terms of stability and uniformity of actions. Our first result provides information about the evolution of herd behavior.

Result 1. There is an upward trend in the degree of uniformity and a high and constant level of stability in all treatments, with the result that, over time, subjects tend to follow a herd more frequently.

Support for Result 1 is presented in Table 3 which shows, turn by turn, the average level of stability and uniformity and the percent of rounds in which subjects followed a herd from that turn on. For comparison purposes, the experimental results presented in Table 3 are given in the same format as the theoretical predictions presented in Table 1. By definition, the number of herds is monotonically non-decreasing over time, but the increase in stability and uniformity is not implied by the definitions. It appears to be the result of learning and information aggregation.

Result 1 confirms that over time subjects are increasingly persuaded by the observed actions and gradually build confidence in the information revealed by their neighbors' actions. In the incomplete networks, some of the information of unobserved subjects is accumulated in the observed actions, so the fact that subjects tend to follow a herd more frequently indicates that they try to extract information of unobserved subjects from the actions they observe. This is consistent with the prediction of the theory that over time more and more information is revealed. The theory, however, also predicts that once agents have chosen the same action and they are not indifferent between the two actions, they have reached an absorbing state and will continue to choose the same action in every subsequent period. In the laboratory, in contrast to this prediction, we sometimes observe deviations from a herd.

Next, we turn to the frequencies of herd behavior in different networks and treatments. Our next results report that, within a given decision turn, in some treatments there is no significant difference between the frequencies of herd behavior, but the situation clearly reverses, particularly in early turns, in other treatments.

Result 2 (Networks). In the complete and star networks, the frequency of herds is highest under full information and lowest under low information; in the circle, the frequency of herds under low information is the same as under high information but lower than under full information.

Result 3 (Information). Under full information, the frequency of herds in the complete network is the same as in the star but higher than in the circle; under high information the frequency of herds in the circle network is the same as in the star but lower than in the complete network; under low information, there are no significant differences between the frequencies of herd behavior in the different networks.

Note that the behavior summarized in these two results is consistent with the theoretical results described in Sections 2 and 3. For example, in the

Table 3. Experimental Results: Stability and Uniformity, Turn by Turn, under the Different Information Structures and Networks (Average Level of Uniformity and Stability and the Percent of Rounds in which Subjects Followed a Herd from that Turn on).

		Decision turn						
Uniformity		1	2	3	4	5	6	Average
Full information	Complete	0.38	0.64	0.74	0.71	0.72	0.71	0.65
	Star	0.32	0.43	0.62	0.64	0.67	0.71	0.57
	Circle	0.31	0.44	0.54	0.53	0.58	0.68	0.51
High information	Complete	0.32	0.55	0.64	0.67	0.75	0.69	0.60
	Star	0.33	0.40	0.61	0.59	0.62	0.52	0.51
	Circle	0.31	0.36	0.52	0.49	0.53	0.49	0.45
Low information	Complete	0.23	0.36	0.46	0.44	0.51	0.63	0.44
	Star	0.20	0.36	0.46	0.38	0.40	0.51	0.38
	Circle	0.25	0.44	0.45	0.51	0.45	0.48	0.43
Stability		1	2	3	4	5	6	Average
Full information	Complete		0.87	0.93	0.94	0.95	0.94	0.92
	Star		0.81	0.83	0.89	0.90	0.96	0.88
	Circle		0.75	0.78	0.77	0.80	0.83	0.79
High information	Complete		0.76	0.84	0.84	0.87	0.86	0.84
	Star		0.71	0.76	0.83	0.83	0.79	0.78
	Circle		0.74	0.74	0.76	0.78	0.77	0.76
Low information	Complete		0.67	0.71	0.72	0.75	0.79	0.73
	Star		0.71	0.73	0.70	0.74	0.77	0.73
	Circle		0.79	0.79	0.84	0.87	0.84	0.83
Herds		1	2	3	4	5	6	Length
Full information	Complete	33.33	58.89	64.44	66.67	66.67	71.11	3.6
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
	Star	28.89	40.00	60.00	62.22	65.56	71.11	3.3
		(0.00)	(0.00)	(2.92)	(8.67)	(8.65)	(11.33)	
	Circle	20.00	31.11	35.56	45.56	51.11	67.78	2.5
		(0.00)	(3.15)	(3.12)	(9.30)	(9.31)	(3.05)	
High information	Complete	25.33	46.67	49.33	58.67	62.67	69.33	3.1
		(10.53)	(5.71)	(5.41)	(4.55)	(6.38)	(7.69)	
	Star	12.22	24.44	32.22	35.56	41.11	52.22	2.0
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(6.68)	
	Circle	13.33	21.11	27.78	32.22	34.44	48.89	1.8
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(10.21)	
Low information	Complete	7.78	14.44	20.00	34.44	43.33	63.33	1.8
		(0.00)	(0.00)	(0.00)	(11.05)	(14.79)	(14.50)	
	Star	6.67	14.44	21.11	22.22	31.11	51.11	1.5
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(11.30)	
	Circle	10.67	18.67	25.33	33.33	38.67	48.00	1.7
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(2.78)	

complete network under full information, a herd must start at the second decision turn for every realization of the private signals. By contrast, under the low-information treatment, subjects do not know who is informed and who is uninformed at the second turn, so learning continues after the second turn and herd behavior is delayed. Thus, herd behavior is more likely and will begin sooner in the full-information treatment. In the circle network, it is optimal for informed subjects to follow their own signals, regardless of the behavior they observe in others, and for uninformed subjects to imitate the behavior they observe. For this reason, we expect herd behavior to develop only if the informed subjects receive identical signals. A herd will develop sooner in the full-information treatment than in the high- and low-information treatments, because it takes time for the uninformed subjects to get on board.

The first evidence about the frequencies of herd behavior is provided in Table 3. The relevant support for Results 2 and 3 comes from Fig. 2 which presents, in graphical form, the data from Table 3 on herd behavior in each network under all information treatments (left panel), and for all networks under each information treatment (right panel). A set of binary Wilcoxon tests indicates that the differences are highly significant.

The right column of Table 3 summarizes, treatment by treatment, the average level of stability and uniformity over all turns and the expected length of herd behavior. Note that in all networks, the expected length of herd behavior is increasing in the probability that an individual subject receives a signal. Also, under full and high information, herd behavior in the complete network is longer than in the star, and in the star is longer than in the circle. Under low information, there are longer herds in the circle than in the star.

Hence, convergence to a uniform action is more rapid in the complete network and under full information because it is common knowledge that all subjects are informed and all actions are common knowledge. In contrast, diversity can continue for a longer time under high and low information and in the star and circle networks. Clearly, the absence of common knowledge makes it harder for subjects to interpret the information contained in the actions of others and requires them to perform complex calculations.

Herd behavior has elicited particular interest because erroneous outcomes may occur despite individual rationality, and they may in fact be the norm in certain circumstances. In the model underlying our experimental design, we note that, except with very small probability in the complete network under high information, herds always adopt an action that is optimal

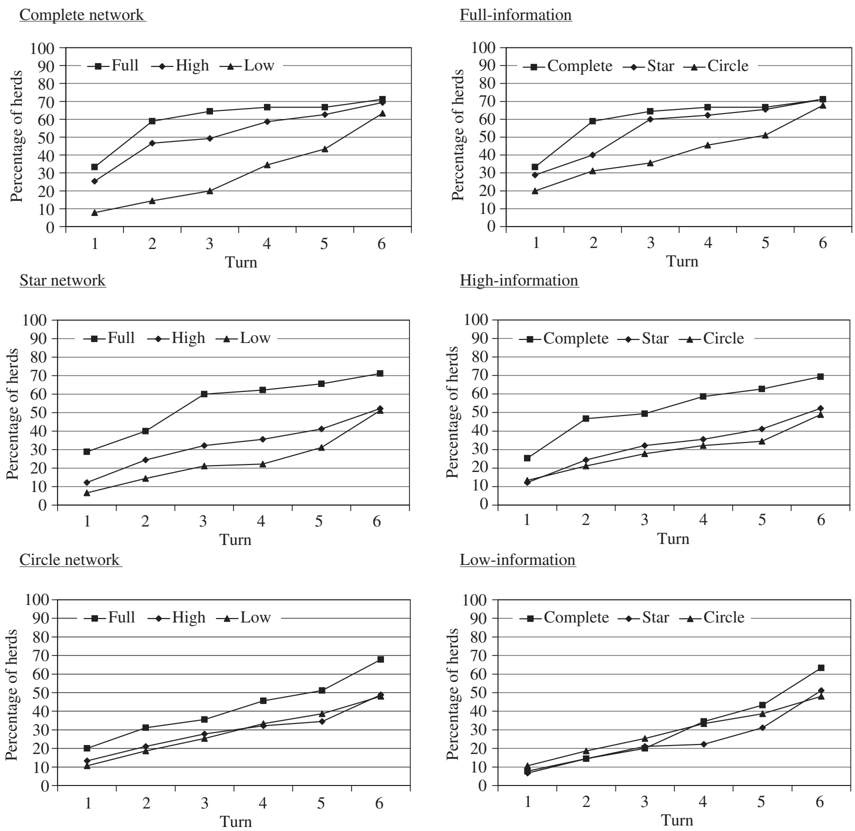


Fig. 2. Herd Behavior in Each Network under All Information Treatments (Left Panel), and for all Networks under Each Information Treatment (Right Panel) (the Percent of Rounds in Which Subsequently All Subjects Acted Alike).

relative to the total information available to agents. Thus, it is particularly interesting that, in the laboratory, almost all herds longer than three decision turns selected the right action, but some differences can be identified in the behavior of different networks. We can report the following result.

Result 4. Relative to the information available, herds entail correct decisions. There are, however, significantly more incorrect herds in the complete network under high information.

Evidence for Result 4 is also provided by Table 3. The numbers in parentheses are the fractions of herds that choose the wrong action, defined relative to the information available. It is noteworthy that herds entail correct decisions even in the star and circle networks in which subjects had imperfect information about the history of decisions.

In summary, we observe several empirical regularities in the experimental data. First, diversity of private information initially leads to diversity of actions, which then gives way to uniformity as subjects learn by observing the actions of their neighbors. Second, convergence to a uniform action can be quite rapid, frequently occurring within two to three periods. Third, herd behavior develops frequently and most herds turned out to be correct.

4.2. Efficiency

We next turn our attention to analyze how efficient our subjects were in using the information revealed by their neighbors' actions. The next results report average actual-efficiency calculations to measure the informational efficiency within a given network, information treatment, and turn.

Result 5 (Network). In the complete network, average actual efficiency is highest under full information and lowest under low information; in the star, average actual efficiency under full information is the same as under high information but higher than under low information; in the circle, the levels of average actual efficiency are the same under all information treatments.

Result 6 (Information). Under full information, average actual efficiency is highest in the complete network and lowest in the circle; under high and low information, there are no significant differences between the levels of average actual efficiency in the different networks.

Table 4, which summarizes, turn by turn, the actual and private-information efficiencies in all networks and information treatments, provides a first indication. Under high and low information, Table 4 also provides the actual efficiency for informed and uninformed individual subjects' decisions. For comparison purposes, the experimental results presented in Table 4 are given in the same format as the theoretical predictions presented in Table 2.

The support for Results 5 and 6 comes from Fig. 3, which presents the data from Table 4 by comparing the total actual efficiency in each network under all information treatments (left panel), and for all networks under

Table 4. Experimental Results: Actual and Private-Information Efficiencies in All Networks and Information Structures for Informed and Uniformed Subjects.

Information	Network		Private-information	Actual						
				Turn 1	Turn 2	Turn 3	Turn 4	Turn 5	Turn 6	Average
Full	Complete		0.716	0.683	0.793	0.829	0.818	0.828	0.839	0.798
	Star		0.746	0.627	0.649	0.707	0.718	0.691	0.681	0.679
	Circle		0.709	0.611	0.526	0.553	0.531	0.580	0.588	0.565
High	Complete	All	0.620	0.487	0.617	0.625	0.713	0.692	0.653	0.631
		Informed	0.877	0.731	0.692	0.744	0.772	0.776	0.720	0.739
		Uninformed	0.000	−0.103	0.435	0.337	0.571	0.489	0.489	0.370
	Star	All	0.688	0.596	0.660	0.707	0.659	0.745	0.580	0.658
		Informed	0.854	0.767	0.765	0.741	0.738	0.816	0.612	0.740
		Uninformed	0.000	−0.118	0.220	0.565	0.329	0.450	0.450	0.316
	Circle	All	0.616	0.613	0.566	0.644	0.604	0.705	0.580	0.618
		Informed	0.831	0.767	0.581	0.637	0.610	0.728	0.561	0.647
		Uninformed	0.000	0.171	0.524	0.662	0.586	0.637	0.632	0.535
Low	Complete	All	0.461	0.271	0.458	0.495	0.414	0.460	0.523	0.437
		Informed	0.975	0.719	0.706	0.783	0.563	0.660	0.701	0.689
		Uninformed	0.000	−0.130	0.236	0.236	0.280	0.280	0.364	0.211
	Star	All	0.440	0.177	0.331	0.389	0.467	0.463	0.545	0.395
		Informed	0.952	0.625	0.543	0.581	0.725	0.673	0.804	0.658
		Uninformed	0.000	−0.207	0.149	0.224	0.245	0.282	0.324	0.169
	Circle	All	0.494	0.452	0.469	0.579	0.590	0.601	0.592	0.547
		Informed	0.957	0.935	0.853	0.836	0.793	0.771	0.689	0.813
		Uninformed	0.000	−0.065	0.060	0.304	0.373	0.419	0.488	0.263

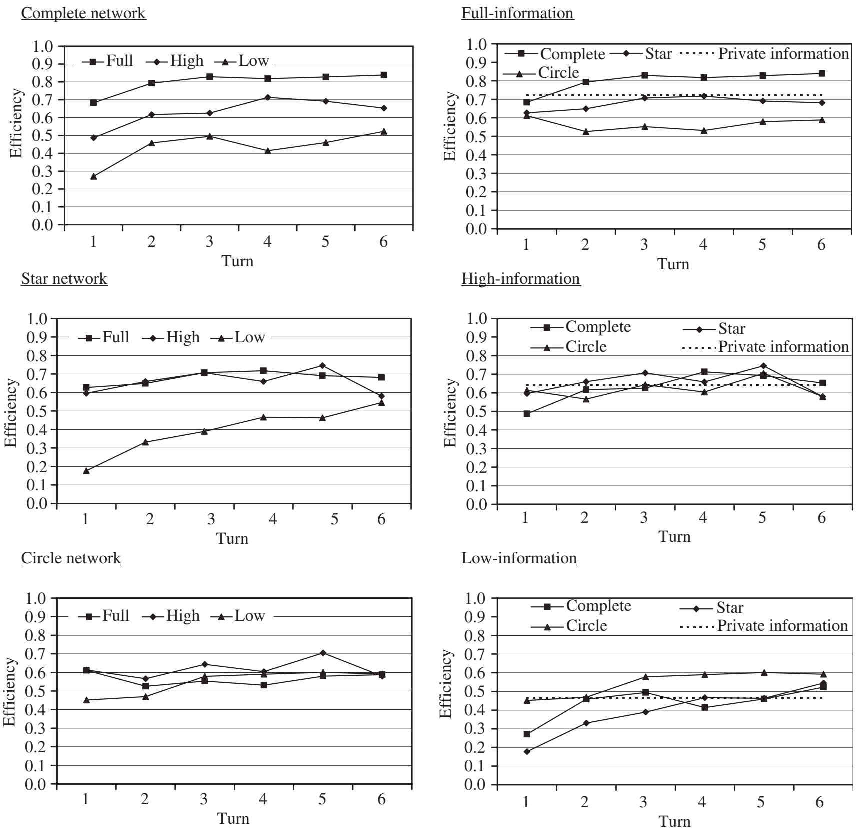


Fig. 3. Actual Efficiency in Each Network under All Information Treatments (Left Panel), and for All Networks under Each Information Treatment (Right Panel), and Average Private-information Efficiency over All Subjects within Each Information Treatment.

each information treatment (right panel). Fig. 3 also depicts the average private-information efficiency over all subjects within each information treatment. A set of binary Wilcoxon tests indicates that all the differences above are highly significant.

Results 5 and 6 emphasize the role of common knowledge in the laboratory. Under full information, it is common knowledge that all subjects are informed. In the complete network, the subjects' actions are also common knowledge, whereas in incomplete networks the actions are not common knowledge. Under full information, the greater efficiency in the complete network compared with incomplete networks can be attributed to the agents' ability to use the greater amount of information available to them. In the other information treatments, subjects are uncertain whether the other participants are informed or uninformed, and this uncertainty appears to prevent them from making use of the additional information available in the complete network.

Our previous results relate to the levels of actual efficiency across treatments, but it appears that efficiency increases over time only in the complete and star networks under low information. Our next result provides information as to why the evolution of actual efficiency depends on the information treatment, i.e. the probability that an individual subject receives a signal. We use the same payoff calculations, but sum π_e, π_p, π_r and π_a over informed and uninformed decision points and not over individual subjects.

Result 7. Whereas the actual efficiency in informed decision points falls slightly from the first to the last decision turn, the actual efficiency in uninformed decision points increases over time. The rate of increase of actual efficiency is greater under low information than under high and full information.

Fig. 4 provides the support for Result 7 by comparing the efficiencies of informed and uninformed decisions in each of the networks under low and high information. Since there are more uninformed subjects in the low-information treatment and actual efficiency is increasing for uninformed subjects and decreasing for informed subjects, the rate of increase will be highest under low information. Fig. 4 also depicts the average private-information efficiency over all informed individual subjects within each information treatment, and reveals that actual efficiency of informed decisions is higher than or roughly the same as private-information efficiency. Note also that Result 7 matches the prediction of the theory that, in many cases, efficiency is not expected to increase in late turns as beliefs reach an absorbing state in which agents either choose the same action or are indifferent between the two actions and no further learning occurs.

Our results so far deal only with average efficiency. We are also interested, however, in the behavioral regularities at the individual level and how they are affected by the network and information treatment. For a better understanding of the decision mechanism of the subjects, we next focus on the data at the level of the individual subject. We find strong evidence that there is a good degree of conformity with the theory in the aggregate data, which we sometimes fail to observe in the individual data.

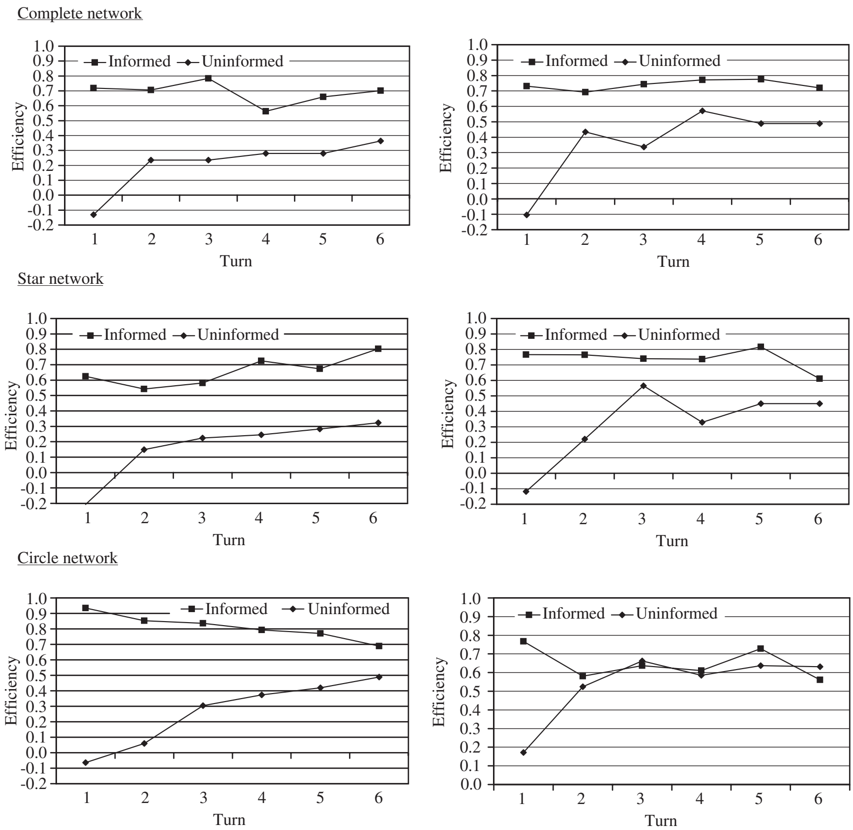


Fig. 4. Actual Efficiencies of Informed and Uninformed Decisions in Each of the Networks under Low- (Left Panel) and High Information (Right Panel).

We look at actual efficiency and ask whether individual behavior is heterogeneous within a given network and information treatment, and across networks and information treatments. To focus on the individual effects, Result 8 deals only with the complete and circle networks under full information since all types have identical sets of neighbors and subjects are equally informed. Result 9 summarizes the behavioral regularities in this regard across networks and information treatments.

Result 8 (Network). Comparing individual efficiency in the complete and circle networks under full information indicates that there is high variation in individual behavior across subjects.

Result 9 (Information). There are significant differences between the distributions of actual efficiency across information treatments in the complete network and across network structures under full information. The distributions of actual efficiency are roughly the same in the other treatments.

Fig. 5 provides the support for Result 8 and Result 9. It presents in the form of histograms the actual-efficiency distributions in each network under all information treatments (left panel), and for all networks under each

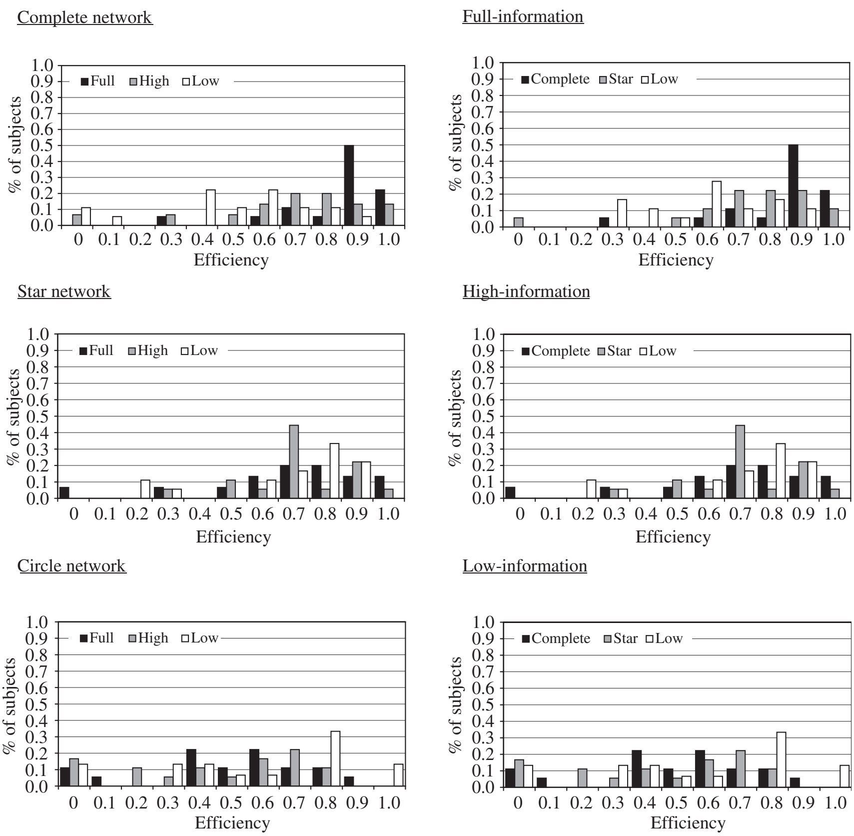


Fig. 5. Actual Efficiency Distributions in each Network under All Information Treatments (Left Panel) and for All Networks under each Information Treatment (Right Panel).

information treatment (right panel). The horizontal axis consists of the intervals of efficiency scores and the vertical axis measures the percentages of subjects corresponding to these efficiency scores.

Note, for example, that under full information the histograms in Fig. 5 show that subject behavior is more efficient in the complete network, because the distribution of efficiency scores shifts considerably to the right when calculated using the complete network data. A Kolmogorov–Smirnov test confirms this observation at the 5% significance level.

4.3. Rationality

In the laboratory, learning in networks is challenging: because of the lack of common knowledge about the history of play, subjects must draw inferences about the actions other subjects have observed as well as about their private signals. Moreover, not all of the decisions in the different networks are comparable in terms of their complexity or sophistication. For example, subjects have larger information sets in some networks than in others, and sometimes there is more information to be gleaned from the actions of others. Also, differences in the amount of common knowledge will require different degrees of sophistication to discern the optimal strategies in different networks.

Hence, even in the three-person case, the difficulty of solving the problem of social learning in networks is sometimes massive. Therefore, one of our main goals is to examine, treatment by treatment, how well the theory approximates the actual behavior observed in the laboratory. We begin by computing the optimal strategies as predicted by the theoretical model and use these to compute the level of rationality in the first and second decision turns. Thus, rationality is simply measured by the percentage of times subjects follow an equilibrium strategy.

At the first decision turn in any treatment, a subject should make a decision based on his private signal (if he is informed) or his prior (if he is uninformed). At the second decision turn, he should make a decision based on his private information and the actions taken at the first turn, and so on. Thus, the complexity of a subject's decision problem increases over time. This leads us to examine how well Bayes rationality approximates the actual behavior observed in the laboratory. At the first and second decision turns, the data support the following result.

Result 10. Over all treatments, only 5.8% of the first-turn actions in each round were inconsistent with the information implicit in the private signal. At the second turn, although there are significant differences across information sets, the error rates are uniformly fairly low.

Evidence for Result 10 is given by the table below, which reports the error rates, i.e. the percentage of times subjects deviate from the equilibrium strategy, at the second turn. The data are grouped according to the number of actions observed, i.e. all types in the complete network and type A (the center) in the star observe $N = 2$ actions, and all types in the circle network and types B and C in the star observe $N = 1$ actions. The numbers in parentheses are the percentage of decisions in which subjects were indifferent between the two actions.

Information	$N = 2$		$N = 1$	
Full	12.5	(0.00)	4.40	(44.4)
High	17.8	(16.2)	16.7	(0.00)
Low	20.3	(36.9)	26.9	(0.00)

The histograms in Fig. 6 show these data across subjects. The horizontal axis measures the error rates for different intervals, and the vertical axis measures the percentage of subjects corresponding to each interval. Note that, in both the first and second decision turns, the distribution is considerably skewed to the left, but the two distributions are significantly different at the 5% significance level using a Kolmogorov–Smirnov test.

To provide further evidence for Result 10, Table 5 summarizes the error rates in the second decision turn in each treatment, including for informed and uninformed decisions. Table 5 clearly identifies some "black spots" in which there are sharp drops in rationality, especially in the star network under low and high information. Therefore, subjects must presumably estimate the errors of others and consider this in processing the information revealed by their neighbors' actions.

It is noteworthy that, under low information, the error rates in uninformed peripherals' decisions and in uninformed decisions in the circle network were very high and roughly the same (33.9% and 34.9%, respectively). Error rates were very high in informed centers' decisions when compared with informed decisions in the complete network (22.2% and 13.8%, respectively). Under high information, the error rates in uninformed peripherals' decisions were much lower than in uninformed decisions in the circle network (26.7% and 16.3%, respectively). These differences are highly significant according to a Wilcoxon matched-pair test.

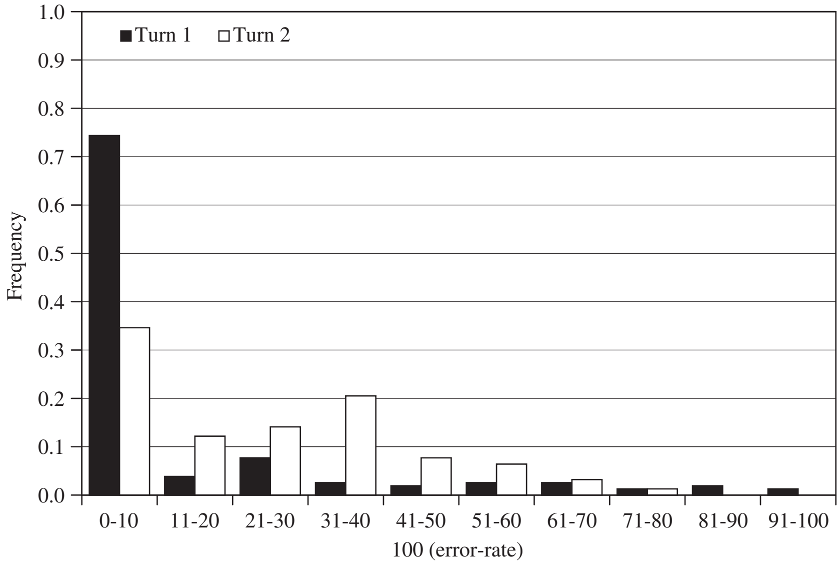


Fig. 6. Error rates in the First and Second Decision Turn.

Table 5. Error Rates at the Second Turn under the Different Information Structures and Networks.

$N = 2$	Complete				Star center		
	Full	High	Low		Full	High	Low
All	13.0	14.2	11.9	All	11.1	16.7	15.6
Informed		17.7	13.8	Informed		18.8	22.2
Uninformed		8.3	10.8	Uninformed		11.5	12.7
$N = 1$	Circle				Star peripherals		
	Full	High	Low		Full	High	Low
All	13.0	15.9	25.3	All	1.1	17.8	28.9
Informed		15.8	7.6	Informed		14.8	18.6
Uninformed		16.3	34.9	Uninformed		26.7	33.9

"black spots"

Note that the complexity of a subject's decision problem increases over time. At the first turn, a subject only has to interpret his private information. At the second turn, he has to interpret his neighbors' actions and try to infer the private information on which it was based. At the third turn, because of the lack of common knowledge about actions in incomplete networks, a subject is forced to think about subjects' knowledge of other subjects' actions and the private information it reveals. Thus mistakes are inevitable and this should be taken into account by evaluating the degree to which the game-theoretic model explains behavior in the laboratory. In the sequel, we discuss a modification of the game-theory model that abandons the assumption of common knowledge of rationality.

5. QUANTAL RESPONSE EQUILIBRIUM (QRE)

Since mistakes are made, especially under low and high information, and this should be taken into account in any theory of rational behavior, in CGK we attempt to formulate this by estimating a recursive model that allows for the possibility of errors in earlier decisions. We adapt the model of Quantal Response Equilibrium (QRE) of Mckelvey and Palfrey (1995, 1998), which enables us to evaluate the degree to which the theory explains behavior in the laboratory. We skip the model development and instead briefly explain the analysis and discuss why the results of the QRE model differ from those of the basic game-theoretic model.

We first extend the basic model of Gale and Kariv (2003) to allow for idiosyncratic preference shocks, which can be interpreted, following Harsanyi and Selten, as the effect of a "trembling hand." The basic model has a natural recursive structure, which suggests a recursive estimation procedure for the logistic random-utility model. In effect, we assume that subjects have rational expectations and use the true mean error rate when interpreting the actions they observe at the first turn. This is the behavioral interpretation of the recursive econometric method.

We begin by estimating the random-utility model using the data from the first turn. Then we use the estimated coefficient to calculate the theoretical payoffs from the actions at the second turn. We then estimate the random-utility model based on the perturbed payoffs and the observed decisions at the second turn. Continuing in this way, we estimate the entire QRE for each treatment. The parameter estimates are highly significant and positive, showing that the theory does help predict the subjects' behavior.

The predictions of the QRE model are different from those of the basic game-theoretic model for two reasons: first, because it allows agents to make mistakes and, secondly, because it assumes that agents take into account the possibility that others are making mistakes when drawing inferences from their actions. The "goodness of fit," as measured by the error rates, is better for the QRE model than for the game-theory model. We also conducted a series of specification tests and found that restrictions of the QRE model are confirmed by the data.

Among our conclusions in this paper, the following are particularly relevant to the evaluation of the QRE model:

- The model appears to fit the data best in the full-information treatments, where every subject receives a private signal. In other information treatments, subjects are randomly informed and do not know whether another subject is informed or not. This asymmetry appears to reduce the efficiency and rationality of the subjects' behavior.
- The model appears to fit the data best in symmetric networks, where each subject observes the same number of other subjects. In asymmetric networks, where one subject is known to have an informational advantage over the others, the behavior of subjects is less close to the predictions of the model and the discrepancy is highest for the subjects with the informational advantage.

In evaluating these departures from the predictions of the theory, we have become aware of several factors that ought to be considered in future work.

- In order to explain subjects' behavior, it is necessary to take into account the details of the network architecture as well as the information treatment. Simple summary characteristics of the network, such as the average distance between subjects, do not account for the subtle and complicated behaviors that we observe.
- Because of the lack of common knowledge in the networks, the decision problems faced by subjects require quite sophisticated reasoning. At the same time, the optimal strategy is simple in some networks, so it is plausible that subjects following a simple heuristic might behave "as if" they were following the optimal strategy, even though it would be quite implausible to expect them to be able to "solve" for the best response.
- In other networks, by contrast, the optimal behavior is sometimes extremely complex, even though the networks themselves are quite simple. This may explain the failure of the QRE to fit some of the data generated by asymmetric networks, for example.

To determine which of these factors are important in explaining subject behavior in a variety of settings, it will be necessary to investigate a larger class of networks in the laboratory. This is perhaps one of the most important topics for future research.

6. CONCLUSION

We have undertaken an experimental investigation of learning in three-person networks, and focus on using the theoretical framework of Gale and Kariv (2003) to interpret the data generated by the experiments. In our experimental design we used three networks, the complete network and two incomplete networks, the circle and the star, each of which theoretically gives rise to its own distinctive learning patterns.

The theory suggests that even in the three-person case the process of social learning in networks can be complicated. In particular, in the incomplete networks the absence of common knowledge makes it harder for agents to interpret the information contained in the actions of others and requires them to perform complex calculations.

Indeed, in the laboratory, we have seen that removing links in three-person networks has a significant effect on social behavior, even if removing links does not have much effect on the degree of separation, i.e. the social distance within the network. The reason is the impact of lack of common knowledge on the dynamics of social learning and the efficiency of aggregation.

The presence of common knowledge makes it easier for subjects to agree on the interpretation of the information contained in the actions of any set of subjects. Lack of common knowledge forces a subject to think about hierarchies of beliefs: for example, subject *A*'s beliefs about subject *B*'s beliefs about subject *C*'s action and the private information it reveals. When links are removed, actions are not common knowledge. This uncertainty appears to prevent subjects from making use of the additional information available to them from others' actions.

We have identified some situations where the theory does less well in accounting for subjects' behavior. We conjecture that the theory fails in those situations because the complexity of the decision problem exceeds the bounded rationality of the subjects. There is convincing theoretical and empirical evidence that some of the decisions faced by subjects are quite "complex"; however, "complexity" is a very difficult idea to conceptualize in a precise or formal way. Also, because of the simplicity of the experimental design, it is difficult to distinguish among different concepts of complexity, and there is a danger that any attempt to measure complexity will be overadapted to this particular application. Progress in this area may require both new theory and new experimental data.

A theory of complexity has to take into account the following points. First, subjects' success or failure in the experiment results from the appropriateness of the heuristics they use as much as the inherent difficulty of the decision-making. Second, the optimal strategy may be intuitively simple even though the analysis is complex, and a simple heuristic may work very well, even though the game or the analysis of the game is complex. Third, complexity is endogenous: in a complex extensive-form game, subjects may be forced to adopt simple strategies; and the decision-making problem for each subject may be simplified as a result.

Another idea of complexity has to do with the equilibrium path: in the star network, for example, the equilibrium path is more complex than in the complete network. Hence, an interesting question is whether we can identify a sufficient statistic for the difficulty or complexity of decisions that will allow us to interpret variation in efficiency and rationality measures. If we can identify more information sets where there are sharp drops in efficiency or rationality (''black spots''), we may be able to come up with some hypotheses about why the decision is complex or difficult and suggest experiments to test this hypothesis.

Our results suggest that the theory adequately accounts for large-scale features of the data. The models and results that we have developed provide a foundation for future theoretical and experimental research, and the techniques can be applied to other setups. For example, we can apply our theoretical model to random graphs, as long as connectedness is satisfied, and it could also be applied to dynamic graphs where the set of neighbors observed changes over time.

There are many more important questions that remain to be explored using our data set. Perhaps, the most important subject for future research is to identify the impact of network architecture on the efficiency and dynamics of social learning. Whether other network architectures will lead to sharply different results is not clear, since all the decision rules will have to be changed to reflect the new environment. It will probably require a more sophisticated analysis to detect the effect of differences in network architectures.

Also, we have not yet explored alternative approaches that might be brought to the interpretation of the data. Clearly, there is much to be done and the uses of this data set are far from exhausted.

