

Are Two Heads Better than One?

Team versus Individual Play in Signaling Games*

David J. Cooper
Department of Economics
Weatherhead School of Management
Case Western Reserve University

John H. Kagel
Department of Economics
Ohio State University

December 22, 2004

Abstract: We compare individuals with two-person teams in signaling game experiments. Teams consistently play more strategically than individuals, and generate positive synergies in more difficult games, beating a demanding “truth wins” norm. The superior performance of teams is most striking following changes in payoffs that change the equilibrium outcome. Individuals play *less* strategically following the change in payoffs than inexperienced subjects playing the same game. In contrast, the teams exhibit *positive* learning transfer, playing *more* strategically following the change than inexperienced subjects. Dialogues between teammates are used to identify factors promoting strategic play.

Key words: teams, signaling games, learning transfer

JEL classification: C72, C92, D82, L12.

* Research support from the National Science Foundation (SBR-9809538) is gratefully acknowledged. Helpful research assistance was provided by Alex Brown, Bethia Cullis, Guillaume Frechette, John Lightle, and Todd Schoellman. We have received helpful comments from Eric Bettinger, John Ham, David Laibson, James Peck, Alvin Roth, seminar participants at Carnegie-Mellon University, Emory University, Harvard University, Ohio State University, University College London, and University of Wyoming as well as the referees and editor of this journal. Any opinions, findings, and conclusions or recommendations in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Many economic decisions are made within a team or group framework with two or more economic agents consulting with each other in deciding what course of action to take. For example, corporate bid teams, often in conjunction with outside consultants, determined bidding strategies in the spectrum (air wave) rights auctions. U.S. monetary policy is determined by the twelve member Open Market Committee of the Federal Reserve Bank. More generally, virtually all significant strategic decisions by corporations are made within a group or team framework. In contrast, much of economic theory and game theory, and most experimental investigations of these theories, make no distinction between strategic decisions made by teams versus individuals. As a result there is, potentially, a significant hole in our understanding of large areas of economic behavior. If there are major differences between individuals and teams in important economic settings, direct extrapolation of individual level research to team performance may be strikingly inconsistent with observed behavior. The results presented here, which find substantial differences between the behavior of individuals and teams in signaling games, indicate these concerns are likely to be well founded.

We report experiments comparing the behavior of individuals versus freely interacting two-person teams in Paul Milgrom and John R. Roberts' (1982) entry limit pricing game. Strategic play in this game revolves around an incumbent monopolist attempting to deter entry by signaling it will be a tough competitor for a potential entrant. Specifically, strategic play takes place through limit pricing, the choice by incumbents of greater quantities (and lower prices by extension) than would prevail in the absence of asymmetric information. Past experiments with signaling games shows that equilibrium play emerges only gradually, requiring a number of replications of the game before anything approaching an equilibrium emerges (Jordi Brandts and Charles A. Holt, 1992; David J. Cooper, Susan Garvin and John H. Kagel, 1997a; Cooper,

Kagel, Wei Lo, and Qing Liang Gu, 1999).

We study three different versions of the limit pricing game which vary in the difficulty of learning to play strategically. We focus on how fast strategic play (e.g. limit pricing) develops for teams versus individuals, as well as identifying possible qualitative differences in the learning/adjustment process between teams and individuals. For all three treatments the equilibrium towards which play converges is the same for teams and individuals, but teams learn to play strategically faster than individuals. Impressively, in the more difficult games teams beat a demanding “truth wins” norm drawn from the psychology literature.

The differences between teams and individuals are most striking following changes in payoffs that affect the equilibrium outcome. In this cross-over treatment, subjects first play a game that supports both pure strategy pooling and separating equilibria, with play converging on a pooling equilibrium, and are then switched to a game in which the only pure strategy equilibria are separating. For individuals, experience in the game with pooling equilibria retards adjustment to the separating equilibria following the change in games (compared to individuals who play the same game but with no prior experience), so that there is *negative* learning transfer between games. However, for teams, experience in the game with pooling equilibria enhances adjustment to the separating equilibrium (compared to teams with no previous experience), yielding *positive* learning transfer between games. Thus, teams not only learn to play strategically faster than individuals but also follow a different learning/adjustment process than individuals.

An important methodological innovation of our procedures is the recording and coding of dialogues between teammates as they coordinate their decisions. These dialogues provide a direct window into teams’ learning processes. Analysis of the dialogues indicates that a critical

step in monopolists' learning to play strategically is putting themselves in the entrant's shoes, reasoning from the entrant's point of view to infer likely responses to their choice as a monopolist. Statements to this effect are (i) among the most commonly coded statements prior to limit pricing, (ii) among the best predictors of whether teams will continue to limit price after their first attempt to do so, and (iii) are much more common immediately following the change in payoffs in the cross-over treatment compared to inexperienced teams playing the same game. Thus, the kind of interactive reasoning that underlies much of game theory has clear empirical validity in our data.

Beyond the importance of our results for economics and game theory, our work also speaks to the large psychology literature on group decision making by consciously "common-purpose" groups seeking consensus on how to solve a specific problem. There are two distinct branches to this literature, one dealing with judgmental questions and another dealing with "eureka" type problems requiring no special information to solve, but having solutions that tend to be self-confirming if discovered. The latter is closest in spirit to our game. For such problems psychologists apply a "truth wins" norm in judging whether or not teams are superior to individuals. Intuitively, a team should be no less likely to solve a problem than its most able member would be *acting alone*. By comparing the performance of freely interacting teams with this norm, psychologists identify the presence of positive, negative, or zero synergies for teams.

The rich psychology literature on team versus individual play in eureka problems consistently finds that teams typically fall well short of the truth wins norm (James H. Davis, 1992). In contrast, for two out of our three treatments teams meet or beat the truth wins norm, indicating strong positive synergies from the teams treatment. The concluding section of the paper identifies important differences in experimental procedures, and between games with

strategic interactions like we study versus games against nature that psychologists typically study, that likely account for the striking differences between our results and those in the psychology literature. We also compare our results to the limited literature on team versus individual play previously reported in the economics literature.

The paper proceeds as follows: Section I characterizes the structure of the limit-pricing game and the equilibria for the games played. Section II defines what we mean by strategic play and relates learning to play strategically in our experiment to the truth wins norm. Section III outlines our experimental procedures. Results are reported in Section IV. Section V reports on the team dialogues for insights into the leaning process and to fully justify our definition of strategic play. The last section of the paper summarizes our main findings and relates our results to the psychology and economics literatures on teams versus individuals and cross-game learning.

I) The Limit Pricing Game: The games studied here are based on Milgrom and Roberts' (1982) entry limit pricing model. For our purposes, the industrial organization implications of the model are of secondary importance. We therefore employ a very stylized version of the model, focusing on the signaling aspects of the game.