NOTES

1. For excellent surveys see Gale (1996) and Bikhchandani, Hirshleifer and Welch (1998), which also provide examples and applications of observational learning in economic contexts. Among others, Lee (1993), Chamley and Gale (1994), Gul and Lundholm (1995), Moscarini, Ottaviani and Smith (1998), and Çelen and Kariv (2004a) provide further extensions of the theory.

2. There is a large and growing body of work which studies the influence of the network structure on economic outcomes. For recent surveys see Goyal (2003) and Jackson (2003).

3. Anderson and Holt (1997) investigate the model of BHW experimentally. Among others, Hung and Plott (2001), Kübler and Weizsäcker (2003), and Çelen and Kariv (2004b, 2005) analyze several aspects of sequential social learning.

ACKNOWLEDGMENTS

We are grateful to John Morgan for his comments and suggestions. This paper has benefited from suggestions by the participants of SITE 2004 Summer Workshop, the New and Alternative Directions for Learning Workshop at Carnegie Mellon University, and seminars at several universities. For financial support, Gale acknowledges the National Science Foundation Grant No. SBR-0095109) and the C. V. Starr Center for Applied Economics at New York University, and Kariv thanks University of California Berkeley (COR Grant).

REFERENCES

Anderson, L., & Holt, C. (1997). Information cascades in the laboratory. *American Economic Review*, *87*(5), 847–862.

Bala, V., & Goyal, S. (1998). Learning from neighbors. *Review of Economic Studies*, *65*, 595–621.

Banerjee, A. (1992). A simple model of herd behavior. *Quarterly Journal of Economics*, *107*(3), 797–817.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascade. *Journal of Political Economy*, *100*(5), 992–1026.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: conformity, fads, and informational cascades. *Journal of Economic Perspective*, *12*(3), 151–170.

Çelen, B., & Kariv, S. (2004a). Observational learning under imperfect information. *Games and Economic Behavior*, *47*(1), 72–86.

Çelen, B., & Kariv, S. (2004b). Distinguishing informational cascades from herd behavior in the laboratory. *American Economic Review*, *94*(3), 484–497.

Çelen, B., & Kariv, S. (2005). An experimental test of observational learning under imperfect information. *Economic Theory*, *26*(3), 677–699.

Chamley, C., & Gale, D. (1994). Information revelation and strategic delay in a model of investment. *Econometrica*, *62*(5), 1065–1085.

Choi, S., Gale, D., & Kariv, S. (2005). *Learning in Networks: An Experimental Study*. Mimo.

Gale, D. (1996). What have we learned from social learning? *European Economic Review*, *40* (3–5), 617–628.

Gale, D., & Kariv, S. (2003). Bayesian learning in social networks. *Games and Economic Behavior*, *45*(2), 329–346.

Goyal, S. (2003). Learning in networks: a survey. In: G. Demange & M. Wooders (Eds), *Group formation in economics: Networks, clubs, and coalitions*. New York: Cambridge University Press.

Gul, F., & Lundholm, R. (1995). Endogenous timing and the clustering of agents' decisions. *Journal of Political Economy*, *103*(5), 1039–1066.

Hung, A., & Plott, C. (2001). Information cascades: replication and an extension to majority rule and conformity-rewarding institutions. *American Economic Review*, *91*(5), 1508–1520.

Jackson, M. (2003). A survey of models of network formation: stability and efficiency. In: G. Demange & M. Wooders (Eds), *Group formation in economics: networks, clubs, and coalitions*. New York: Cambridge University Press.

Kübler, D., & Weizsäcker, G. (2003). Limited depth of reasoning and failure of cascade formation in the laboratory. *Review of Economic Studies*, *71*(2), 425–441.

Lee, I. H. (1993). On the convergence of informational cascades. *Journal of Economic Theory*, *61*(2), 396–411.

Mckelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for extensive form games. *Games and Economic Behavior*, *10*, 6–38.

Mckelvey, R. D., & Palfrey, T. R. (1998). Quantal response equilibria for extensive form games. *Experimental Economics*, *1*, 9–41.

Moscarini, G., Ottaviani, M., & Smith, L. (1998). Social learning in a changing world. *Economic Theory*, *11*, 657–665.

Smith, L., & Sørensen, P. (2000). Pathological outcomes of observational learning. *Econometrica*, *68*(2), 371–398.

This page intentionally left blank

COMMUNICATION AND EFFICIENCY IN COORDINATION GAME EXPERIMENTS

Anthony Burton, Graham Loomes and Martin Sefton

ABSTRACT

We examine the effects of pre-play communication in an experimental game with conflicting risk-dominant and payoff-dominant equilibria. We find that most players condition their choices on the messages received, and do so in an intuitive way, announcing an intention to play the payoff-dominant action, and choosing the payoff-dominant action if the opponent expresses the same intention. However, a significant minority of players misrepresent their intentions. In some sessions where these players appear, behavior converges to an equilibrium in which subjects misrepresent their intentions and play the risk-dominant equilibrium.

1. INTRODUCTION

In this chapter, we explore the role of communication in selecting among equilibria of a simple coordination game. The game of interest is shown in Fig. 1, and has two pure strategy equilibria: (S, S), which is risk-dominant, and (R, R), which is payoff-dominant. There are conflicting views about

Experimental and Behavioral Economics
Advances in Applied Microeconomics, Volume 13, 63–85

ISSN: 0278-0984/doi:10.1016/S0278-0984(05)13003-X

		Column Player	
		R	S
Row	R	1000, 1000	0, 900
Player	S	900, 0	700, 700

Fig. 1. Payoff Matrix for CKS Game.

how pre-play communication will affect outcomes: one view is that opportunities for players to communicate with one another prior to playing the game will have no effect, while another suggests that communication will enable players to attain the payoff-dominant equilibrium. Aumann (1990) makes a logical argument that any message should carry no information, while Farrell and Rabin (1996) feel that cheap talk will help players coordinate on the efficient equilibrium (we discuss these views later). Since formal equilibrium analysis of an extended game that precedes the game in Fig. 1 by cheap talk admits equilibria where actions do depend on messages, as well as equilibria where actions are independent of messages, we view the effect of messages on outcomes as essentially a behavioral issue.

There is divergent experimental evidence on the extent to which communication helps players coordinate on efficient outcomes. Clark, Kay, and Sefton (2001) compared the outcomes of the game in Fig. 1 (henceforth CKS Game) with and without pre-play communication. In their experiment with communication they allowed players to simultaneously send non-binding messages about which action they intended to play and found this significantly increased the amount of efficient play. However, a substantial amount of inefficient play remains and, moreover, a significant number of subjects make choices that support Aumann's position. Charness (2000), in contrast, finds not much support for Aumann's argument when only one player sends a signal about their intention prior to making their choice. Under these conditions he observes near complete efficiency.[1] The present chapter attempts to reconcile these empirical results and further inform the theoretical discussion.

A number of potentially important procedural differences may account for these conflicting experimental results. We explore the effect of one of these procedural differences by comparing 'one-way' and 'two-way' communication technologies. Broadly, we find no evidence that differences between the Clark et al. and Charness results can be accounted for by the communication technology: we fail to find any evidence of a communication technology effect in our data.

Also, noting that the disagreement between Aumann and Farrell is essentially about whether decisions are conditioned on messages, we conduct a treatment that elicits richer information about subjects' strategies. In this treatment, subjects submit a plan that specifies their message and their decision conditional on messages that could be sent by their opponent. This design allows us to observe how subjects would have responded to alternative messages, in contrast to treatments in which we only observe how subjects respond to the message they actually received.

We find that most subjects do condition their choices on their opponent's message, but that there is an important minority of subjects who misrepresent their intentions. Their importance is based on two factors. First, these players have a strong impact on the profitability of an opponent's strategy that does condition on messages. Second, and related to this point, there is a dynamic pattern in our data suggesting that when such misrepresentation occurs, the propensity to condition on messages declines and there is a movement toward the inefficient equilibrium.

This interaction between initial heterogeneity across subjects and changing behavior over time generates substantial session effects *within* treatments. In some sessions, behavior converges to one of the equilibria, whereas in other sessions with the same treatment combination, behavior either does not converge or converges to a different equilibrium.

The remainder of this chapter is organized as follows. In the next section, we provide a brief review of the relevant theoretical and experimental literature, and in Section 3 we describe in detail our experimental treatments. The results of the experiment are presented in Section 4, and Section 5 concludes this chapter.

2. COMMUNICATION IN COORDINATION GAMES

Previous experimental studies have examined symmetric 2×2 coordination games where there is a tension between payoff-dominant and risk-dominant equilibria.[2] Cooper, DeJong, Forsythe, and Ross (1992) conducted an experiment with the game in Fig. 2 (henceforth CDFR game). Eleven subjects participated in each of their sessions, with each subject playing each other subject twice (one player sitting out in each round), providing data on 110 games per session. In sessions where no communication was allowed, subjects predominantly chose S: in the last eleven rounds R was chosen only 5/330 times.[3]

		Column Player	
		R	S
Row	R	1000, 1000	0, 800
Player	S	800, 0	800, 800

Fig. 2. CDFR Game.

In their 'one-way communication' sessions, the row player completed the message "I plan to play …" and sent it to the column player before the two made their choices simultaneously: R was chosen 227/330 times. In their 'two-way communication' sessions, both players simultaneously sent messages prior to making choices: R was chosen 315/330 times. From this Crawford (1998) concludes that when communication enables desirable outcomes by providing *reassurance*, two-way communication is more effective than one-way communication.

A natural interpretation of these results is that players use non-binding communication to agree on the efficient outcome, and, since this outcome is an equilibrium, players have individual incentives to keep this agreement. Indeed, the argument that communication will lead to selection of efficient equilibria is implicit in numerous equilibrium refinement concepts (e.g. see the discussion in Fudenberg & Tirole, 1991, Section 5.4). However, whether an agreement to play an efficient equilibrium can be viewed as self-enforcing is a controversial issue.

Aumann (1990) has argued that communication will be ineffective in games such as the CKS game. He points out that each player prefers that his opponent choose R; that is, regardless of his own choice a player gets a higher payoff if his opponent chooses R. He then argues that since each player will realize this, messages have no power to convey what action will follow and will therefore be disregarded. To be clear here, Aumann is not suggesting rational players will never choose R, "but only that *agreeing* to do so won't lead them to *do* it."[4] He suggests that "impulsive and optimistic" players will choose R while "careful and prudent" players will choose S.[5] Neither type of player will make their decision on the basis of the message they receive; rather it is their characteristics as individuals that matter.[6]

A different view is expressed in Farrell (1988) and Farrell and Rabin (1996). They acknowledge the force of Aumann's argument, but nevertheless express a suspicion that in practice communication will lead to efficiency. Farrell (1988) argues that whether one agrees with Aumann "is a matter of whether one thinks of player 1 deciding on his move at stage 2 'after' he chooses his stage-1 message, or deciding on his move first and then on his message. If the latter, then Aumann's criticism is compelling; if the former, then matters are rather unclear."[7]

Clark et al. examined Aumann's argument by comparing the effectiveness of communication in the CDFR and CKS games. For each game four sessions were conducted: two sessions without communication and two with two-way communication.[8] In sessions without communication there were relatively few R choices: 68/400 choices in the CDFR game and 76/400 choices in the CKS game. With two-way communication, announcements were similar in the two games: "I intend to choose R" was announced 342/400 times in the CDFR game and 324/400 times in the CKS game. However, responses to announcements were different across the two games. R was chosen 96% of the time when both subjects announced R in the CDFR game, but only 50% of the time in the CKS game. Thus, although communication did increase the amount of efficient play in both games, the effect was much weaker in the CKS game. They concluded that the effectiveness of communication was sensitive to payoffs in a manner consistent with Aumann's conjecture.

Charness (2000) also tested Aumann's conjecture. In his sessions without communication, using a variety of payoff specifications, only 16% of the games resulted in the payoff-dominant outcome.[9] He then conducted sessions using a one-way communication structure in which, prior to making choices, one of the players sent a signal about their intention to the other. Under this form of communication 86% of games resulted in the payoff-dominant outcome. Charness then conducted sessions with the CKS and CDFR games, using his procedures and subject pool. In his CKS games 120/120 messages involved signaling an intention to choose R. In all but 13 cases, the sender followed up on this stated intention by choosing R. It also appears that receivers believed the message at face value: in all but 10 cases the receiver chose R. Thus, in the Charness experiment one-way communication was extremely effective in the CKS game: 84% of cases resulted in the payoff-dominant outcome. In fact, one-way communication was slightly less effective in his CDFR replication. The payoff-dominant outcome was observed in 72% of the games.

Both studies show that communication increases the amount of efficient play. Where they differ is in the extent of this effect. There are a large number of design differences that could, in principle, account for the difference between Clark et al. and Charness results. We focus here on differences in communication technologies. Clark et al. used the two-way communication technology that can result in 'disequilibrium messages'; i.e. one subject announces an intention to choose R while the other announces an intention to choose S. In this situation a natural reaction for both players may be to choose S. In the context of the CKS game it is then interesting to think about how the observation of a player announcing R and then choosing S would be interpreted. One possibility is that this would undermine the confidence in truthfulness in messages. Charness, on the other hand, used one-way communication.[10]

Charness also conducted a treatment in the order in which the sender's action and signal decisions were reversed. That is, the sender made a choice, and then completed the message "I indicate that my choice is … ." The receiver then observed this message, but not the actual choice, before making his/her own choice. In this action-signal treatment two groups converged on the payoff-dominant outcome, two on the risk-dominant outcome, while two did not converge. Although 147 of 180 messages (82%) indicated an R choice, only in 84 of these 147 cases (57%) did the sender's message match their choice. This order-reversal result appears to support Farrell's comment that the temporal order of decisions matters. While this interpretation is very plausible, we note that controlling the timing with which actions and signals are submitted does not necessarily control the timing with which subjects decide upon their actions and signals. This discussion suggests another behavioral question: do subjects condition their choices on the messages they send and receive?

3. DESCRIPTION OF EXPERIMENT

3.1. Design

Our experiment consists of 22 sessions. In each session, six players were randomly rematched over ten rounds, each round consisting of a single play of a game.[11] This game was based on the CKS game, although it was implemented in three different forms depending on whether the session employed the '1-WAY', '2-WAY', or 'PLAN' treatment.

One of the goals of our new experiment was to investigate the effect of different communication technologies. Thus we conducted some sessions using a one-way communication technology (1-WAY treatment) and some with a two-way communication technology (2-WAY treatment). A disadvantage of these treatments is that it is impossible to know what choices subjects would have made, had they received a different message.

For a clearer assessment of Aumann's conjecture, the rest of our sessions employed a PLAN treatment that was designed to investigate whether subjects condition their choices on the messages they send and/or receive. This treatment employed a strategy method in which subjects submitted a plan for playing the CKS game with two-way communication. In this plan subjects entered both their message and how they would respond to each of the messages they might receive. Actual choices were then determined by playing out these plans. Therefore this treatment allows us to examine directly whether, and how, players condition their action on their opponent's message.[12]

Finally, our sessions were differentiated according to whether players were represented by individual subjects ('INDIVIDUAL' treatment) or by a pair of subjects who collaborated in making decisions ('TEAM' treatment). The former is the standard procedure in experimental economics, and is used, for instance, by Cooper et al., Clark et al., and Charness. The sessions using the TEAM treatment were intended to investigate how players make decisions in these types of laboratory environment by asking or encouraging subjects to explain and discuss their decisions with one another, with these discussions being recorded.

This design is summarized in Fig. 3, where the number of sessions with each treatment combination is indicated in each cell. The imbalance is due to cancellation of two sessions because of no-shows.

3.2. Procedures

All sessions were conducted at the University of Newcastle in spring 1998 and in total involved 204 subjects, each participating in no more than one session. Subjects were undergraduates who had responded to an e-mail message sent to a university-wide subject pool. The message recruited subjects for sessions "lasting approximately 55 minutes" and offering an opportunity to "earn up to £10."[13]

	1-WAY	2-WAY	PLAN
INDIVIDUAL	3 sessions	4 sessions	3 sessions
TEAM	4 sessions	4 sessions	4 sessions

Fig. 3. Experimental Design.

In the INDIVIDUAL treatment, upon arrival subjects were seated at visually isolated computer terminals. A set of instructions was then read aloud and subjects were led through a practice round to familiarize them with the computer screens they would be seeing.[14] They were then given an opportunity to ask questions, and were asked to complete a quiz to verify that they understood how the various possible message and choice combinations translated into earnings.[15] Then the decision-making part of the experiment began.

In sessions employing the TEAM treatment, twelve subjects were seated in one room and a set of instructions was read aloud. The subjects were then randomly divided into six pairs, where each pair was to take the role of a single player. The pairs were taken to separate rooms where they could discuss their joint decisions without being overheard by other subjects. Subjects in each pair were led through the practice round, given an opportunity to ask questions, and asked to complete the quiz, by a monitor. The subjects were encouraged to discuss with one another their (joint) decision, and these discussions were recorded (openly) by the monitor. However, the decision-making part of the experiment was just as in the INDIVIDUAL treatment.[16]

The decision-making part of the experiment consisted of ten identical rounds in which players earned points as described in the CKS game payoff matrix. In each round players were randomly and anonymously matched, although the matching scheme was determined prior to the first session and then used in all sessions. Also, the matching scheme was designed so that each player met each of the other five players exactly twice.

At the beginning of each round of the 2-WAY, treatment players were prompted to complete the sentence: 'I intend to choose …'. When all players had done this, the messages were transmitted to the relevant opponent. Players were then prompted to make their choices (the instructions explicitly stated that they were not necessarily required to make the choice they had announced). When all players had made their choices, the players were informed of their own choice and their opponent's choice, and their point earnings for the round. Subjects kept a record of announcements, choices, and earnings by filling out a record sheet. Players were not allowed to communicate with other players except via their formal decisions, which were transmitted across the computer network. At the end of round ten, subjects were paid in private £1 for every 1000 points earned.[17]

The 1-WAY treatment was identical to the 2-WAY treatment except that after all players had sent their messages, players were randomly designated either a "Sender" or a "Receiver." Senders' messages were then delivered to

I intend to choose ____
If the other person says 'I intend to choose R', I choose _____
If the other person says 'I intend to choose S', I choose _____

Fig. 4. PLAN Template.

the appropriate receivers, while receivers' messages were discarded. Senders were informed that their message had been delivered, while receivers, in addition to seeing their opponent's message, were informed that their own messages had not been delivered. Players then made their choices, just as in the 2-WAY treatment.

This design ensured that all subjects entered keystrokes at the same stage during a round. If only senders were to type in and enter a message, attentive subjects in the INDIVIDUAL treatment could infer that their opponent was one of three people in the room, creating a slight, but unwanted, difference from the 2-WAY treatment. The design in which all players submit messages also has the added advantage that it yields more information about intended messages.

The PLAN treatment implements the 2-WAY treatment via a strategy method. At the beginning of a round, each player is prompted to submit a plan in which they indicate their message, and their choice conditional on their opponent's message. An example PLAN template is shown in Fig. 4.

Players were not required to complete the various parts of the template in any particular order and were allowed to revise any part of their plan until they were satisfied with it, at which point they submitted the desired plan. When all players had submitted their plans, each player was informed of their own message, their opponent's message, their own action, and their opponent's action. The important difference between the PLAN and the 2-WAY treatment is that in the plan the entries in the second and third fields correspond to contingent choices that will only become actual choices if the appropriate message is sent by the other player.

4. RESULTS

Our procedures and subject pool differ from those used by Charness or Clark et al. in numerous respects and, perhaps unsurprisingly, this leads to some differences between their results and ours. In our treatment that most closely resembles that of Charness (INDIVIDUAL, 1-WAY), subjects announced an intention to choose R 77% of the time, whereas subjects made this announcement 100% of the time in Charness' experiment.[18] (The proportion of times senders followed up on this announcement by actually choosing R is strikingly similar: 89% in Charness and 85% here.) In our treatment that most closely resembles Clark et al. (INDIVIDUAL, 2-WAY) R was announced 73% of the time, and R was chosen 80% of the time when both subjects had announced R. The proportion of R announcements in Clark et al. was similar, 81%, but a combination of R announcement resulted in R choices only 50% of the time.

In this section, we first present the results on communication technologies, and second present evidence on the Aumann conjecture. For our analysis of communication technology, we first present results from sessions using 1-WAY communication, comparing the INDIVIDUAL, TEAM, and Charness' treatments, and then present results from sessions using 2-WAY communication, this time comparing INDIVIDUAL, TEAM, and the Clark et al. treatments. We then complete our analysis of communication technologies by comparing signaling or choice behavior under our 1-WAY and 2-WAY treatments.

Similarly, before evaluating Aumann's conjecture we first compare the results from INDIVIDUAL and TEAM sessions. We then examine the extent to which our PLAN treatment elicits behavior comparable to the 2-WAY treatment. We close the section with a detailed assessment of the plans submitted by subjects, paying particular attention to the degree to which subjects condition choices on messages and the development of this over the course of a session.

4.1. Comparing Communication Technologies

Results from our seven 1-WAY sessions, and, for purposes of comparison, the relevant data from Charness' sessions, are presented in Table 1. The first column shows that players usually announce R; in fact more than 90% of the messages announce R in seven of eleven sessions. However, choice behavior is more variable across sessions. In the next column the proportions of "R" choices can be seen to range from 12 to 100%, with a median value of 63%. The next four columns show how choices are related to messages and it is clear that senders usually, but not always, follow an R signal with an R choice. Across all 11 sessions, 23% of senders who signal R choose S (although again, there is a high degree of variability between sessions). The

Table 1. Results from Sessions Using One-Way Communication.

Treatment: Session	R Messages	R Choices	R Choices by				Efficient equilibria	Inefficient equilibria
			'Senders' who sent		'Receivers' who received			
			R message	S message	R message	S message		
Individual								
1	57/60 (95%)	53/60 (88%)	25/29 (86%)	1/1 (100%)	26/29 (90%)	1/1 (100%)	23/30 (77%)	0/30 (0%)
2	44/60 (73%)	32/60 (53%)	14/19 (74%)	2/11 (18%)	14/19 (74%)	2/11 (18%)	10/30 (33%)	8/30 (27%)
3	38/60 (63%)	37/60 (62%)	17/18 (94%)	3/12 (25%)	14/18 (78%)	3/12 (25%)	14/30 (47%)	7/30 (23%)
Team								
1	51/60 (85%)	32/60 (53%)	15/25 (60%)	0/5 (0%)	17/25 (68%)	0/5 (0%)	10/30 (33%)	8/30 (27%)
2	45/60 (75%)	7/60 (12%)	2/20 (10%)	0/10 (0%)	5/20 (25%)	0/10 (0%)	1/30 (3%)	24/30 (80%)
3	55/60 (92%)	32/60 (53%)	15/28 (54%)	0/2 (0%)	17/28 (61%)	0/2 (0%)	9/30 (30%)	7/30 (23%)
4	59/60 (98%)	56/60 (93%)	27/29 (93%)	0/1 (0%)	29/29 (100%)	0/1 (0%)	27/30 (90%)	1/30 (3%)
Charness								
1	30/30 (100%)	60/60 (100%)	30/30 (100%)	0/0 (–)	30/30 (100%)	0/0 (–)	30/30 (100%)	0/30 (0%)
2	30/30 (100%)	38/60 (63%)	18/30 (60%)	0/0 (–)	20/30 (67%)	0/0 (–)	12/30 (40%)	4/30 (13%)
3	30/30 (100%)	59/60 (98%)	29/30 (97%)	0/0 (–)	30/30 (100%)	0/0 (–)	29/30 (97%)	0/30 (0%)
4	30/30 (100%)	60/60 (100%)	30/30 (100%)	0/0 (–)	30/30 (100%)	0/0 (–)	30/30 (100%)	0/30 (0%)
Kruskal–Wallis Statistic (p-value)	7.477 (0.024)	5.386 (0.068)	4.587 (0.101)	N/A	2.557 (0.279)	N/A	4.869 (0.088)	4.415 (0.110)

last two columns show the impact of this behavior on the attainment of equilibrium. Players attain the efficient equilibrium in 59% of games, but this varies between 3% (actually, 1 game out of 30) and 100% (30 games out of 30).

The last row of Table 1 reports the Kruskal–Wallis statistic for testing the hypothesis that the INDIVIDUAL, TEAM, and Charness sessions are drawn from the same distribution.[19] We reject at the 5% level in the case of proportions of R announcements; 10% level in the cases of proportion of R choices and proportion of games attaining efficient equilibrium. Pairwise comparisons between our INDIVIDUAL and TEAM treatments reveal significant differences only in the cases where subjects choose R in response to sending S messages ($p = 0.034$), or choose R in response to receiving S messages. This type of behavior only occurred, and very infrequently, in our INDIVIDUAL treatment. In the other cases, the Kruskal–Wallis test is detecting differences between the Charness and Newcastle treatments.[20]

Specifically, R was announced in all 120 of Charness' games, while in both of our 1-WAY treatments we get a non-negligible number of S announcements (23% and 12% in our INDIVIDUAL and TEAM treatments, respectively). Given this difference, the higher percentage of R choices in Charness' sessions (90%) than in the Newcastle sessions (59%) is not too surprising. In turn, the percentage of games resulting in the efficient equilibrium is higher in Charness' sessions (84%) than the Newcastle sessions (45%).

Table 2 presents the results from our 2-WAY treatments and from the Clark et al. sessions. There is a great deal of variability across sessions. The percentage of R announcements ranges from 62 to 92% across sessions. The percentage of R choices ranges between 27 and 87%, and the percentage of games resulting in efficient equilibria varies between 10 and 77%.

There are apparent differences between treatments: for example, aggregating across sessions within a given treatment, the proportion of times an (R, R) message combination led to an R choice was 80%, 67%, and 50% in the INDIVIDUAL, TEAM, and Clark et al. treatments, respectively. However, a Kruskal–Wallis test fails to detect any significant differences between the three treatments.

Given the similarity between Charness' procedures and those used in our 1-WAY treatments – each session used six players, each playing ten games in a random re-matching scheme sometimes as a sender and sometimes as a receiver – the differences between our results and his are puzzling. In fact, because we wished to maintain comparability between our 1-WAY and 2-WAY treatments, there are more procedural differences between the

Table 2. Results from Sessions Using Two-Way Communication.

Treatment: Session	R Messages	R Choices	R Choices by				Efficient equilibria	Inefficient equilibria
			R senders who received R	S senders who received S	R senders who received S	S senders who received R		
Individual								
1	39/60	20/60	13/22	2/4	1/17	4/17	4/30	14/30
	(65%)	(33%)	(59%)	(50%)	(6%)	(23%)	(13%)	(47%)
2	55/60	52/60	48/50	0/0	1/5	3/5	23/30	1/30
	(92%)	(87%)	(96%)	(–)	(20%)	(60%)	(77%)	(3%)
3	37/60	16/60	14/22	0/8	0/15	2/15	3/30	17/30
	(62%)	(27%)	(64%)	(0%)	(0%)	(13%)	(10%)	(57%)
4	45/60	31/60	26/32	0/2	3/13	2/13	11/30	10/30
	(75%)	(52%)	(81%)	(0%)	(23%)	(15%)	(37%)	(33%)
Team								
1	50/60	19/60	17/40	0/0	1/10	1/10	4/30	15/30
	(83%)	(32%)	(43%)	(–)	(10%)	(10%)	(13%)	(50%)
2	47/60	20/60	19/40	0/6	1/7	0/7	3/30	13/30
	(78%)	(33%)	(48%)	(0%)	(14%)	(0%)	(10%)	(43%)
3	54/60	51/60	46/48	0/0	3/6	2/6	23/30	2/30
	(90%)	(85%)	(96%)	(–)	(50%)	(33%)	(77%)	(7%)
4	48/60	32/60	30/38	0/2	1/10	1/10	12/30	10/30
	(80%)	(53%)	(79%)	(0%)	(10%)	(10%)	(40%)	(33%)
Clark et al.								
1	150/200	81/200	55/108	2/8	13/42	11/42	15/100	34/100
	(75%)	(41%)	(51%)	(25%)	(31%)	(26%)	(15%)	(34%)
2	174/200	87/200	74/152	0/4	5/22	8/22	22/100	35/100
	(87%)	(44%)	(49%)	(0%)	(23%)	(36%)	(22%)	(35%)
Kruskal–Wallis Statistic (*p*-value)	1.52	0.014	2.195	N/A	2.475	3.355	0.014	0.014
	(0.468)	(0.993)	(0.334)		(0.290)	(0.187)	(0.993)	(0.993)

2-WAY and Clark et al. procedures than there are between the 1-WAY and Charness procedures.

Charness offers three potential explanations for the differences between his results and those of Clark et al.: communication technology, computerization, and subject pool differences. We are in a position to formally test the first of these by comparing our 1-WAY and 2-WAY sessions. Using the Kruskal–Wallis test to compare the seven 1-WAY and eight 2-WAY sessions, we find no significant differences.

Of course, the other factors, computerization and subject pool, remain as potential candidates for explaining the difference. In any case, we argue that a striking aspect of our results is the variability across sessions within a given treatment. One implication of this is that it may be difficult to detect underlying treatment or procedural effects – the power of such tests will be low. A second implication is that, independent from such effects, there are other important determinants of session outcomes. A full account of behavior would address the initial heterogeneity of decisions within a treatment, and how this evolves over the course of a session.

4.2. Evaluating Aumann's Conjecture

Our PLAN treatment was designed to give a direct insight into whether subjects condition their actions on the messages they receive in the 2-WAY game. Before examining the PLAN decisions in detail, however, we note that although there is a game-theoretic sense in which the PLAN and 2-WAY games are equivalent, they may not be behaviorally equivalent. In particular, by having subjects submit plans we ask them to think about the game in a very different way from when we ask them to submit messages and choices sequentially. Moreover, the notion of signaling intentions is quite natural in a sequential setting, but somewhat contrived and difficult to grasp in a simultaneous move setting. Thus, despite the strategic similarity, the subjects' task in the PLAN treatment is, in our view, more complex than in the 2-WAY treatment, and so we begin this sub-section by first evaluating whether the PLAN treatments elicits decisions that are consistent with those from the 2-WAY treatment.

The results from sessions using our PLAN treatment are summarized in Table 3. As in all other treatments, subjects predominantly send R signals. However, and again as in other treatments, there is a lot of variability in other measures of behavior across sessions. For example the percentage of games resulting in the efficient equilibrium varied from 3% (1 out of 30) in

Table 3. Results from Sessions Using PLAN Treatment.

Treatment: Session	R Messages	R Choices	R Choices by				Efficient equilibria	Inefficient equilibria
			R senders who received R	S senders who received S	R senders who received S	S senders who received R		
Individual								
1	51/60	38/60	34/44	0/2	0/7	4/71	12/30	4/30
	(85%)	(63%)	(77%)	(0%)	(0%)	(57%)	(40%)	(13%)
2	28/60	56/60	51/52	0/0	1/4	4/4	26/30	0/30
	(93%)	(93%)	(98%)	(–)	(25%)	(100%)	(87%)	(0%)
3	59/60	59/60	58/58	0/0	0/1	1/1	29/30	0/30
	(98%)	(98%)	(100%)	(–)	(0%)	(100%)	(97%)	(0%)
TEAM								
1	50/60	34/60	30/42	0/2	0/8	4/8	11/30	7/30
	(83%)	(57%)	(71%)	(0%)	(0%)	(50%)	(37%)	(23%)
2	42/60	11/60	8/28	0/4	0/14	3/14	1/30	20/30
	(70%)	(18%)	(29%)	(0%)	(0%)	(21%)	(3%)	(67%)
3	51/60	12/60	11/42	0/0	0/9	1/9	2/30	20/30
	(85%)	(20%)	(26%)	(–)	(0%)	(11%)	(7%)	(67%)
4	52/60	48/60	46/46	0/2	0/6	2/6	23/30	5/30
	(87%)	(80%)	(100%)	(0%)	(0%)	(33%)	(77%)	(17%)
Kruskal–Wallis	2.531	3.125	1.531	N/A	0.500	4.500	3.125	4.500
Statistic (*p*-value)	(0.112)	(0.077)	(0.216)		(0.480)	(0.034)	(0.077)	(0.034)

one of the TEAM sessions to 97% (29 out of 30) in one of the INDIVIDUAL sessions. A closer examination of the table suggests that an important source of these session effects lies in how subjects responded to (R, R) message combinations (see the third column in the body of the table). In some sessions R choices always followed (R, R) message combinations, while at the other extreme there was one session in which "R" was chosen only 26% of the time in response to such message combinations.

The last row of Table 3 indicates significant differences between the choice behavior of subjects in the INDIVIDUAL and TEAM treatments. In particular there are significantly fewer R choices in the TEAM treatment, resulting in significantly fewer efficient equilibrium outcomes.

In all respects but one, differences between the TEAM 2-WAY (summarized in Table 2) and TEAM PLAN treatments are insignificant. The exception is that in the TEAM 2-WAY treatment there is a significantly higher proportion of R choices by subjects who sent an R-message but received an S-message ($p = 0.021$). In terms of announcements, choices made, and actual equilibria attained, differences between the TEAM 2-WAY and TEAM PLAN are insignificant.

In contrast, a comparison of the INDIVIDUAL 2-WAY and INDIVIDUAL PLAN treatments reveals significant differences according to most measures of behavior. Most importantly there are significant differences in the propensities to announce R ($p = 0.077$), choose R ($p = 0.077$), and attain the efficient equilibrium ($p = 0.077$). R play is more likely in the INDIVIDUAL PLAN than in the INDIVIDUAL 2-WAY treatment.

Our tentative explanation for the difference between the INDIVIDUAL PLAN and other treatments lies in the difficulties the subjects face in dealing with the contingent choices required under the PLAN treatment. Unlike the 2-WAY treatment – where players make two decisions, in each selecting one of two options – the PLAN treatment has players who make a single decision, in which they select one of eight options. Moreover, the consequences of selecting an option are more difficult to see in the PLAN treatment. As a result, understanding how different entries in different fields of the PLAN template translate into outcomes is a more cognitively demanding task. In such a situation a simple heuristic may be to "go for the highest possible payoff." Of course, if enough subjects in a session follow this heuristic, the heuristic is successful and there is no reason for subjects to modify their behavior. In the TEAM treatment two factors may have helped overcome these cognitive difficulties. First, subjects in a TEAM were encouraged to discuss how they arrived at their decision, inducing them to put more effort

Table 4. Number of Plans Submitted in Sessions Using PLAN Treatment.

Treatment: Session	RRS	RSS	SSS	RRR	SRS	SRR	RSR	SSR
Individual								
1	38	12	4	1	4	1	0	0
2	43	0	0	12	4	0	1	0
3	50	0	0	9	0	1	0	0
All	131	12	4	22	8	2	1	0
Team								
1	35	14	4	1	6	0	0	0
2	9	32	15	1	3	0	0	0
3	14	37	8	0	1	0	0	0
4	52	0	5	0	3	0	0	0
All	110	83	32	2	13	0	0	0
Total	251	95	36	24	21	2	1	0

Note: Plan XYZ refers to: announce X, choose Y if opponent announces R, choose Z if opponent announces S.

into making their decision.[21] Second, subjects in a TEAM were sometimes able to help each other think through the consequences of alternative plans.

However this difference between the INDIVIDUAL PLAN and 2-WAY treatments is explained, the difference implies that some caution should be exercised in using the results of the INDIVIDUAL PLAN treatment to interpret strategic thinking in the 2-WAY treatment.

We turn now to a detailed examination of the submitted plans. Table 4 gives a breakdown of the frequency of different plans across sessions. There are notable differences between treatments (to which we will return), but overall players predominantly choose one of the two plans. Of the eight possible plans, the most frequent is announcing R, choosing R if the opponent announces R, and choosing S if the opponent announces S (RRS), accounting for 241 of 420 (57%) plans. The next most frequent consists of announcing R and choosing S regardless of the opponent's announcement (RSS). This accounts for 95 of 420 (23%) plans.[22]

All plans can be rationalized in the sense that they can be a best response to some belief. However, the two most popular plans can also be easily interpreted in terms of Aumann's conjecture that communication will not lead to efficiency, and Farrell and Rabin's argument that in practice it may. A player who submits the RRS plan announces that they intend to play their part of the (R, R) equilibrium, and will actually do so if the other player announces the same intention. Here actions are conditioned on messages, and for a population of such players, messages do coordinate actions on the efficient equilibrium as Farrell and Rabin suspect. On the other hand, a player who submits the RSS plan is acting in a manner consistent with Aumann's conjecture. Such a subject has decided to choose S regardless of her opponent's message, but, presumably for strategic reasons, to announce an intention to play R.

The strongest interpretation of Aumann's conjecture is that no player ever conditions their choice on the messages they receive. Our PLAN data allow us to test this hypothesis. Since 63% of submitted plans condition on the opponent's message, this hypothesis can be overwhelmingly rejected. However, the minority of plans corresponding to RSS is very significant for a number of reasons. First, as described above, they are naturally interpreted as consistent with Aumann's conjecture. Second, they have a strong impact on the success of the RRS strategy. Third, and related to the previous point, the presence of a small number of subjects submitting the RSS plan heavily influences subsequent play in the sessions in which they appear.