A. Structure of the Game: The limit pricing game is played between an incumbent monopolist (M) and a potential entrant (E). The game proceeds as follows: (1) M observes its type, high cost (MH) or low cost (ML). The two types are realized with equal probability with this being common knowledge. (2) M chooses one of seven output levels. M's payoff, shown in Table 1a, is contingent on its type, the output level chosen, and E's response. (3) E sees M's output, but not M's type, and either plays IN or OUT. The asymmetric information, in conjunction with the fact that it is profitable to enter against MHs, but not against MLs, provides an incentive for

strategic play (limit pricing) by Ms. E's payoff depends on M's type and on E's decision, not on M's choice. As a treatment variable, two different payoff tables, Tables 1b and 1c, were used for Es. These represent "high cost" and "low cost" Es respectively. Only one of these tables was in use at any given time.¹

[Insert Table 1 here]

Three features of Table 1a capture the main strategic elements confronting Ms. First, all else being equal, Ms are better off if Es choose OUT rather than IN. Second, reflecting their lower marginal costs, MLs generally prefer higher outputs than MHs. This can be seen in Ms' payoffs should they ignore the effect of their choices on E's behavior -- MLs would choose 4 as opposed to 2 for MHs. We refer to choice of 2 by MHs or 4 by MLs as the "myopic maxima." Third, 6 and 7 are dominated strategies for MHs, but not for MLs. At these outputs MLs can, in theory, perfectly distinguish themselves from MHs.

For either the high or low cost entrant payoff table, it always pays for Es to play IN when M is known to be an MH and OUT against an ML. However, given the prior probability of the different M types, the expected value of OUT is greater than IN in Table 1b (250 vs. 187) and the expected value of IN is greater than OUT in Table 1c (350 vs. 250).

B. Equilibrium Predictions: For Tables 1a and 1b (games with high cost Es), there exist multiple pure strategy pooling, as well as separating, (sequential) equilibria. Pure strategy pooling equilibria occur at outputs 1-5. To understand how these work, consider the pooling equilibrium where both MHs and MLs choose 3. Given the prior probabilities over M's type, E's expected value of OUT is greater than IN following a choice of 3. Sequential equilibrium puts no

¹ Payoffs are given in the experimental currency "francs." Francs were converted to dollars with one franc equal to \$.0025. Headings in Tables 1 and 2 have been changed to match the exposition in the text.

restrictions on E's beliefs following M's choice of an output level that is supposed to be chosen with zero probability in equilibrium. Therefore, it is admissible for E to believe that any deviation from 3 involves an MH type with sufficiently high probability to induce choice of IN. Given these beliefs and the resulting actions by Es, both MHs and MLs achieve higher profits at 3 rather than deviating to some other output level. The other pooling equilibria are constructed in a similar fashion. The pooling equilibria at outputs 3-5 involve limit pricing by MHs - choosing higher outputs (and hence lower prices) than would prevail under full information about Ms type.

Two pure strategy separating equilibria also exist. In both of these MHs choose 2 and are always entered on; MLs either always choose 6 or 7 and never incur entry. With MLs choosing 6 or 7, MHs cannot profitably imitate them since 2 dominates 6 and 7 for MHs. Out-of-equilibrium beliefs supporting these equilibria are that any deviation involves an MH type with sufficiently high probability to induce entry. These separating equilibria involve limit pricing by MLs.

For Tables 1a and 1c (the limit pricing game with low cost Es) the expected value of IN is greater than OUT if both types choose the same output level. This destroys any pure strategy pooling equilibrium, with the only pure strategy equilibria being the two separating equilibria just described.

As is typical of signaling games, the limit pricing game suffers from an overabundance of equilibria. To obtain sharper predictions, one must apply equilibrium refinements. The intuitive criterion (In-Koo Cho and Kreps,1987) reduces the equilibria in games with high cost Es to pooling at 4 or 5, and the efficient separating equilibrium with MLs choosing 6. For games with low cost Es, only the efficient separating equilibrium with MLs choosing 6 survives the intuitive

criteria.

Our experiments employ three different versions of this limit pricing game: (1) games with only high cost Es so that both pure strategy separating and pooling equilibria exist, (2) games with only low cost Es for which the only pure strategy equilibria are separating, and (3) a crossover treatment in which subjects play the game with high cost Es first, in which play reliably converges to the pooling equilibrium at 4, after which they are switched to the game with low cost Es. This last treatment tests for learning across games.

C. Defining strategic play: In the analysis of the experimental data that follows we focus on the development of strategic play for Ms. Strategic play is defined as MHs choice of outputs 3-5 in games with high cost Es and MLs choice of outputs 5-7 in games with low cost Es. These are clearly not the only possible definitions of strategic play we could employ since, in terms of the theory, strategic play depends critically on Ms' beliefs about Es' responses to their actions.

However, based on the dialogues from the team treatment, and substantial circumstantial evidence from other experiments, it is clear that Ms initial choices (which are overwhelmingly at output 2 for MHs and 4 for MLs) involve attempts to maximize their payoffs *ignoring* the effect of their choices on Es potential responses. Once Ms begin to consider the effect of their choices on Es responses, their choices almost exclusively involve MHs choosing 3-5 in games with high cost Es and MLs choosing 5 or 6 in games with low cost Es.

For example, consider games with high cost Es where MHs initial choices are focused on 2 for both teams and individuals (see Figures 1 and 2 reported in Section IVa below). The team dialogues, analyzed in detail in Section V below, make it clear that MHs are essentially not thinking about Es' responses when making these choices. The following dialogue is a prime example of this:

“I think we should pick 2. What about you? It is the highest for both [meaning the highest payoff for both choices that Es have].” [brackets added]

“I agree.”

“OK then. 2 it is.”

“It’s the highest for both.”

In the coding scheme we have developed for quantifying the dialogues between teammates, dialogues like the preceding are coded under the category “Myopic choice as a monopolist.” There is nothing in this dialogue (and other dialogues in the “myopic choice” category) where MHs account for the fact that their choice will influence Es’ beliefs about their types and/or impact on Es actions. Even though the high entry rate on 2, as compared to 3, 4 or 5, makes it profitable for MHs to limit price from the earliest rounds of the experiments, the myopic choice category is the second highest coded category prior to a first attempt of MH teams to play strategically (choose 3, 4 or 5).² In other words, MHs aren’t initially choosing 2 because they (incorrectly) anticipate that it will lead to favorable responses by Es, but rather because they fail to consider that their choice will impact on E’s response.

Additional circumstantial evidence pointing to the fact that *initially* most Ms fail to consider Es responses to their choices comes from several sources. Most tellingly, in both earlier experiments with the limit pricing games as well the current dataset, there are no significant differences between MLs early plays of the game regardless of whether they faced high or low cost Es, even though entry rates on 4 are far higher when playing against low cost Es. Further, there is substantial evidence from a number of other experiments that subjects fail to correctly account for their rivals actions in formulating their initial choices, and only gradually learn to do

² The highest coded category involves taking notice of the feedback information about the population choices of Ms as well as Es responses to same.

so.³

II) Connections to Team Play in the Psychology and Economics Literatures

A. Connections to the psychology literature on teams: The substantial psychology literature on team versus individual subject play distinguishes between judgmental tasks and so called “eureka” problems.⁴ Judgment tasks involve settings where either there isn’t a “correct” action, or the correct action is highly unlikely to be discovered by an untrained subject. For example, there are numerous studies comparing the attitudes of teams and individuals towards risk. While the choices of teams systematically differ from those of individuals, they cannot meaningfully be termed more or less correct.⁵ By way of contrast, eureka problems have a correct solution (or solutions). While this solution may be difficult to discover, it is self-confirming once discovered and can easily be demonstrated to others. Logic problems like the Tower of Hanoi are good examples of eureka problems. This classic puzzle, while challenging for a novice, is solved by a simple algorithm that can be explained in a few sentences. In other words, once the puzzle has been solved it is easy to show others that it has been successfully solved.