Figs. 5 and 6 display the proportions of RRS and RSS plans across rounds for each session. In the INDIVIDUAL PLAN sessions (Fig. 5) the proportion of RRS plans is high and stable. In these sessions there is little interaction between RRS and RSS players simply because only two subjects submit the latter plan (and one of these not until round nine). In fact against the observed distribution of plans, the optimal plan is RRS.[23]

Fig. 6 tells a different story. In three of the TEAM PLAN sessions both plans are evident and the relative proportion changes as the sessions progress. In particular, the proportion of RSS plans grows while the proportion of RRS plans diminishes. In fact, in sessions 2 and 3 play converges to an inefficient equilibrium by the last round (since in session 2 the two plans not graphed are SSS). In the fourth TEAM PLAN session, no RSS plans are observed and by the end of the session all plans correspond to RRS. Thus, in the last TEAM session play converges to an efficient equilibrium. Against the observed distribution of plans in the TEAM sessions, the optimal plan is RSS.

5. CONCLUSION

The motivation for our experimental design was to explore whether communication technologies could explain differences between the Charness and

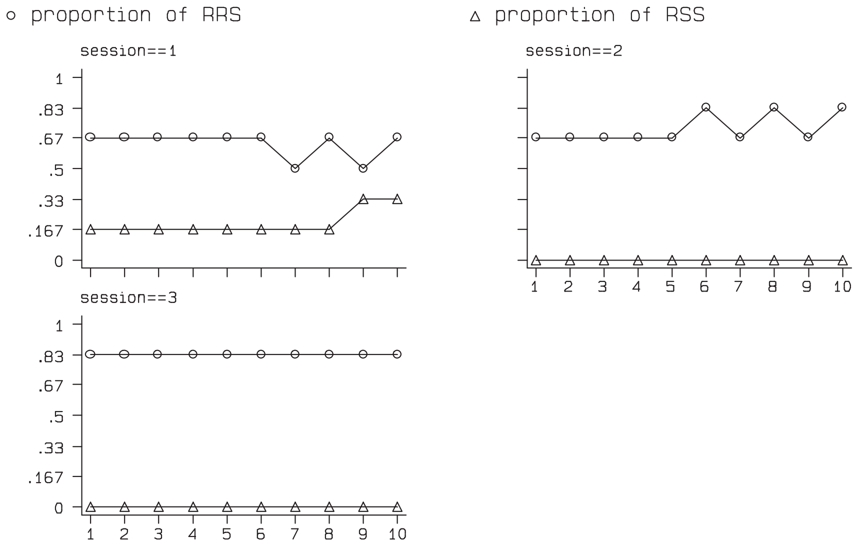


Fig. 5. Proportions of RRS and RSS Plans by Round in INDIVIDUAL Sessions.

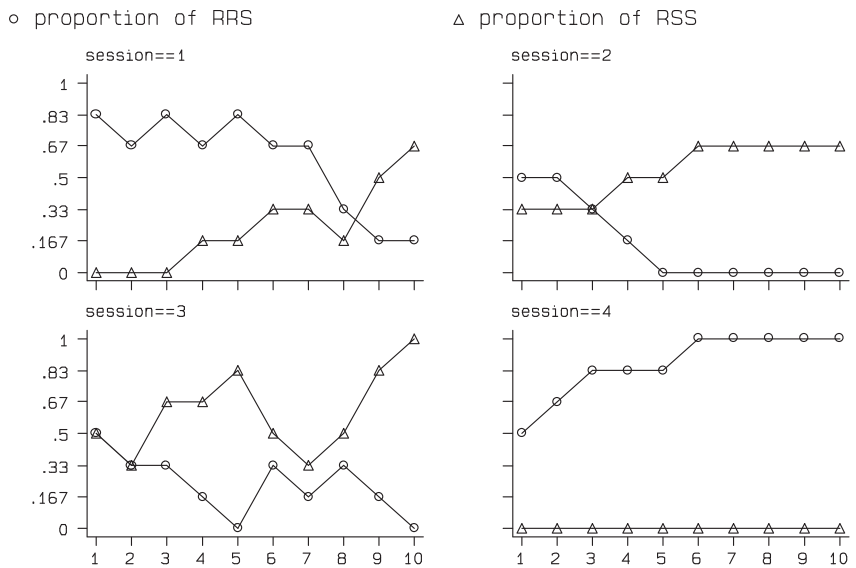


Fig. 6. Proportions of RRS and RSS Plans by Round in TEAM Sessions.

Clark et al. experimental results, and to provide a more direct test of whether and how subjects condition choices on messages.

We do not find evidence of a communication technology effect. Our most striking finding, in our view, is the considerable amount of variability between sessions using the same treatments. In fact, at the session level, there is as much variability within a treatment as between treatments. One way of putting this is that, except for the INDIVIDUAL PLAN treatment, if the treatment labels were removed then it would be extremely difficult from inspecting the choice data to know which session had been conducted under which treatment.

Notably, there are related experiments in which this kind of variability is not evident. For example, the payoffs for the CDFR and CKS games lie within a range for 2×2 coordination games where absent communication results are readily reproducible. Different procedures and subject pools may produce quantitative differences, as Battalio et al. (2001) note, but the essential findings are quite robust. Initially, R is frequently chosen, but R choices diminish with repetition, so that the risk dominant equilibrium is eventually observed. However, Battalio et al. also note that there is a range of payoffs of 2×2 coordination games where the 'results are mixed'. This happens when the strategies leading to the payoff-dominant equilibrium are only slightly riskier than the strategies leading to the risk-dominant equilibrium. One possibility is that communication opportunities transform the CKS game into one, where R is only slightly riskier than S.

One task for future research is to identify the links between initial play and convergence dynamics. We believe that the qualitative data from our TEAM treatment offers a promising avenue for understanding how subjects respond to differing histories of play. We also believe there is wide scope for designing experiments that separate learning and signaling influences on subject decisions.

Turning to our evaluation of Aumann's conjecture, the PLAN treatment provides information about how choices condition on messages under 2-WAY communication. We find an interesting mix of behaviors in the TEAM PLAN treatment. Even after the opportunity to discuss and reflect, most subjects do condition on messages. They signal an intention to play R, and go through with this intention if and only if their opponent also signals an intention to play R. A profile of two such plans does in fact constitute an equilibrium, and in sessions where there are sufficiently many of such subjects, this plan survives.

However, there is a substantial minority of subjects who do not condition, but instead announce R and then play S regardless of their opponent's message. Again, a profile of two such plans forms an equilibrium. In sessions where first round behavior includes both types of plan, changes in subsequent rounds are to be expected since the two plans are not best responses to one another. In all PLAN sessions where the latter plan appears, it increases in frequency over time.

A direction for further research is to examine these dynamic processes in more detail. An obvious approach is to consider subjects as adopting strategies by choosing the most attractive of a set of options, and evaluating the relative attractiveness of strategies with reference to their accumulated experience. Central to such an approach is how subjects evaluate attractiveness. It is interesting to think about how players would evaluate the attractiveness of their strategies after an RRS and RSS player meet. For example, the RSS player may view his payoff of 900, rather than 700, as including an unexpected bonus of 200 points reinforcing the attractiveness of his strategy. The RRS player may view the opponent's payoff of 900 as including a 100 point loss relative to the 1000 points obtainable from playing RRS, and may continue to believe RRS to be a better strategy.

Finally, results from our INDIVIDUAL sessions suggest that the PLAN and 2-WAY treatments may differ in unintended ways. The first conclusion to draw from this is about a limitation of the strategy method for eliciting information about some sequential games. Not only does the strategy method require subjects to consider all possible contingencies, but it also removes the temporal ordering of decisions. The first feature generates a decision task that is more demanding of subject's thought and decision processes, while the second may make some labels that are naturally interpreted in a sequential context more difficult to interpret. The second conclusion we draw from this is that individuals and teams approach the plan in different ways. We suspect that subjects manage the cognitive demands of the plan better in the TEAM treatment, either because team members help each other directly by sharing knowledge, or indirectly by providing an environment that encourages subjects to think more carefully about decisions.

NOTES

1. Interestingly, he finds less efficient outcomes when a player sends a message *after* making his own choice, but before his opponent makes their choice.

2. See Harsanyi and Selten (1988, p. 80).

3. Battalio, Samuelson, and Van Huyck (2001) study a range of 2×2 coordination games and find that, without opportunities for communication, the risk-dominant equilibrium appears to have strong attracting properties. Evolutionary analysis

of such games also often predicts convergence to risk-dominant equilibria (e.g. see Kandori, Mailath, & Rob, 1993).

4. Aumann (1990, p. 203). Italics in original.

5. Aumann (1990, p. 203).

6. Aumann's argument does not apply to the CDFR game because a player intending to choose S has no incentive to misrepresent his intention.

7. Farrell (1988, p. 213).

8. Twenty subjects participated in each session, with each subject playing a sequence of ten one-shot games.

9. Charness found no substantive differences between the specifications.

10. There are a number of other differences between the two designs. See the relevant papers for a detailed description of their respective procedures.

11. As we shall discuss presently, what constitutes a 'player' differs across sessions.

12. The strategy method was introduced by Selten (1967), and has been used in numerous studies. Since subjects indicate a response to a contingency that may not actually arise, their response may be less considered than if they were actually facing that contingency, especially if they view the contingency as ex ante unlikely. Some studies have found differences between choices elicited via strategy and direct response methods (e.g. Brosig, Weimann, & Yang, 2003) while others have not (e.g. Brandts & Charness, 2000). See Roth (1995) for a discussion of the strategy method.

13. Actual earnings ranged from £3.30 to £10 and averaged £7.57. TEAM sessions averaged slightly longer than 1 h, while INDIVIDUAL sessions were substantially shorter.

14. Full copies of the instructions are available from the authors on request.

15. For expository purposes we will continue to refer to the strategies as R and S, although they were labelled "1" and "2" in the experiment.

16. Except for obvious changes in the wording of messages/choices, e.g. "We" was used in place of "I".

17. In the TEAM treatment both subjects in the team were paid in this manner. At the time of the experiment £1 was worth approximately $1.60.

18. All comparisons with the Charness data refer to his CKS sessions. See his original paper for results from other payoff matrices.

19. The statistic is based on the percentages listed in each column. In cases where some percentages are undefined the statistic is not computed.

20. The INDIVIDUAL and Charness treatments deliver significant differences in proportions of R announcements ($p = 0.034$) and proportions of R choices ($p = 0.077$). The TEAM and Charness treatments deliver significant differences in proportions of R announcements ($p = 0.021$), R choices ($p = 0.043$), R choices by R senders ($p = 0.061$), efficient equilibrium outcomes ($p = 0.043$) and inefficient equilibrium outcomes ($p = 0.043$).

21. One indication of this is that the decision-making phase of the TEAM sessions lasted longer than the INDIVIDUAL sessions.

22. The next most common plan accounted for only 9% of all those submitted.

23. That is, RRS is optimal taking the proportion of each plan across all rounds and sessions as the constant probability of meeting that plan.

ACKNOWLEDGEMENT

We thank John Bone, Gary Charness, and two anonymous referees for useful comments and advice. We also thank Sue Chilton, Judith Covey, Paul Dolan, Martin Jones, Roxanna Radulescu and Rebecca Taylor for assistance in running the experiment. Research support from the Leverhulme Foundation is gratefully acknowledged.

REFERENCES

Aumann, R. J. (1990). Nash equilibria are not self-enforcing. In: J. J. Gabszewicz, J.-F. Richard & L. A. Wolsey (Eds), *Economic decision making: Games, econometrics and optimisation*. Amsterdam: North-Holland.

Battalio, R., Samuelson, L., & Van Huyck, J. (2001). Optimization incentives and coordination failure in laboratory stag hunt games. *Econometrica*, *69*, 749–764.

Brosig, J., Weimann, J., & Yang, C.-L. (2003). The hot versus cold effect in a simple bargaining experiment. *Experimental Economics*, *6*, 75–90.

Brandts, J., & Charness, G. (2000). Hot vs. cold: Sequential responses and preference stability in experimental games. *Experimental Economics*, *2*, 227–238.

Charness, G. (2000). Self-serving cheap talk: A test of Aumann's conjecture. *Games and Economic Behavior*, *33*, 177–194.

Clark, K., Kay, S., & Sefton, M. (2001). When are nash equilibria self-enforcing? An experimental analysis. *International Journal of Game Theory*, *29*, 495–515.

Cooper, R., DeJong, D. V., Forsythe, R., & Ross, T. W. (1992). Communication in coordination games. *Quarterly Journal of Economics*, *53*, 739–771.

Crawford, V. (1998). A survey of experiments on communication via cheap talk. *Journal of Economic Theory*, *78*, 286–298.

Farrell, J. (1988). Communication, coordination, and nash equilibrium. *Economics Letters*, *27*, 209–214.

Farrell, J., & Rabin, M. (1996). Cheap talk. *Journal of Economic Perspectives*, *10*, 103–118.

Fudenberg, D., & Tirole, D. (1991). *Game theory*. Cambridge, MA: MIT Press.

Harsanyi, J. C., & Selten, R. (1988). *A general theory of equilibrium selection in games*. Cambridge, MA: MIT Press.

Kandori, M., Mailath, G., & Rob, R. (1993). Learning, mutation and long run equilibria in games. *Econometrica*, *61*, 29–56.

Roth, A. (1995). Bargaining experiments. In: J. Kagel & A. Roth (Eds), *Handbook of experimental economics*. Princeton, NJ: Princeton University Press.

Selten, R. (1967). Die strategiemethode zur erforschung des eingeschrankt rationalen verhaltens im rahmen eines oligopolexperiments. In: H. Sauermann (Ed.), *Beitrage zur experimentellen wirtschaftsforschung*. Tubingen: J.C.B. Mohr.

This page intentionally left blank

TRUST BUT VERIFY: MONITORING IN INTERDEPENDENT RELATIONSHIPS

Maurice E. Schweitzer and Teck H. Ho

ABSTRACT

For organizations to be effective, their employees need to rely upon each other even when they do not trust each other. One tool managers can use to promote trust-like behavior is monitoring. In this chapter, we report results from a laboratory study that describes the relationship between monitoring and trust behavior. We randomly and anonymously paired participants ($n = 210$) with the same partner, and had them make 15 rounds of trust game decisions. We find predictable main effects (e.g. frequent monitoring increases trust behavior) as well as interesting strategic behavior. Specifically, we find that anticipated monitoring schemes (i.e. when participants know before they make a decision that they either will or will not be monitored) significantly increase trust behavior in monitored rounds, but decrease trust behavior overall. Participants in our study also reacted to information they learned about their counterpart differently as a function of whether or not monitoring was anticipated. Participants were less trusting when they observed trustworthy behavior in an anticipated monitoring period, than when they observed trustworthy behavior in an unanticipated monitoring period. In many cases, participants in our study systematically anticipated their

Experimental and Behavioral Economics
Advances in Applied Microeconomics, Volume 13, 87–106

ISSN: 0278-0984/doi:10.1016/S0278-0984(05)13004-1

counterpart's untrustworthy behavior. We discuss implication of these results for models of trust and offer managerial prescriptions.

1. INTRODUCTION

Managers, employees, and customers routinely rely on others to choose trustworthy actions. Managers expect employees to complete their work, employees expect to be paid, and customers expect goods and services to be delivered on time. In some cases, people choose trustworthy actions because they are genuinely trustworthy people. In other cases, however, people choose trustworthy actions because they are concerned with the consequences of being caught engaging in untrustworthy actions. In this article, we examine the influence of monitoring on trustworthy and trusting behavior. We conceptualize monitoring as a tool that can promote trust-like behavior, and we investigate the relationship between different monitoring systems and the trust-like actions people choose.

In general, trust reduces transaction costs and improves the efficiency of economic transactions (Bromiley & Cummings, 1995; Hirsch, 1978; Ring & Van de Ven, 1992). At the managerial level, trust enables managers to negotiate more efficiently (Bazerman, 1994) and lead more effectively (Atwater, 1988). Managers, however, often have difficulty judging others (Wu, Heath, & Knez, 2003), and in many organizational settings, including negotiations (O'Connor & Carnevale, 1997; Schweitzer & Croson, 1999), sales (Santoro & Paine, 1993), and accounting (Chang & Schultz, 1990; Degeorge, Patel, & Zeckhauser, 1999), people routinely engage in untrustworthy and unethical behavior (Carr, 1968). Even in settings where people cannot or do not trust each other, however, people often act in trusting and trustworthy ways and reap economic and social benefits from exchanges.

In practice, across many organizational settings managers need to be both trusting and cautious. In many cases, people audit or check the claims of others, and prior work has analyzed how principles and agents can structure optimal contracts and auditing arrangements (Townsend, 1979; Mookherjee & Png, 1989). In this chapter, we use experimental methods to investigate how managers should monitor the actions of others when trust is low.

We report results from a laboratory study using a repeated version of the trust game (Berg, Dickhaut, & McCabe, 1995). In our experiment, we measure behavior that reflects trusting and trustworthy behavior, and we examine the influence of different monitoring systems on this behavior.

1.1. Trust and monitoring

Prior work has considered a wide range of trust relationships (see Ross & LaCroix, 1996 for a review). In this chapter, we focus on repeated interactions in emerging relationships. Lewicki and Bunker (1996) and Lewicki and Wiethoff (2000) define early stage trust relationships as calculus-based relationships. In these relationships, people "calculate" the costs and benefits of keeping or breaking trust. Relative to well-developed or mature relationships, calculus-based trust is easily broken and (relatively) easily repaired.

Several recent studies have explored trust behavior in experimental settings. This work has identified a number of individual and contextual factors that influence trust. These include solidarity, familiarity, a common nationality (Glaeser, Laibson, Scheinkman, & Soutter, 2000), and cultural orientation (Buchan & Croson, 1999), as well as contextual factors such as non-task communication (Buchan & Croson, 1999) and the stage of the game (Ho & Weigelt, 2001). In fact, even the labels used to describe a counterpart influences trust. Labeling a counterpart as a partner increases trust, while labeling a counterpart as an opponent decreases trust (Burnham, McCabe, & Smith, 2000).

While much of the experimental work investigating trust has used a version of the trust game (Berg et al., 1995), related work has investigated cooperation using paradigms such as repeated prisoners dilemma games (Gibson, Bottom, & Murnighan, 1999) and repeated ultimatum games (Boles, Croson, & Murnighan, 2000). This work has identified important dynamic changes in behavior such as a link between revealed deception and retribution. For example, in repeated ultimatum games responders are more likely to reject offers of proposers when they used deception in the past (Boles et al., 2000).

No prior experimental work, however, has investigated the interplay between monitoring and trust behavior. In fact, most prior experiments have used full information conditions under which participants always learn about the actions of their counterpart. The relationship between trust and monitoring, however, is clearly an important one for both theoretical and practical reasons. Managers make important decisions to trust and to monitor the actions of others. In many cases, these decisions are made selectively (e.g. using a mixed strategy; Amaldoss & Jain, 2002). For example, some managers randomly drug test employees, conduct audits, and even listen in on employee phone calls (e.g. in call centers). In one study, Aiello (1993) documented the purchase of surveillance software between 1990 and 1992.

He found that over 70,000 U.S. companies made at least one such purchase, at a total cost of more than $500 million. The goal of this work is to examine the dynamics of monitoring and trust behavior.

2. HYPOTHESIS

We conceptualize monitoring as a tool to produce trust-like behavior, and we consider ways, in which monitoring changes incentives for engaging in trust behavior. We adopt a functional view of trust by assuming individuals calculate the costs and benefits of engaging in trust behavior in a manner consistent with Ajzen's (1985, 1987) theory of planned behavior. In particular, we focus on the role of monitoring systems in altering the expected costs and benefits of choosing trust-like actions. This conceptualization matches our experimental design, because participants in our study are paid for their outcomes and remain anonymous to their counterpart. There is no way for participants to find out who their partners are, and as a result, participants in our study face a well-defined "shadow of the future" defined by the future rounds of a repeated trust game.

In developing our hypotheses we make two assumptions about trust. First, we assume that individuals maximize their long-term profits when they are trusted by others, but maximize their short-term profits when they choose untrustworthy actions. In our experiment, the repeated trust-game framework matches this incentive structure. Second, we assume that people reciprocate trust-like actions. That is, people are more likely to (dis)trust their counterpart when they observe their counterpart engaging in (un)trustworthy behavior (Gibson et al., 1999).

Consider monitoring regimes that range from one extreme, complete monitoring, to another extreme, no monitoring. Under complete monitoring, participants have an incentive to exhibit trust-like behavior, especially in early rounds, because their counterpart will observe their actions and can reciprocate in future rounds. On the other hand, under no monitoring, participants have little incentive to engage in trust-like behavior, because their counterpart cannot observe their actions. In this case, the expected benefits of choosing untrustworthy actions exceed the expected benefits of choosing trust-like actions. If players are self-interested, opportunism is likely to prevail. The more frequent the monitoring, the greater the expected benefits of choosing trust-like actions. Consequently, we hypothesize that monitoring frequency will be positively correlated with trust-like behavior. With more frequent monitoring, untrustworthy behavior is more likely to be detected and subsequently punished or reciprocated. As a result, if players are calculative and maximize their total payoff for the entire interaction, they will exhibit more trust-like behavior when they experience more frequent monitoring.

H1. *Frequent monitoring will increase overall trust-like behavior.*

We also expect anticipated monitoring to significantly influence behavior. Specifically, if a decision-maker anticipates that his or her actions will be monitored during a specific period of time, we expect the decision-maker to be more likely to engage in trust-like behavior for that period. In anticipated monitoring periods, the decision-maker knows that his or her actions will be observed, and the expected costs of engaging in untrustworthy actions in these periods are particularly high. Observed untrustworthy actions harm trust, and prior work demonstrates that the trust recovery process is both slow and costly (Schweitzer, Hershey, & Bradlow, 2004; Tomlinson, Dineen, & Lewicki, 2004). Consequently, we hypothesize that players will be particularly likely to engage in trust-like behavior in anticipated monitoring rounds.

H2. *Anticipated monitoring will increase trust-like behavior for anticipated monitoring rounds.*

Although an anticipated monitoring scheme is likely to increase trust-like behavior in periods when monitoring is anticipated, anticipated monitoring schemes may decrease trust-like behavior in periods when no monitoring is anticipated for two reasons. First, anticipated monitoring schemes lower the costs of engaging in untrustworthy actions in anticipated, non-monitored periods. In anticipated monitoring schemes, participants know when their actions will be observed – and when their actions will not be observed. In non-monitoring periods, participants face no adverse economic consequences for choosing untrustworthy actions.

Second, anticipated monitoring schemes decrease trust-like behavior in periods of no monitoring by harming trust development. Players who observe others choosing trustworthy actions in anticipated monitoring periods may attribute the trustworthy behavior they observe to the monitoring scheme rather than the trustworthiness of the individual. As a result, players who observe trustworthy behavior in anticipated monitoring periods may be less likely to assume that their counterpart will choose trustworthy actions when they are not monitored. As a result, trustworthy actions are less diagnostic of true trustworthiness and less effective in building trust when the observed trust behavior occurs in an anticipated period of monitoring.

For both of these reasons, we hypothesize that anticipated monitoring will decrease trust-like behavior in non-monitored rounds.

H3. *Anticipated monitoring will decrease trust-like behavior in rounds when participants are not monitored.*

3. EXPERIMENTAL METHODS

We recruited participants for an experiment in decision-making from class announcements. Prospective participants were told that they would have the opportunity to earn money in the experiment and that the amount they earned would depend partly upon their own decisions, partly upon the decisions of others, and partly upon chance.

Upon arrival to the experiment, participants were randomly assigned to either the Odd or the Even role. Participants in the two roles were separated and anonymously paired with a member of the opposite role.

In this experiment participants played 15 rounds of the trust game depicted in Fig. 1. In each round the Odd player begins with an endowment of 5 points. The Odd player can choose to take some portion of the 5 points or pass the 5 points. If the Odd player chooses to take some of the points, the round ends and the Odd and Even player earn the division of points selected by the Odd player. If the Odd player chooses to pass the 5 points, the amount of points doubles and the Even player decides how to divide 10 points between the two players.

We used the strategy method in this experiment. In each round both Odd and Even players make decisions, even though the Even player's decision may not influence the outcome of the round. Participants played the same game with the same partner for 15 rounds. Participants remained in their role throughout the experiment, and received limited information (monitoring) about their partner's decisions.

Dyads were randomly assigned to one of four between-subject monitoring conditions. The four conditions result from a 2×2 design. We assigned dyads to one of two "frequency of monitoring" conditions (frequent monitoring or infrequent monitoring) and to one of two "anticipated monitoring" conditions (anticipated monitoring or unanticipated monitoring).

When participants were able to monitor their partner's actions, they only learned what their partner's choice for that particular round was. That

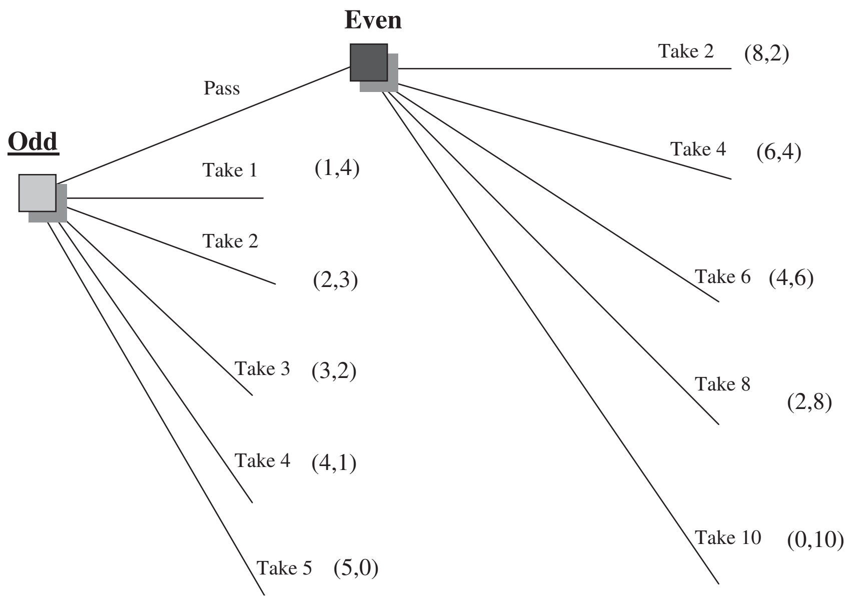


Fig. 1. The Trust Game Participants Played.

is, Odd players learned what their Even partner had selected for that round, and Even players learned what their Odd partner had selected for that round.

We manipulated the frequency of monitoring by giving participants either frequent monitoring (10 of 15 rounds of monitoring) or infrequent monitoring (5 of 15 rounds of monitoring). We randomly and differently selected a set of 5 rounds for each dyad to be either monitoring or non-monitoring rounds. Within each dyad, Odd and Even players had the same monitoring and non-monitoring rounds.

We also manipulated whether or not monitoring rounds were anticipated. In the anticipated condition we indicated on each participant's decision sheet whether or not a round was a monitoring round *before* they made their decision for that round. In the unanticipated condition we indicated whether or not a round was a monitoring round only *after* they had made their decision for that round. As a practical matter, participants knew the total number of monitoring rounds they would encounter and could update their probability estimates that an upcoming round would be a monitoring round.

At the conclusion of the experiment we gave participants feedback (monitoring) for every round and paid them based upon the total number of points they earned. We paid participants \$1 for every 5 points they earned.

4. RESULTS

4.1. Model

We fit two related logit models to our data. These models represent the likelihood a participant will choose a trusting or trustworthy action as a function of the treatment variables and the amount of trust-like behavior she/he has observed. We define trusting behavior as the Odd player decision to pass, and we define trustworthy behavior as the Even player decision to return at least 6 (Model 1) or to return at least 4 (Model 2) of the 10 points. Both models take the standard logistic functional form as follows:

$$P_i(r) = \frac{e^{\alpha_i+\beta_{iA}A+\beta_{iH}H+\beta_{iMA}M(r)A+\gamma_{in(r)}n(r)+\gamma_{in(r)A}n(r)A}}{1+e^{\alpha_i+\beta_{iA}A+\beta_{iH}H+\beta_{iMA}M(r)A+\gamma_{in(r)}n(r)+\gamma_{in(r)A}n(r)A}}$$

In this model, $P(r)$ represents the likelihood of choosing a trust-like action (to pass or to return at least 6 or 4) in round r. We use i to indicate whether the participant is Odd (O) or Even (E), and we include α as a model intercept to allow for differences in the propensity to engage in trust-like behavior between Odd and Even players.

We represent the experimental conditions with A and H. We set $A = 1$ for the anticipated monitoring conditions and 0 for the unanticipated monitoring conditions, and we set $H = 1$ for the frequent monitoring conditions (10 rounds of monitoring), and 0 for the infrequent monitoring conditions (5 rounds of monitoring).

$M(r)$ represents whether or not round r is a monitoring round. We set $M(r) = 1$ for a monitoring round, and 0 for a non-monitoring round. In our model the parameter estimate β_{iMA} (when $M(r) * A = 1$) measures how trust-like behavior might be different in monitored rounds in the anticipated monitoring condition.

The model also includes a variable, $n(r)$, to represent prior observations of trust-like behavior. Specifically, $n(r)$ represents the fraction of observed (monitoring) rounds that include trustworthy or trusting behavior in rounds 1 through $r-1$. This fraction represents trust-building from observed behavior. We expect the importance of observed behavior to differ across anticipated and unanticipated conditions, so we also include an interaction term $n(r) * A$ to account for the potential moderating effect of anticipated monitoring.

4.2. Participants

A total of 210 participants completed the experiment. These participants created 105 dyads. Each dyad was randomly assigned to each of the four conditions; a total of 24 dyads completed the unanticipated infrequent condition, 23 dyads completed the unanticipated frequent condition, 26 dyads completed the anticipated infrequent condition, and 32 completed the anticipated frequent condition. We describe the data and report results from our model across these conditions.

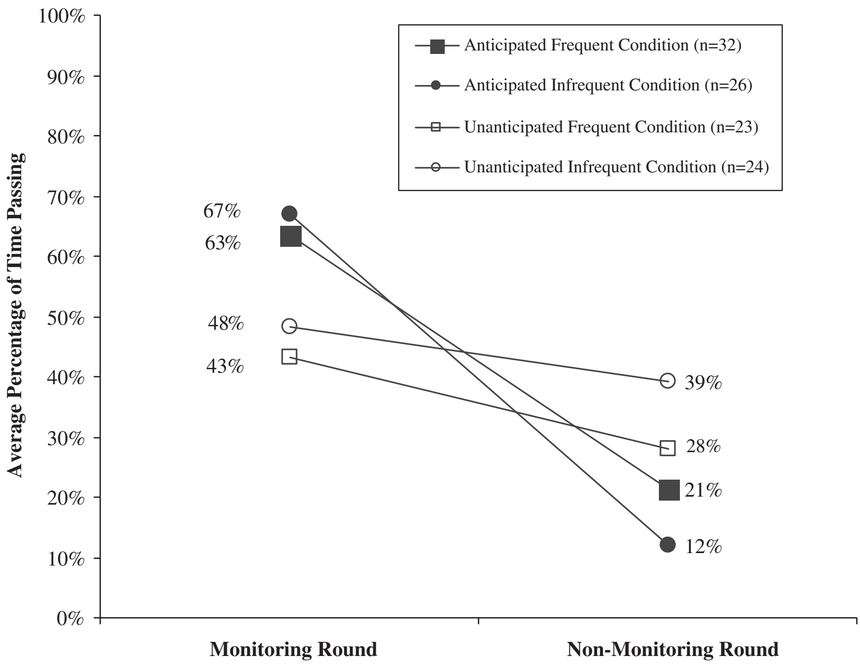


Fig. 2. Odd Player Decisions across Conditions by Type of Monitoring Round.

4.3. Trusting Results

On average, Odd players in our experiment passed 40.1% of the time and Even players returned at least half of the points 52.9% of the time. This behavior contrasts with trust game behavior in prior experiments with full monitoring and different growth multiples for passing. In prior work, participants were generally more trusting and more trustworthy than they were on average in our experiment. For example, Berg, Dickhaut, and McCabe (1995) found that Odd players passed money almost 94% of the time and that Even player returned money 75% of the time.

Importantly, in our study we find that Odd and Even player decisions were significantly influenced by whether or not monitoring was anticipated in the upcoming round. Odd players were most likely to choose trusting actions and Even players were most likely to choose trustworthy actions in rounds with anticipated monitoring. Conversely, Odd players were least likely to choose trusting actions and Even players were least likely to choose trustworthy actions in rounds when they anticipated no monitoring. We depict the average Odd player decisions to pass, and the average Even player decisions to return in Figs. 2 and 3. In Fig. 4, we depict these patterns of results for a representative dyad in the frequent anticipated monitoring

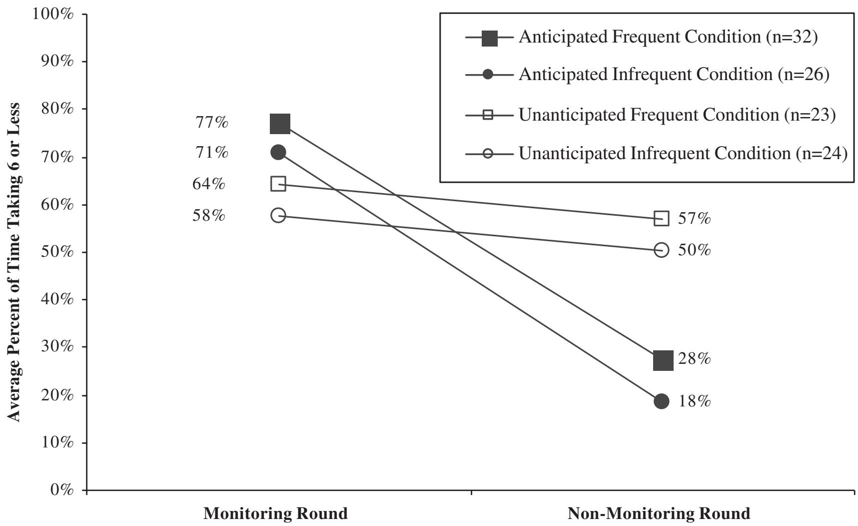


Fig. 3. Even Player Decisions across Conditions by Type of Monitoring Round.

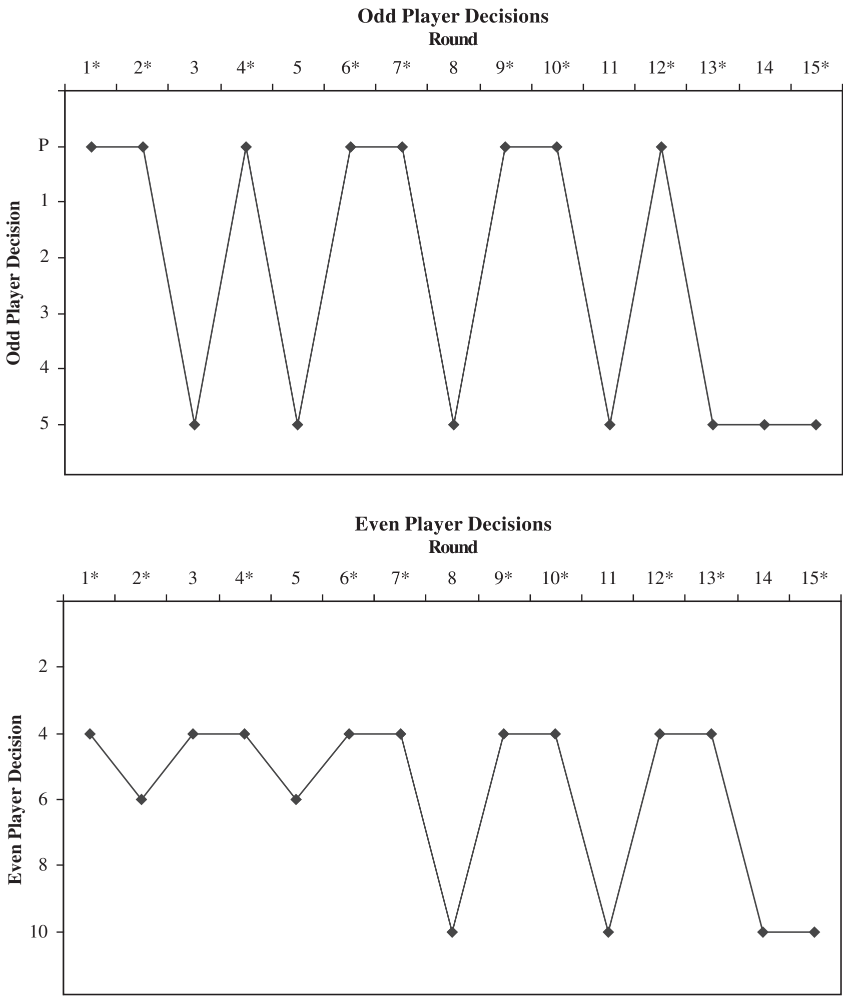


Fig. 4. Example Odd and Even Player Actions (One Dyad in the Anticipated, Frequent Monitoring Condition). *Indicates a monitoring round.

condition. In this dyad, both the Odd and Even player behave strategically: they chose trusting and trustworthy actions in rounds 7, 9, and 10 when they anticipated monitoring, but chose untrustworthy and distrusting actions in rounds 8, 11, and 14 when they anticipated no monitoring.

Table 1. Model 1 Predicting Trust-Like Behavior, Defining Trustworthy Actions as Taking 6 or Less.

	Parameter Estimate
Odd Player (*Predicting Probability of Passing*)	
Intercept, α	-1.22^{***}
Anticipated A, β_{OA}	-0.91^{***}
High monitoring H, β_{OH}	0.18^{**}
Interaction $M(r) * A$, β_{OMA}	2.28^{***}
# Trustworthy/# feedback, $\gamma_{\mathrm{On}(t)}$	0.59^{***}
Interaction $A * n(t)$, $\gamma_{\mathrm{OAn}(t)}$	-0.24^{*}
Even Player (*Predicting Probability of Taking 6 or Less*)	
Intercept, α	-0.16 n.s.
Anticipated A, β_{EA}	-1.67^{***}
High monitoring H, β_{EH}	0.54^{***}
Interaction $M(r) * A$, β_{EMA}	2.07^{***}
# Trustworthy/# feedback, $\gamma_{\mathrm{En}(t)}$	0.21^{***}
Interaction $A * n(t)$, $\gamma_{\mathrm{EAn}(t)}$	0.26^{t}

Log-likelihood = −1868

Across both model 1 (Table 1) and model 2 (Table 2), we find that participants were generally more trusting and trustworthy when monitoring was not anticipated and frequent, but that participants were more trusting and trustworthy in specific rounds with anticipated monitoring. Participants were also significantly influenced by the behavior they observed.

$^{*}p<0.05$,

$^{**}p<0.01$,

$^{***}p<0.001$,

$^{\mathrm{t}}p<0.10$

We next examine results from our model. We depict these results in Tables 1 and 2. We define trustworthy behavior differently across the two models: taking 6 or less in model 1 and taking 4 or less in model 2. Both models yield very similar results.

In each round Odd players could either pass, a trust-like action, or take, and Even players could choose to return either a substantial amount, a trustworthy action, or a small amount. The first set of parameter estimates in Tables 1 and 2 describe the influence of monitoring on the likelihood that Odd players will pass.

Our first hypothesis predicts that frequent monitoring will increase trust-like behavior. We find support for this hypothesis across both models; the parameter estimates for β_{OH} are positive and significant for models 1 and 2, 0.18 (0.17), $p = 0.01$ and 0.19 (0.10), $p = 0.05$, respectively.

Table 2. Model 2 Predicting Trust-Like Behavior, Defining Trustworthy Actions as Taking 4 or Less.

	Parameter Estimate
Odd Player (*Predicting Probability of Passing*)	
Intercept, α	-1.18^{***}
Anticipated A, β_{OA}	-1.05^{***}
High monitoring H, β_{OH}	0.19^{*}
Interaction $M(r) * A$, β_{OMA}	2.31^{***}
# Trustworthy/# feedback, $\gamma_{\mathrm{On}(t)}$	0.91^{***}
Interaction $A * n(t)$, $\gamma_{\mathrm{OAn}(t)}$	-0.29^{*}
Even Player (*Predicting Probability of Taking 4 or Less*)	
Intercept, α	-0.81^{***}
Anticipated A, β_{EA}	-1.61^{***}
High monitoring H, β_{EH}	0.31^{***}
Interaction $M(r) * A$, β_{EMA}	2.10^{***}
# Trustworthy/# feedback, $\gamma_{\mathrm{En}(t)}$	0.24^{***}
Interaction $A * n(t)$, $\gamma_{\mathrm{EAn}(t)}$	0.20 n.s.

Log-likelihood = -1824
$^{*}p<0.05$, $^{**}p<0.01$,
$^{***}p<0.001$, $^{t}p<0.10$

Our second and third hypotheses predict that anticipated monitoring will increase trust-like behavior for anticipated monitoring rounds, but decrease trust-like behavior overall. The β_{OMA} parameter estimates are positive and significant for both models, 2.28 (0.15) and 2.31 (0.12), $p<0.01$, and the β_{OA} parameter estimates are negative and significant for both models, -0.91 (0.20) and -1.05 (0.19), $p<0.01$. That is, while anticipated monitoring increases trust-like behavior for specific anticipated monitoring rounds, it harms trust-like behavior overall.

We next consider the influence of experience on trust-like behavior. Specifically, we examine whether Odd players will be more trusting when they observe trustworthy behavior and whether Odd players will be less influenced by trustworthy observations when monitoring is anticipated. We find evidence for both. When Odd players observed trustworthy actions they were more likely to pass to their Even player counterpart; parameter estimates for $\gamma_{\mathrm{n}(t)}$ are 0.59 (0.07) and 0.91 (0.08), $p<0.01$ for models 1 and 2. In addition, Odd players discounted the trustworthy behavior they observed when that behavior occurred in an anticipated monitoring round; parameter estimates for $\gamma_{\mathrm{An}(t)}$ are -0.24 (0.11), $p = 0.03$ and

-0.29 (0.14), $p = 0.04$ for models 1 and 2. That is, the trustworthy behavior Odd players observed influence their subsequent actions, but the same behavior influences their subsequent actions less if it occurs in anticipated monitoring rounds.

4.4. Trustworthy Behavior

We find a similar set of results for trustworthy behavior. Even players decided how much to return to their Odd player counterpart. Even players could choose to return either a substantial amount (at least 4 or at least 6 of the potential 10 points) or a small amount (e.g. 2 or 0). The second set of parameter estimates in Tables 1 and 2 describe the influence of monitoring and observed behavior on the likelihood that Even players will choose trustworthy actions by choosing to return a substantial amount of the potential 10 points (at least 6 in Model 1 and at least 4 in Model 2).

Our first hypothesis predicts that frequent monitoring will increase trustworthy behavior. We find support for this hypothesis across both models; the parameter estimates for β_{EH} are positive and significant for models 1 and 2, 0.54 (0.10) and 0.31 (0.08), $p < 0.01$ for both models.

Our second and third hypotheses predict that anticipated monitoring will increase trustworthy behavior for anticipated monitoring rounds, but decrease trustworthy behavior overall. Across both models the β_{EMA} parameter estimates are positive and significant, 2.07 (0.13) and 2.10 (0.16), $p < 0.01$ for both models, and the β_{EA} parameter estimates are negative and significant, -1.67 (0.18) and -1.61 (0.29), $p < 0.01$ for both models. That is, while anticipated monitoring increases trust-like behavior for specific, anticipated monitoring rounds, it harms trust overall.

We next consider the influence of experience on trust-like behavior. We conjecture that Even players will be more trustworthy when they observe trusting behavior, and that Even players will be less influenced by these observations when monitoring is anticipated. We find that, when Even players observe trusting actions they are more likely to return a substantial amount to their Odd player counterpart; parameter estimates for $\gamma_{\mathrm{En}(t)}$ are 0.21 (0.07) and 0.24 (0.07), $p < 0.01$ for both models. Even players, however, do not significantly change their behavior as a function of whether or not the trusting behavior they observe occurred in an anticipated monitoring round; parameter estimates for $\gamma_{\mathrm{EAn}(t)}$ are 0.26 (0.14) $p = 0.06$ and 0.20 (0.14) $p = 0.16$, for models 1 and 2.

4.5. Violations of Trust

We next consider violations of trust. We define extreme untrustworthy behavior as Even player decisions to take 10. We define a trust violation as a round in which the Odd player passes to an Even player who takes 10. We report differences in these behaviors across conditions in Table 3.

For each dyad, we counted the number of times Even players chose to take 10. This behavior occurred most often in the infrequent and anticipated monitoring conditions. In analysis of variance, the amount of extreme untrustworthy behavior was significantly influenced by the amount of monitoring, $F(1,101) = 18.66$, $p<0.001$, and by anticipated monitoring, $F(1,101) = 6.19$, $p = 0.014$, but not significantly influenced by an interaction between the two, $F(1,101) = 1.91$, $p =$ n.s.

As we depict in Table 3, however, Odd players generally anticipated Even player attempts to take 10. From an analysis of variance model of cases in which Even players attempted to take 10, we find that trust violations occurred most often when monitoring was infrequent, $F(1,101) = 18.66$, $p<0.001$, and anticipated, $F(1,101) = 6.19$, $p = 0.01$. We find no significant interaction between the two conditions, $F(1,101) = 1.91$, $p =$ n.s.

Overall, successful violations of trust occurred during monitoring rounds about half the time. Although Even players chose to take 10 more often in non-monitoring rounds than monitoring rounds, Odd players passed much less often in non-monitoring rounds. Successful violations of trust during a

Table 3. Violating Trust.

	Monitoring regime			
	Anticipated		Unanticipated	
	Frequent	Infrequent	Frequent	Infrequent
Number of attempted violations (Even chose to take 10)	4.84	8.96	4.04	6.17
Number of actual violations (Odd passed and Even took 10)	0.88	1.50	0.61	0.63
Percent of violations in monitoring rounds	51.3%	51.8%	93.3%	53.4%

Strategic attempts to violate the trust of others varied significantly across conditions. In general, however, participants often anticipated the strategic actions of their counterpart.

monitoring round were significantly more likely to occur when monitoring was unanticipated and frequent. We depict this pattern in Table 3.

5. DISCUSSION

While prior work has argued that trust is an essential ingredient for managerial effectiveness (Atwater, 1988), many relationships within organizations lack trust. For example, managers within large organizations with high turnover may not have sufficient time to develop trust among all of their employees. Even in the absence of actual trust, however, managers need to induce trust-like behavior. In this chapter, we conceptualize monitoring as a substitute for trust, and we demonstrate that monitoring systems significantly influence trust-like behavior.

We report results from a laboratory experiment that describes important relationships between different monitoring conditions and behavior. First, we find that frequent monitoring increases overall trust-like behavior. Second, we find that anticipated monitoring harms overall trust-like behavior, but significantly increases trust-like behavior for periods in which monitoring is anticipated; in our study, Even players were particularly trustworthy in anticipated monitoring rounds, but were particularly untrustworthy when they anticipated no monitoring.

Our results also demonstrate that people anticipated the strategic trust-like actions of their counterpart. For example, Even players often attempted to take all of the money (points) when they could not be observed (e.g. anticipated no monitoring rounds). Anticipating this, Odd players rarely passed in these rounds.

The decisions our participants made were also influenced by the set of actions they observed. Specifically, Odd players were more trusting when they had observed trustworthy behavior. Prior work has found that people often under-attribute behavior in strategic games to contextual factors (Weber, Camerer, Rottenstreich, & Knez, 2001; Weber & Camerer, 2003), but in our study, Odd players were very sensitive to whether or not the behavior they observed occurred when monitoring was anticipated or unanticipated. Odd players were less trusting when the trustworthy behavior they observed occurred when monitoring was anticipated than when it was unanticipated.

Although we find significant differences in behavior across anticipated and unanticipated monitoring conditions, these differences are probably understated by the design of this experiment. In each condition, participants knew the total number of monitoring rounds they would encounter. As a result, in later rounds of the experiment participants in the unanticipated conditions could update their beliefs about the likelihood that an upcoming round would be a monitoring round. In these cases, behavior in unanticipated monitoring conditions is likely to mirror behavior in anticipated monitoring conditions.

By design we paired participants randomly and anonymously. Our participants had no communication and they had no history with their partner. This aspect of our design enables us to focus our attention on our treatment conditions (i.e. to gauge the effects of monitoring in a calculus-based trust environment) by controlling for relationship and communication effects. Future work, however, should extend our investigation of the dynamics between monitoring and trust-like behavior to richer environments. For example, communication may facilitate trust-like behavior. Communication could help trustors promote trustworthy behavior with promises (Schweitzer et al., 2004), facilitate relationship development, and articulate consequence of observed trust violations. As a result, we expect communication to moderate the relationship between monitoring and trust-like behavior.

Another important direction for future research is the interplay between monitoring and trust development itself. By implementing a monitoring system managers may actually impede trust development. Cialdini (1996) notes that when employees are monitored it communicates to them that they are not trusted. This can create psychological reactance, as employees develop beliefs about their own behavior consistent with low expectations. Employees may come to attribute their own trust-like behavior to the monitoring system rather than to underlying feelings of trust. Ultimately, the use of a monitoring system may lead otherwise trustworthy employees to act in untrustworthy ways when they are not monitored.

Monitoring may also harm performance more broadly. The mere presence of a monitoring system may create anxiety and change behavior in unintended ways. For example, Hochschild (1983) documented how the fear of monitoring among Delta Airlines flight attendants harmed their performance. An additional concern about the unintended consequences of monitoring is the effect of monitoring on those who do the monitoring. Kruglanski (1970) demonstrates that people who conduct surveillance become less trusting.

Future work should also examine the role of monitoring with respect to organizational culture. Organization culture impacts a number of organizational behaviors related to trust-like behavior (Wimbush & Shepard, 1994), and recent experimental work has begun to identify important dynamics of organizational culture (Weber & Camerer, 2003). Some organizational cultures may require more active monitoring than others and some organizational cultures are likely to be more receptive to monitoring than others. In addition, the use of monitoring itself is likely to impact organizational culture by communicating a set of standards (Tenbrunsel, Wade-Benzoni, Messick, & Bazerman, 2000) and shifting the focus of managerial attention (Sims & Brinkmann, 2002).

Results from our work inform a number of important prescriptions. For example, our results suggest that announced inspections can significantly increase trustworthy behavior – for that period. Prescriptively, inspection programs should include anticipated inspections for key events. One option may be to couple anticipated monitoring with unannounced (unanticipated) inspections. Taken together, anticipated and unanticipated monitoring may yield benefits that either system alone cannot.

Our results also suggest an approach for building trust. Participants in our study discounted trustworthy behavior they observed when monitoring was anticipated. As a result, we expect unanticipated monitoring schemes to build greater trust within organizations than anticipated monitoring schemes. For example, managers should allow unanticipated monitoring of their operations to engender trust.

Despite the importance of trust in organizational life, surprisingly little is known about how managerial tools, such as monitoring, influence trust-like behavior. Our results identify a number of important relationships between monitoring and trust and reveal the complex nature of this relationship with respect to strategic behavior and observed behavior. Managers can certainly use monitoring to influence trust-like behavior, but they must be careful. Paradoxically, monitoring can both increase and decrease trusting and trustworthy behavior.

REFERENCES

Aiello, J. (1993). Computer-based monitoring: Electronic surveillance and its effects. *Journal of Applied Social Psychology*, *23*, 499–507.

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In: J. Kuhl & J. Beckman (Eds), *Action-control: From cognition to behavior* (pp. 11–39). Heidelberg: Springer.

Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology. In: L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 1–63). New York: Academic Press.

Amaldoss, W., & Jain, S. (2002). David vs. Goliath: An analysis of asymmetric mixed-strategy games and experimental evidence. *Management Science*, *48*, 972–991.

Atwater, L. (1988). The relative importance of situational and individual variables in predicting leader behavior. *Group and Organization Studies*, *13*, 290–310.

Bazerman, M. (1994). *Judgment in managerial decision making*. New York: Wiley.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*, 122–142.

Boles, T., Croson, R., & Murnighan, J. (2000). Deception and retribution in repeated ultimatum bargaining. *Organizational Behavior and Human Decision Processes*, *83*, 235–259.

Bromiley, P., & Cummings, L. L. (1995). Transaction costs in organizations with trust. In: R. J. Bies, B. Sheppard, & R. J. Lewicki (Eds), *Research on negotiations in organizations* (Vol. 5, pp. 219–247). Greenwich, CT: JAI.

Buchan, N., & Croson, R. (1999). Gender and culture: International experimental evidence from trust games. *American Economic Review, Papers and Proceedings*, *89*, 386–391.

Burnham, T., McCabe, K., & Smith, V. (2000). Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior & Organization*, *43*, 57–73.

Carr, A. (1968). Is business bluffing ethical? *Harvard Business Review*, *46*, 143–153.

Chang, O., & Schultz, J. (1990). The income tax withholding phenomenon: Evidence from TCMP data. *The Journal of the American Taxation Association*, *12*, (Fall) 88–93.

Cialdini, R. (1996). The triple tumor structure of organizational behavior. In: D. Messick & A. Tenbrunsel (Eds), *Codes of conduct* (pp. 44–58). New York: Russell Sage Foundation.

Degeorge, F., Patel, J., & Zeckhauser, R. (1999). Earnings management to exceed thresholds. *Journal of Business*, *72*, 1–33.

Gibson, K., Bottom, W., & Murnighan, K. (1999). Once bitten: Defection and reconciliation in a cooperative enterprise. *Business Ethics Quarterly*, *9*, 69–85.

Glaeser, E., Laibson, D., Scheinkman, J., & Soutter, C. (2000). *The Quarterly Journal of Economics*, *115*, 811–846.

Hirsch, F. (1978). *Social limits to growth*. Cambridge, MA: Harvard University Press.

Ho, T., & Weigelt, K. (2001). *Trust Building Among Strangers*. Wharton Marketing Department Working Paper 99-008.

Hochschild, A. (1983). *The managed heart: Commercialization of human feeling*. Berkeley: University of California Press.

Kruglanski, A. (1970). Attributing trustworthiness in supervisor–worker relations. *Journal of Experimental Social Psychology*, *6*, 214–232.

Lewicki, R. J., & Bunker, B. B. (1996). Developing and maintaining trust in work relationships. In: R. M. Kramer & T. R. Tyler (Eds), *Trust in organizations: Frontiers of theory and research*. Thousand Oaks, CA: Sage.

Lewicki, R. J., & Wiethoff, C. (2000). Trust, trust development, and trust repair. In: M. Deutsch & P. T. Coleman (Eds), *Handbook of conflict resolution: Theory and practice*. San Francisco, CA: Jossey-Bass.

Mookherjee, D., & Png, I. (1989). Optimal auditing, insurance, and redistribution. *The Quarterly Journal of Economics*, *104*, 399–415.

O'Connor, K., & Carnevale, P. (1997). A nasty but effective negotiation strategy: Misrepresentation of a common-value issue. *Personality and Social Psychology Bulletin*, *23*, 504–515.

Ring, P. S., & Van de Ven, A. (1992). Structuring cooperative relationships between organizations. *Strategic Management Journal*, *13*, 483–498.

Ross, W., & LaCroix, J. (1996). Multiple meanings of trust in negotiation research: A literature review and integrative model. *International Journal of Conflict Management*, *7*, 314–360.

Santoro, M., & Paine, L. (1993). Sears Auto Centers, Harvard Business School Case 9-394-010. Boston, MA: Harvard Business School Publishing.

Schweitzer, M., & Croson, R. (1999). Curtailing deception: The impact of direct questions on lies and omissions. *The International Journal of Conflict Management*, *10*, 225–248.

Schweitzer, M., Hershey, J., & Bradlow, E. (2004). Promises and lies: Restoring violated trust. Working paper, Wharton School, University of Pennsylvania.

Sims, R. R., & Brinkmann, J. (2002). Enron ethics (or, Culture matters more than codes). *Journal of Business Ethics*, *45*, 243–256.

Tenbrunsel, A., Wade-Benzoni, K., Messick, D., & Bazerman, M. (2000). Understanding the influence of environmental standards on judgments and choices. *Academy of Management Journal*, *43*, 854–866.

Tomlinson, E., Dineen, B., & Lewicki, R. (2004). *The road to reconciliation: The antecedents of reconciled trust following a broken promise*. Working paper, Fisher College of Business, Ohio State University.

Townsend, R. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, *22*, 265–293.

Weber, R., & Camerer, C. (2003). Cultural conflict and merger failure: An experimental approach. *Management Science*, *49*, 400–415.

Weber, R., Camerer, C., Rottenstreich, Y., & Knez, M. (2001). The illusion of leadership: Misattribution of cause in coordination games. *Organization Science*, *12*, 582–598.

Wimbush, J., & Shepard, J. (1994). Toward an understanding of ethical climate: Its relationship to ethical behavior and supervisory influence. *Journal of Business Ethics*, *13*, 637–647.

Wu, G., Heath, C., & Knez, M. (2003). A timidity error in evaluations: Evaluators judge others to be too risk averse. *Organizational Behavior and Human Decision Processes*, *90*, 50–62.

DO LIBERALS PLAY NICE? THE EFFECTS OF PARTY AND POLITICAL IDEOLOGY IN PUBLIC GOODS AND TRUST GAMES

Lisa R. Anderson, Jennifer M. Mellor and Jeffrey Milyo

"Jesus was a liberal"—Alan Colmes, political commentator for Fox News[1]

ABSTRACT

We test whether party affiliation or ideological leanings influence subjects' behavior in public goods experiments and trust games. In general, party is unrelated to behavior, and ideology is not related to contributions in the public goods experiment. However, there is some evidence that self-described liberals are both more trusting and more trustworthy.

INTRODUCTION

Democrats and liberals are generally understood to be more caring and kind than Republicans and conservatives; for example, even the conservative

Experimental and Behavioral Economics
Advances in Applied Microeconomics, Volume 13, 107–131

ISSN: 0278-0984/doi:10.1016/S0278-0984(05)13005-3

author and media personality Ben Wattenberg has acknowledged that "the word 'conservative' conjures up images of the miserly Ebenezer Scrooge, while 'liberal' brings to mind kindly Santa Claus." (PBS Think Tank, 1995). This perception of Democrats and liberals as more other-regarding, while not universal,[2] is pervasive enough that George W. Bush, while campaigning for the Republican nomination for president, adopted the moniker of a "compassionate conservative" to counter such stereotypes. But are left-leaning individuals really more generous and trusting?

We put conventional wisdom to the test by examining differences in the behavior of liberal versus conservative subjects in two classic experimental settings: the public goods game and the bilateral trust game. In the first set of experiments, we test whether Democrats or liberals are more likely to contribute to a group account when such actions are contrary to self-interest. In the bilateral trust game, we test whether Democrats and liberals choose to trust strangers or to behave in a trustworthy fashion, despite monetary incentives to the contrary. The question of whether political attributes influence behavior is interesting both in itself and because experimental subjects are often drawn from a pool of atypically liberal college students. To the extent that political attitudes are an important determinant of behavior, experimental researchers must take added caution in applying findings to contexts outside the lab. Similar concerns regarding the "indoctrinating effect" of economic instruction received by the pool of likely experimental subjects have generated an extensive literature,[3] but no previous published work systematically examines the effects of political party or ideology.[4] Finally, our study complements recent experimental work on the validity of survey measures of trust (Glaeser et al., 2000 and Anderson, Mellor, & Milyo, 2004a), in that we test whether the self-proclaimed generosity of Democrats and liberals is just cheap talk.

While an experimental test of the proposition that liberals "play nice" has many advantages over non-experimental techniques, one drawback is that the true association between ideology and generosity may be a response to perceived inequities in the social environment. Therefore, liberals may not behave more compassionately in the artificially egalitarian setting of the laboratory. In order to address this concern, we induce inequality among subjects by manipulating the show-up fee paid to all participants.[5] This experimental design allows us to test whether liberals play more nicely in non-egalitarian versus egalitarian settings.

In the next section, we describe evidence from national surveys on the relationship between political attitudes and support for government spending and trust. In turn, we describe the study design, survey results, and experimental results. In short, we find that despite conventional wisdom and survey evidence, there is no tendency for adherents of either major party to play nice, nor do self-described liberals have a greater tendency to make contributions in a public goods experiment. However, in keeping with conventional wisdom (but not necessarily national survey results), we find some evidence that self-described liberals behave in a more trusting and trustworthy manner.

POLITICAL PARTY AND IDEOLOGY: EVIDENCE FROM NATIONAL SURVEYS

The popular perception that Democrats and liberals are more kindhearted stems in part from the consistent finding in opinion surveys that left-leaning individuals tend to support increased public spending on social programs. Such associations in survey data are well known and are stock material in popular textbook accounts of American politics (e.g. Wilson & DiIulio, 2004). However, it is less obvious that left-leaning individuals are more trusting, even though trust is often considered part and parcel with generosity. In fact, recent research shows that minorities, lower-income individuals and women (i.e. the core constituency of the Democrat party) tend to exhibit less trusting attitudes (Alesina & La Ferrara, 2002). On the other hand, generalized trust is also well known to be lowest in the South, a bastion of ideological conservatism (Putnam, 2000). Consequently, survey evidence lends at best mixed support for the conventional view that Democrats and liberals are more trusting. As we will illustrate shortly, the patterns found in national survey data are also observed among our experimental subjects.

In order to provide a baseline for comparing the representativeness of the opinions of our experimental subjects, we examine the correlation of political party and ideology with support for public goods and a common attitudinal measure of trust. For this exercise, we employ data from two national opinion surveys, the 1972–2002 General Social Survey (GSS) and the 2000 National Election Survey (NES).

We first consider the evidence from the GSS; this cumulated data set has the advantage that we can separately analyze the opinions of college-aged individuals (18- to 22-year-olds). We measure opinions about spending on public programs by constructing a count of the number of times an individual states that there is currently "too little government spending" on any of eight major programmatic areas (arms, education, foreign aid, health, social security, transportation, police and prisons and welfare); in a similar fashion, we also measure support for the subset of social spending categories (education, health, social security and welfare). The results presented in Table 1 are quite consistent with textbook accounts: Democrats and liberals are more supportive of government spending, particularly on social programs. These proclivities are likewise present among 18- to 22-year-olds.

Table 1. Correlations in Survey Responses in General Social Survey, 1972–2002.

	Mean	Number of 8 Programs With "Too Little" Government Spending	Number of 4 Social Programs With "Too Little" Government Spending	Agree That "Most People Can Be Trusted"
Panel A: All respondents ($n = 24,198$)				
Democrat or Democrat leaning	0.49	0.08***	0.10***	−0.04***
Republican or Republican leaning	0.36	−0.09***	−0.11***	0.08***
Independent	0.13	0.01*	0.01*	−0.05***
Liberal (normalized to 0–10 point scale)	5.08	0.07***	0.11***	−0.01*
Panel B: 18- to 22-year-olds only ($n = 1,555$)				
Democrat or Democrat leaning	0.45	0.06**	0.09***	0.00
Republican or Republican leaning	0.36	−0.06**	−0.08***	0.04**
Independent	0.18	0.01	0.02	−0.05**
Liberal (normalized to 0–10 point scale)	5.60	0.03	0.08***	0.06**

Notes: ***Significant correlations for $p<0.01$, **for $p<0.05$, *for $p<0.10$.

Table 2. Correlations in Survey Responses in National Election Study, 2000.

($n = 686$)	Mean	Agree That "Government Should Increase Spending and Services"	Agree That "Most People Can Be Trusted"
Democrat or Democrat leaning	0.51	0.39***	−0.07**
Republican or Republican leaning	0.40	−0.43***	0.08**
Independent	0.09	0.03	0.01
Liberal (normalized range is 0 to 10 points)	6.28	0.21***	−0.05
Liberal relative to Bush (normalized range is −10 to +10)	2.69	0.16***	0.01

Notes: ***Significant correlations for $p<0.01$, **for $p<0.05$, *for $p<0.10$.

Next, we follow Putnam (2000) and others in measuring trust by agreement with the statement "in general, most people can be trusted." Contrary to common stereotypes, panel A of Table 1 shows that Republicans are more trusting of others, while Democrats are less so.[6] In addition, there is a weak negative correlation between liberal ideology and trust. Panel B reveals that young Republicans are likewise more trusting, but in contrast to the general population, 18- to 22-year-olds tend to be more liberal, and liberalism among this group is associated with *increased* trust in others.

We attempt to resolve this inconsistent relationship between ideology and trust by adopting a more meaningful measure of self-rated ideology. Using the 2000 NES, we construct an alternative ideological score that is simply the difference between a respondent's self-rating and their rating of George W. Bush. We prefer this measure since the meaning of the term "liberal" may be sensitive to context; for example, the "liberal" label may have more positive connotations among college-aged individuals than in the general population.

The results in Table 2 demonstrate that the 2000 NES respondents are somewhat more partisan and liberal than the 1972–2002 GSS respondents (this is understandable given the different time periods examined). In addition, Democrats and liberals (however defined) prefer increased government spending. Further, Republicans tend to be more trusting and Democrats less so. However, liberal ideology is only weakly and negatively associated with trust, while our alternative measure of liberal ideology is weakly and positively correlated with trust. We observe a broadly similar pattern for the subset of 18- to 22-year-olds in the NES, although the sample size is quite limited ($n = 30$) so nearly all of these correlations are insignificant (not shown).

Taken together, these national survey results confirm that Democrats and liberals are more likely to favor spending on public programs, while Republicans are more likely to profess trust in others. However, the association between ideology and trust is more ambiguous, particularly when considering the college-aged population or our alternative measure of relative liberalism.

STUDY DESIGN

Subjects participated in either a public goods experiment or a trust experiment. For each session, a group of eight subjects was recruited from undergraduate classes at the College of William and Mary. The games, which are described below, were repeated for 30 rounds with feedback about others' decisions provided at the end of each round. At the completion of the 30 rounds, one round was randomly chosen to determine earnings. Earnings average \$19.57 in the public goods game and \$22.21 in the trust game. Finally, at the end of each experimental session we administered a survey with 42 questions covering demographic characteristics, political attitudes and social capital measures.

We conducted six sessions of the public goods experiment designed by Marwell and Ames (1979). In particular, each person in a group of eight was given ten tokens to divide between a private account and a group account (i.e. the public good). The private account earned a return of \$1 to the individual. Each token contributed to the group account earned \$0.25 for all eight members of the group. This public goods design is linear, in the sense that the return to the group account is a linear function of the total number of tokens in that account. Note that it is individually optimal to put all tokens in the private account (since $\$1 > \$0.25$). Additionally, it is socially optimal for everyone to put all tokens in the public account (since $8^{*}\$0.25 = \$2 > \$1$), making this a prisoner's dilemma game.

We conducted 12 sessions of the trust (investment) game designed by Berg, Dickhaut and McCabe (1995). In the trust game one subject (the first mover) was given \$10 and offered the opportunity to pass some, all or none to a partner (the second mover). All passed money was tripled before being received by the second mover. Finally, the second mover had the opportunity to pass some, all or none of the money he or she received back to the

Table 3. Experimental Design.

Session	Experiment	Block 1 (10 rounds)	Block 2 (10 rounds)	Block 3 (10 rounds)	No. of Subjects
1–2	Public Goods	Egalitarian	Skewed	Symmetric	16
3–4	Public Goods	Skewed	Symmetric	Egalitarian	16
5–6	Public Goods	Symmetric	Egalitarian	Skewed	16
Total subjects in the public goods experiment					48
7–10	Trust	Egalitarian	Skewed	Symmetric	32
11–14	Trust	Skewed	Symmetric	Egalitarian	32
15–18	Trust	Symmetric	Egalitarian	Skewed	32
Total subjects in the trust experiment					96

Notes: Egalitarian show-up payments = (8 at $7.50).
Skewed show-up payments = (1 at $20, 4 at $7, 3 at $4).
Symmetric show-up payments = (3 at $10, 2 at $7.50, 3 at $5).

first mover. Using backward induction, it is straightforward to show that the Nash equilibrium for this game is that no money will be passed in the first stage, since second movers have no incentive to return money in the second stage.[7] Subjects were randomly assigned to be a first mover or a second mover in the game. Roles remained constant throughout the experimental session, but subjects were randomly re-paired at the beginning of each new round.

The experimental design is described in Table 3. Note that each session was divided into three blocks of ten rounds. Each block represented a different distribution of fixed show-up payments.[8] The purpose of this variation in fixed payments was to create a less egalitarian environment that would allow us to test whether the association between liberal ideology and support for public goods or trust is conditioned by the perceived fairness of the social environment.[9] We considered two inequality treatments, which are described as "skewed" and "symmetric" in Table 3. Note that the average fixed payment is $7.50 in all three treatments.

SURVEY RESULTS

We next examine the correlations of measures of political party and ideology with attitudinal measures regarding spending and trust among subjects who participated in our public goods and trust experiments. Our survey included four questions pertaining to political ideology. The first asked subjects to indicate the political party to which they belong, and a follow-up question asked subjects to choose the political party that best represents their interests, where available responses included Democrat, Republican, other, and none. Because one-third of the subjects did not report a response to the party membership question, we used the second question to define party interests. As shown in Table 4, 40.3% of respondents reported that the Democrat party best represents their interests, 37.5% reported Republican, with the remainder divided among the other party and no party categories. This formulation of party affinity is similar to questions about Democrat or Republican leanings in the GSS and NES. Compared with national opinion then, our experimental subjects were somewhat less partisan, but among party-leaners, slightly more Republican.[10]

To measure political ideology, the survey first asked subjects to rate their ideological leanings on a scale from 0 to 10, with 0 defined as extreme conservative, 5 as moderate and 10 as extreme liberal. Subjects averaged slightly "left" of moderate with a 5.46 on this scale. The survey next instructed subjects to rate President Bush in the same manner; subtracting this rating from the subject's own rating yields our second measure of ideology. On average, subjects in our experiments perceived themselves as more liberal than President Bush. Our subjects were also more liberal than the average of all GSS respondents, but very comparable in this dimension with NES respondents and the subset of college-aged GSS respondents.

Table 4 describes the correlation between political views and either support for government spending or generalized trust among our experimental subjects. One immediate difference between these results and those observed with the national survey data is that party and ideology are strongly correlated with views on government spending.[11] However, the relative support for spending by Democrats and liberals compared to Republicans and conservatives is more in keeping with the GSS, and to a lesser extent, the NES. Alignment with the Democrat party was positively and significantly correlated with views that overall government spending and on social programs in particular is too low.[12] Significant negative correlations exist between both measures of spending attitudes and subject affinity for the Republican party. Further, the sizes of the correlations between party and social program spending are almost identical to those reported in Table 2 using the NES respondents, a much more recent sample than the pooled GSS sample. The findings reported in Table 4 also indicate that being more liberal in either absolute terms or relative to Bush is positively associated with views that government spending is too low, as was the case in both the GSS and NES.

Table 4. Correlations in Survey Responses among Subjects in Public Goods and Trust Experiments.

	Mean	No. of 8 Government Programs in Which Spending is "Too Little"	No. of 4 Social Programs in Which Spending is "Too Little"	Agrees That "Most People Can Be Trusted"	Describes Oneself as "Trustworthy"
Democratic Party best represents interests	0.403	0.33***	0.38***	−0.07	0.02
Republican Party best represents interests	0.375	−0.36***	−0.44***	0.18**	−0.05
Other Party best represents interests	0.069	0.24***	0.26***	0.06	0.08
No Party best represents interests	0.153	−0.13	−0.11	−0.19**	−0.02
11-point ideology scale (0 = extreme conservative, 10 = extreme liberal)	5.46	0.45***	0.55***	−0.08	0.14*
Ideology difference (own rating less Bush rating)	2.51	0.45***	0.53***	−0.12	0.20**
	Mean:	3.09	2.37	0.30	0.92

Notes: Sample means and correlations are based on 144 subjects, except for three variables with a few missing values: spending preferences for government and social programs ($n = 140$ and 141, respectively) and ideological difference with President Bush ($n = 143$).
***Significant correlations between variables for $p < 0.01$, **for $p < 0.05$, *for $p < 0.10$.

Consistent with the national surveys, we also find that Republicans are significantly more trusting, and that liberal attitudes negatively correlate with perceptions that most people can be trusted, although the latter is not statistically significant.[13] Finally, Table 4 reports the correlations between political party and ideology and a question not examined in the GSS or NES – self-reported trustworthiness.[14] While 92% of our subjects view themselves as trustworthy, there are positive and significant correlations between self-reported trustworthiness and both measures of liberal political leanings. Interestingly, Republicans are unlikely to describe themselves as trustworthy, despite articulating a greater amount of trust in others.

In summary, our survey of the opinions of our experimental subjects exhibit similar patterns to those found in national surveys. Consistent with popular belief, Democrats and liberals support increased spending on public programs. However, in contrast to popular wisdom (but consistent with evidence from national surveys), Republicans exhibit more trust and political ideology, is only weakly and inconsistently associated with trust. However, our survey of the experimental subjects reveals a significant positive correlation between liberal ideology and self-reported trustworthiness. We now turn to the question of whether party and ideology affect behavior in public goods and trust games.