We argue that in signaling games, as subjects actually play the game, the discovery of strategic play corresponds closely to solving a eureka type problem. The “aha” type insight in signaling games is for Ms to realize that their actions affect Es’ beliefs and, by extension, choices of Es. Once this insight has been reached, it is relatively straight forward to realize that

³ For example, in describing the failure of the intuitive criterion in signaling games, Brandts and Holt (1992, p. 1357) note that, “a type-L’s deviation to decision A might be motivated by the belief that the ‘signal,’ A or B, will have no effect on the respondent’s decision. These beliefs are contradicted by the actual decisions of the respondents . . .”

⁴ Much of our discussion of the psychology literature is based on Davis’s (1992) review article (also see Gayle W. Hill, 1982).

⁵ For a general survey of the psychology literature dealing with group versus individual decisions for which there are no demonstrably correct answers, or difficult statistical decisions, see Norbert L. Kerr, Robert J. MacCoun, and Geoffrey P. Kramer (1996).

MHs can gain by imitating MLs (for games with high cost Es) or MLs can benefit by distinguishing themselves from MHs (for games with low cost Es). Further, as we will show, these insights are largely self-confirming since by the time they occur, there is massive evidence that strategic play will have the desired effect of reducing entry.

For many years the prevailing wisdom among psychologists for eureka type problems was consistent with the folk wisdom that teams outperformed individuals. For example, in a classic experiment Majorie Shaw (1932) observed ad hoc, freely interacting, four-person groups working on word puzzles.⁶ Shaw found that in most problems both the proportion of solutions and the time to find a solution was superior for groups than for a comparable sample of individuals working privately. This is a typical result when comparing groups and individuals directly, a result attributed to the ability of group members to catch others' errors, reject incorrect solutions, and to generally stimulate more thoughtful work (Davis, 1992).

By the mid-1950s findings of superiority for groups in eureka type problems came to be viewed with some suspicion. The stimulus for reevaluation was work by Irving Lorge and Herbert Solomon (1955) who proposed the following "truth wins" (TW) standard against which to evaluate the superiority of group performance: Assuming that group interactions are neutral, the group should be able to achieve a correct answer if at least one member proposes it. Therefore, if the probability of an individual working alone solving the problem equals p , the probability P of a randomly selected group with r members solving the problem is the probability that this random sample contains at least one individual who can solve the problem: $P = 1 - (1 - p)^r$. The Lorge-Solomon baseline provides a quantifiable measure of synergy. If teams exceed

⁶ For example, in one problem there are three cannibals and three missionaries on one side of a river. The puzzle is to get them across to the other side by means of a boat that holds only two people at a time. Further, all missionaries but only one cannibal can row, and never under any circumstances can the cannibals outnumber the missionaries.

this standard, the inference is that the interaction among group members generates better performance than the individuals could achieve acting independently. Likewise, failure to reach the Lorge-Solomon baseline indicates that interactions actually make teams perform worse than their members would do acting alone.

Lorge and Solomon reanalyzed Shaw's data, with other researchers doing the same for a number of other data sets. The results of this reevaluation establish that "... freely interacting groups *very rarely exceed, sometimes match, and usually fall below the Lorge-Solomon [TW] baseline.*" (Davis, 1992, p. 7, italics in the original).⁷ Psychologists attribute this relative "inefficiency" of groups to social process losses such as reduced effort due to free-riding or coordination problems involved in combining team members' contributions, problems familiar to economists. The first relates to the economics literature on shirking in team production processes (Armen A. Alchian and Harold Demsetz, 1972). The second relates to diminishing marginal productivity and to common coordination issues involved in team production.⁸

Although there is far from a one-to-one mapping between strategic play in the signaling games studied here and the individual choice problems that psychologists use to compare teams with individuals, we argue that learning to play strategically is sufficiently close to a eureka problem as to make comparisons relevant. The comparison is of interest both because beating the TW norm provides evidence for the presence of positive synergies in team play and because, given the frequent evidence from the psychology literature that this norm is *not* satisfied, beating

⁷ Davis (1992) does not discuss the effects of team size in his survey, suggesting that it is at best of second-order importance.

⁸ There are, of course, reasons other than process loss as to why teams might fail to meet or beat the TW norm. Communication itself is a costly enterprise that reduces the time subjects have available for solving the problem at hand. Also, if a team randomly selects a dictatorial decision maker who does not consult with team members prior to choosing, average results for these teams would not beat the TW norm. Neither of these factors are likely to have played a role in our experiment.

the TW norms is impressive evidence of the efficacy of team play.⁹ This is not to say that the TW norm is the only relevant reference point against which to evaluate the desirability of teams from an economic perspective. That, of course, depends on the context and on the additional costs of using teams as opposed to whatever additional benefits they confer.¹⁰

B. Connections to the economics literature: There have been a handful of studies of group versus individual performance in the economics literature. We focus on the subset of these papers that are most relevant to our study, those that study behavior in games.

Gary Bornstein and Ilan Yaniv (1998) study individual versus team behavior in a standard, one-shot ultimatum game experiment. Their main result is that 3-member teams are more game-theoretically rational players than individuals as they demand more than individuals as proposers and are willing to accept less in the role of responders. In contrast, in a dictator game experiment Timothy N. Cason and Vai-Lam Mui (1997) find that team choices tend to be dominated by the more other regarding member of the team.¹¹

Cox and Stephen C. Hayne (2002) explore differences between group and individual bids in common value auctions for once- and twice-experienced bidders. They focus on “rational” bidding, defined as bidding low enough to avoid falling prey to the winner’s curse. With signal sample size of 1, there are no material differences between groups (of size 5) and individuals in bidding. In contrast, with a signal sample size of 5, groups tend to be *less* rational than individuals (reported in two of four treatments, with no differences found in the two other

⁹ One might also object to the terminology - “truth wins” - in the context of games. We employ it primarily to connect with the relevant psychology literature

¹⁰ Further, as one of our referees suggests, outside the lab team size is likely to be endogenous and may turn out to be optimal for the problem at hand.

¹¹ James C. Cox (2002) on the other hand finds that teams return significantly smaller amounts in the trust game than do individuals.

treatments.)

Martin G. Kocher and Matthais Sutter (in press) (also see Sutter, 2004) compare individual and team behavior in “beauty-contest” games where decision makers compete for a fixed prize by simultaneously guessing a number in a given interval. The winner is the decision maker whose number is closest to a pre-determined fraction of the average of all the numbers that everyone picks. If this fraction is less than one, the unique serially undominated strategy is to guess zero. In practice, optimal guesses depend both on one’s own insight into the equilibrium outcome as well the degree to which one believes that others have the same insight. There are no differences between teams (of size 3) and individuals in the first round of the game, but teams learn faster than individuals as they choose lower numbers in subsequent rounds.