EXPERIMENTAL RESULTS

As a first step, we present descriptive statistics and conduct non-parametric tests on the average play of each individual in our study (see Table 5). From the public goods experiment, we report mean values of the number of tokens contributed by subjects to the group account. This action is thought to reflect the value subjects place on the welfare of other subjects, much like the survey questions on government spending on social programs. From the trust experiments, we report mean amounts sent by first movers, a measure of the level of trust that a player has in his or her randomly matched partner. Finally, we also examine the mean ratio of amount returned to amount available among the second movers, which can be interpreted as the trustworthiness of these subjects. For each subject, we calculate the average decision over 30 rounds, and we then average those values over the 48 subjects who participated in that feature of the experiment. The means for the full sample are reported in Table 5, along with means of the subjects' party and ideology.

Unlike the survey responses on public spending, trust, and trustworthiness, the mean decisions of our subjects do not show marked differences by political party and ideology. For example, in the public goods experiment, subjects with Democratic leanings contributed an average of 2.8 tokens to

Table 5. Mean Subject Decisions in Public Goods and Trust Experiments, By Political Party and Ideology.

	Public Goods Experiment	Trust Experiment	
	Mean group account contribution	Mean tokens sent to second mover	Mean ratio of tokens returned to first mover to tokens available
All subjects	2.75 (1.58) $n = 48$	4.97 (2.60) $n = 48$	0.35 (0.17) $n = 48$
Democratic Party best represents interests	2.78 (1.57) $n = 18$	4.52 (2.20) $n = 15$	0.35 (0.17) $n = 25$
Republican Party best represents interests	2.60 (1.55) $n = 21$	4.51 (2.71) $n = 18$	0.35 (0.19) $n = 15$
Other Party best represents interests	3.12 (0.45) $n = 2$	6.40 (2.77) $n = 4$	0.33 (0.10) $n = 4$
No Party best represents interests	3.01 (2.10) $n = 7$	5.82 (2.83) $n = 11$	0.38 (0.28) $n = 4$
Liberal (equal to 1 for values of 6 or higher on 11-point scale, 0 otherwise)	2.62 (1.51) $n = 19$	5.14 (2.76) $n = 20$	0.34 (0.17) $n = 32$
More liberal than Bush (equal to 1 if ideology difference exceeds sample mean, 0 otherwise)	2.74 (1.53) $n = 20$	5.49 (2.46) $n = 24$	0.33 (0.17) $n = 27$

Notes: Mean values of subject choices taken over 30 decision-making rounds in the experiment.

the group account, and those with Republican leanings contributed only slightly less, or 2.6 tokens. This difference between Democrats and Republicans does not prove to be significant according to a Mann–Whitney test carried out in a sample of only those two groups. Moreover, in a series of

Mann–Whitney tests conducted for each party and ideology subgroup, we found no significant differences in group account contribution for any one group of subjects relative to the remaining subjects. This same pattern of results is exhibited for both the amount sent decision and the return ratio decision.[15] There is little discernable difference in trusting behavior between Democrats and Republicans (i.e. amounts sent average 4.52 and 4.51, respectively). While the differences between liberals and non-liberals are somewhat larger at 5.14 versus 4.85 tokens, or 5.49 versus 4.45 tokens, these are not significant according to Mann–Whitney tests. Trusting behavior is also higher among subjects with other party or no party interests compared with all others, but again these differences are not statistically significant. In the analysis of mean return ratio, conjectured to reflect the subject's trustworthiness, there is a clear tendency among all groups to return roughly one-third of the tokens available. Recall that in the experiment, tokens sent to the second mover were tripled before the return decision was made; thus, the return ratio is strongly associated with the multiplier used in the experimental design. The minor fluctuations by subgroup are not significant according to Mann–Whitney tests.

In contrast to the analysis above, we now conduct multivariate regression analysis of the round-by-round decisions made by the experimental subjects. We analyze these subject decisions in each round of either game using a GLS regression model with random subject-specific effects; this allows us to control for the influence of subject race and gender in testing for the effects of political leanings. This approach also allows us to test the effects of the inequality treatment in our design. Because the effects of ideology may only be triggered by perceived inequities not visible in an egalitarian laboratory setting, we varied the show-up payments that our subjects received within each session. As described earlier, of the 30 decisions made by subjects in each game, ten were made in settings where all subjects received the same show-up payment of $7.50, another ten were made in settings where show-up payments were symmetrically distributed around a mean of $7.50, and ten decisions were made with show-up fees that were skewed (one subject received $20, but the mean fee remained $7.50). In both of the non-egalitarian settings, payments were randomly distributed among the subjects. In the regression analyses that follow, we discuss how these inequality treatments affected subject decisions directly, and moreover, we discuss whether inequality affected the associations between political leanings and our subjects' behaviors.

As a check on the appropriateness of the subject-specific random effects model, we test the hypothesis that the estimated coefficients are not systematically different from consistent estimates obtained via a subject-specific fixed effects model. The p-values for these Hausman tests are reported in the last row of all subsequent tables; in no case can we reject the null hypothesis. While this exercise reinforces the appropriateness of random versus fixed effects, it does not rule out alternative models. For example, we could address the potential dependence of decisions made by a given subject by adjusting standard errors for clustering at the level of the individual in OLS estimation. An even more conservative method would be to use subject means across rounds of each experiment as the dependent variable. We discuss the sensitivity of our findings to these two alternative specifications below.

Starting with our public goods experiments, we report results from the random effects GLS regressions of round-by-round contributions to the group account in Table 6 (for political party) and Table 7 (for ideology). Model 1 in each table reports results from a regression controlling for subject gender and race, as well as fixed payment amount and the round of play. The omitted category in the party analysis is Republican. These results are substantively similar to our descriptive statistics; that is, none of the political party or ideology measures has a significant association with mean contributions to the group account. In Model 2 of each table, we report results that take into account the unequal nature of the show-up fees; in this case, the indicator variable for an unequal distribution of fixed payments is negative and significant. Elsewhere, we analyze this effect of inequality on public good provision in greater detail (Anderson, Mellor & Milyo, 2004a,b). Once again, the measures of political party (in Table 6) are not significantly associated with group account contributions individually or jointly, nor are the measures of political ideology (in Table 7).

To allow the effects of political leaning to vary with the inequality treatment, we interact each political measure with the inequality indicator. These results are reported as Model 3 in our tables. We continue to find no effect of major political party attachment or liberal ideology on public goods contributions, nor do we find evidence that such political measures interact with the inequality treatment. However, the interaction between other party and inequality is negative and significant in Table 6. Next, Model 4 includes indicators for sessions;[16] this leads to only one difference from the previous model: now, other party is significantly associated with increased contributions ($p<0.10$). The effects of minor party affiliation are only reinforced in Model 5 when we add interaction effects between political party and fixed payment (jointly significant at $p<0.05$). However, a good deal of caution is in order, as only two individuals in the public goods experiment identify

Table 6. Effects of Political Party on Group Account Contributions.

	Model 1	Model 2	Model 3	Model 4	Model 5
Fixed payment	0.024	0.024	0.026	0.025	0.004
	(0.99)	(1.00)	(1.07)	(1.06)	(0.11)
Democratic Party	0.365	0.365	0.400	−0.043	0.023
	(0.65)	(0.65)	(0.67)	(0.07)	(0.03)
Other Party	0.510	0.510	1.853	2.161*	4.809**
	(0.42)	(0.42)	(1.43)	(1.66)	(2.17)
No Party	0.408	0.408	0.770	0.063	−1.167
	(0.57)	(0.57)	(1.01)	(0.08)	(1.21)
Unequal treatment		−0.386***	−0.203	−0.203	−0.204
		(3.10)	(1.05)	(1.05)	(1.05)
Unequal* Democratic Party			−0.053	−0.053	−0.059
			(0.18)	(0.18)	(0.21)
Unequal*Other party			−2.016***	−2.015***	−1.618**
			(3.12)	(3.12)	(2.36)
Unequal* No Party			−0.542	−0.543	−0.481
			(1.40)	(1.40)	(1.24)
Fixed payment* Democratic Party					−0.006
					(0.12)
Fixed payment* Other Party					−0.357
					(1.50)
Fixed payment* No Party					0.170**
					(2.44)
Includes Session Dummies	No	No	No	Yes	Yes
Hausman test p value	1.000	1.000	1.000	1.000	1.000

Notes: All models are based on a sample of 1440 observations. Coefficients from random effect GLS models are reported, with absolute values of t-statistics in parentheses. All models include controls for the round of play, and race and gender of the subject.
*for the 0.10 level, **for the 0.05 level, ***Statistical significance for the 0.01 level.

with minor parties. Therefore, these anomalous findings for other party should not distract from the overall finding that while major party and ideological preferences are strongly associated with *attitudes* about spending on public programs, there is no evidence that such political leanings explain subject *behavior* in our public goods sessions.

Table 7. Effects of Political Ideology on Group Account Contributions.

	Model 1	Model 2	Model 3	Model 4	Model 5
Panel A: Ideology scale ($n = 1,440$)					
Fixed payment	0.024 (0.99)	0.024 (1.00)	0.025 (1.03)	0.025 (1.07)	0.066 (0.83)
Ideology scale (0 = extreme conservative, 10 = extreme liberal)	−0.015 (0.12)	−0.015 (0.12)	0.018 (0.13)	−0.091 (0.67)	−0.035 (0.21)
Unequal treatment		−0.386*** (3.10)	−0.133 (0.36)	−0.132 (0.36)	−0.121 (0.33)
Unequal∗Ideology scale			−0.050 (0.74)	−0.050 (0.74)	−0.051 (0.76)
Fixed payment∗ Ideology scale					−0.007 (0.53)
Includes session Dummies	No	No	No	Yes	Yes
Hausman test p value	1.000	1.000	1.000	1.000	1.000
Panel B: Ideology difference ($n = 1,410$)					
Fixed payment	0.023 (0.94)	0.023 (0.95)	0.023 (0.95)	0.023 (0.95)	−0.007 (0.23)
Ideology difference (own rating less Bush rating)	−0.043 (0.40)	−0.043 (0.40)	−0.016 (0.15)	−0.096 (0.89)	−0.210 (1.61)
Unequal treatment		−0.400*** (3.16)	−0.318* (1.83)	−0.318* (1.83)	−0.300* (1.72)
Unequal∗ Ideology difference			−0.040 (0.69)	−0.040 (0.69)	−0.048 (0.83)
Fixed payment∗ Ideology difference					0.015 (1.57)
Includes session Dummies	No	No	No	Yes	Yes
Hausman test p value	1.000	1.000	1.000	1.000	1.000

Notes: Coefficients from random effect GLS models reported, with absolute values of t-statistics in parentheses. All models include controls for the round of play, and race and gender of the subject.
***Statistical significance for the 0.01 level, **for the 0.05 level, *for the 0.10 level.

Turning to our series of trust experiments, we next examine the results of random effects GLS regressions of mean amounts sent by the first mover, controlling for subject specific random effects. These results are reported in Table 8 (for political party) and Table 9 (for ideology). We examine the

Table 8. Effects of Political Party on Tokens Sent.

	Model 1	Model 2	Model 3	Model 4	Model 5
Fixed payment	0.024 (0.90)	0.028 (1.03)	0.031 (1.16)	0.028 (1.04)	0.037 (0.99)
Democratic Party	0.302 (0.32)	0.306 (0.33)	−0.041 (0.04)	0.215 (0.20)	0.836 (0.69)
Other Party	2.215 (1.50)	2.221 (1.51)	1.888 (1.24)	−0.155 (0.09)	3.064 (1.52)
No Party	1.589 (1.56)	1.589 (1.56)	1.226 (1.17)	0.892 (0.81)	0.387 (0.33)
Unequal treatment		−0.353** (2.45)	−0.686*** (2.90)	−0.683*** (2.89)	−0.696*** (2.94)
Unequal* Democratic Party			0.527 (1.48)	0.523 (1.47)	0.511 (1.44)
Unequal*Other Party			0.510 (0.92)	0.502 (0.90)	−0.083 (0.14)
Unequal*No Party			0.545 (1.42)	0.545 (1.42)	0.480 (1.24)
Fixed payment* Democratic Party					−0.086 (1.11)
Fixed payment* Other Party					−0.436*** (2.76)
Fixed payment* No Party					0.068 (1.04)
Includes session Dummies	No	No	No	Yes	Yes
Hausman test p value	1.000	1.000	1.000	1.000	1.000

Notes: All models are based on a sample of 1440 observations. Coefficients from random effect GLS models are reported, with absolute values of t-statistics in parentheses. All models include controls for the round of play, and race and gender of the subject.
***Statistical significance for the 0.01 level, **for the 0.05 level, *for the 0.10 level.

Table 9. Effects of Political Ideology on Tokens Sent.

	Model 1	Model 2	Model 3	Model 4	Model 5
Panel A: Ideology scale ($n = 1,440$)					
Fixed payment	0.026 (0.95)	0.029 (1.08)	0.037 (1.36)	0.033 (1.23)	−0.082 (1.40)
Ideology scale (0 = extreme conservative, 10 = extreme liberal)	0.284* (1.65)	0.285* (1.66)	0.206 (1.16)	0.081 (0.42)	−0.136 (0.62)
Unequal treatment		−0.353** (2.45)	−0.996*** (2.63)	−0.988*** (2.61)	−0.877** (2.30)
Unequal∗Ideology scale			0.122* (1.83)	0.120* (1.82)	0.107 (1.62)
Fixed payment∗ Ideology scale					0.029** (2.22)
Includes session Dummies	No	No	No	Yes	Yes
Hausman test p value	1.000	1.000	1.000	1.000	1.000
Panel B: Ideology difference ($n = 1,440$)					
Fixed payment	0.025 (0.93)	0.028 (1.06)	0.033 (1.24)	0.030 (1.11)	−0.003 (0.10)
Ideology difference (own rating less Bush rating)	0.209* (1.94)	0.209* (1.94)	0.144 (1.29)	0.000 (0.00)	−0.197 (1.38)
Unequal treatment		−0.353** (2.45)	−0.596*** (3.35)	−0.594*** (3.34)	−0.545*** (3.07)
Unequal∗Ideology difference			0.098** (2.33)	0.098** (2.32)	0.091** (2.17)
Fixed payment∗ Ideology difference					0.026*** (3.11)
Includes session Dummies	No	No	No	Yes	Yes
Hausman test p value	1.000	1.000	1.000	1.000	1.000

Notes: Coefficients from random effect GLS models reported, with absolute values of t-statistics in parentheses. All models include controls for the round of play, and race and gender of the subject.

*for the 0.10 level, **for the 0.05 level, ***Statistical significance for the 0.01 level.

same set of five model specifications as we did in analyzing the public goods data. Results reported in Table 8 suggest that political party has no significant effect on trusting behavior, although the point estimates for other party and no party are large and positive in Model 1 (again, the cell sizes for these cases are small, $n = 4$ and 11, respectively). Other than a persistent dampening effect of inequality on trusting behavior, the only statistically significant coefficient is for the interaction of fixed payment and other party (there is also a large and marginally significant effect of other party). In all, these results are quite consistent with those in the public goods experiment, in that there is no evidence that preferences for the two major parties are associated with greater generosity.

However, in some of the models reported in Table 9 we find that subjects who rate themselves as more liberal (or more liberal than President Bush) send significantly higher amounts to their randomly matched partner. This result is in contrast to the survey responses reported in Table 4, which reveal a negative correlation between liberal views and an attitudinal measure of trust. Looking across the estimates in Models 1–3, the association between liberalism and tokens sent in the trust game is driven by behavior in the inequality treatment; as conjectured, liberals behave differently when the artificial egalitarian structure of the experiment is perturbed through variations in fixed payments. This effect persists even when session indicator variables and interactions between ideology and the individual's fixed payment are included as controls (Models 4 and 5, respectively); however, for self-reported ideology the interaction between ideology and inequality is only marginally significant (see panel A of Table 9).

The substantive importance of the association between liberal ideology and tokens sent in the trust game can be assessed by comparing the effects of a one standard deviation change in either self-reported liberalism (2.20 units) or relative liberalism (3.46 units). Using the estimates from Model 4, a one standard deviation increase in liberal ideology produces an increase in mean tokens sent of about 5–7%, but only in the inequality treatment. The results of Model 5 again demonstrate that this effect is not an artifact of the interaction between ideology and fixed payment; that is, the ideology-inequality effect persists even after controlling for the statistically significant ideology-fixed payment interaction. Finally, the latter interaction is also substantively non-trivial. For example, the estimates in Table 9 imply that moving from the low fixed payment of \$4 to the high of \$20 will increase tokens sent by 20–29% for a subject who is one standard deviation more liberal than the average (by either measure of liberalism).

Our final series of regressions examines the second movers' decisions in the trust experiments. In Tables 10 and 11 we report the results of random effects GLS models of the number of tokens returned, a behavioral measure analogous to trustworthiness. The independent variables in Models 1–5 are

Table 10. Effects of Political Party on Tokens Returned.

	Model 1	Model 2	Model 3	Model 4	Model 5
Fixed payment	−0.015 (0.39)	−0.019 (0.48)	−0.015 (0.38)	−0.014 (0.36)	−0.060 (0.65)
Democratic Party	0.276 (0.31)	0.275 (0.30)	−0.110 (0.12)	0.295 (0.24)	−0.045 (0.03)
Other Party	−0.407 (0.27)	−0.406 (0.27)	−1.348 (0.83)	−1.987 (0.90)	−6.982** (2.53)
No Party	0.684 (0.42)	0.686 (0.42)	1.138 (0.67)	1.087 (0.51)	0.921 (0.42)
Unequal treatment		−0.324 (1.64)	−0.685* (1.91)	−0.687* (1.92)	−0.711** (1.97)
Unequal* Democratic Party			0.578 (1.28)	0.578 (1.28)	0.606 (1.33)
Unequal*Other Party			1.413* (1.79)	1.414* (1.79)	1.800** (2.23)
Unequal*No Party			−0.682 (0.88)	−0.679 (0.87)	−0.644 (0.82)
Fixed payment* Democratic Party					0.041 (0.39)
Fixed payment* Other Party					0.670*** (2.56)
Fixed payment* No Party					0.025 (0.18)
Includes session Dummies	No	No	No	Yes	Yes
Hausman test p value	1.000	1.000	1.000	1.0000	0.999

Notes: All models are based on a sample of 1440 observations. Coefficients from random effect GLS models are reported, with absolute values of t-statistics in parentheses. All models include controls for the amount available to return, the round of play, and race and gender of the subject.

*for the 0.10 level, **for the 0.05 level, ***Statistical significance for the 0.01 level.

Table 11. Effects of Political Ideology on Tokens Returned.

	Model 1	Model 2	Model 3	Model 4	Model 5
Panel A: Ideology scale ($n = 1{,}440$)					
Fixed payment	−0.015 (0.38)	−0.019 (0.47)	−0.020 (0.51)	−0.020 (0.51)	−0.026 (0.17)
Ideology scale (0 = extreme conservative, 10 = extreme liberal)	0.104 (0.51)	0.104 (0.51)	−0.028 (0.13)	−0.056 (0.21)	−0.064 (0.20)
Unequal treatment		−0.324 (1.64)	−1.507** (2.33)	−1.510** (2.34)	−1.512** (2.34)
Unequal* Ideology scale			0.198* (1.92)	0.198* (1.93)	0.198* (1.93)
Fixed payment* Ideology scale					0.001 (0.04)
Includes session Dummies	No	No	No	Yes	Yes
Hausman test p value	1.000	1.000	1.000	1.000	1.000
Panel B: Ideology difference ($n = 1{,}440$)					
Fixed payment	−0.015 (0.38)	−0.018 (0.47)	−0.021 (0.55)	−0.021 (0.55)	−0.070 (0.92)
Ideology difference (own rating less Bush rating)	−0.054 (0.38)	−0.054 (0.38)	−0.160 (1.06)	−0.208 (1.18)	−0.314 (1.37)
Unequal treatment		−0.324 (1.64)	−0.797*** (2.71)	−0.799*** (2.72)	−0.853*** (2.82)
Unequal* Ideology difference			0.158** (2.17)	0.159** (2.17)	0.173** (2.29)
Fixed payment* Ideology difference					0.014 (0.75)
Includes session Dummies	No	No	No	Yes	Yes
Hausman test p value	1.000	1.000	1.000	1.00	0.929

Notes: Coefficients from random effect GLS models reported, with absolute values of t-statistics in parentheses. All models include controls for the amount available to return, the round of play, and race and gender of the subject.
*for the 0.10 level, **for the 0.05 level, ***Statistical significance for the 0.01 level.

identical to those described above, except that we also control for the amount of tokens available for the second mover to return. Recall that our analysis of survey responses shown in Table 4 found that both measures of liberal ideology were positively and significantly correlated with self-reports of trustworthiness. These results are very similar to those already observed in the analysis of tokens sent. In short, we observe a persistent negative effect of inequality on tokens in both Tables 10 and 11; major party leanings are not associated with the return ratio, but minor party is associated with a lower return ratio; and liberal ideology is associated with a significant increase in tokens returned, but only in the inequality treatment. However, the effects of liberalism on tokens returned are substantively much larger; for example, the estimates for Model 5 imply that a one standard deviation increase in liberal ideology (by either measure) will yield an increase in the number of tokens returned of about 7–8% for liberal second movers in the inequality treatment.[17]

So far, we have described only the results obtained from estimating random effects GLS models; this method allows for unobserved heterogeneity across individuals, but assumes that any correlation in the individual-specific disturbance terms do not vary systematically by round. We also estimated two alternative specifications that incorporate different assumptions about the independence of individual-specific disturbance terms across rounds of the experiments. First we re-estimated our models by OLS and adjusted the errors for clustering at the level of the individual experimental subject. Second, we treated the individual means across all rounds as the dependent variable and estimate our models by OLS. Both of these methods adopt more conservative assumptions about the independence of the individual specific disturbance terms than the random effects GLS, but sacrifice efficiency. Not surprisingly, in both cases, we obtain similar point estimates as in the random effects model, but with larger standard errors; as a result, the finding that liberals are significantly more trusting and trustworthy in the inequality treatment is not robust to these alternative specifications. However, because the point estimates on liberal ideology are in some cases substantively large, we opt to report in detail the model that yields the most efficient estimates, which is the random effect GLS.

CONCLUSION

There exists a common perception that Democrats and liberals are inherently more other-regarding than Republicans and conservatives. National survey evidence on attitudes toward public spending and redistribution strongly supports (and likely perpetuates) this stereotype, although survey evidence on generalized trust is not always consistent with the popular wisdom. However, despite the plethora of experimental research conducted using public goods and trust games, no previous published study explicitly tests whether Democrats and liberals do indeed "play nice." We address this lacuna in the scientific literature and put conventional wisdom to the test.

In general, we find that Democrats behave no differently than Republicans in either a canonical public goods game or trust experiment. This stands in contrast to survey evidence in which Democrats describe themselves as more supportive of public goods and Republicans describe themselves as more trusting. As is the case with major party affinity, self-described liberals do not contribute more in the public goods experiment than conservatives. Further, this evidence is robust to alternative model specifications and estimation methods. Surprisingly, minor party affiliation is associated with greater contributions and trustworthiness in these experiments, but the small number of individuals with such leanings and the lack of robustness to the estimation method makes us wary of making too much of this finding. Nevertheless, future work could examine this result more closely by over-sampling subjects with minor party affiliations and distinguishing between specific minor party affiliations.

In contrast to the above, we do find some evidence from our trust experiment that liberals appear both to trust more (i.e. send more tokens) and to be more trustworthy (i.e. return more tokens). These differences are not observed in the usual egalitarian setting of equal fixed payments for participation; only when we induce inequality through differential fixed payments do we observe a significant effect of liberal ideology on behavior in trust games. Further, this effect is not solely due to increased generosity by liberals who happen to get the high fixed payment. While such "lucky liberals" do send more tokens, the effect of liberal ideology on trust and trustworthiness persists even after controlling for the interaction of fixed payment amount and liberal ideology. Even so, some caution is in order before concluding that liberals play nice.

When we adopt either of the more conservative estimation procedures, we find no significant differences in play by political party or political ideology in either experiment. Consequently, before concluding that liberals are more trusting or trustworthy, more experimental evidence is in order. If this finding holds up to further scrutiny, then we will have identified two possible conundrums: (1) why do liberals behave differently than Democrats? and, (2) why do liberals behave differently in public goods games versus trust games? The first of these is perhaps less puzzling than it would appear at first glance, since there are many liberal Republicans and conservative Democrats (e.g. Rudy Giuliani and Zell Miller, respectively). Therefore, the existence of some difference in the apparent importance of political party and political ideology in determining subject behavior in experiments is perhaps not so surprising. However, we leave the political ideology puzzle to be addressed by future research.

Our results also offer some lessons for experimental research. The disproportionate presence of Democrats and liberals in the most common pools of potential experimental subjects (i.e. college students) does not appear to affect the results of public goods or trust games in the typically egalitarian setting of the experimental lab. However, liberal-leaning subjects appear to behave differently once we induce inequality in our trust experiment. This finding, together with the persistent importance of induced inequality itself, suggests that more attention may be needed regarding the importance of the heterogeneity of subjects on experimental results. For example, while our undergraduate subjects were mostly white, middle-class and native citizens, it is not inconceivable that more diverse groups of subjects may influence individual behavior in some experimental settings. Finally, the egalitarian conditions of the experimental lab itself may be a confounding factor that obscures some behavioral regularities.

NOTES

1. From a chapter heading in Colmes (2003).
2. Some would argue that liberals are indeed generous, albeit with others' money. This opposing view is most succinctly articulated by the conservative author and provocateur Ann Coulter: "…there is only one thing wrong with liberals: They're no good" (Coulter, 2002).
3. Findings on the "indoctrination effect" are quite mixed; for a recent review and novel evidence contra the existence of such an effect, see Frey and Meier (2003).
4. There are only two related studies of which we are aware: Mestelman and Feeny (1988) report some suggestive evidence that political ideology influences free-riding in a public goods game, while Fehr et al. (2003) show that major party affiliation among experimental subjects in Germany is associated with trust and trustworthiness in a one-shot trust game. However, the Fehr study does not estimate the effects of left-leaning versus right-leaning party affiliation or ideology.
5. Elsewhere, we have demonstrated that such manipulations influence contribution levels (Anderson, Mellor, & Milyo, 2004a,b).
6. Of course, party alignment has changed over the last 30 years or so, especially in the South, where generalized trust tends to be lower. As a check on this finding, we also examined the GSS data pooled only over the last 15 or 5 years. While we observe

similar patterns in the data, there are fewer significant differences due to the smaller sample size, particularly for the sub-group of 18- to 22-year-olds.

7. This analysis applies to a one-shot game, but can also be extended to a repeated game with a known endpoint.

8. It is a standard practice to pay subjects a fixed fee for showing up for an experiment. This payment supplements what subjects earn based on their decisions and serves as a lower bound on their compensation for participating in the experiment.

9. The behavioral effect of heterogeneity in the fixed payments is discussed in Anderson, Mellor and Milyo (2004a,b).

10. Our survey did not distinguish among minor parties; however, a contemporaneous survey of 190 undergraduates at the College of William and Mary found party identification to break down as follows: 50.6% Democrat, 32.7% Republican, 10.9% Independent, 2.9% Green, 2.3% Libertarian and 0.6% Other (www.wm.edu/government/content/spring%202003%20student%20survey.html).

11. This suggests that our sample of college students are more consistent than the general public when it comes to matching their political views with preferences about government spending; we speculate that this may be a function of youthful idealism and the increasing polarization of the American electorate.

12. The survey question for government spending was phrased: "Use the following scale to indicate your opinion about government spending on each program (1 = Too little, 2 = About right, 3 = Too much): National defense, foreign aid, welfare, education, transportation, Social Security, Medicare, police and prisons." When looking at attitudes about social programs we used only the responses regarding welfare, education, Social Security, and Medicare.

13. The survey reads "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" with available responses of "most people can be trusted", "it depends" and "you can't be too careful in dealing with people." We formed a dichotomous variable equal to 1 for responses of "most people can be trusted" and 0 otherwise.

14. This question is worded "Do you agree or disagree with the statement: 'I am trustworthy'?". For responses of "strongly agree" and "agree, but not strongly," we coded a trustworthiness indicator variable to 1; for responses of "uncertain, or it depends," "disagree, but not strongly," and "strongly disagree," we coded the indicator variable as 0.

15. We define the "return ratio" as the amount of tokens returned to the first mover relative to the amount available to be returned by the second mover.

16. The session indicators approach joint significance, with $p = 0.12$.

17. In the sample used to estimate the amount returned models, self-reported liberalism has a standard deviation of 1.98 and relative liberalism has a standard deviation of 2.76.

REFERENCES

Alesina, A., & La Ferrara, E. (2002). Who trusts others? *Journal of Public Economics*, *85*(2), 207–234.

Anderson, L. R., Mellor, J. M., & Milyo, J. (2004a). Social Capital and Contributions in a Public Goods Experiment. *American Economic Review Papers and Proceedings*, *94*(2), 373–376.

Anderson, L. R., Mellor, J. M., & Milyo, J., (2004b). Inequality, group cohesion and public good provision. Working Paper, Department of Economics, The College of William and Mary.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*, 122–142.

Colmes, A. (2003). *Red, white and blue*. New York: Regan Books.

Coulter, Ann. (2002). Battered Republican Syndrome, AnnCoulter.com (August 28, 2002); viewed at www.anncoulter.org/columns/2002/082802.htm on May 25, 2004.

Fehr, E., Fischbacher, U., von Rosenbladt, B., Schupp, J., & Wagner, G., (2003). *A nation-wide laboratory: examining trust and trustworthiness by integrating behavioral experiments into representative surveys. IZA Discussion Paper No. 715* (February), Bonn, Germany.

Frey, B. S., & Meier, S. (2003). Are political economists selfish and indoctrinated? Evidence from a natural experiment. *Economic Inquiry*, *41*(30), 448–462.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring Trust. *Quarterly Journal of Economics*, *115*(3), 811–846.

Marwell, G., & Ames, R. (1979). Experiments on the provision of public goods I: Resources, interest, group size and the free-rider problem. *American Journal of Sociology*, *84*(6), 1335–1360.

Mestelman, S., & Feeney, D. (1988). Does ideology matter? Anecdotal experimental evidence on the voluntary provision of public goods. *Public Choice*, *57*(2), 287–294.

PBS Think Tank, Can conservatives be compassionate? Transcript from December 21, 1995; viewed at www.pbs.org/thinktank/transcript239.htm on May 25, 2004.

Putnam, R. (2000). *Bowling Alone*. New York, NY: Simon and Schuster.

Wilson, J. Q., & DiIulio, J. J., Jr. (2004). *American Government* (9th ed.). Boston: Houghton Mifflin.

This page intentionally left blank

AN ECONOMICS WIND TUNNEL: THE SCIENCE OF BUSINESS ENGINEERING

Kay-Yut Chen

ABSTRACT

Companies are starting to capitalize on the potential of experimental economics as a decision-making tool. Hewlett-Packard (HP) is one of such pioneering companies. Experiments, conducted at HP Labs, were used to test retailer contract policies in three areas: return, minimum advertised-price (MAP), and market development funds. The experimental design models the multifaceted contemporary market of consumer computer products. While the model is quite complex, participants were found to be effective decisions-makers and that their behavior is sensitive to variations in policies. Based on the experimental results, HP changed its policies; for example, it made the consequences for minimum advertisement price violations forward-looking as well as backward-looking. This line of research appears promising for complex industrial environments. In addition, methodological issues are discussed in the context of differences between business and academic economics experiments. Finally, the author speculates about potential future business applications.

Experimental and Behavioral Economics
Advances in Applied Microeconomics, Volume 13, 133–167

ISSN: 0278-0984/doi:10.1016/S0278-0984(05)13006-5

> Companies are always trying to predict the future. These days, the field of experimental economics – which replicates market and business scenarios in the lab – is giving the crystal ball an upgrade.
>
> (''A New 'Wind Tunnel' for Companies'', Newsweek, Oct. 6, 2003.)

1. INTRODUCTION

It is inevitable that the business world will take notice of experimental economics. Laboratory experiments have become an increasingly important tool in economics to study behavior, test policies and to provide information to design better mechanisms. On the one hand, pioneers such as Vernon Smith and Charles Plott have provided the link between economics theory and actual economics behavior in areas such as markets and auction processes. On the other, experimental methods have been used to study policy in areas such as emission trading, natural-gas pipelines, transportation, and allocation of precious NASA deep space resources. Relevant papers include McCabe, Rassenti, and Smith (1990, 1991), Rassenti, Smith, and McCabe (1994), Cason (1995), Cason and Plott (1996), Plott (1997, 1999), and Brewer and Plott (2002). HP has long recognized the potential of this methodology as a business decision-making tool. HP Labs, the research arm of HP, has started an experimental economics program in 1994. Its strategy is to develop experimental models that closely mirror specific businesses. Human subjects then participate in experiments based on these models, and the results are used to isolate and evaluate the effects of policies. Although this program is located inside HP Labs, its impact has been felt in many of its business units from consumer businesses to procurement.

It may come as a surprise to some readers that the easiest part of this endeavor is to convince decision-makers to use experiments to answer policy questions. Since companies are used to test marketing, surveys and policy analysis, it is not too much of a leap of faith to consider the use of laboratory experiments. Furthermore, in our experience interacting with business partners, it is more often that they will end up with more ideas of how to use the laboratory than what can be done practically within a reasonable time frame.

The challenge lies in the need to distill the questions down to a scope that economists can design experiments to answer and to capture the important elements of the business environment into an economics model, while keeping complexity to a limited degree so that experimentation is still possible. Take the HP consumer business as an example. Most of the $18 billion business is sold through retailers. There are many retailers, although the major seven or eight cover 80% of the market. There is substantial heterogeneity in terms of size, business objectives, and supply chain characteristics among the retailers. Each of them manages thousands, if not tens of thousands, of products. They do business with HP and HP's competitors. Their strategic space is complex. Their decisions include: pricing, advertising, ordering, returns, and many others. On top of this, already very complicated environment is a set of equally complicated policies governing the relationship between any pair of retailer and manufacturer. Apart from the crucial wholesale pricing policy, which can take forms of discount and special rebates, there are minimum advertising price policies that limit how retailers can advertise products, return policies that govern what they can do with unsold products, soft-funds policies that provide marketing development money and many others. Furthermore, there are nontrivial interactions among these policies. Thus, it may not be possible to isolate and study them separately.

Ideally, a firm would like to have a test market to determine the effects of changing its policies toward its retailers, since blindly adopting new policies in billion-dollar markets may seem to be less than optimal. However, test markets are, at best, expensive. Furthermore, test markets are usually small and would not be able to show market-wide changes caused by a policy change. Laboratory experiments, on the other hand, do not have these limitations at the expense of a less realistic environment.

Over the span of a little more than 2 years, laboratory experiments were used to study policies in several areas: return, minimum advertised-price (MAP) (Charness & Chen, 2002), and market development funds (MDF). This sequence of projects was a collaborative work with HP Consumer Business Organization, which was in charge of the $12 billion[1] consumer business in 2001. The first study of return policy was a pilot project to show HP business that experimental methods can be an effective mean to gather information about retailer behavior. The goal of the experiments was to study how structural and parametric variations in the policy would affect HP's performance in the marketplace. In the return policy study, a policy with stocking fees was evaluated, with respect to multiple business measurements such as HP's market share, retailers stocking levels, and etc., as an alternative to an unlimited return policy.

Minimum advertised price is a common industry practice of setting a lower bound on the price a retailer can advertise for a particular product. The accepted business reasoning behind this practice argues that a minimum advertised price reduces competition and a retailer can make more profit on the product. Since there are multiple manufacturers for retailers to choose from, retailers will incline more to promote products from manufacturers who offer MAP policies. Previous work has shown that an MAP and an advertising subsidy together are sufficient to enable profit maximization (Kali, 1998). However, issue of multiple manufacturers has not been examined.

HP was interested in finding the optimal policy to enforce MAP since the alternative, enforcement through the court system, was very expensive. Complex characteristics of the environment such as product life-cycles and heterogeneity of retailers made pure theoretical analysis impractical. In addition, equilibrium analysis is not the best predictor of human behavior. Thus, it is not wise to base policy decisions solely on theoretical analysis, even if a realistic model can be created. Laboratory experiment seems to be the only alternative to field test for obtaining relevant information. In the study, we uncovered an exploit where subjects employed strategies of cleverly timing MAP violations around the end of product life-cycles. Based on this information, a new MAP policy was designed, tested, and implemented in HP's consumer business. This new policy employed a new timing mechanism of the penalties to prevent retailers to carry out the exploitation strategy. It was difficult to predict which feature(s) in the model would lead to this result before the experiment was conducted. Any model that does not include all these features might not be able to identify this exploit.