To synthesize the results of these experiments, teams do the same or somewhat better than individuals (with the possible exception of the 5 signal treatment in Cox and Hayne). Our experiment differs substantially from the preceding literature on team play in games. In no case do earlier investigators compare teams against individuals using the demanding “truth wins” norm employed here. This is appropriate given that the games being studied rely less on a “eureka” type insight than on subjects’ judgements. For example, in the ultimatum game rejection or acceptance of offers closer to the subgame perfect equilibrium outcome are tied to whether or not own income is the only argument of players' utility functions, a matter of preferences rather than logic.¹² In contrast, we argue that for signaling game experiments there is a “eureka” insight to be gained, and that the truth wins norm therefore provides a relevant

¹² In the beauty contest game, a subject’s judgement about how logical other players will be is as important as their own ability to reason about the game. In common value auctions, formulating a strategy to avoid the winner’s curse is clearly a task that is demonstrable. However, it involves avoiding a decision making bias that virtually all subjects, both students and those presumably practiced in industries subject to the curse, fall prey to (Kagel and Dan Levin, 2002).

benchmark for the presence of synergies associated with team performance.

A further key innovation of our study is that we have analyzed the team dialogues to obtain insights into subjects' learning and reasoning processes. This is as much a purpose of the present paper as making comparisons between teams and individuals.

III) Experimental Design and Procedures: In what follows we refer to individuals who participated in our experiment as “subjects” while “players” refers to agents in the limit pricing game. A player is a single subject in the individual player (1x1) sessions, but consists of two subjects in the team (2x2) sessions.

A. General Procedures: Subjects were recruited through announcements in undergraduate classes, posters placed throughout the Ohio State University campus, advertisements in the campus newspaper, and direct e-mail contact with students. This resulted in recruiting a broad cross section of mostly undergraduate students and some graduate students. Experienced subject sessions generally took place about a week after the inexperienced subject sessions. Subjects from different inexperienced sessions were mixed in the experienced subject sessions, but subjects were *not* switched between the 1x1 and 2x2 treatments.¹³

Inexperienced 2x2 sessions lasted approximately two hours; inexperienced 1x1 sessions lasted approximately one and a half hours. Experienced subject sessions were substantially shorter than this as short, summary instructions were used and subjects were familiar with the game. Subjects were paid \$6 for showing up on time with total earnings averaging between \$26-\$27 per subject in inexperienced subject sessions. Earnings were higher in experienced subject sessions, averaging a little over \$32 person (including the \$6 show up fee), largely as a

¹³ Econometric analysis indicates that there are no systematic differences between choices in the inexperienced sessions for subjects who later returned for an experienced subject session and those who did not.

result of playing more games.

The 1x1 sessions were designed to employ between 12 and 16 subjects, with the 2x2 sessions employing between 20 and 28 subjects.¹⁴ This results in a somewhat smaller number of players in the team sessions (since each player requires two subjects), but was dictated by the difficulty of assembling larger numbers of subjects, as well the fact that the 2x2 treatment must be run in multiples of 4 subjects (two pairs) and the lab only has 30 work stations. All but two sessions included an even number of subjects so that all subjects participated in every round. The two exceptions were two inexperienced 2x2 sessions with 23 subjects, where the experimenter served as the 24th subject to avoid discarding 3 subjects. In these cases the solo player was told that her teammate was the experimenter who would agree to all of her choices without any further communication.

Upon arrival subjects were randomly assigned to computer terminals. A common set of instructions were read out loud, with each subject having a written copy.¹⁵ All sessions employed abstract terms throughout. For example, Ms were referred to as "A players," with types "A1" and "A2" respectively, and potential Es were described as "B players." Other terms were given similarly meaningless labels. Subjects had copies of both Ms' and Es' payoff tables and were required to fill out a short questionnaire to insure their ability to read them. After reading the instructions, questions were answered out loud and play began with a single practice round followed by more questions.

Before each play of the game the computer randomly determined each M's type and

¹⁴ One experienced 1x1 session was run with 10 subjects and one experienced 2x2 session was run with 16 subjects to avoid losing hard to obtain experienced subject data. The first inexperienced 2x2 session had 12 subjects. Normally, a session with such a low turnout would have been canceled, but this was run in part as a final test of the software.

¹⁵ A copy of the instructions is available at <http://www.econ.ohio-state.edu/kagel/teaminst.pdf>.

displayed this information on Ms' screens. The screen also showed the payoff tables for both types with the table for that player's type displayed on the left. Ms chose by clicking the output level on the payoff table displayed on their screens. The program automatically highlighted Ms' possible payoffs and required that the choice be confirmed. After *all* Ms had confirmed their choices, each M's choice was sent to the E they were paired with. Es then decided between IN and OUT by clicking the appropriate choice on their payoff table. Here too possible payoffs were highlighted and subjects were required to confirm their choices.

Following each play of the game subjects learned their own payoff and Es were told the type of M they were paired with. In addition, the lower left-hand portion of each player's screen displayed the results of all pairings: M's type, M's output, and E's response ordered by output levels (pooled over *all* M types) from highest to lowest.¹⁶ The screen automatically displayed the three most recent periods of play, with a scroll bar available to see all past periods.

The following rotation procedures were generally employed: subjects switched roles with Ms becoming Es and vice versa every 6 games for inexperienced sessions, and every 4 games for experienced sessions. We refer to a block of 12 (8) games in an inexperienced (experienced) session as a "cycle." Within each half-cycle, each M was paired with a *different* E for each play of the game. Inexperienced subject sessions had 24 games divided into two 12 game cycles. Experienced subject sessions had 32 games, divided into four 8 game cycles. The number of games in a session was announced in the instructions. The exceptions to these general procedures occurred in the process of discovering how many replications we could achieve within two hours for the 2x2 treatment.¹⁷ The first two 1x1 inexperienced low cost E sessions

¹⁶ The use of public information in the 1x1 treatment might be expected to crowd out some of the beneficial effect of discussions between team members as it provides a large amount of group information.

¹⁷ The two hour constraint on experimental sessions is designed to avoid subject fatigue.

had three 12 game cycles (36 games) as we were still hopeful that we would be able to complete a comparable number of games in the team sessions. The first inexperienced 2x2 session switched roles every 4 games, with a second inexperienced 2x2 session switching roles every 5 games. A third inexperienced 2x2 session only completed 18 games (1.5 cycles) due to time constraints (exacerbated by a computer crash). Note that these few deviations from the norm (5 out of 33 sessions) all resulted in 2x2 subjects having less experience than 1x1 subjects, which would tend to favor more rapid adjustment to equilibrium in the 1x1 sessions, contrary to the outcomes observed.

B. Team (2x2) Procedures: Team pairings were determined randomly by the computer at the beginning of each session. Matches could not be preserved between inexperienced and experienced subject sessions due to attrition and mixing of subjects from different sessions.¹⁸ Subjects were not told the identity of the person they were matched with, and were asked not to identify themselves.