Another common industry practice is MDF, sometimes also known as soft-funds. MDF is given to retailers for the purpose of promoting a manufacturer's product. It can be used for advertising, promotional campaign, printing marketing literature, and anything related to marketing. Leveraging on the economics model created for the MAP study, experimental methods were used to test two new ideas being considered for a new HP marketing program. At the time of the study, each retailer receives a fixed percentage of wholesale payment to HP as MDF. HP was considering a new quota system in which each retailer would receive MDF equal to a higher percentage of its whole payment to HP if its sales revenue met or exceeded its quota. A competing idea was to use a ranking system. Retailers would be ranked based on a pre-specified scoring rule and their MDF percentages would be based on their ranking.

Experiments found that neither of the new system offered any significant benefits in the relevant business measurements, such as revenue or market share of HP products. As a result, HP Consumer Business Organization decided to cancel their plans. The major lesson here is that a negative result can also be extremely useful in a business. In this case, a substantial amount of money was saved because a new marketing plan no longer was to be implemented. Notice that the experiments did not, and probably could not, prove conclusively that the new quota system is always inferior to the existing fixed percentage MDF rule. They merely showed that, in properly calibrated settings, the new system failed to generate significant benefits with a high probability. This, much weaker, result was all HP consumer business needed to decide against implementing the new program because the potential benefits did not outweigh the costs.

The primary constraint on developing a business application is time. Business decisions need to be made in a timely fashion. A perfect analysis will still be worthless if it takes too long to complete. In the concluding remarks, the issues of developing the right tools to streamline the experimental process were discussed in the context of accommodating the timing requirements of business projects. There is still no getting around designing good experiments. However, investment has been made to make sure that the execution of experiments will be as painless and as speedy as humanly possible.

This chapter tries to lay out the important issues regarding using experiments as a business decision-making tool. Three applications, all drawn from work done at HP, are used as examples. Sections 2, 3, and 4 provide detailed case studies of these three business applications in the area of return policy, MAP policy, and MDF policy, respectively. Section 5 concludes, hopefully provides some insights about business economics experiments and discusses about the future of using experimental economics as a business tool.

2. CAN EXPERIMENTS BE REALISTIC TO BUSINESSES? RETURN POLICY EXPERIMENTATION

2.1. Background

HP conducts much of its consumer business through retail channels. The distribution channels for its products include national retailers, regional retailers, mass merchant firms, clubs, and Internet retailers. Each type of retailer may have its own success metrics or business goals, which may or may not be consistent with those of HP. For example, at the time of our experimental sessions, during the big Internet bubble of the late 1990s, many observers felt that Internet retailers were not concerned with profitability, as these retailers often sold to consumers at or below cost in an attempt to increase their market share. HP is concerned with the financial viability and market share of its retailers, if only because they affect its own market share and profitability. Contract policies were the primary tools HP uses to manage its relationship with its retailers. Some examples are return policies that govern the terms of which unsold products can be returned to HP, price-protection policies that provide credit to the retailers that offset manufacturers' price fluctuations, and benefits or penalties contingent on retailer compliance with MAP policies. To design effective policies consistent with its business objectives, HP must understand the implications of these policies on retailer behavior.

2.2. The Experimental Model

Standard methodology of experimental economics is adhered to in all experiments. All the experiments were conducted in the HP Experimental Economics Laboratory. Subjects were recruited from the Stanford student body. Participants were given accurate information about the game, and were told how their actual monetary rewards depended on their aggregate performance over the course of the experiments. Anonymity was preserved with respect to roles and payment, and no deception was used. Instructions were posted on the web several days before the actual experiments. Due to the complex nature of some experiments, subjects were required to pass a web-based quiz before they were allowed to participate. This has the dual benefits of weeding out potential subjects who cannot understand the mechanics of the experiment and reducing training time during an experimental session. Sessions lasted for about 3 hours. Each person was seated at a computer in a carrel separated from others by dividers, so that participants could not observe others' information.

In our laboratory market, we attempted to model the natural setting of HP retailers. Models for different policy studies are similar with variations in the number of retailers, number and characteristics of the products, and the set of policies in effect during the experiments. Therefore, the description of this design is not repeated in the discussion of the other two studies.

Each participant played the role of a retailer, while demand was computer-simulated. Retailers were heterogeneous and they interacted repeatedly in competition for consumer demand for products differentiated by price and manufacturer. Retailers made decisions about stocking, advertising, and pricing. Each simulated consumer considered the best option available when deciding whether to buy a product but was only aware of the products and prices to which it was exposed. A retailer's demand could also be sensitive to its reputation for pricing, relative to other retailers. Manufacturers were not represented by subjects in this formulation. The characteristics of each manufacturer were kept constant throughout an experiment and each manufacturer employed the same set of policies throughout an experiment. Alternative policies were studied by comparing experiments. This can be justified because companies typically take a few months, at the least, to respond to policy changes.

Seven differentiated retailers were modeled. They were intended to represent realistic classifications such as national firms, PC Direct/Mail Order companies, mass merchants, clubs, and Internet retailers. Each retailer chose a price for each product in each period and competed for some percentage of the potential market for the products. Most firms could increase this percentage by advertising, although each type of retailer had a maximum exposure percentage and advertising yielded diminishing marginal returns. Most retailers also had to make inventory decisions, with the cost of holding excess inventory balanced against a negative reputation if a retailer failed to meet most of the demand for a product. The timing of deliveries to the retailers was stochastic. The MAP related experiments were conducted with all seven types of retailers. The experimental model of the MDF study uses only five types.

We computer-simulated consumer demand using a random utility multilevel logit model (Dubin, 1998; McFadden, 1976) adapted to the HP environment by Steven Gjerstad and Jason Shachat. This model treats each product as a collection of attributes (such as price, brand, retailer, print speed, and memory). When assessing a potential product choice, each consumer assigns a weight to the value of each attribute, and the model adds these values together to determine that consumer's score for the product. The probability that the consumer purchases a product increases with this score, and the probability that any one product is selected is the estimated market share of that product. The stochastic market size lies within a range known to the retailers, who also receive a signal that further limits this range at the beginning of a period.

Retailers can sell products offered by HP and competing manufacturers. These products vary by retailer costs and manufacturer policies on product returns and advertising. We evaluate different retailers using diverse measures that reflect the contemporary business goals of the different categories of retailer. These measures include various combinations of gross profit, net income, revenue, and Gross Margin Return On Investment, which is a specific way of measuring return on investment. The model incorporates product obsolescence through a life-cycle assumption – some products get phased out and others take their place, with retailers receiving notice five periods in advance.

Inventory control is a crucial aspect of the natural retailer environment. Most retailers, although not all, need to stock products to be able to sell them. However, it is expensive to carry excess inventory. In addition, while a retailer may place an order for products, the actual shipment date is uncertain. Further, supplies may be short at any particular time. Retailers must consider all of these factors when making stocking decisions; a retailer who cannot meet existing demand develops a negative reputation for service, which negatively impacts subsequent demand.

Finally, advertising clearly affects demand and must be considered, particularly because advertising policy is the control variable in one of the studies. A retailer has some minimum level of market exposure even without any advertising. However, advertising increases market exposure in a nonlinear fashion, until it saturates the market for the retailer. While a firm may be free to advertise any price it likes, violating manufacturer mandates concerning MAP jeopardizes the advertising funds potentially available from the manufacturer. Manufacturers employ several schemes to punish violations. This aspect of the model also interacts with the MDF policies. In some experiments, subjects could only use MDF to advertise their products, and this would be forfeited if not used.

2.3. Return Policy Issues and Experimental Treatments

Return policies applied to retailers are different from the well-known consumer version, which usually allow a consumer to return a product within a pre-determined period of time. Return policies for retailers, sometimes known as stock rotation policies, specify the amount of unsold products, usually as a percentage of the amount shipped to the retailer, that can be returned to the manufacturer and the amount, if any, of restocking fee that will be charged. The terms can be as simple as a fixed limit of 15% of the amount shipped to the retailer, or they can be more complicated such as the following example taken from an actual contract between a retailer and a major computing equipment manufacturer: the amount of return is unlimited. However, a 6% restocking fee will be charged for any amount beyond a 4% return rate, calculated for a certain period of time. Usually, manufacturers will also allow retailers to return defective products without any restrictions, subjected to some monitoring.

The most obvious function of a return policy is to allow the sharing of demand risks between a manufacturer and its retailer(s). In the case of more than one retailer, it may also be efficient for the manufacturer to pool the demand risks of all the retailers. However, if we assume some information is asymmetric (e.g. manufacturer does not know the retailer's existing stock level or its willing to pay for the next unit), return policy can also be used in a strategic manner (Bali, Callandar, Chen, & Ledyard, 2001).

At the time of the project, one of the product divisions of HP was already in the process of changing its return policy. It was decided to use this opportunity to compare laboratory experiments to actual policy changes in the real world. The goal was modest when compared to a typical scientific test because only crude aggregate data was available for comparison.

One issue complicating the modeling is that HP does not enforce its return policies. This led us to believe that competitors also do not consistently enforce their policies. Thus, the model had to allow for inconsistent enforcement of return policies. Inconsistent enforcement of return policies was implemented in the following manner: A policy would be enforced with a fixed probability in each period. Subjects did not know whether a policy would be enforced when they determined the amount of units to return. A subject would only know how likely the policy will be enforced. If a policy was not enforced, no limit would be in effect and no restocking fees would be collected. It was as if the policy allowed unlimited returns with full refund in that particular period. If the policy was enforced, restocking fees, if any, would be collected and return limit would be in effect. If the subject specified an amount over the limit, he would only be able to return up to that limit.

Two treatments were used. Since it was assumed that HP did not enforce its existing return policy, its status quo policy was modeled as an unlimited return policy. The proposed return policy has a restocking fee of 21% but with no return limit. A 1% (of net-shipment) bonus credit will be given to any reseller as an incentive to reduce returns.

2.4. Experimental Results

Three conclusions about the overall policy can be drawn from the experimental data. Table 1 summarizes[2] the resulting statistics:

First, the proposed policy, because of the 21% restocking fee, induced a lower return rate. This should not come as a surprise to anyone. The average

Table 1. The Alternative Return Policy Resulted in Significantly Lower Return Rates, while Maintaining Similar Service Levels and Market Share at an Expense of Slightly High but Non-significant Inventory Costs.

	Base	Alternate Return Policy
Reseller average return rates to HP (%)	15.2 (8.8)	11.8 (5.7)
Reseller average stocking levels (in weeks of supply)	5.57 (3.41)	6.29 (3.78)
Reseller average service levels (%)	93.2 (3.21)	91.4 (8.31)
Average HP market share (units) (%)	46.4 (8.0)	46.4 (8.7)
Reseller average inventory cost[a] (% of gross revenue)	0.56 (0.3)	0.60 (0.3)

Note: Standard errors are in parentheses.
[a]Inventory costs were modeled as a fixed percentage (0.5%) of end of period inventory levels. It includes all related costs such as depreciation, handling, and storage costs. Each period represents a week in real life.

return rate was reduced by around 28% when the proposed return policy was applied. The more encouraging aspect of the results was that the size of the reduction was consistent with the reduction observed in the real markets when the new policy was applied. Unfortunately, no data were provided to test the statistical difference between experimental observations and real world data. The conclusion was drawn based on anecdotal evidence provided by business experts internal to HP.

Second, instead of responding to the proposed return policy with lower inventory levels and lower service levels, the subjects chose to maintain similar service levels at the expense of higher inventory costs. The average of the service levels is 93.2% in the base model and 91.4% in the alternate return policy model. The differences are not significant under any reasonable statistical test. Notice that the average stocking levels and inventory costs are higher under the alternate return policy, although not by much. Also worth mentioning is that actual reseller stocking levels at the time of the experiments were about 6 weeks of supply. The observed stocking levels in the experiments were close enough to the real world levels to add weight to the belief that the experiments were calibrated adequately to support business decisions.

Third, in the experimental environment, which simulates two HP products and two competitor products, the market shares of HP products did not change significantly when the alternate return policy was introduced. The average market shares (based on units) for the base model is 46.4%. The corresponding number in the alternate return policy model is also 46.4%. There is no change in HP's market share when the new policy was introduced in the experiments. It is our belief that the existence of demand for HP products and competition in the marketplace override the incentive to switch to other products when resellers are faced with a tougher HP return policy. Once again, this was consistent with anecdotal observations with real-world roll-out of the alternate policy.

This project is the first significant HP contract policy experimental study with business sponsorship. HP Consumer Business Organization agreed that the results were consistent with what they observe (unfortunately there was no formal statistical comparison between experimental data and real world business results) and decided to sponsor continuing efforts to use experiments as a tool to evaluate future policies.

3. IDENTIFY EXPLOITATION: MINIMUM ADVERTISED PRICE POLICY EXPERIMENTATION

3.1. Background

In a series of experiments, the behavior of retailers was studied with respect to the common industry practice of setting a MAP, a lower bound on the price a retailer can advertise for a particular product.

The focus of this research is not to determine whether MAP is a worthwhile policy, but to decide the best way of implementing a MAP. Since tens of thousands of products are involved, MAPs are usually not enforced by the courts because of expense involved. Instead, manufacturers will impose penalties on retailers that violate MAP. Punishment can range from refusing to ship a product to the retailer to eliminating or reducing the amount of MDF provided. It is not clear which form of MAP (if any) is best and which enforcement policies are effective. In addition, HP wants to know what effect eliminating or modifying MAP policies would have on its market share and its retailers' profitability. An effective policy should also take into account such factors as the short life-cycle of products in this market. Because it is not feasible to isolate a test market of retailers, the laboratory is an attractive alternative for investigating the impact of various policies. We conducted laboratory experiments to investigate the effects of various MAP policies on retailers' behavior and profitability, and on HP's market share.

Table 2. In the MAP Experiments, Participants Played the Roles of Seven Very Different Types of Retailers.

Retailer #	Must Stock?	Can Advertise?	Minimum % Market Exposure	Maximum % Market Exposure	Business Objective (%)
1	Yes	Yes	30	100	70 GMROII, 30 Net Income
2	Yes	Yes	30	100	Gross profit
3	Yes	Yes	30	70	70 GMROII, 30 Net Income
4	Yes	Yes	30	70	70 GMROII, 30 Net Income
5	Yes	Yes	30	50	100 GMROII
6	Yes	No	40	40	70 GMROII, 30 Net Income
7	No	Yes	10	30	Revenue

Note: They differed in many aspects from their reach in the market (min/max market exposure %), whether they stocked and held their own inventories (some retailers fulfilled orders through a third party and hold no inventory), and their business objectives which can range from profit maximization to optimizing return on investment (GMROII) or a linear combination of different objectives.

Product life-cycle was found to be an important factor in this study. Each participant managed 8 products in the experiment. Some of the products would become unavailable in the middle of the experiments simulating ends of product life-cycles. In addition, not all of the products were available from the beginning of the experiment. Some of the products would become available to replace the ones that ended their life-cycle.

Reseller heterogeneity also played an important role. Seven different types of retailers were modeled. Table 2 provides a summary.

The retailers were very different. For example, a club retailer (Number 6) does not advertise, while is retailer (no advertising) and an Internet retailer (Number 7) has a small potential market share, keeps no stock, and has only one performance metric – revenue. Retailers may also have different payoff functions. For example, retailer 5 is measured on return on investment (GMROII), while retailer 2 is measured on profit. Participants were paid according to these measures. However, different rates were used to convert participants' experimental earnings into the actual dollars, reflecting the heterogeneity of types of retailers. Products were also differentiated with respect to cost and levels of demand.

3.2. Policy Alternatives and Experiment Treatments

A total of seven treatments were used in two sequences of experiment. The sequences are identified by the months (September and February) the experiments were conducted in. At the time of the study, if a retailer violated MAP, HP would refuse to sell this particular product to the reseller again (referred to as the "pulled" penalty below). Obviously, this might not be sub-game perfect. After much discussion, HP Labs and its partners in the consumer business decided to explore penalties that were financial in nature. The focus of the September experiments was to study several financial penalties as alternatives to the "pulled" penalty.

The September experiments identified an exploit in the structure of all of the alternatives as well as the "pulled" penalty. As a result, a new policy was design (referred to as "3% + 3%" below) and was tested in the February experiments. In addition, we explored the possibility of eliminating MAP penalty from some or all of the products.

The following two tables summarize the treatments. Most of the alternative penalties were financial in nature and were calculated based on MDF. Typically, a retailer would earn 3% of the shipment value on each product as MDF. This fund was paid to the retailer in the period after the particular shipment was received. (Tables 3 and 4)

3.3. Results (from Charness & Chen, 2002)

In the September sessions, we used penalties that primarily applied to future periods. We varied these penalties for products 1 through 4, the HP products. We kept penalties for the remaining products constant across treatments; for products 5 and 8, a violation meant losing 4 periods of MDF, while for products 6 and 7, a violation meant being fined the current period's ad expense.

Since we observed that forward-looking penalties became less effective as we neared the end of product life cycles in each September session, for the February sessions we made the penalties also retroactive for some number of periods. Again, we varied the penalties for products 1 through 4, and held the penalties constant for products 5–8. In one session, we imposed multi-period penalties on MAP violations for products 1 through 4. In the other two sessions, we removed price restrictions for either products 1 through 4 or for only products 1 and 4 [product 1 (or product 4, its life-cycle replacement) has the largest market share]. We ran 20 periods in one session; eight in second session, and 12 in third session.

Table 3. There were Four Types of Treatments in our September Experiments Labeled Sep 1 through Sep 4.

Product	Treatment: Sep 1	Treatment: Sep 2	Treatment: Sep 3	Treatment: Sep 4
	Penalty	Penalty	Penalty	Penalty
1	Pulled	4 periods MDF	4 periods MDF group	12 periods MDF
2	Pulled	4 periods MDF	4 periods MDF group	12 periods MDF
3	Pulled	4 periods MDF	4 periods MDF group	12 periods MDF
4	Pulled	4 periods MDF	4 periods MDF group	12 periods MDF
5	4 periods MDF	4 periods MDF	4 periods MDF	4 periods MDF
6	Current period ad expense	Current period ad expense	Current period ad expense	Current period ad expense
7	Current period ad expense	Current period ad expense	Current period ad expense	Current period ad expense
8	4 periods MDF	4 periods MDF	4 periods MDF	4 periods MDF

Note: Products 1 through 4 were HP products, while the others were competitors' products. Therefore, only products 1 through 4 had different penalties across treatments. Here is the explanation of the penalties:
"Pulled" – product would never be sold to the retailer again after a MAP violation.
"12 periods MDF" – no MDF in the next 12 periods for the product after the violation.
"4 periods MDF" – no MDF in the next 4 periods for the product after the violation.
"4 periods MDF group" – same as above but applied to products 1–4, not just one.
"Current period ad expense" – penalty equaled to the advertised expense of the product for the period.

The September sessions (Table 5) differed with respect to the penalty for a MAP violation for products 1 through 4, with the violation penalty for products 5 through 8 kept constant across sessions. In the February sessions (Table 6), we imposed the restriction that a retailer could advertise at most two products in any one period.

Before moving to our analysis, we caution against imputing statistical significance to our results, due to the dynamic nature of the experiment. Since there was inventory carry-over, observations from different periods were not independent. Thus, any statistical tests that compared period measurements may not be accurate. Individual sessions varied considerably, further weakening statistical comparisons. Nevertheless, we see some patterns in the data.

Table 4. There are Three Types of Treatments in our February Experiments.

Product	Treatment: Feb 1	Treatment: Feb 2	Treatment: Feb 3
	Penalty	Penalty	Penalty
1	3% + 3%*	No penalty	No penalty
2	3% + 3%	No penalty	3% + 3%
3	3% + 3%	No penalty	3% + 3%
4	3% + 3%	No penalty	No penalty
5	4 periods MDF	4 periods MDF	4 periods MDF
6	Current period ad expense	Current period ad expense	Current period ad expense
7	Current period ad expense	Current period ad expense	Current period ad expense
8	4 periods MDF	4 periods MDF	4 periods MDF

Note: Once again, products 1 through 4 represented HP products. Here are explanations of the penalties:
"3% + 3%" – penalty is 3% of net shipment value for the past 4 periods + 3% of revenue for the current and next 3 periods. This is the newly designed policy based on results from the Sep experiments.
"4 periods MDF" – no MDF in the next 4 periods for the product after the violation.
"Current period ad expense" – penalty equaled to the advertised expense of the product for the period.

Table 5. In the September Experiments, "Pulled" Penalty was the most Effective with Highest Margin Observed when Compared to "MDF" Penalties.

Treatment and MAP Violation Penalty (Products 1–4)	HP Products (1–4)		Other Products (5–8)	
	Average margin	Market share (%)	Average margin	Market share (%)
Sep 2: 4 periods MDF	0.10 (0.02)	55	0.16 (0.03)	45
Sep 3: 4 periods MDF group	0.11 (0.04)	55	0.16 (0.03)	45
Sep 4: 12 periods MDF	0.08 (0.04)	41	0.11 (0.03)	59
Aggregated ad funds penalties	0.10 (0.03)	50	0.14 (0.04)	50
Sep 1: Pulled	0.12 (0.01)	54	0.17 (0.04)	54

Table 6. In the February Experiments, the New "3% + 3%" Penalty was found to be as Effective as the Previously Tested (September Experiments) Penalties. It also Maintained Substantially and Significant Higher Margins and Market Share when Compared to Scenarios where it was not Applied.

Treatment and MAP Violation Penalty (Products 1–4)	HP Products (1–4)		Other Products (5–8)	
	Average margin	Market share (%)	Average margin	Market share (%)
Feb 2: No penalty	0.03 (0.01)	56	0.06 (0.03)	44
Feb 3: No penalty (only products 1 and 4)	0.00 (0.09)	49	0.04 (0.10)	51
Aggregated No MAP	0.02 (0.06)	53	0.05 (0.07)	47
Feb 1: (3% + 3%)	0.11 (0.04)	59	0.14 (0.05)	41

In the September experiments, the overall market share of the HP products was only slightly reduced (54% vs 50%) by using MDF based penalties instead of the more severe "pulled" penalty. In the February experiments, HP market share is 59% with the new "3% + 3%" penalty, which is higher than all the September experiments as well as the other treatments in the February experiments when there were no MAP penalties. HP does better with harsher penalties but only slightly.

In both September and February, we found that retailer margins were higher with the more severe penalty ("pulled" and "3% + 3%"). This was true for both HP and non-HP products, even though we held the penalties for the non-HP products constant across treatments. This finding suggests that the retailers' pricing decisions for all goods are sensitive to the nature of the penalties for violating MAP on only the HP products.

In September, the average margins were about 20% higher when a violation led to products being permanently pulled from the retailer (for reference, we set the price restrictions so that the average margin at the restricted price was 10–13% for the control products and 17–20% for the other products). If we were to assume the independence of each observation, this difference would be statistically significant at $p = 0.04$ (one-tailed test).

Figs. 1 and 2 shows that the margins are always lower for every retailer when the penalty for violating MAP was the product becoming unavailable (black) when compared to MDF penalties (white). Figs. 3 and 4 shows the margins comparisons in the February sessions. The focus on this session is

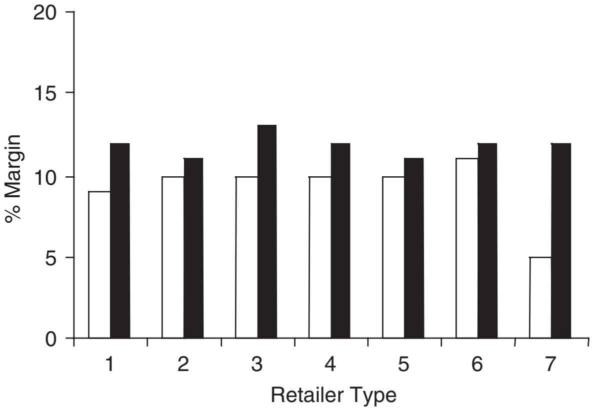


Fig. 1. September Experiments Show that Retailer Margins in HP Products are Higher with Pulled Product Penalty (black) than MDF Penalties (white).

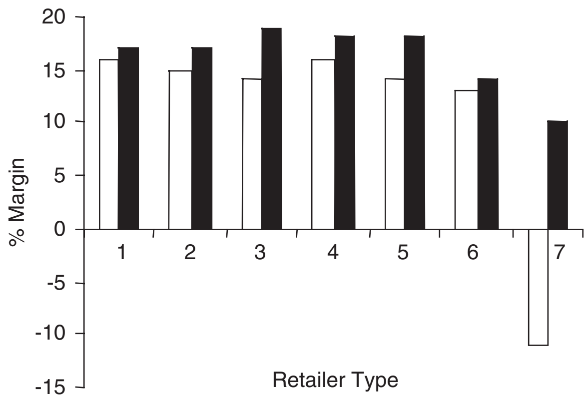


Fig. 2. September Experiments Show that Retailer Margins in Other Products are Higher with Pulled Product Penalty (black) than MDF Penalties (white).

the comparison between the "3% + 3%" policy (black) to the no penalty policy (white). The difference is significant at $p = 0.002$ (one-tailed test). It is apparent that the margin for individual retailers on all products is robustly higher with strict penalties for MAP violations.

In all the experiments, the differences between the policies are consistent across retailer types. Moreover, the same differences, driven by different

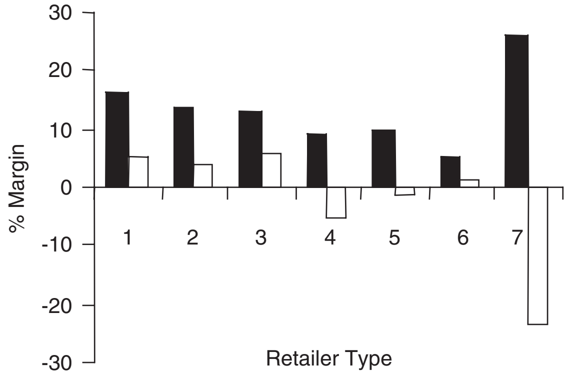


Fig. 3. February Experiments Show that Retailer Margins are Substantially Higher in HP Products with MAP Penalties (black) than with No Penalties (white).

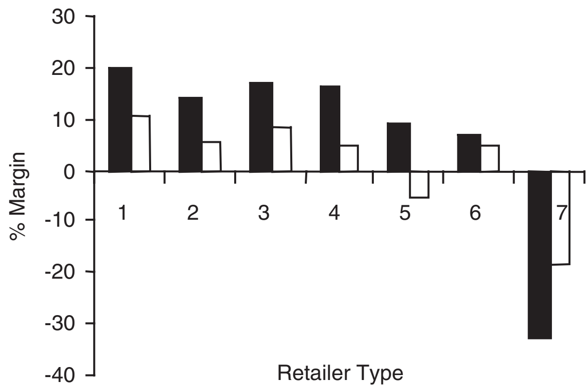


Fig. 4. February Experiments Show that Retailer Margins are Substantially Higher in Other Products with MAP Penalties (black) than with No Penalties (white).

policies, were observed not only in HP products but also in the other products which had constant policies across treatments. Intuition suggests that different MAP policies change the competitive pressure in the market for HP products. A stricter policy, such as pulling the products, lowers competition since a retailer not only has less incentive to lower his prices, he would expect his competitors have lower incentive to lower theirs and thus his incentive to lower prices is further reduced. Furthermore, in the model, all the products are substitutes (though imperfect) and compete in the same market. Thus, if a policy increases the competitive pressure for HP products, the effect will spread to all the products.

Readers may notice that the margins of the type 7 retailers were consistently lower than the other type when the MAP policy was less strict (white). This was driven primarily by the different business objectives of the different types. The objective of the type 7 retailers was revenue maximization. They were representing Internet retailers, who, in 1999, were still primarily concerned with expanding their markets without any regard of profitability.

In the September sessions, we used an exclusively forward-looking violation. We observed a pattern in the violation rate over time: Close to the end of product life cycles, every retailer violates MAP substantially more. A forward-looking penalty should (and did) have diminishing effectiveness as a product is approaching the end of its life cycle. (Figs. 5 and 6).

T represents the final period of a product's life, $T-1$ the penultimate period, etc.

We observed a positive time trend in the number of violations per period, as there are more violations as the end of the lifecycle approaches. In the February sessions, we introduced a violation penalty with both forward-looking and an additional retroactive component (3% + 3%).

Here we see no real time trend. This approach seems to have been effective in reducing the violation rate near the end of product lifecycles. We also found that the frequency of violations was related to the form of MAP imposed. Retailers (particularly mid-sized retailers) did not fare as well without MAP, as their margins were distinctly smaller; interestingly, removing the MAP on some products affects the margins for both those

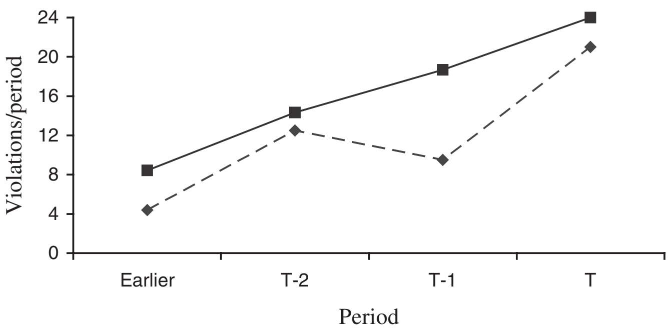


Fig. 5. MAP Violations per Period in the September Experiments Show an Upward Trend under Both "Pulled Products" Penalty (dashed line) and "MDF" Penalties (solid line).

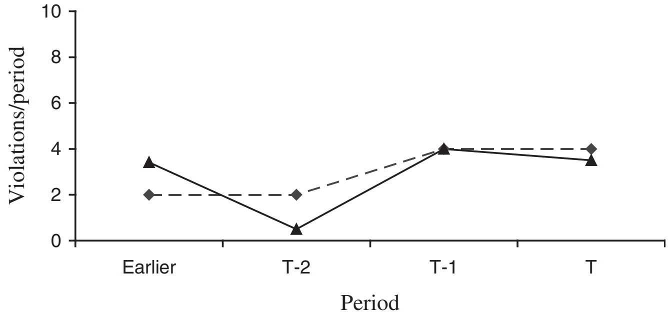


Fig. 6. MAP Violations per Period in the February Experiments no Longer Show an Upward Trend after switching to the "3% + 3%" policy.

products and for the others. This calibration suggests that equilibrium prices may well be below the price floor. Based on our results, HP felt it would be best to continue some form of MAP.

We were also able to detect other weaknesses in the design and enforcement of several advertised-price policies; this led HP to revise the policies it implemented. For example, retailers may carry several different HP products. One proposed enforcement policy would link these products, so that a violation on any individual product would trigger penalties on all of them ("4 period MDF group"). When we tested this policy, we found that retailers who decided to violate the MAP on one product would often violate the MAP on all the linked products. As a result, HP decided not to implement a linked-product MAP design.

However, the most important results of this study are the identification and verification of the escalating violations at the end of product life cycles and the validation of the need of MAP. Initially, we tied MAP penalties to future shipments and future market-development funds for the product at issue. Retailers correctly perceived that forward-looking penalties would have little effect late in a product's life. Furthermore, they could game the system by cleverly timing shipments to avoid penalties close to the end of product life cycles. As a result of this work, HP decided to adopt a completely different enforcement policy, which we validated in the February experiments. The new policy is retroactive as well as forward-looking, so that retailers cannot escape penalties even if they violate MAP at the end of a product's life. As of the writing of this chapter, this is still the standard MAP policy for HP consumer businesses.

4. NO CHANGE IS GOOD NEWS: MARKET DEVELOPMENT FUNDS POLICY EXPERIMENTATION

Another industry practice is MDF, sometimes also known as soft-funds. MDF is given to retailers for the purpose of promoting a manufacturer's product. It can be used for advertising, promotional campaign, printing marketing literature, and anything related to marketing. Generally, enforcement of the uses of MDF is not perfect. Even in cases where enforcement is perfect, retailers can shift their already planned marketing expenses to MDF. Thus, MDF is a close substitute to discounts on the wholesale prices.

Usually, MDF is around 3% of the wholesale payment to the manufacturer. In the computing industry where the retailers' margins are in the 5–15% range, MDF can be a significant part of the profit equation.

A sequence of 13 experiments was conducted in 2001 to study the three types of MDF scenarios. The existing HP policy gave out MDF based on a fixed percentage of the value (the amount HP charged for) of the shipments of products. The standard rate for retailers was 3%. The goal of MDF is to provide an incentive, in addition to the profit retailers are already making, to promote HP's products.

At the time of the project, HP sought to examine two new ideas. The first is to give out MDF based on a quota system. A growth quota is calculated based on past sales. Each retailer receives MDF equal to a higher percentage of their net shipment value if their sales revenue exceeds their quota. The second idea is to use a ranking system to determine MDF. Scores are calculated based on revenue growth. Retailers are ranked based on their scores. The MDF percentages are based on their ranking.

4.1. Results

Table 7 summarizes the experimental treatments and results.

The experiments were conducted in two sequences. The first sequence consisted of experiments one through seven. A strong learning effect was observed. Subjects were making substantially more money in the last 4 of the 7 experiments. As shown in the table, their average revenue increased from around 250–350 thousand to more than 400 thousand, while the margins remained roughly the same. The second sequence of six experiments (8–13 in the table) was conducted mostly with experienced subjects to

Table 7. No Alternative Policies Showed Significant Improvement over the "Base" Treatment.

Expt	Date	Treatment	Revenue HP Received	Market Share (%)	Reseller Margins (%)	Ad $ Spent on HP
1	1/17/2001	Base	254,786	48.7	13.2	15,635
2	*1/18/2001*	*Quota*	217,123	*50.3*	*11.9*	*14,643*
3	1/25/2001	Rank 1	306,850	49.6	12.5	23,093
4	2/1/2001	Rank 1	391,797	50.5	13.3	36,457
5	2/1/2001	Rank 2	354,038	48.3	13.6	25,345
6	2/15/2001	Base	361,043	48.5	12.8	41,809
7	*2/15/2001*	*Quota*	315,466	*51.2*	*11.3*	*16,304*
8	3/20/2001	Base	359,319	44.5	12.6	13,466
9	*3/20/2001*	*Quota*	*379,258*	*50.8*	*12.5*	*17,084*
10	3/22/2001	Rank 1	338,156	45.9	13.7	11,129
11	3/22/2001	Rank 1	263,212	48.6	9.6	6,586
12	3/23/2001	Base	292,833	46.1	11.8	5,233
13	3/23/2001	Rank 1	389,948	48.4	12.6	17,345

minimize learning effects. One of the subjects in experiment 11 exhibited an exceptional level of price competition. As a result, the average margins of the market were lower than similar experiments. Since we observed this behavior only in one of the five experiments with the same treatment, the result of this experiment was less emphasized in formulating business recommendations.

Summary statistics suggest that the experiments were calibrated reasonably well with the markets being studied. For example, advertising spending in the experiments ranged from 3% to 10% of resellers' total revenue, which were consistent with a 2–8% figure in the real world. Unfortunately, it was not possible to acquire detailed reseller advertising spending data to further strengthen the results with a statistical test.

The quota policy sparked the fiercest competition among the reseller. Its resulting margins were lower than all the baseline experiments, except in one case. It may seem paradoxical that the resulting revenue for HP was lower in both experiment 2 and 7. Closer examination revealed that advertising spending on HP products was also lower in those experiments. This also may seem paradoxical. If the quota policy was causing an increase level of competition, why did the subjects only competed on price while reducing their advertising spending? The answer lied in the fact that advertising budgets were strongly influenced by the quota policy. The quota was set exogenously. In experiments 2 and 7, on average, very few subjects met quota. As a result, they all received lower amount (on the order of 2% of HP's revenue) of MDF. Since their margins were on the order of 10% and their advertising spending was on the order of 6% of HP revenue, this reduction had a significant impact on the amount of advertising the subjects were willing to produce on HP's behalf. This phenomenon was reversed in experiment 9 where most subjects were able to meet the quota. This result highlighted the importance of setting a "correct" quota that can result in more revenue.

The task of setting quota, was difficult even in a controlled laboratory environment. The task becomes almost impossible in a real business setting because additional forces such as supply or demand shock can easily render any quota unbeneficial. A business plan to implement a quota system was halted because of this study.

The ranking policies did not provide any significant increase in revenue for HP. They did better only when compared to experiment 1. Since experiment 1 was the only experiment with only inexperienced subjects, we believe that learning is the primary explanation of that difference. As oppose to the quota policy, which provided the subjects with a fixed target to hit, the ranking policies required the subjects to respond to their competitors' future performance if they wanted to receive more MDF. Most subjects seemed to ignore the ranking policies because it would be difficult to assume at and hit a moving target. HP has decided not to pursue a program of ranking policies based on the experimental results.

This study highlights a difference between academic research work and business research. Academic experiments were designed to illustrate the effects caused by different treatments. A negative result that shows little effect of the treatment variable is usually less desirable. In the case of business experiment, even a negative result, such as the case in this study, can have an important impact on the business.