Teammates were able to communicate and coordinate their decisions using an instant messaging system with full knowledge that these messages would be recorded, but with no other team having access to their messages.¹⁹ In addition to instructing subjects that the instant messaging system was intended to be used for coordinating their decisions, subjects were told to be civil to each other and to not use any profanity. Otherwise, subjects were given no instruction

¹⁸ In three of the team sessions the software had to be restarted, which necessitated new team pairings. Two of these restarts were due to software crashes; the third was due to the session running beyond its advertised time, necessitating the release of some subjects.

¹⁹ The team effect is, of course, inherently confounded with the effect of the particular communication channel the teams use. We have no reason to suppose that written communication is any different from verbal communication especially for subjects who have grown up with e-mail and instant messaging on the Internet. Two key advantages of instant messaging over face-to-face discussions are: (i) transcripts of dialogues are created automatically, and (ii) it would be impossible to have team discussions in the lab while preserving confidentiality between team members.

about what messages to send.²⁰ The message system was open almost continuously, and messages were time stamped with the period of the game being played.

When teams made choices the relevant payoff table on the screen had a column labeled “partner’s choice” on the left and a column labeled “my choice” on the right. When a subject entered a choice, the possible payoffs were highlighted in blue. When a subject’s partner entered a choice, the possible payoffs were highlighted in pink. Once choices coincided, possible payoffs were highlighted in red, at which point teammates had 4 seconds to change their choice before it became binding. Team play started with 3 minutes to coordinate choices, with a countdown clock shown on the computer screens.²¹ If teams failed to coordinate within this time constraint, the dialogue box was closed and one teammate was randomly selected as “leader” with his choice implemented unilaterally. There were virtually no disagreements of this sort.

C. Experimental Design and Hypotheses: Table 2 summarizes the four types of experimental sessions conducted: sessions with inexperienced subjects and high cost Es, sessions with inexperienced subjects and low cost Es, sessions where subjects who had played in inexperienced subject sessions with low cost Es were recruited back for more games with low cost Es, and crossover sessions where subjects who had played in games with high cost Es were recruited back. Cross-over sessions started with a full cycle (8 games) with high cost Es. This was followed by 24 plays (3 cycles) with low cost Es. When Es’ payoffs were changed subjects were given written copies of the new payoff tables, with a brief set of instructions read out loud indicating that the only change in procedures involved new payoffs for Es.

²⁰ Beyond these instructions, there was no attempt to prevent subjects from sending any message they desired. Virtually all of the discussions were civil. Many teams discussed topics in addition to the experiment and there was some use of profanity. There is little evidence that subjects were inhibited by the knowledge that their messages were being recorded.

²¹ This time was reduced by 30 seconds between cycles except for cycles following a cross-over.

For each cycle of the games played we examine two related hypotheses:

Hypothesis 1: There will be more strategic play in each of the 2x2 games than in the corresponding 1x1 games.

Hypothesis 2: The level of strategic play in the 2x2 games will meet or beat the TW norm, as described in Section IIA, based on the level of play observed in the 1x1 games.²²

Hypothesis 1 constitutes the minimal requirement to justify team decision making on purely economic grounds.²³ Hypothesis 2, if validated, indicates that interactions between teammates generate positive synergies, giving them a greater chance of success working together than the most able member of the team would have working independently.

Hypothesis 2 is based on the argument that strategic play in signaling games, as it occurs in the lab, has strong similarities to the puzzle solving branch of the psychology literature dealing with team versus individual play. However, there are important differences between the sort of individual choice problems studied by psychologists and a strategic environment like the limit pricing game. The most critical of these rests on the inherently stochastic nature of Es' choices. It is well-established that subjects' ability to best respond in games is sensitive to the payoff premium (Raymond C. Battalio, Larry Samuelson, and John Van Huyck, 2001). While playing strategically almost always maximizes expected payoffs in our games, the premium for playing strategically varies quite a bit. To the extent that this premium varies between the 2x2 and 1x1 treatments, it could cause the 2x2 treatments to either over or underperform relative to the TW norm. We use probit regressions to address this issue since this permits comparisons of the level

²² More technically, suppose p_i is the probability that, given the opportunity, an individual plays strategically in cycle t and P_t is the corresponding probability for a two person team. Beating the TW norm in cycle t requires $P_t \geq 1 - (1 - p_i)^2$.

²³ Team decision making can, of course, be justified on other grounds including legal, administrative or political reasons.

of strategic play between treatments while controlling for possible differences in the incentives Ms face for playing strategically.²⁴ These probits are reported in detail in the appendix to the paper, with the results summarized in the text. They are important to our analysis of whether or not Hypotheses 1 and 2 are supported by the data, but the details can be somewhat tedious and distracting from the overall analysis.

We pose a third hypothesis specific to the cross-over sessions. In addition to giving us another venue for comparing the decision making of teams versus individuals, this treatment also allows us to address the question of learning generalizability - the ability to take experience with one game and apply it in a related game. Learning generalizability is an important issue in psychology, and important to economic arguments that rely on learning processes to justify equilibrium outcomes.²⁵ The usual result from psychology experiments is that there is little or no learning transfer (and sometimes even negative learning transfer) unless subjects are explicitly queued to draw on their previous experiences (see, for example, Gavriel Solomon and David N. Perkins, 1989).²⁶ One could argue that positive learning transfer is likely to be particularly difficult here since strategic behavior following this cross-over requires substantially different actions than before the cross-over. Prior to the cross-over MHs learn to imitate MLs, but MLs

²⁴ An alternative, and superior, method for dealing with this issue would be to conduct sessions in which there are both teams and individuals playing at the same time. Unfortunately our software, which took some time to develop, cannot accommodate such a design. This is one of several issues to be investigated in the future.

²⁵ Virtually all studies of learning in economics employ an environment in which learning takes place within an essentially stationary environment. Studies of cross-game learning are important since it's unreasonable to expect the exact same game to be repeated over and over, so that if one can only justify convergence to equilibrium in such situations, there would not be much reason to have faith in the widespread applications that are found in the literature. Rather, faith in such applications can be greater if players infer how their opponents will act in one situation from how they acted in other, related situations.

²⁶ Here too, psychology experiments deal primarily with puzzles, or individual skills, and not strategic situations such as those involved in signaling games.

must learn to distinguish themselves following the cross-over.²⁷

Hypothesis 3: There will be little or no learning transfer in the cross-over treatments. That is, levels of strategic play will be the same, or possibly even lower, following the cross-over compared to inexperienced subjects in games with low cost Es. This will hold for both the 2x2 and the 1x1 treatments.

IV. Experimental Results:

A. Limit Pricing in Games with High Cost Entrants: Figure 1 aggregates data from the 1x1 sessions for games with high cost Es. This data provides a baseline for how strategic play evolves for MHs. In the first cycle of play, choices of each type are concentrated at their myopic maxima; 2 is the modal choice for MHs and 4 is the modal choice for MLs. The lack of strategic play by MHs cannot be attributed to a lack of incentives; even in cycle 1 entry rates for 2 are 46.3% higher than for 4. A difference of only 13% is needed to make strategic play profitable for MHs.

In the second cycle the difference in entry rates between 2 and 4 becomes even more pronounced, rising to 59.4%. Responding to these strong incentives, MHs begin to play strategically with greater frequency, with 4 becoming their modal choice. At the same time, play of MLs becomes even more concentrated at 4. The first cycle of experienced subject play continues these trends: the entry rate differential between 2 and 4 rises slightly, MHs play strategically even more frequently, and MLs choose 4 almost exclusively.

Figure 2 reports the results for the 2x2 sessions for each cycle of play. Comparing Figure 1 with Figure 2, it's clear that the general dynamics of play are similar, as MHs in the 2x2 sessions

²⁷ A fictitious play learning model that has worked well in tracking play from previous signaling games (Cooper *et al*, 1997b) predicts that MLs' strategic play immediately following the change in Es' payoffs will be *less* than in inexperienced control sessions (negative transfer), and will remain so until behavior converges to the equilibrium outcome.

only gradually move from myopic play at 2 to strategic play at 4. Within the first cycle of play there is a distinctly higher frequency of strategic play by MHs in the 2x2 treatment, which continues into later cycles as well. Moreover, strategic play is relatively more concentrated on 4 (rather than 3) in the 2x2 treatment.

Figure 3 directly compares levels of strategic play (MHs' choices of 3, 4 and 5) between the 1x1 sessions and the 2x2 sessions, along with the TW norm (shown as filled diamonds with error bars for the 90% confidence interval).²⁸ Strategic play emerges somewhat more rapidly for MHs in the 2x2 treatment, being some 12.5% greater than in the 1x1 sessions for the first cycle of inexperienced subject play. The difference between treatments in the proportion of strategic play narrows over time, with 7.9% greater strategic play for 2x2s in the second cycle of inexperienced subject play, and 4.8% greater strategic play in the first cycle of experienced subject play. Quite clearly Hypothesis 2 fails to be satisfied, as the level of strategic play in the 2x2 sessions is well below the 90% confidence interval for the TW norm. Thus, differences between the two treatments largely reflect a greater proportion of MHs who immediately play strategically in the 2x2 treatment rather than faster learning, with the teams failing to generate positive synergies as measured by the TW norm.

The results reported in Figure 3 suggest that Hypothesis 1 holds, albeit weakly, for games with high cost Es. To examine this more formally, we ran probit regressions for MHs' choices,

²⁸ Because of clustering in the data, simulations are needed to correctly calculate the error bars. The simulated 2x2 data is based on 100,000 simulated 2x2 data sets for each cycle of play, with the same number of teams in each data set as in the experiment. Simulated 2x2 play is based on randomly drawing two individuals (with replacement) from the 1x1 sessions. The likelihood of any individual being drawn is proportional to the number of times that individual was an MH in that cycle, with the probability of playing strategically based on the observed frequency of strategic play as an MH in that cycle. A simulated team was considered to have played strategically if either of its members played strategically. The error bars then display the 5th and 95th of the distribution of percentages of strategic play in a simulated 2x2 data set. If the percentage of strategic play in the actual 2x2 data is below the error bar for the simulated data, as in all three cycles of the data for games with high cost Es, this indicates that over 95% of the simulated data sets yield more strategic play than the actual data.

with the dependent variable being whether or not MHs played strategically (i.e., chose output 3, 4 or 5).²⁹ This analysis finds that (1) without any controls for differential entry rates, there is significantly more strategic play in the 2x2 treatment for the first cycle of inexperienced subject play ($p < .10$), with no significant differences in later cycles, (2) adding controls for Es' choices shows that Ms only respond weakly to differences in entry rates, as the parameter estimate for this variable is not significant at conventional levels, and (3) the entry rate controls have little impact on the estimated differences between 1x1 and 2x2 play.

The results for games with high cost Es may be summarized as follows:

Conclusion 1: In games with high cost Es, MHs play more strategically with teams than with individuals, with maximal differences observed in the first cycle of inexperienced subject play. However, teams do not meet or beat the TW norm.

B. Limit Pricing in Games with Low Cost Entrants: Figure 4 aggregates data from 1x1 sessions for games with low cost Es.³⁰ Play in the first cycle for inexperienced subjects is similar to games with high cost Es, with Ms' play clustered at the myopic maxima for both types. There are strong incentives to play strategically as an MH in this first cycle of play, but only weak incentives for MLs to play strategically. One notable difference in early play between these games and those with high cost Es is that entry rates are much higher here for outputs 2-4, consistent with the substantially higher payoffs for IN versus OUT. By the second cycle of inexperienced play there exist strong incentives to play strategically for both MHs and MLs.

²⁹ Running random effects probits with the 2x2 treatment poses some non-standard statistical issues. We employ very conservative assumptions regarding the degree of independence between team members particularly with respect to experienced players which, if anything, biases our results against finding statistical significance between the two treatments. Robustness tests for the probit specifications employed are reported at <http://www.econ.ohio-state.edu/kagel/onlineappendix.twoheads.pdf>.

³⁰ Data from the cross-over treatments are *not* included here but are discussed in the next section.

Steady movement towards strategic play takes place for both types, but this movement is smaller for MLs than for MHs, even though MLs have stronger incentives than MHs to play strategically in the second cycle.

Experienced sessions with low cost Es are largely a continuation of patterns from inexperienced subject play. MHs increase their play of 4 going from cycle 1 to 2, with little retreat back to 2 thereafter, as it is incentive-compatible for them to choose 4 over 2 throughout (albeit somewhat less so in later cycles of play). MLs' level of strategic play in cycle 1 is roughly what it was in the ending cycle of inexperienced subject play, but steadily increases thereafter, with strategic play largely directed to output 6. This movement is quite slow – only in the third experienced cycle does 6 become the modal choice for MLs. In contrast, in the games with high cost Es, strategic play by MHs is the modal outcome by the second cycle of the *inexperienced* sessions. This difference is even more striking since the incentives for MLs to play strategically here are roughly the same as for MHs to play strategically in the previous treatment. There is clearly something quite different, and more difficult, about learning to play strategically as an ML than as an MH.

Figure 5 aggregates data from the 2x2 sessions for games with low cost Es. The general pattern of play is the same as in the 1x1 sessions, but convergence to the efficient separating equilibrium is much more rapid and complete. Comparing Figure 4 with Figure 5, the more rapid development of strategic play for teams is clear by the second cycle of inexperienced subject play. More striking yet is the near complete convergence to the efficient separating equilibrium in the last two cycles of inexperienced subject play in the 2x2 games as compared to the 1x1 sessions. Not only is there far more strategic play by MLs in the 2x2 games for these final cycles, but it is much more heavily concentrated on 6 compared with 5 in the 1x1 sessions.