5. DISCUSSIONS

5.1. Design Philosophy of Business Experiments

There are obviously many very broad classes of business questions that experimental methodology can provide answers. There is not enough room to cover them all. Instead, this chapter focuses on the low hanging fruit: areas that experimental economics has contributed immediately and significantly.

Contracting between manufacturer and retailers is one such area. Manufacturers operating in the contemporary market for technology products face a daunting task in designing effective incentives for their retailers. Channels of distribution are diverse, with new channels emerging, and demand fluctuations, market exposure, advertising, stocking, and product life-cycles are uncertain. The behavior of retailers is a critical element in whether a manufacturer achieves its business goals. Experimental economics offered a unique method to study and predict such behavior. This chapter described three studies, each was used to evaluate policies, in the following areas: return, minimum advertised price (Charness & Chen, 2002) and MDF. HP Consumer Business Organization, which was in charge of a $18 billion business, changed its contract policies based on the results of these experiments. These studies not only show how experiments can be used to help business decision making, but also illustrate how experimental designs and goals were affected by business needs.

In this class of problems, experimental method is perhaps the only reasonable alternative to field test. The complexity of a real retailer business environment makes a full theoretical analysis impractical. Even discounting the importance of scale (in the number of products, the number of retailers, and the number of manufacturers competing), there are non-trivial interaction between numerous economics structure and forces. For example, features in a typical retailer model used in one of these studies include retailers' heterogeneity in multiple dimensions, a consumer demand environment where the retailers competed in, an advertising model interacting with the underlying consumer model, product life-cycles, supply chain structures and multiple interacting policies. Although theoretical analysis of a simplified model can be useful in providing some insight for policy design, it is obviously not practical to create a theory model to encompass all the features that are deemed important. The major value of the experimental methodology is the ability to study the interactions of these features.

HP Labs developed in-house experimental economics capabilities instead of relying on academic institutions for consultants because business considerations make such consultation impractical. Business decisions must be made in a timely fashion, even if they are made with less than perfect information. HP typically develops its potential business in 3 to 6 months, depending on the cycle of contract and policy decisions. Experiments are often designed with the expectation that redesign and repetitions are unlikely, except in the most critical situations. Academic researchers generally want to establish statistical significance, necessitating replications and increasing the turnaround time.

Furthermore, time limitations often mean that exploring the parametric space fully is impractical. As a result, complexity of the field environment is preserved in many projects. For example, in the retailer experiments discussed above, many features such as stochastic supply, demand and delivery times, residual advertising effectiveness, and price reputation are included. The experimental environment was therefore quite complex. This design philosophy runs counter to standard academic experimental practice, where researchers prefer the simplest design that can encompass the modeling issues at hand. However, the goals of such studies are also modest. Although desirable, full identification of causes and effects are not required. It is more important to know whether a policy works than to know why it works. It is also important to identify possible exploitation of policies. From a business point of view, identifying such exploitation is unquestionably useful. It is less of a concern whether or not it is the equilibrium strategy for a retailer. In effect, we are employing subjects to find flaws in proposed policies.

This research strategy replies upon the accuracy of experimental models. Multiple safe-guards are built into the process to ensure experiments that are accurate reflection of the real business environment. Business experts are asked to evaluate all the features and assumptions included in the experimental models. Real business data, if possible, are used for calibration. All the models also go through a validation process in which business experts will play the game and offer feedback.

The reader needs to keep in mind that experimental results will at best provide an accurate evaluation of what will happen if something is done. They will be of limited utility as guidance of what to try. In the past, most policy alternatives were created in meetings where intuition and reasoning were the primary tools. An additional research strategy was developed to address this issue. When possible, we will develop a simpler model, theoretical, simulation-based and/or experimental, to provide some insights of what policies would be effective. These policies would then be evaluated along side with policies created by other means in full experimental models. This tandem strategy has only been applied on a limited basis due to resource and time constraints. The study of durable goods markets described below is a good example of how multiple methodologies were used to address the same problem.

5.2. Application Areas

It is not due to chance that all of the internal HP applications are dealing with issues of contracting between supply chain partners. There are very compelling reasons, business and intellectual, that make contracting a "sweet spot" for experimental economics work. The importance of these policy decisions is the most obvious one. It is worth noting that these decisions are important because both the size of the business is big and that behavior can be substantially changed by them. As importantly, experimental results, such as the exploitation identified in the MAP experiments, can lead to actionable recommendations because a large manufacturer such as HP has substantial control over its contracts. A third reason is that while the environment is complex, it is not beyond the scope of a reasonable experiment design. In this issue, we are blessed by the fact that important characteristics of the environment can be captured by several rules that can be taught to subjects in a short period of time. For example, although retailers were heterogeneous, the number of types was quite manageable. All the models only considered the immediate business of the retailers surrounding several key products. It would be much more difficult to consider the interaction of different businesses (such as printers *and* PCs which have a different set of competitors) or the interactions of multiple points (e.g. including suppliers) in the supply chain. Furthermore, contracting is also an intellectually interesting area where game theory, micro economics theory, marketing science, behavioral economics, and even psychology can play a major role. It is more satisfying to researchers at HP Labs because this research theme was not planned but just emerged naturally driven by a set of consistent business needs. At present, experimental work is continuing to help HP explore contracting options with its reseller partners in multiple business areas.

There are other obvious application areas. Aside from the primary focus of contracting, HP Labs have also developed applications in other areas. The most noteworthy example is a study of durable goods market, collaboration between HP and another Fortune 20 company. This was an attempt to ascertain whether experimental methodologies could be successfully applied to a different industry. Thus, we selected a problem that was as different to the retailer studies as possible. Instead of focusing on game theory and behavioral effects, such as how a retailer can manipulate the rules of the contract, we focused on aggregate market behavior brought on by many small players. As a result, the durable goods market experiments (20+ subjects per experiment) are substantially larger than the retailer experiments (5–7 subjects per experiment). In Chen and Huang (forthcoming), an experimental model was developed to study the behavior of the secondary market for durable goods. This research created a general framework to address some of the unique issues in automobile marketing. Based on this work, experiments were developed to study whether the additional of a new channel of sales to the used-goods market would be beneficial or not? This manufacturer sold around 1 million used vehicles to dealers annually through life auctions. There was significant transportation costs associated with life auctions. Furthermore, life auctions did not necessarily capture all the potential demand. A new marketing program was designed to address these issues. Most of these used vehicles were "returns" from both consumers and commercial customers, such as rental companies, declining to purchase at the end of lease. The usual process was for the customer to return his end-of-lease vehicle to a nearby dealer. This dealer would then notify the manufacturer and send the unit to the auctions. The proposed marketing program works in the following manner. When a dealer notified the manufacturer the return of an end-of-lease vehicle, the manufacturer, based on some pre-established decision rules, may offer to sell this vehicle to the dealer at a dynamically generated fixed price. If the dealer declined, the vehicle would be shipped off to auction.

The proposal was to use auction prices of the previous month to determine the fixed prices for the current month. Previous internal research of this manufacturer has identified variables, such as mileage, color and an established way of measuring the condition of a vehicles, which are good predictors of auction prices under a specific regression model. This model, referred to as the floor price model, formed the basis of the pricing process. For any vehicle that might be offered, an "average" price was calculated based on the floor price model estimated from auction data of the previous month. The final fixed price offered to the dealer was the average price with a fixed percentage marked-up or marked-down. This process created a feedback loop. Yesterday's auction determined today's fixed prices. These prices would affect the acceptance rate of the offers. The acceptance rate would change the supply, as well as the demand, of today's auctions since the vehicles that were rejected would be sent to the auctions. Today's auctions would then again decide the pricing of tomorrow's offers. And the cycle continued.

This particular pricing feedback structure, in conjunction with the new marketing program, produced a significant more revenue to the manufacturers in the experiments. As a result, this manufacturer has implemented this new program.

Another example is the evaluation of the software procurement agent, AutONA (Byde, Chen, Bartolini, & Yearworth, 2003). Technology advancements have offered dramatically new ways of doing business. However, these technologies are often thrust into the business world with little understanding of how the economics would be affected. Sometimes, the effect can be wonderful (i.e. eBay) and sometimes it is a waste of capital (eToys, Webvan…). In situations where there is very little past experience, experimental economics offer a way to provide some guidance of the economics effects of applying new technologies. AutONA is a software that HP has developed for multiple one-to-one negotiations in a procurement setting. AutONA is a rule-based system that negotiates price and quantities with multiple suppliers on a one-to-one basis. The belief is that it will reduce the operational aspects of procurement costs by automating a significant part of the procurement functions. HP procurement was seriously considering turning over its negotiations to this software. The Achilles' heel of this argument is that it may not be true that this system could provide deals that are comparable to those made by human negotiators. Laboratory experiments were conducted to test whether human suppliers can take advantage of the robot buyer. Results show that although the robot buyer passed a simple Turing test, that is no human player could identify who the robot player was, it exhibited significant behavioral biases that resulted in worse prices when compared to those negotiated by human subjects. As a result, HP procurement decided not to deploy this system. It would be disastrous for HP if this system is deployed across an organization that procures around $4 billion worth of DRAM.

5.3. Implementation Issues: Software and Experimental Procedures

The carefully constructed research strategy will come crashing down if experiments cannot be implemented and modified in a timely basis. In the early days before the arrival of mass computing, simple games such as the prisoners' dilemma and simple auctions were implemented with pen and paper. With the advent of personal computers and networking technologies, more sophisticated games such as smart markets, combinatorial auctions, and information markets are possible. The complex nature of business experiments takes this evolution one step further. The needs for fast implementation and modification of economics scenarios have to be addressed if business decisions are to be made in a timely fashion. The strategy is to create a software tool that simplifies and automates as much of the implementation and experimental procedures as possible.

HP Labs developed a software platform, called MUMS, for the purpose of implementing economics experiments (Chen, 2003). The design called for two guiding principles. First, it needs to support many types of games, since there are many different economics models and business processes that are of interests. It is more cost-effective to invest upfront in a more sophisticated system, which can be used to support a wider range of models than to incur the costs of programming for each individual project from scratch. Second, the programming interface needs be simple so that researchers with little programming experience would be able to use it effectively. The ease of programming is the determining factor governing how efficient researchers can implement their models and execute their experiments. The main design challenge is to maintain the balance between ease of use and the flexibility of the system. A very flexible system such as the C++ programming language would be a terrible choice for someone without the right computing experience. At the other extreme, we can develop systems for particular games, such as auctions, but would lose the ability to use the system for any other types of games. We have decided on the approach of a script-language based system. The design of the language allows the user to define a game as a collection of high-level concepts: a set of players, inputs and outputs from players, and sequential logic that govern the rules of the game. Basic computing functions such as elements of interface design, networking, and database functions are taken out of the hands of the users.

The idea of script languages for particular games is not new. In chess, for example, there are several languages, which have been developed to simplify the knowledge acquiring process and to help creating better AI. (George et al., 1990; Donninger, 1996). The MUMS script language is a general purpose language, and have the common features found in other languages, such as data types, multi-dimension arrays, variables, functions, control statements, and so on. It has its root in the C programming language, although many complications have been eliminated. The syntax is similar although the lower level functions such as pointers are completely eliminated. The language hides all the details of the physical computing environment, such as a distributed network, where the game will be executed. For example, "players" are defined as elements in a script. A special global array variable "Player" is used to reference players. All input and output functions are called with reference to a player. The language treats players universally in definitions of games. During actual execution, the system determined dynamically how to map "players" to physical computers, which can be local or remote. It is even possible to map a "player" to a software agent.

Here are some examples of the types of games that have been implemented: retailers (oligopoly) games with business policies, multiple interacting durable goods markets, various types of auctions (one and two-sided), multiple one-to-one bilateral negotiation, information markets with Arrow-Debreu securities. All the research projects described below were also implemented in this system.

Although the primary focus of this tool is to support business experimental research, its design lends itself to research in other fields. This philosophy is not very different from goals of the Berkeley Xlab, which was created to serve multiple disciplines that study human behavior.

As a result, our development on computer technologies has gone beyond the issue of implementation of economics games. Scientists in both fields have started to exploit the synergies in economics and computer AI to ask the question of whether computers can be good economics agents either in place of or as support tools of human beings. At the focal point of these fields, which are different in nature but similar in goal, is an obvious need of a generalized platform to support games and agents. The MUMS system, which designed primarily to run economics experiments, has an additional feature that accommodates artificial behavior. The game scripts are agnostic to whether a player is going to be human. During run time, a human or a robot can be assigned to any role in a game. Switching between all human, all robot or partially human, partially robot experiments requires very little work.

In addition to developing software to implement economics scenarios, effort has been made to streamline other aspects of experimental operations. A standard set of procedures was developed from the point when recruitment of subjects is initiated to when the subjects are paid and escorted out of the laboratory. These procedures ensure continuity and consistency in laboratory operations when there are personnel changes in laboratory administration.

Training was also integrated into the procedures. Written instructions are posted on a special website created for the HP experimental economics program at a minimum three days before an experiment. All subjects are recruited either through an email list or notices posted on electronic bulletin boards. Subjects who are invited to participate are required to pass a web-based quiz, usually consists of multiple choice questions, before they are allowed to take part in experiments. It is particularly important in business experiments to recruit subjects that understand the mechanics of the games.

Web-based training and quiz helps to ensure the qualities of the subjects in this aspect. This process raises the issue of self selection because potential subjects can choose to sign up for experiments that they believe they can do well. This may be an issue for behavioral experiments that require a representative sample from the general population. However, self-selection usually is desirable for business experiments where we want the subjects to do well. Additional reference material and training time will be provided in the beginning of experiments.

Standard spreadsheet based database is used to keep track of the subject's profile, payment information and the history of participation. These data allow us to control samples of subjects as the need arise. A special checking account was created and all the subjects were paid by checks. In addition, a financial framework, internal to HP, was set up enabling the charging of experimental expenses to the proper accounts, since multiple business organizations are engaging in collaborative projects with the experimental economics program.

These mundane considerations, often ignored by researchers, can be determining factors of whether a business project is successful or not. This system ensures a consistent capacity of producing experimental data. In a world where timeliness and predictability can be as important as the validity of modeling assumptions, to be able to predict, plan and deliver experiments are crucial to whether the research will be useful and create significant impact in businesses.

5.4. Looking toward the Future

There is an obvious parallel to be drawn between physical sciences and social sciences. If you want to build a quantum computer, you hire theoretical physicists to work out the underlying science. The experimentalists test whether the theoretical models are accurate in a laboratory. The engineers then create prototypes based on the experimental results. Economics, and in particular experimental methodologies, in my opinion, has developed to a point where a similar process of engineering has become possible for business processes. Furthermore, the advancement of software technologies will start to blur the line between an economics experiment and a business prototype in the future. In most of the research projects discussed above, business executives participated in mock experiments to get a feel for the experimental environment so that they could offer feedback. In some ongoing studies, plans have been made to go one step further and to use the experimental model as a war-game for business executives to test and develop their skills in an interacting environment. The day will come that software used in laboratory experiments can be scaled up and become prototypes of actual business processes.

There are two extremes of business engineering. The first is the tinkering of existing well-known business processes. This is akin to upgrading the design of an existing automobile. The engineers already know all the major elements to be included in a car. The work is to tweak each element so that they can work better together. There also may be an upgrade of a certain component such as the engine. The objective is incremental improvement or validating existing methods. The retailer experiments fall into this category. There might be a proposal of policy changes in one or more areas, but the fundamental way of doing business remained the same. The focus is to find out what works and what does not. The author would claim that there is a great need to use rigorous scientific experiments for this purpose. The need is driven by the way of how all these business processes came to existence. Rarely, they were designed from scientific understandings or even intuition. Most of them emerge from a kind of business evolution where any process that keeps a business from being unprofitable will survive. This is a far cry from optimality. In addition, experimental analysis is valuable to business decision-makers even in the cases where it validates the optimality of the status quo. Simply knowing there is no need to change saves all the costs of new marketing program and field tests. The author believes that this kind of business engineering, tinkering with existing businesses, will blossom in the next few years.

The second extreme is similar to trying to build nano machines, something completely new to the world. Information aggregation technologies such as prediction markets, combinatorial auctions, and quantum economics fall into this category. A cottage industry surrounding prediction markets has already mushroomed overnight.

Laboratory experiment is obviously not the only scientific method to answer policy questions. Theoretical (game theory or market theory) analysis and software simulation are the other two methods. These methods have advantages and disadvantages. Theory can provide invaluable understanding and guideline for actions but it becomes intractable quickly in a complex environment. Furthermore, theory based on rational assumptions sometimes is not the best predictor of human behavior. Computer simulation can scale up to large complex system easily, but their applicability depends on assumptions about the decisions human agents make in the field. There are obvious complementarities among these methodologies. My belief is that there will be a convergence of all these methodologies, including experimental economics, as arrows in a quiver to target business decision-making problems in the future.

The research discussed in this chapter has only begun to scratch the surface of potential business applications. Policy analysis, obviously not the only application area, can be extremely lucrative. Since most of the decisions made with the help of experimental economics scaled with the size of the business, the monetary impact of these projects was enormous. For example, the MAP policy designed based on experimental results has become the standard policy in a $12 (now $18) billion retail business for the last 4 years. Before the project in 1999, it was widely known inside HP businesses that one incident connected with the failure of the MAP policy at that time has cost HP around ten million dollars.[3] The belief is that if experimental methodology was available prior to this incident, HP would have been able to avoid it. Obviously, it cannot be known how many more of this types of problems were avoided after the new policy was implemented. Similarly, the MDF study stopped the implementation of a new policy that distributed on the order of $300 million worth of funds. These numbers seem to indicate a very bright future for experimental business research.

The nature question is how HP can develop a coherent strategy to integrate experimental economics into its decision-making process. On this issue, HP suffers slightly from its size, which ironically, also makes experimental economics very compelling. HP can be viewed as a loose federation of business units with semi-autonomy.[4] HP Labs is the central research arm of HP, which often acts as an internal consultant to the businesses. Despite all the past successes, it remains an ongoing process to educate and communicate to the businesses about the potential applications. In the past, all the projects were initiated because specific business needs, for example, the need to change the MAP policy, was brought to the attention of HP Labs. Initially, these contacts were brought about by regular HP Labs to HP business information exchange. HP Labs obviously will accept or reject based on its assessment of the importance of the problem as well as its scientific value. Due to the increasing visibility of the program in the past few years, more and more business problems are brought to us because one has heard that experimental economics may offer a new solution. This also means that, unlike in the beginning of the experimental economics program, substantial effort was no longer necessary to "market" the technology. Nevertheless, the resulting choices of experimental projects may not intentionally adhere to a strong theme and can scatter across many different types of problems and businesses.

On the other hand, as mentioned earlier, the "sweet spot" of these applications seems to be in the area of managing contracts with supply chain partners, particularly downstream resellers. HP Labs is developing a more efficient strategy for the application of experimental economics given the fact that (a) managing the reseller channel is important (b) contracts and policies need regular updating due to the changing nature of business environment, and (c) HP Labs experimental economics has the best track record in this area. The idea is to integrate regular laboratory testing of policies into selected business units, which are responsible for contracts and reseller policies. The experimental process will need to synchronize with the typically annual cycle of contract updating and runs in parallel to the real business. Experimental models will be updated according to changing business environment. Potential policy changes, or just the status quo policy, should be tested in the regularly updated environment before implementation. A system like this will not only be reactive to the policy problems that HP businesses identify and want to fix but also preventive to potential problems. To institutionalize experimental economics into business processes will be a major endeavor. At present, HP Labs is still at a very early stage of exploration although several business units have expressed interests in such a vision. Finally, HP Labs is also working with HP consulting to explore whether it is possible to create a consulting business. It is obvious that if such research is valuable to HP, it will also be valuable to many other companies.

NOTES

1. HP's consumer business grew to around $18 billion in 2004.
2. There are 12 subjects in each experimental scenario. They were organized into groups of three. 40 periods were conducted for each group. The statistics reported are based on period 5 to 35. The reason to truncate part of the data is to eliminate end-game and start-game effects.
3. Note that although $10 million was a small percentage of a $12 billion business, it was still more than large enough to fund experimental economics research for the next 20 years.
4. In the recent years, HP has been moving toward a more centralized model. However, at the time of this chapter, HP is still divided into many business units, although they are encouraged to cooperate with one another.

REFERENCES

Bali, V., Callandar, S., Chen, K.Y., & Ledyard, J. (2001). *Contracting between a retailer and a supplier,* Working paper.

Byde, A., Chen, K.-Y., Bartolini, C., & Yearworth, M. (2003). AutONA: A system for automated multiple 1-1 negotiation. In: *Proceedings of IEEE international conference on E-commerce*, June, (pp. 59–67).

Brewer, P., & Plott, C. R. (2002). A decentralized, smart market solution to a class of back-haul transportation problems: Concept and experimental test beds. *Interfaces, Special Issue on Experimental Economics*, *32*(5), 13–36.

Cason, T. (1995). An experimental investigation of the seller incentives in EPA's emission trading auction. *American Economic Review*, *85*(4), 905–922.

Cason, T., & Plott, C. (1996). EPA's new emissions trading mechanism: A laboratory evaluation. *Journal of Environmental Economics and Management*, *30*(2), 133–160.

Charness, G., & Chen, K.-Y. (2002). Minimum advertised price policy rules and retailer behavior: An experiment. *Interfaces, Special Issue on Experimental Economics*, *32*(5), 62–73.

Chen, K.-Y., Huang, S. (forthcoming). Durable goods lease contracts and used-goods market behavior: An experimental study. In: A. Rapoport & R. Zwick (Eds), *Experimental business research, Vol. 2: Economic and managerial perspectives*. Norwell, MA and Dordrecht, The Netherlands: Kluwer Academic Publishers.

Chen, K., & Wu, R. (2003). Computer games and experimental economics. In *Proceedings of ICEIS*, April.

Donninger, C. (1996). CHE: A graphical language for expressing chess knowledge. *ICCA Journal*, *19*, 234–241.

Dubin, J. (1998). *Studies in consumer demand – econometric methods applied to market data*. Boston, MA: Kluwer Academic.

George, M., & Schaeffer, J. (1990). Chunking for experience. *ICCA Journal*, *13*, 123–132.

Kali, R. (1998). Minimum advertised price. *Journal of Economics and Management Strategy*, *7*(4), 647–668.

McCabe, K., Rassenti, S., & Smith, V. (1990). Designing 'Smart' Computer-assisted Markets. In: V. Smith (Ed.), *Papers in experimental economics* (pp. 678–702). New York, NY: Cambridge University Press.

McCabe, K., Rassenti, S., & Smith, V. (1991). Experimental research on deregulated markets for natural gas pipeline and electric power transmission networks. *Research in Law and Economics*, *13*, 161–189.

McFadden, D. (1976). Quantal choice analysis: A survey. *Annals of Economic and Social Measurement*, *5*, 363–390.

Plott, C. (1997). Laboratory experimental testbeds: Application to the PCS auction. *Journal of Economics and Management Strategy*, *6*(3), 605–638.

Plott, C. (1999). Policy and the use of experimental methodology in economics. In: L. Luini (Ed.), *Uncertain decisions bridging theory and experiments* (pp. 293–315). Boston, MA: Kluwer Academic.

Rassenti, S., Smith, V., & McCabe, K. (1994). Designing a real time computer-assisted auction for natural gas resources. In: W. Cooper & A. Whinston (Eds), *New directions in computational economics* (pp. 41–54). Boston, MA: Kluwer Academic.

This page intentionally left blank

EXPERIMENTS ON AUCTION VALUATION AND ENDOGENOUS ENTRY

Richard Engelbrecht-Wiggans and Elena Katok

ABSTRACT

We present results of several experiments that deal with endogenous entry in auctions and auction valuation. One observation that is constant across all the experiments we report is that laboratory subjects have a difficult time evaluating potential gains from auctions. Even after they are given some experience with particular auctions, the uncertainty inherent in the auctions (the probability of winning as well as the potential gains from winning) makes it difficult for subjects to compare different auction mechanisms. This highlights the need for new experimental procedures to be used for testing theories that involve endogenous auction entry in the laboratory.

INTRODUCTION

Ascending bid (or "oral") auctions,[1] in various forms, are used to sell and purchase a variety of goods and services. These auctions are especially popular when the objects are auctioned and the bidders change from auction

Experimental and Behavioral Economics
Advances in Applied Microeconomics, Volume 13, 169–193

ISSN: 0278-0984/doi:10.1016/S0278-0984(05)13007-7

to auction. Oral auctions are used to price art, used cars, cattle, real estate, estate contents, used machinery, and miscellaneous junk. Various internet auction sites – most notably eBay – use ascending bid auctions. Various charities conduct ‘‘silent auctions’’ in which individuals write their name and bid – which must be at least some specified increment than the previous bid – on a ‘‘bid sheet’’ for each of the several simultaneously auctioned items.

The studies we present in this paper attempt to shed some light on the popularity of oral auctions. Why are ascending bid auctions – versus, for example, sealed bid auctions[2] – so commonly used? One explanation might be that auction mechanisms compete for sellers…and perhaps sellers prefer ascending bid auctions. Sellers might prefer ascending bid auctions if they generate higher revenues…but do they? Another explanation might be that auction mechanisms compete for bidders (for example, eBay makes its money based on transaction volume, and therefore eBay may choose a mechanism that maximizes buyer traffic – the number of transactions rather than the average value of individual transaction) and bidders prefer ascending bid auctions.

Auction theory attempts to address the question of which auction mechanism generates the higher revenue. Vickrey (1961) pioneered the theory by defining the commonly used independently drawn privately-known values model (IPV) with risk neutral bidders, and we presume this model in what follows. Within this model, Vickrey argued that, at equilibrium, ascending bids would result in the same average price as the sealed bid auction. Subsequently, Myerson (1981) generalized this model to other auctions, and established that sellers typically profit from setting a reservation price greater than their own value for the object.

One critical assumption in this standard theory is that the number of bidders is exogenously fixed. When this assumption is not satisfied, some standard results change. For example, Engelbrecht-Wiggans (1987), McAfee and McMillan (1987) and Engelbrecht-Wiggans (1993) show that with endogenous entry, the optimal reservation price equals the seller's value. In other words, when we view different auction mechanisms as competing for bidders, a seller who sets the reservation price equal to his own value will attract more bidders and obtain a better price than one who sets a higher reservation price.

Engelbrecht-Wiggans (2001) suggests that, everything else equal, bidders might prefer ascending to sealed bid auctions. Unlike sealed bid auctions, ascending bid auctions have a dominant strategy that is both transparent and independent of the number of competitors or the competitors' strategies. If there is some cost associated with estimating the number of competitors, then bidders may prefer ascending to sealed bid auctions. Since the expected revenues to the seller increases with the number of bidders, ascending bid auctions could generate higher revenues in the IPV setting. Engelbrecht-Wiggans (2001) then provides a formal, illustrative example in which the entry, decision to discover the number of competitors, and bidding itself, are all endogenous. The expected equilibrium price for ascending bid auctions exceeds that of the sealed bid auctions by approximately 20% – a substantial increase.

The Engelbrecht-Wiggans (2001) model uses three major assumptions that have been tested in the laboratory by various researchers:

1. Participants are able to discover the dominant strategy in ascending bid auctions. The fact that participants learn to bid up to their value in ascending bid auctions is fairly quick and has well been established (see Kagel, 1995 and the references therein).
2. Participants bid close to the risk-neutral Nash equilibrium (RNNE) in sealed bid auctions. Most of the laboratory evidence suggests that in fact, laboratory participants bid substantially above the RNNE (see Kagel, 1995 and the references therein). Some of the explanations for this "overbidding" that have been suggested are risk aversion (Cox, Smith & Walker, 1988), saliency (Harrison, 1989), learning (Ockenfels & Selten, 2004), and the desire to avoid regret (see Engelbrecht-Wiggans, 1989 who first introduced the notion of regret in auctions and Engelbrecht-Wiggans & Katok, 2004 who find empirical evidence for it).
3. All else equal, participants prefer the ascending bid to the sealed bid mechanism. Ivanova-Stenzel and Salmon (2004a,b) find evidence that participants overwhelmingly prefer ascending bid auctions when entry fees are equal, and are willing to pay to participate in the ascending bid instead of a sealed bid auction, but the amount they are willing to pay is substantially less than the ex-post difference in expected profits from the two auctions.

These experimental results have several important implications. Specifically, while experimental subjects tend to bid very close to what the theory predicts in the case of ascending bids, they tend to bid significantly more in sealed bids than the theory presumed when establishing the revenue equivalence between the two types of auctions. Therefore, for a fixed number of experimental subjects, the seller would obtain a higher price using sealed rather than ascending bids. However, subjects prefer ascending to sealed bids. Therefore, given a choice, ascending bid auctions may attract more bidders than sealed bid auctions, and these additional bidders may drive up the price for ascending relative to sealed bids.

This brings us to our central question: Is the preference for oral auctions strong enough to offset the higher-than-equilibrium bids that have been empirically observed in sealed bid auctions? More specifically, if bidders are given a choice, would the number of bidders who chose ascending over sealed bid auctions be high enough to drive the average price in ascending bid auctions above the average price in sealed bid auctions? In short, which type of mechanism would generate higher prices if bidders had a choice?

To explore these questions we conducted a series of studies that we describe in the following sections. What unites these studies is that in all of them participants compare, in one form or another, the ascending and the sealed bid auctions. In the first study they compare the two mechanisms directly, and we find it much like Ivanova-Stenzel and Salmon that although participants prefer ascending bid auctions, it is usually not strong enough to offset the overbidding in the sealed bid auctions. In subsequent studies we simplified the auctions and converted them to lotteries in order to understand better why participants seem to "undervalue" the ascending bid relative to the sealed bid auction. We find that experience, in the form of information about auction outcomes, does not change behavior, and specifically, the willingness to pay and to participate in ascending instead of sealed bid auctions is substantially smaller than the difference in the ex-post average profits that participants earn in the two auctions. In other words, experimental subjects have a difficult time estimating the expected earnings from auctions, and systematically underestimate the expected earnings from ascending bid auctions relative to the expected earnings from the sealed bid auctions.

STUDY 1: FREE CHOICE OF AUCTION MECHANISM

To start with, we considered an experiment in which the two mechanisms explicitly compete for bidders. This setting is different from the one investigated by Ivanova-Stenzel and Salmon, who measured the participants' willingness to pay for being in a two-person ascending bid auction instead of a two-person sealed bid auction. Our setting allows us to directly observe the interaction between the number of bidders, bidding behavior, entry decisions, and the resulting revenues. In a group of N bidders, two identical objects are auctioned off simultaneously, one via a sealed bid, and the other via an ascending bid auction. Each bidder must bid in only one of the auctions, and must decide in which of the two auctions to bid prior to finding out his value.[3] We conducted three different treatments in this basic setting: $N = 4$, 9, and 20 (to examine behavior in small, medium, and large groups of potential bidders). In the $N = 4$ and 9 treatments, each group participated in 40 rounds, and in the $N = 20$ treatment the group participated in 60 rounds. See the Appendix for a complete set of instructions and laboratory protocol.[4] Table 1 summarizes the results.

The basic conclusion is that even without any prior experience with either auction mechanism, bidders have a slight preference for ascending bid auctions – more bidders enter ascending bid auctions, and the differences are always highly significant. Fig. 1 provides information about the distribution of bidders between the two auctions.

This preference for ascending bid auctions translates to larger groups of bidders in ascending bid auctions. However, the difference in group sizes in not large enough to offset the propensity to overbid relative to the RNNE in sealed bid auctions, and therefore does not generally translate to higher revenues in ascending bid auctions. The average overbidding relative to RNNE decreases as the number of bidders increases, which is not surprising, since RNNE bid for large groups is very close to the value, and thus there is little room to overbid. The average overbidding (measured as the average bid/RNNE bid) is 1.253 for $N = 4$, 1.249 for $N = 9$, but is only 1.019 for $N = 20$, and consequently two things happen as N increases: (1) the overall variability associated revenues decreases due to the decrease in the variability associated with bidding behavior,[5] and (2) the differences in revenue shift from strongly favoring the sealed bid auction when $N = 4$ to slightly favoring the ascending bid auction when $N = 20$. Note from Table 1 that the standard deviation of revenue in the $N = 20$ treatment is about 40% of that in the $N = 4$ treatment.

Fig. 2 shows average bid/value over time in sealed bid auctions in the three treatments. There is a slight trend in some of the data. In the $N = 4$ treatment bid/value increases over time, and this trend is significant (ordinary least squares (OLS) estimate of the slope is 0.0037, $p = 0.0273$), in the $N = 9$ treatment bid/value does not change over time in any significant way (OLS slope estimate is 0.0008, $p = 0.3401$), and in the $N = 20$ treatment bid/value decreases over time (OLS slope estimate is -0.0018, $p = 0.0264$). The significant negative trend in the $N = 20$ treatment is entirely due to the two outliers in periods 7 and 9 – after the initial noise the bidding behavior is quite stable and very close to RNNE. Overall, there is no evidence of systematic learning.

Given the actual auction entry and bidding behavior observed, we conclude that revenues are significantly higher in sealed bid auctions than in ascending bid auctions when the number of potential bidders is small ($N = 4$). For the medium number of potential bidders ($N = 9$) the power of

Table 1. Entry, Revenue, and Profit Comparisons in the Free Choice Study.

Group size (N)	Number of sessions	Entry Percentage			Average Revenue (standard deviation)			Average Profit (standard deviation)		
		Ascending bid	Sealed bid	Difference	Ascending bid	Sealed bid	Difference	Ascending bid	Sealed bid	Difference
4	6	0.53	0.47	13.06%**	29.78 (25.22)	37.18 (25.71)	−19.92%**	11.64 (18.43)	8.81 (12.84)	32.15%
9	2	0.58	0.43	35.29%**	66.01 (22.80)	67.48 (16.80)	−2.17%	2.70 (7.26)	2.35 (4.97)	15.19%
20	1	0.54	0.46	17.79%**	83.00 (11.62)	82.92 (10.35)	0.10%*	0.50 (3.04)	0.45 (2.47)	11.43%
Overall		0.54	0.46	17.34%**	43.31	48.66	−10.99%**	6.16	4.76	29.44%*

*Statistically significant at 10% level.
**Statistically significant at 5% level.

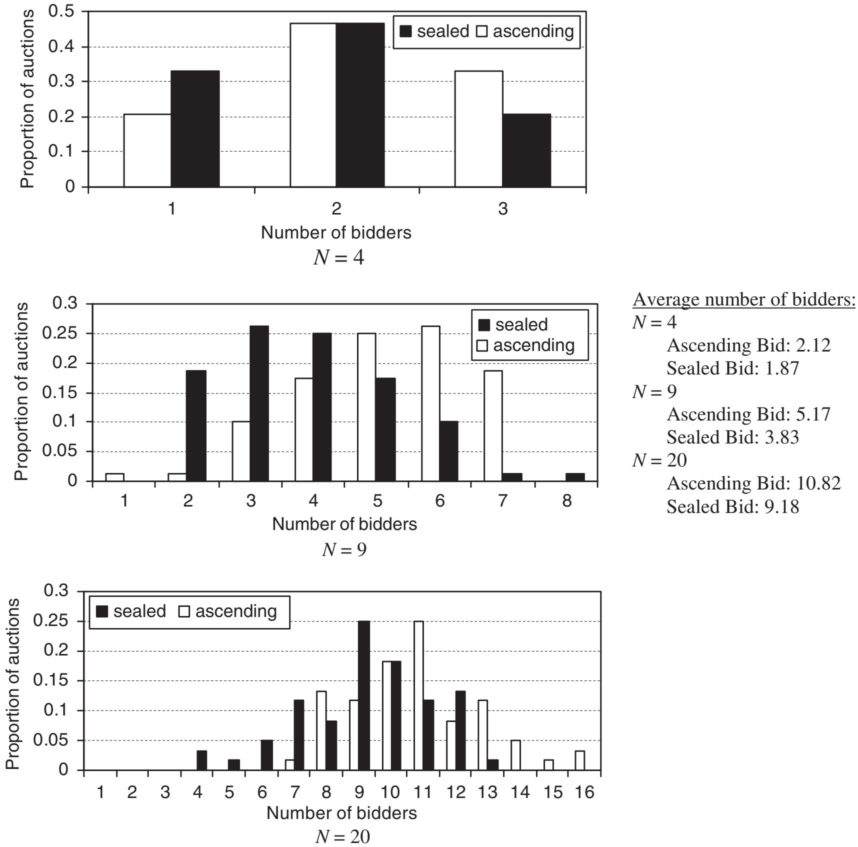


Fig. 1. Distribution of the Number of Bidders Between the Two Auctions.

our experiment to distinguish revenue is extremely low. The revenue in the two auctions is very close and a power test indicates that we would require around 80 independent observations (80 sessions of 9 bidders participating in 40 rounds) to be able to conclude that sealed bid auctions generate more revenue than ascending bid auctions. The story is different for the large number of potential bidders ($N = 20$), however. Here, the average revenue in the ascending auctions is slightly higher than the revenue in the sealed bid auctions, and moreover, the variability of this revenue is substantially lower. A power test estimates that we would have needed about 4 independent observations (in other words, 4 sessions of 20 potential bidders playing for

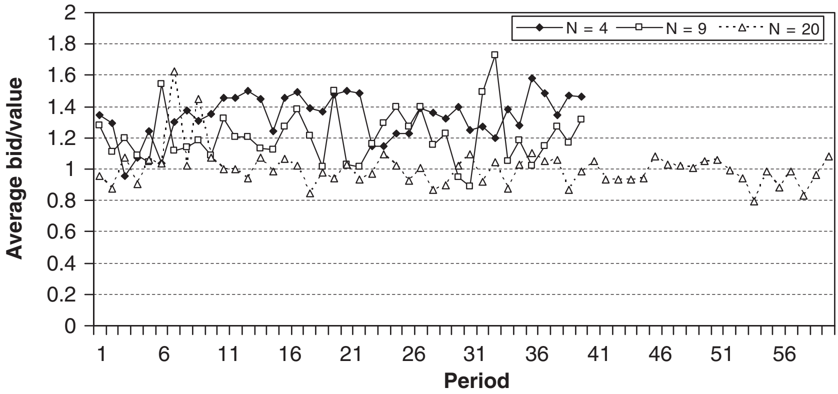


Fig. 2. Bidding Behavior Over Time.

60 rounds) to conclude that ascending bid auctions generate better revenue than sealed bid auctions do. This power test is of course based on the assumption that the one session-worth of data we collected in the $N = 20$ treatment is representative of the general population in terms of bidding behavior and auction preferences.

Bidder profits in ascending bid auctions are slightly higher, although typically not significantly so. They are about 30% higher overall, and this difference is weakly significant. These results are based on small sample sizes (24 participants in $N = 4$, 18 participants in $N = 9$, and 20 participants in $N = 20$; yielding very few truly independent observations). Possibly, bidder profits are in fact not different in the auction of two different types. This would be consistent with the hypothesis that bidders enter roughly to equalize expected utility. However, bidders in sealed bid auctions bid significantly more than expected profit maximizers should. This appears to be a stable pattern, consistent with previous experimental studies; bidders in sealed bids auctions do not bid as if they were maximizing expected profit. Therefore, we are uncomfortable suggesting that these same bidders base their entry decisions solely on expected profit.

STUDY 2: VALUING AUCTIONS AND LOTTERIES

The second study examines the question of how well the laboratory participants are able to evaluate and compare expected profits from different auctions. After all, for the same number of bidders, the ascending bid auction generates higher profits than it does to the sealed bid auction, and yet, participants are unwilling to pay close to this expected amount for the privilege to bid in the ascending bid auction instead of the sealed bid auction. Since the profit in the ascending bid auction is substantially more variable than in the sealed bid auction,[6] it is not clear whether participants are making an error in undervaluing the ascending bid auction, or whether the higher profit variability makes this mechanism truly less attractive.

In this study, bidders were asked to specify the minimum fixed payment amount they prefer to an auction. We used the Becker deGroot Marschak (BDM) procedure to elicit these valuations. After bidders specified, for each possible fixed payment amount, whether they prefer an auction or the fixed payment amount. The actual fixed payment amount was drawn randomly for each individual in a group of N individuals. The actual participation was then determined in accordance with the stated preferences. As a result, each auction included some unspecified number of bidders between 0 and N, and this number was not announced to the bidders. Consequently, bidders competed against an unknown number of opponents; making ascending bid auctions even more attractive (because ascending bid auctions have a dominant bidding strategy that is independent of the number of bidders). We used within-subject design, where half the subjects played 20 rounds of the ascending bid auctions followed by 20 rounds of sealed bid auctions, and the sequence was reversed for the other half of the subjects (see instructions in Appendix A.2). We summarize treatment 1 results in Table 2.

In treatment 1 we used $N = 5$, and to our surprise we found that valuations for the two auctions were virtually identical and not statistically different (one-sided t-test, $p = 0.2679$). This is in spite of the fact that bidder profits in ascending bid auctions were almost double the profits than in sealed bid auctions (one-sided t-test, $p = 0.0001$). It appears that subjects are valuing the ascending bid auction accurately, but are severely overvaluing the sealed bid auction, and there are several potential explanations for these results. One potential explanation may have to do with the overconfidence bias (see e.g. Lichtenstein and Fischhoff, 1977), a well-established phenomenon in the psychology literature. It may be that people overestimate their probability of winning the sealed bid auction. Another potential explanation may have to do with the fact that subjects do not fully understand the BDM procedure (although this would not explain especially the poor calibration of the sealed bid auctions). Plott and Zeiler (2003) show that the "willingness-to-pay or willingness-to-accept" gap disappears when the subjects are given extensive training in the use of the BDM procedure. We did not go as far as Plott and Zeiler, 2003 in training subjects in the use

Table 2. Results summary of Treatment 1 in Study 2.

Subject No.	Ascending Auction		Sealed Bid Auction	
	Valuation	Profit	Valuation	Profit
1	11.83	10.73	2.98	5.13
2	7.58	9.67	10.18	2.65
3	12.88	16.26	12.00	6.33
4	6.85	9.91	15.00	9.11
5	10.23	13.75	7.73	5.25
6	9.85	8.67	10.48	6.65
7	10.80	8.25	7.05	7.04
8	9.20	5.74	8.30	3.71
9	10.63	13.46	10.05	7.73
10	10.30	14.38	9.08	4.33
Overall	10.01	11.08	9.28	5.79

of the BDM procedure, but in an attempt to minimize the noise due to the misunderstanding of the BDM procedure, we did provide an example in the instructions (see Example 1 in Appendix A.2) designed to illustrate why truthful revelation of preferences is the correct way to proceed. We worked through this example (as well as all other examples in the instructions) using PowerPoint slides, and additionally provided participants with the following advice, both in written instructions and in the presentation:

> Many people prefer the auction if the fixed amount is below some level, and prefer the fixed amount if it is above that level. If you agree that this is logical, then make sure your indicated preferences follow this pattern.

Although some confusion probably remained, we do not think that it can explain the data in Table 2.

In treatment 2, subjects were provided with a history of the outcome of "100 auctions from a session we conducted earlier." This history included 100 actual sets of value, bid or profit combinations from treatment 1, as well as summary statistics that included average profit. This made little difference to subjects' ability to correctly evaluate the auctions, however. Evaluations of the two mechanisms were still quite close and not statistically significant (one-tail t-test, $p = 0.1722$). The lack of significance could be due to a small sample size (10 subjects), and at least qualitatively the movement is in the right direction. Of the 10 subjects, 4 valued the ascending auction significantly higher, 3 valued the sealed bid auction significantly higher, and the remaining 3 valued them the same.

In treatment 3 we replaced the two auctions by lotteries that match the auctions' expected profits and the variance of profits (calculated at the RNNE). The sealed bid auction was replaced by a lottery with a 25% chance of winning 20 tokens, and the ascending bid auction was replaced by a lottery with a 25% chance of winning 50 tokens. Subjects were again provided with the history of 100 outcomes of each lottery that included summary statistics. In this treatment, subjects were able to distinguish the values of the two lotteries, and valued the lottery with a 50-token prize at about two times the value of a lottery with a 20-token prize. Summary of treatments 2 and 3 is displayed in Table 3. The column labeled *t*-test contains the one-sided *p*-value testing whether the subject's valuations for the ascending bid and the sealed bid auctions (or the high and low prize lotteries) were different. Note that not a single subject valued the high prize lottery lower than the low prize lottery, although 3 subjects valued them the same. Overall, the valuations of the two lotteries are significantly different (one tail *t*-test $p = 0.0021$).

Taken in isolation, the results from treatment 3 are not surprising, but this treatment was meant to provide a bridge between auctions and lotteries. The outcomes of the two auction mechanisms differ in two ways: (1) ascending bid auctions are more profitable than sealed bid auctions, and (2) the variance of profit conditional on winning is higher in ascending bid auctions than in the sealed bid auctions. The probability of winning itself is the same

Table 3. Summary of Treatments 2 and 3.

Subject No.	Auction			Lottery		
	Ascending	Sealed bid	*t*-test	High prize	Low prize	*t*-test
1	2.30	2.00	0.3763	13.85	6.95	0.0000
2	6.15	5.0	0.0240	11.25	5.80	0.0000
3	6.30	8.95	0.0000	14.00	9.80	0.0000
4	13.90	11.80	0.0000	0.00	1.00	0.5000
5	12.05	8.05	0.0000	14.10	9.15	0.0000
6	5.10	8.65	0.0000	1.00	1.00	0.5000
7	8.40	7.25	0.0840	8.65	6.50	0.0051
8	11.00	11.85	0.0055	4.70	5.35	0.1664
9	9.30	9.45	0.2405	12.90	7.50	0.0000
10	16.00	2.50	0.0000	12.45	8.45	0.0000
11				11.65	7.95	0.0000
12				11.65	12.2	0.1375
Overall	9.05	7.55	0.1722	9.68	6.80	0.0021

in the two auctions. We chose the specific lotteries in this treatment because they make it transparent that (1) the probability of winning is the same, and (2) the prize (expected profit conditional on winning) is higher in one than in the other. The fact that subjects are able to correctly value the two lotteries, but not the corresponding auctions, it is suggestive of the fact that the difficulty in evaluating and comparing auctions is related to the uncertainty about profit conditional on winning rather than about the probability of winning.[7]

Note that our result cannot be easily explained by ambiguity aversion. The lottery treatment removes the ambiguity about the size of the prize (profit conditional on winning) as well as about the probability of winning, so ambiguity aversion would suggest that both lotteries should be valued higher than the corresponding auctions. But Table 2 shows that the average value for the sealed bid auction is 7.55 and, contrary to the ambiguity aversion explanation, it is higher than the average reported value for the low prize lottery, which is 6.80 (the difference is not statistically significant; two-sided t-test $p = 0.6113$). The reported average value for the ascending bid auction is 9.05, and it is lower than the reported average value for the high prize lottery, which is 9.68, but the difference is also not statistically significant (two-sided t-test $p = 0.7522$).

The sample size in our study is small, (10 observations in treatment 1 and 12 in treatment 2) so we have to be cautious about drawing conclusions. To the extent that our data is suggestive, it seems to point towards the uncertainty about profit conditional on winning as being an important variable that cause difficulties for people in evaluating auctions. We can speculate, that the outcome of a sealed bid auction is the result of an explicit single decision, a subject makes the overconfident may cause people biased to over-value sealed bid auctions. Computing expected profit from an auction is not an easy task for a typical subject in an experiment, and having access to 100 past auction outcomes does not lead to the insight that if the future is like the past, then the ascending bid auction is twice as profitable as the sealed bid auction and consequently it should be valued higher.

DISCUSSION

The main conclusion from the studies reported in this paper is that it is extremely difficult for people to evaluate expected profits from auctions. Without understanding the auction theory and possessing the insight that this understanding provides most people are unable to correctly evaluate an auction, let alone compare two different auction mechanisms. The standard laboratory setting, where participants see a randomly drawn value, each period makes the problem even more difficult by impeding learning because subjects do not see outcomes of their actions in any systematic way. In some sense, the problem of how to bid with a high value may be very different from the problem of how to bid with a low value because, for example, with high values subjects may be focusing on earning the highest possible profit from winning the auction, while with low values subjects may be primarily focusing on winning the auction if possible, without much regard for the amount they win. Given the amount of variability in auction outcomes, it may well be that the standard procedure, where participants bid in 20–40 auctions during a session, with a different randomly-determined value for every period, does not provide sufficient experience for participants to understand how to value an auction.

In spite of the difficulties in evaluating gains from auctions, participants do exhibit a preference for ascending bid auctions, albeit a small one (Study 1). When asked to directly evaluate the two auction types, they are unable to discern the difference, although when historical profitability information is given they do move towards favoring the ascending bid auctions (Study 2).

In summary, the main difficulty in studying endogenous entry directly in the laboratory is the fact that in the standard setting, subjects experience great difficulties in discerning differences among auction mechanisms. This difficulty is likely due to (1) subjects' inability to understand the auction structure sufficiently to compute expected earnings a priori, and (2) the lack of systematic feedback standard laboratory auction settings afford, that makes it difficult for subjects to learn from experience. Our results suggest, that to be able to effectively test some theories in the laboratory (theories based on endogenous entry in auctions, specifically) researchers need to develop different laboratory environments – with more systematic feedback and more transparent strategies.

The studies presented here motivated a study in Engerlbrecht-Wiggans and Katok (2004) that was designed to further examine bidding behavior in sealed bid auctions, with a specific focus on looking at reasons why participants tend to overbid relative to the RNNE. In this study, one human participant bids against several automated competitors who have been programed to bid according to the RNNE, and the session lasted for 100 rounds. We found that letting participants keep their values for 20 periods substantially reduces the amount of overbidding, to the extent that, contrary to the risk aversion theory, this overbidding disappears at very high values. In fact, regardless of whether participants see the same values repeatedly or not, we found strong evidence that the bid functions are non-linear in values – participants bid a larger proportion of their values when the values are low-to-moderate than when they are high. The consistent findings of Ockenfels and Selten (2005) and Isaac and Walker (1985), is also inconsistent with the risk aversion model, publicly announcing that the second highest bid causes bids to go down overall.

NOTES

1. In ascending bid auctions, there is a tentative price that increases until no bidder is willing to bid any higher; the last bidder wins and pays an amount equal to the final price.
2. By ''sealed bid'' we mean ''sealed-bid first-price,'' namely an auction in which each bidder submits a sealed bid, the bids are opened, and the bidder who submitted the highest bid wins the object and pays the amount that he bid.
3. This entry mechanism is similar to Meyer, Van Huyck, Battalio, and Saving, (1992).
4. All sessions for studies described in this paper were conducted in the laboratory for economic management and auctions (LEMA) in the Smeal College of Business at Penn State during the Summer and Fall of 2003. The software was built using the zTree system (Fischbacher, 1999). Sessions lasted approximately 90 min and average earnings were \$25.
5. When N is large the average number of bidders in sealed bid auctions is large, and since most of the variability in bidding behavior is due to bidding above RNNE, this variability decreases simply due to the proximity of the RNNE to bidders' valuation.
6. Proposition 2.4 of Krishna (2002) predicts this. Table 1 shows that this happens in our data. Also note that as the number of players increases, one might expect the price in both auctions to converge to the competitive price, and indeed the difference in variability in our data shrinks as N increases.
7. Dorsey and Razzolini (2003) find that showing the subjects the probability of winning the auction causes the bids at high values to decrease and become closer to the RNNE bids. On the other hand, Engelbrecht-Wiggans and Katok (2004) find that when given extensive experience, subjects bid very close (and in some cases below) RNNE at high values. This observation is suggestive of a learning explanation for the Dorsey and Razzolini (2003) result.

REFERENCES

Cox, J. C., Smith, V. L., & Walker, J. N. (1988). Theory and individual behavior of first-price auctions. *Journal of Risk and Uncertainty*, *1*, 61–99.

Dorsey, R., & Razzolini, L. (2003). Explaining overbidding in first price auctions using controlled lotteries. *Experimental Economics*, *6*(2), 123–140.

Engelbrecht-Wiggans, R. (1987). Optimal reservation prices in auctions. *Management Science*, *33*, 763–770.
Engelbrecht-Wiggans, R. (1989). The effect of regret on optimal bidding in auctions. *Management Science*, *35*(6), 685–692.
Engelbrecht-Wiggans, R. (1993). Optimal auctions reconsidered. *Games and Economic Behavior*, *5*, 227–239.
Engelbrecht-Wiggans, R. (2001). The effect of entry and information costs on oral versus sealed-bid auctions. *Economics Letters*, *70*, 195–202.
Engelbrecht-Wiggans, R., & Katok, E. (2004). *Learning, regret, and risk aversion in first price auctions*, Working Paper, Penn State.
Fischbacher, U. (1999). *z-Tree – Zurich Toolbox for Readymade Economic Experiments – Experimenter's Manual*, Working Paper 21, Institute for Empirical Research in Economics, University of Zurich.
Harrison, G. W. (1989). Theory and misbehavior of first-price auctions. *American Economic Review*, *79*, 749–762.
Isaac, R. M., & Walker, J. M. (1985). Information and conspiracy in sealed bid auctions. *Journal of Economic Behavior and Organization*, *6*, 139–159.
Ivanova-Stenzel, R., & Salmon, T. C. (2004a). Bidder preferences among auction institutions. *Economic Enquiry*, *42*(14), 223–236.
Ivanova-Stenzel, R., & Salmon, T. C. (2004b). Entry fees and endogenous entry in electronic auctions. In: A. Herrmann, A. Kambil, B. F. Schmid, E. van Heck, Y. Xu, *EM – Electronic Markets* (Vol. 14, No. 3, 9).
Kagel, J. H. (1995). Auctions: a survey of experimental research. In: J. H. Kagel & A. E. Roth (Eds), *The Handbook of Experimental Economics* (pp. 501–585). Princeton, NJ: Princeton University Press.
Krishna, V. (2002). *Auction Theory* (pp. 22–23). London: Academic Press.
Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? The calibration of probability judgments. *Organizational Behavior and Human Performance*, *20*, 159–183.
McAfee, R. P., & McMillan, J. (1987). Auctions with entry. *Economics Letters*, *23*, 343–347.
Meyer, D., Van Huyck, J., Battalio, R., & Saving, T. (1992). History's role in coordinating decentralized allocation decisions. *Journal of Political Economy*, *100*, 292–316.
Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, *6*, 58–63.
Ockenfels, A., & Selten, R. (2005). Impulse balance equilibrium and feedback in first price auctions. *Games and Economic Behavior*, *51*, 155–170.
Plott, C. R., & Zeiler, K. (2003). *The willingness to pay/willingness to accept gap, the 'endowment effect' and experimental procedures for eliciting valuations*. Social Science Working Paper 1132, California Institute of Technology, February.
Vickrey, W. (1961). Counterspeculation, Auctions, and Competitive Sealed Tenders. *Journal of Finance*, *16*, 8–37.

APPENDIX A

Laboratory protocol in all studies went as follows: Subjects arrived, were seated at computer terminals and given a set of written instructions (see below). They were given approximately 10 min to read the instructions, after which point the instructions were read to them aloud by the experimenter. Examples were illustrated using PowerPoint slides that were displayed on the computer screens. Subjects were specifically told that the numbers in the examples have been chosen with the sole purpose of illustrating the rules of the game and are not to be taken as bidding advice. After instructions were read and participants had a chance to ask any clarifying questions, they signed the informed consent forms, the forms were collected, and the games commenced. At the end of the session, participants filled out receipts with their earnings, were paid their earnings in private and in cash, and left.

A.1. Study 1 Instructions ($N = 4$ treatments; differences with $N = 9$ and 20 treatments shown in italics)

A.1.1. Introduction

(The following two paragraphs are in the beginning of all instructions in these studies and differ only by the exchange rate. This is the standard beginning.)

> *You are about to participate in an experiment in the economics of market decision-making. You will earn money based on the decisions you make. All earnings you make are yours to keep and will be paid to you IN CASH at the end of the experiment. During the experiment, the unit of account will be experimental dollars. At the conclusion of the experiment, the amount of experimental dollars you earn will be converted into dollars at the conversion rate of 3 cents per experimental dollar. Your converted earnings plus a lump sum of five dollars ($5) will be paid to you in private.*
>
> *Do not communicate with the other participants except according to the specific rules of the experiment. If you have a question, feel free to raise your hand. I will come over to you and answer your questions in private.*

In this session you will participate in a sequence of 40 *trading periods* in which you and the other participants in the room will compete for a fictitious asset. This asset will be sold in two ways, the *Sealed Bid* auction, and the *Oral* auction, both of which are explained below. At the start of each trading period, each participant will select the auction type they prefer. Both auctions will then be separately conducted. You are allowed to participate in a different auction type during different trading periods.

There are 3 (*8 or 19 in the $N = 9$ and 20 treatments*, *respectively*) other participants in this session. During each trading period, you will only be matched with those participants who have chosen to participate in the same auction as you. You will not be told which of the other participants in the room are in the same auction during each trading period, and they will not be told that you are in the same auction with them. However, you will be told the number of bidders in your auction. Since participants are allowed to choose either auction in each trading period, it is possible that you may interact with another participant in more than one trading period. *What happens in your auction during any trading period has no effect on the participants in the other auction or in future trading periods.*

A.1.2. How Your Earnings are Determined

Your "resale value" for the asset will be assigned to you at the beginning of each trading period. Resale values may differ among individuals and *do not depend* on which auction you choose. Your resale value will be randomly drawn from a uniform distribution between 1 and 100. Each integer between 1 and 100 has an equal chance of being chosen. *You are not to reveal your resale values to anyone. It is in your best interest to keep this information private.* During each trading period your earnings from the asset purchase are equal to the difference between your *resale value* for the asset and the price you paid for the asset.

That is:

YOUR EARNINGS = RESALE VALUE – PURCHASE PRICE.

For example, if your resale value is 64 and you pay 30, then your earnings are

EARNINGS FROM THE ASSET = 64 – 30 = 34 experimental dollars.

If you purchase the asset at a price that is higher than your resale value, your earnings will be negative and you will lose money.

Your earnings from the current period as well as from all previous periods will be displayed on your screen at the end of each period. In any trading period, if you do not win your auction, you do not buy the asset and your earnings are zero for that period.

A.1.3. Sealed Bid Auction

Each buyer in the sealed bid auction submits one bid during a trading period by entering their bid amount into the computer and clicking the "Place Bid" button. You may only place integer valued bids (i.e. whole numbers) from 1–100.

After all bidders in the sealed bid auction have submitted their bids, the period is closed, and the asset is sold to the bidder with the highest bid for the asset. All bidders will see the outcomes of the trading period on their screen. If yours was the highest bid you will be told that you won, and you will be informed of your earnings for the period. If yours was not the highest bid, you will be told that you did not win, and your earnings for that period are 0.

To calculate your profit associated with a certain bid amount *provided* that you win the auction, click on the "Calculate Profit" button *after* you have entered a bid amount in the box provided. Clicking the "Calculate Profit" button does *not* submit your bid. You will need to re-calculate this potential profit amount whenever you enter a new bid amount. The running tally of the items you purchased, the price you paid, your value, and your profit will be displayed on your screen at the end of each auction.

Important Note. If more than one bidder places the same winning bid, then a sole winner of the asset will be randomly selected from them. (Examples were scaled up appropriately for the $N = 9$ and 20 treatments)

Example 1.1. In a given trading period, suppose the Bidders 2 and 3 have chosen to participate in the sealed bid auction, and that they have the following resale values for the asset:

Bidder 2 has the resale value of 85

Bidder 3 has the resale value of 80

If the following bids are entered:

Bidder 2 bids 72

Bidder 3 bids 65

Then the asset is sold to Bidder 2 for 72 experimental dollars.

Bidder 2 earns $85-72 = 13$ experimental dollars for this trading period, and bidder 3 earns 0 experimental dollars for this trading period.

A.1.4. Oral Auction

In this auction you bid on the asset by doing nothing, and you indicate when you wish to STOP bidding by clicking a button. The unit price of the asset will start at 0 and will increase by 1 experimental dollar every 0.5 sec.

When the price is as high as what you are willing to pay, click the 'Stop' button. Clicking this button guarantees that you will not win the asset if the price increases further. The auction ends when only one bidder remains active. The number of active bidders at any point of time is displayed on your screen. The last remaining bidder wins the asset at a price that is *1* *experimental dollar higher* than the price at which the second to last bidder stopped bidding.

The profit that you could expect to earn if you were to be the last active bidder will be displayed on your screen. This amount is updated each second, as the price of the asset increases.

Important Note. If there is more than one eligible bidder and all of the eligible bidders stop bidding at the same price, then a sole winner of the asset will be randomly selected from them. In this case, the sole winner will pay a price *equal* to the price at which he/she chose to stop.

Example 2.1. Suppose your resale value for the asset is 70 and you click the 'Stop' button when the price is at 60. At this point there is only one other eligible bidder (that is, there is only one bidder remaining who did not click the 'Stop' button). At this point, the auction ends and the only remaining bidder wins the asset at a price of 61. Your earnings are zero in this trading period since you did not win the asset.

Example 2.2. Suppose your resale value for 1 unit is 80, and the auction ends when the price reaches 74 without you clicking the 'Stop' button. In this case you win the asset, and you pay 75 for the asset. Your earnings are $80-75=5$.

A.1.5. Ending the Experiment

(This paragraph, up to the exchange rate figure and the number of rounds, is identical in all our instructions, with a standard ending)

> *At the end of the experiment, your earnings from all 40 auctions will be totaled and converted to dollars at the rate of 3 cents per experimental dollar. You will be paid this total amount plus an additional five dollar participation fee, in private and in cash. Your total earnings will be displayed on your computer screen at the end of the session.*

A.2. Instructions for study 2 treatment 1

A.2.1. Introduction (Standard beginning)

This session will consist of 40 periods. In the beginning of every period you and the other participants will decide whether to bid in an auction where bidders compete for a fictitious asset, or to receive a fixed payment.

There will always be 5 participants in each group, but the composition of the groups may change every period, so auctions may consist of different people in different periods. You will not be told which of the other participants in the room are in your group, and they will not be told that you are in their group. *What happens in your group during any period has no effect on the participants in the other groups or in the future periods.*

A.2.2. How Your Earnings are Determined

A.2.2.1. Deciding whether to bid in an auction. At the start of each period, you will be asked to decide between bidding in an auction and receiving a fixed payment. Your fixed payment amount is chosen randomly before you make your decision and is between 1 and 20 tokens. To make your decision, you will indicate whether you prefer the auction or the fixed payment for each possible fixed payment amount. After you submit your decisions, your actual fixed payment amount will be announced, and you will receive either that fixed payment or you will go on to bid in the auction, in accordance with your preference. Fixed payment amounts will be different for all participants.

Note that it is in your best interest to be careful to decide what you actually prefer, and to state those preferences truthfully.

> **Example 1.** Suppose you would actually prefer a fixed payment of 18 tokens to bidding in the auction, but you report that you prefer the auction over a fixed payment of 18 tokens. Then, if the fixed payment happened to be 18 tokens, you would not get what you actually preferred. Similarly, suppose that you actually prefer the auction to a fixed payment of 2 tokens, but you report that you prefer the fixed payment. Then, if the fixed payment happened to be 2 tokens, you would again not get what you actually preferred.

Hint: Many people prefer the auction if the fixed amount is below some level, and prefer the fixed amount if it is above that level. If you agree that this is logical, then make sure your indicated preferences follow this pattern.

A.2.2.2. Earnings from an auction. If you bid in an auction during a particular period, then your "resale value" for the asset will be assigned to you at the beginning of this period. Resale values differ among individuals. Your resale value will be randomly drawn from a uniform distribution between 1 and 100. Each integer between 1 and 100 has an equal chance of being chosen. *You are not to reveal your resale values to anyone. It is in your best interest to keep this information private.* During each auction your earnings from the asset purchase are equal to the difference between your *resale value* for the asset and the price you paid for the asset.

That is:

YOUR EARNINGS = RESALE VALUE − PURCHASE PRICE.

For example, if your resale value is 64 and you pay 55, then your earnings are

$$64 - 55 = 9 \text{ tokens.}$$

If you purchase the asset at a price that is higher than your resale value your earnings will be negative and you will lose money.

Your earnings from the current period as well as from all previous periods will be displayed on your screen at the end of each auction. At any period, if you do not win your auction, you do not buy the asset and your earnings are zero for that period.

A.2.2.3. How the auction works. This session will include two different auction mechanisms: the Sealed Bid Auction and the Ascending Bid Auction. Half the participants in this room will bid in the Sealed Bid Auctions in rounds 1–20 and the Ascending Bid Auction in rounds 21–40. The other half of the participants in this room will bid in the Ascending Bid Auction in rounds 1–20 and in the Sealed Bid Auction in rounds 21–40.

Sealed Bid Auction: Each bidder in the auction submits one bid during a trading period by entering the bid amount into the computer and clicking the "Place Bid" button. You may only place integer valued bids (i.e. whole numbers) from 1 to 100.

After all bidders have submitted their bids, the period is closed, and the asset is sold to the bidder with the highest bid for the asset. All bidders will see the outcomes of the trading period on their screen. If yours was the highest bid you will be told that you won, and you will be informed of your earnings for the period. If yours was not the highest bid, you will be told that you did not win, and your earnings for that period will be 0.

To calculate your profit associated with any bid amount *provided* that you win the auction, click on the "Calculate Profit" button *after* you have entered a bid amount in the box provided. Clicking the "Calculate Profit" button does *not* submit your bid. You will need to re-calculate this potential profit amount whenever you enter a new bid amount.

Example 2. Suppose your value for the asset is 71, and there are two other bidders in the auction who have values of 60 and 68. Suppose you enter the bid of 65 and your two competitors enter the bids of 52 and 63. Then you win the auction because your bid of 65 is higher than the other two bids of 62 and 63. You win the asset, pay 65 for it, and earn $71 - 65 = 6$ tokens.

Example 3. Now suppose that the auction is the same as in Example 2, but you bid 60. In this case the bidder who bid 63 wins the asset, and you do not, because your bid of 60 is below the highest bid of 63. The winning bidder earns $68 - 63 = 5$ tokens, and you earn 0 tokens.

Ascending Bid Auction: The price in each auction starts at 1, and gradually increases. Every time the price goes up by 1 you need to make a decision on whether you wish to continue to bid in the auction, or to stop. To continue to bid you do not need to do anything. When you wish to stop bidding because the price has reached the highest price you are willing to pay, click on the "Stop Bidding" button.

After all but one bidder has stopped bidding, the last bidder wins the auction and pays the current price for the asset. Note that you will not be told how many members of your group have chosen to compete in the auction with you.

Example 4. Suppose there were 3 bidders in the auction and your value for the asset is 71. The price starts at 1 and gradually increases by 1. Suppose one of your competitors stop bidding when the price gets up to 52, and another of your competitors stop bidding when the price gets to 65, then the auction ends at the price of 66. You win the asset, pay 66 for it, and earn $71 - 66 = 5$ tokens.

Example 5. Now suppose that instead in the same auction, after one of your competitors stopped bidding at 52, you stop bidding when the price gets to 60, then the auction stops at the price of 61 and you do not win the auction (even though your value was 71). The other bidder wins the auction and pays 61.

In the event of a tie, a winner will be determined randomly from the highest bidders.

A.2.3. Additional Information to Help You Make Your Decision

We have provided you a table that shows the actual outcomes of 100 auctions from a session we conducted last week. The auctions were identical to the ones in which you will participate today. For each of the 100 auctions we show the value the participant had in this auction, and the profit for this auction. Also at the bottom of the page we show the average value and the average profit for this auction. Note that there is one table for the Sealed Bid Auction and a different table for the Ascending Bid Auction *(Standard ending)*.

A.3. Instructions for Study 2 Treatments 2 (Auctions) and 3 (Lotteries)

These are the instructions used for the lottery treatment. The auction treatments differed in the description of the games, where instead of "Lottery A" description we had a description of a sealed bid auction and in place of "Lottery B" description we included a description of an ascending bid auction (see Appendix A.1 for descriptions of these auctions used). The differences in treatment 2 are marked in *italics*.

A.3.1. Introduction (Standard beginning)

This session will consist of 40 periods. In the beginning of every period you will decide whether to participate in a lottery, or to receive a fixed payment.

A.3.2. How Your Earnings are Determined

A.3.2.1. Deciding Whether to Participate in the Lottery (auction in treatment 2). At the start of each period, you will be asked to decide between participating in a lottery and receiving a fixed payment between 1 and 20 tokens. To make your decision, you will indicate whether you prefer the lottery or the fixed payment *for each possible fixed payment amount*. After you have made this decision, we will spin a roulette wheel that has numbers 1–100 on it. All participants will observe the roulette wheel and the number that comes up will determine the actual fixed payment amount as follows: Numbers 1–5 correspond to 1, 6–10 correspond to 2, 11–15 correspond to 3, and so on. The complete conversion chart is displayed on the white board at the front of the room.

At this stage, one of two things will happen:

(1) If the random number spun on the wheel corresponds to a fixed payment amount for which you have indicated that you prefer the fixed payment to the lottery, then you will receive the fixed payment that corresponds to the random number spun on the wheel. You will not enter the lottery for that period.
(2) If the random number spun on the wheel corresponds to a fixed payment amount for which you have indicated that you prefer the lottery to the fixed payment, then you will not receive any fixed payment but will be entered into the lottery. The details of the lottery are described later.

Note that it is in your best interest to be careful to decide what you actually prefer, and to state that preference truthfully. The reason for this is illustrated in the following example.

Example 1. Suppose you would actually prefer a fixed payment of 18 tokens to the lottery, but you report that you prefer the lottery over a fixed payment of 18 tokens. Then, if the fixed payment happened to be 18 tokens, you would not get what you actually preferred. Similarly, suppose you actually prefer the lottery to a fixed payment of 2 tokens, but you report that you prefer the fixed payment. Then, if the fixed payment happened to be 2 tokens, you would again not get what you actually preferred.

Hint: Many people prefer the lottery if the fixed amount is below some level, and prefer the fixed amount if it is above that level. If you agree that this is logical, then make sure your indicated preferences follow this pattern.

A.3.3. The Lottery (in treatment 2 this section was replaced with descriptions of the two auctions).

This session will include two different lotteries: Lottery A and Lottery B. Half of the participants in this room will play Lottery A in rounds 1–20 and Lottery B in rounds 21–40. The other half of the participants in this room will play Lottery B in rounds 1–20 and Lottery A in rounds 21–40.

Lottery A: In this lottery, there is a 25% chance of winning 20 tokens and a 75% chance of winning 0 tokens.

The outcome of the lottery is decided by spinning the wheel. If the number spun is between 1 and 25, then you will earn 20 tokens. Otherwise, you earn zero tokens.

Lottery B: In this lottery, there is a 25% chance of winning 50 tokens and a 75% chance of winning 0 tokens.

The outcome of the lottery is decided by spinning the wheel. If the number spun is between 1 and 25, then you will earn 20 tokens. Otherwise, you earn zero tokens.

Example 2. After everyone has entered their preferences into the computer, the wheel is spun. Suppose that the number spun is 82, which corresponds to a fixed payment of 18 tokens. Suppose you indicated that you prefer a fixed payment of 18 tokens to participating in the lottery, then you will earn a fixed payment of 18 tokens in this period. You will not participate in the lottery in this period.

Example 3. Suppose the number spun on the wheel is 26, which corresponds to a fixed payment of 5 tokens. Suppose you indicated that you prefer the lottery to a fixed payment of 5 tokens, then you will enter the lottery and you will not receive a fixed payment in this period.

In the lottery, suppose the number spun on the wheel is 24. If you are in Lottery A, you will earn 20 tokens in this period. If you are in Lottery B you, will earn 50 tokens in this period.

Example 4. Suppose in example 3, the number spun on the wheel in the lottery is 56. Then regardless of which lottery is being run, your earnings in the period will be 0 tokens.

A.3.4. Additional Information to Help You Make Your Decision

We have provided you a chart that shows 100 past lotteries, their outcomes, the total number of wins (out of 100) and the average lottery profits. Note that there is one chart for Lottery A and a different chart for Lottery B.

A.3.4.1. How the Session will be Conducted. In the beginning of every period all participants will report for fixed payments of 1–20, whether they prefer this fixed payment or the lottery. You will see on your screen the description of the lottery for the current period (the probability of winning the prize, which is always 25%, and the size of the prize, which will be either 20 or 50). You will always see on your screen the average profit for your lottery.

After everyone enters their preferences into the computer, we will spin the wheel, and everyone will observe the resulting number between 1 and 100. We will then convert this number to the fixed payment of 1 to 20, as indicated on the white board at the front of the room. If for the actual fixed payment drawn you have indicated that you prefer the lottery, you will enter the lottery, otherwise, you will receive the fixed payment. The outcome of your decision will be displayed on your computer screen. After you observe this outcome, please click the "Continue" button.

We will then conduct the lottery by spinning the wheel again. All participants will observe the second random number. If this number is 25 or below, the lottery pays the prize of either 20 or 50 tokens, depending on its type. If the number is above 25, the lottery pays 0. All participants, regardless of whether they are participating in the lottery or receiving the fixed payment, will observe the lottery and will see the outcome of the lottery on their computer screens.

After the first 20 periods your lottery will change. If you played Lottery A in periods 1–20, you will play lottery B in periods 21–40. If you played Lottery B in periods 1–20, you will play lottery A in periods 21–40. The size of the lottery prize, 20 or 50 tokens, the probability of winning the lottery of 25%, and the average lottery profit, will always be displayed on your screen at the beginning of the period *(Standard ending)*.

This page intentionally left blank