Further, there is 100% entry on 4, and MHs have almost completely retreated back to choosing 2 in the 2x2 games. In contrast, in the corresponding cycles of 1x1 sessions, the entry rate on 4 relative to 2 is still sufficiently low that choice of 4 is incentive compatible for MHs.³¹

Figure 6 compares the level of strategic play by MLs between the 2x2 and 1x1 treatments as well as simulated outcomes for the TW norm. The results are striking: teams play more strategically than individuals throughout, with the difference growing steadily with experience. They match the TW norm in the two cycles of inexperienced subject play, and then either beat it, or fall into the upper tail of the 90% confidence interval, for all four cycles of experienced play.³²

Probits reported in the appendix show that (1) the differences in the level of strategic play for MLs between the team and 1x1 treatments are statistically significant in all cycles of play, (2) entry rates have a strong impact on MLs choosing to play strategically, and (3) controlling for entry rate differentials, MLs' level of strategic play is still significantly higher in the teams treatment for four of the six cycles, although the magnitude of the effect is reduced.

As the preceding suggests, the ability of teams to meet or beat the TW norm may be due to greater entry rate differentials in the 2x2 treatment. The results of the probit analysis allow us to explore this issue in more detail. We calculate for each cycle what the proportion of strategic play by MLs would have been in the 2x2 sessions, *if the entry rate differential had been*

³¹ It might be argued that individuals are likely to keep learning over time so that with enough experience they will catch up with teams. However, we have no evidence for this to date in spite of, in one instance, bringing back twice experienced subjects (see, for example, Cooper et al. (1997a), Figure 5), so the jury is still out on this question. The closest we have seen to anything like the teams' convergence to the efficient separating equilibrium for individuals occurred in games where MHs were prohibited from choosing 6 or 7, and this was announced as part of the instructions (see Cooper et al., 1997b).

³² While strategic play by MHs (choice of 3, 4, or 5) is clearly an important feature of the data in games with low cost Es, we do not compare the development of this behavior across treatments. Unlike strategic play by MLs, strategic play by MHs in the low cost entrant game does not cleanly fit our definition of a eureka type problem. Recall that a eureka type problem has a *demonstrably* correct solution. While the entry rate differential between 2 and 4 starts out large, it shrinks steadily over time. So although there is scope, early on, for an "aha" type insight for MHs (in that they can profitably imitate MLs), at some point (unknown to us) individual MHs become increasingly disabused of the profitability of choosing 4 over 2.

identical, in that cycle of play, to the 1x1 sessions, and compared these adjusted percentages with the TW norm. These calculations show that after the first cycle of inexperienced subject play teams meet the TW norm, but only beat it in experienced cycle 3. For example, if teams in the 2x2 treatment had faced the entry rates observed for the 1x1 treatments in experienced cycle 2, the estimated proportion of strategic play by teams in cycle 2 would have been 63.4% rather than the observed frequency of 88.4%. This adjusted figure is slightly lower than the TW norm of 70.6%, so that teams meet (within the 90% confidence interval), but do not exceed, the estimated value of the TW norm.³³ Thus, we cannot reject a null hypothesis that the ability of teams in the low cost entrant game to *beat* the TW standard is due to greater incentives to play strategically. However, it is clear that teams *meet* the TW norm after controlling for any differences in MLs incentives to play strategically between the two treatments.

Conclusion 2: MLs have higher levels of strategic play in the 2x2 treatment, meeting or beating the TW norm in all cycles of play. Accounting for entry rate differences between treatments, MLs generally meet, but do not beat, the TW norm. Further, play is much closer to the efficient separating equilibrium in the last two cycles of experienced subject play in the 2x2 sessions than in the 1x1 case.

C. Limit Pricing in the Crossover Treatment: Figure 7 reports data for the 1x1 and 2x2 treatments (top and bottom panels respectively) by cycle following the cross-over. The

³³ As an alternative, we have adjusted the 1x1 proportions for the difference in entry rate differentials, used the adjusted 1x1 proportions to calculate an adjusted TW norm, and compared this adjusted TW norm with the 2x2 data. Not surprisingly, this alternative approach yields similar conclusions to adjusting the 2x2 data. Specifically, the adjusted TW norm moves up sufficiently to make it difficult to distinguish the adjusted TW norm from the 2x2 data in most cycles. More precise statements cannot be made since calculating error bars for this adjusted TW norm is problematic. Because of these difficulties, this alternative approach is inferior to adjusting the 2x2 data for entry rate differentials. Note, however, that the probit regressions use an extremely conservative approach in controlling for clustering in the data. As described in Section B.1 of the appendix, approaches that make fuller use of the information in the data consistently yield smaller estimates for the marginal effect of entry rate differentials on strategic play by MLs. As such, the adjusted proportions of strategic play for the 2x2 data are probably too low, understating how well teams are performing versus the TW norm after controlling for entry rate differentials.

differences are quite striking on a number of dimensions.

1. Right from the start there are substantially higher levels of strategic play for MLs in the 2x2 sessions. Figure 8 directly compares MLs' level of strategic play between the two treatments, along with the TW norm. The frequency of strategic play in the 2x2 sessions is well above the 90% confidence interval for the TW norm for the first two cycles, and slightly exceeds its upper bound in the last cycle. Further, using the probit regressions to control for entry rate differentials, strategic play for MLs in teams remains significantly above the 90% confidence interval of the TW norm in the first two cycles of play. Thus, not only do teams *beat* the TW norm in the cross-over treatment, it *cannot* be attributed solely to entry rate differentials between the two treatments.
2. In Figure 7, it's clear that in the last cycle following the cross-over play has fully converged to the efficient separating equilibrium in the 2x2 treatment, but is still very much in transition in the 1x1 treatment. In the 1x1 treatment MLs' strategic choices are fairly evenly split between 5 and 6 in the final cycle, with 4 still attracting a large number of choices for both MLs and MHs. In contrast, 6 accounts for 89.2% of MLs strategic choices for teams and the vast majority of MH choices are at 2.
3. Comparing the levels of strategic play in Figure 7 with Figures 4 and 5 gets at the issue of cross-game learning. Subjects in Figures 4 and 5 are playing the same game as those in Figure 7, but have no prior experience with the entry limit pricing game. Subjects in Figure 7 have prior experience with the entry limit pricing game, but in games with high cost Es where play converges on the pooling equilibrium at 4. Does prior experience in the game with high cost Es help with the development of strategic play for MLs in the game with low cost Es? The answer, shown in Figure 9, is clearly yes for 2x2 treatment, but no for the 1x1

treatment. In fact, as probits reported in the appendix verify, there is significantly *less* strategic play for MLs in the first two cycles of the 1x1 treatment following the cross-over than in the first two cycles of inexperienced subject play, both with and without including entry rate measures in the regressions.³⁴ In contrast, there is significantly *more* strategic play in the first cycle of play following the cross-over than in the first cycle of inexperienced subject play in the 2x2 sessions. These results are summarized in the following two conclusions.

Conclusion 3: The cross-over treatment shows substantially higher levels of strategic play for MLs in the 2x2 treatment than in the 1x1 case for all cycles following the crossover, beating the TW norm throughout. This conclusion is robust to accounting for different incentives to play strategically between the two treatments in the first two cycles of play. Further, a clean separating equilibrium has emerged in the 2x2 treatment by the last cycle of play, while the 1x1 treatment is still very much in a state of flux.

Conclusion 4: There is *negative* learning transfer in the 1x1 treatment, as prior experience in the game with high cost Es provided subjects with a slight, but statistically significant, *disadvantage* following the introduction of low cost Es relative to subjects with no prior experience. However there is *positive* learning transfer under the same conditions in the 2x2 treatment, as prior experience in games with high cost Es *facilitates* the development of strategic play relative to subjects with no prior experience in games with low cost Es.

³⁴ The negative learning transfer reported here for the 1x1 treatment differs from the positive learning transfer found in the 1x1 treatment reported in Cooper and Kagel (2004). Beyond the use of different subject populations, there are a number of methodological differences between the present experiment and the one reported in Cooper and Kagel (2004). Preliminary results in a follow-up study indicate that the most prominent of these differences has to do with the use of an abstract context here versus the use of meaningful context in the earlier experiment (see Cooper and Kagel, 2003). We plan a full report on the effects of context on learning transfer in games in another paper.

There is one final point worth making before analyzing the team dialogues. It is based on the notion that the speed with which strategic play develops in the 1x1 treatments is a measure of the difficulty subjects have with learning to play the game. Assuming this is the case, we can rank the games in terms of their degree of difficulty, with the games with high cost Es and the pooling equilibrium being the easiest, and the cross-over treatment being the most difficult. This leads to the following conjecture:

Conjecture: The more difficult it is to learn to play strategically, the greater the advantage of two-person teams over individuals.

The possible reasons for such an inverse relationship are discussed in the concluding section of the paper.

V. Insights into the Learning Process: Analysis of the Team Dialogues

Our original motivation for conducting the teams treatment was to obtain direct insight into the learning process underlying the development of strategic play through analyzing the team dialogues. These dialogues are a natural part of the experimental task, and are clearly relevant for the task at hand, thereby providing an unbiased, albeit noisy, window into the underlying learning process.³⁵

The primary goal of this section is to provide a brief overview of the major contents of the team dialogues. In doing so we provide further evidence that Ms' initial choices (2 for MHs and 4 for MLs) reflect a basic failure to think strategically, and that the development of strategic play is associated with Ms thinking from the point of view of Es, thereby allowing Ms to anticipate

³⁵ Post-experiment surveys gather data retrospectively, thereby relying on subjects' possibly shaky and/or biased memories of what they were thinking in earlier stages of the experiment. "Talk out loud" techniques allow for the gathering of real time data, but the dialogues generated are not integral to the task at hand.

Es' responses to their choices.³⁶ The dialogues also provide clear evidence that the positive cross-game learning reported *within* the teams treatment is associated with increased numbers of subjects who anticipate that the change in Es' payoffs will promote increased entry, and who understand the appropriate response to the increased entry. This implies that one of the key things teams have learned from prior experience with the game with high cost Es is to think strategically, and this thinking readily generalizes to games with low cost Es.

To analyze the team dialogues, we developed a coding system for types of statements as follows: first, the authors separately read through a common sample of the dialogues, establishing a set of preliminary codings, which were then reconciled, establishing the single list shown in Table 3. This coding scheme is designed to capture all statements that are relevant to subjects' learning to play strategically. Two undergraduate RAs were then trained (separately) to do the coding. Although there is a fair amount of variance between the two coders (the average cross-coder correlation for the 24 categories is .388), much of the variance comes from categories that were not coded very frequently. (The average cross-coder correlation across the five most frequent categories is .570, with a minimum of .517.)³⁷ Results reported are based on the average of the two independent codings, unless otherwise stated. In averaging across coders, we are implicitly assuming that errors are independent across coders so that averaging reduces

³⁶ Given the differences in team versus individual play documented here, these insights do not necessarily extend to individual play. The latter will require obtaining monologues directly from individual play, a project that plays a prominent role in our future research agenda. Our educated guess is that the development of strategic play by individuals also relies on learning to think from the E's point of view.

³⁷ As an alternative measure of agreement between coders we calculated the proportion of times that coders agreed on their categorization of a dialogue relative to the total number of observations where at least one coder categorized the dialogue. Averaging over the five most frequent categories, the proportion of agreement is 46.4%. Two sources of noise hold down agreements between coders. First, the two coders systematically disagreed on the exact codings for several categories. Second, given the unstructured nature of the dialogues, discussions frequently extend over multiple rounds of the experiment. The agreement between coders is therefore reduced by cases in which the coders agree on what should be coded but disagree on exactly which round it should be coded for.

the total error.

A. What Teams Are Talking About Prior to First Strategic Play: Table 4 reports coding frequencies for those categories that were coded 10% of the time or more for inexperienced teams up to and including the first time they play strategically.³⁸ The columns labeled “No” and “Yes” in Table 4 distinguish between teams who played strategically at the next opportunity following their first strategic play (the Yes column) and those who failed to do so (the No column). The latter is not uncommon for inexperienced subjects, occurring for 22.8% of the teams that play strategically in games with high cost Es and 28.9% of the time in games with low cost Es. The “All” column ignores this distinction, averaging across the Yes and No columns.³⁹ The column labeled “Fisher Exact” reports the results of a Fisher exact test for whether there is a statistically significant relationship between continuing to play strategically and the row variable.

Our analysis begins with focusing on the All frequencies. Category 10 refers to teams using the feedback data provided after each game which records the results of all pairings: M’s type, M’s output, and E’s response. This is, by far and away, the most frequently coded category prior to strategic play in both the high and low cost entrant games.⁴⁰ The frequent use of feedback regarding others' choices suggests that pure reinforcement learning, in the spirit of Alvin E. Roth

³⁸ Experienced subject sessions are less informative as many teams include individuals who have previously figured out how to play strategically, which can lead to very cryptic conversations. Table A.8 in <http://www.econ.ohio-state.edu/kagel/onlineappendix.twoheads.pdf> provides a complete set of codings along with the between coder correlations. We employ all comments up to and including the period a team first played strategically because teams were in constant communication, but only had the opportunity to play strategically at randomly determined intervals. Thus, in many cases a team had figured out how to play strategically when they were an E or were playing in the other role as an M.

³⁹ The frequencies in the All column are not the weighted average of the frequencies in the Yes and No columns as some teams never got the chance to play strategically again.

⁴⁰ A typical dialogue here for an MH team: “Look at the table. Every 4 was sent an X [IN].” “So what then.” “Have no clue. You pick a number.”

and Ido Erev (1995), is likely to be a poor model, by itself, of team learning.

Category 1 refers to myopic choice as an M player. This is the second most frequently coded category prior to strategic play in games with high cost Es, and a close third in games with low cost Es. It corresponds to MHs choosing 2 and MLs choosing 4 on account of its having the highest payoff *ignoring Es potential response to these choices*. To quote from a ML team in a session with low cost Es:

“i think we should pick 4 because no matter what they pick, the highest payoff will be for either X [In] or Y [Out].” “cool.”

Teams, in their role as Es, make similar statements in looking forward to how they will play as Ms.⁴¹ To quote from a low cost E team:

“when we’re team A2 [ML] we should always choose #4 cause that makes the most money no matter what the B [E] team chooses.” “good point.” “and with A1 [MH] I think its option 2.”

The high frequency of coding for category 1 supports our contention that early play of the myopic maxima (2 for MHs, 4 for MLs) is non-strategic in nature.

The categories of greatest interest to us are those which relate to how subjects justify strategic play. Category 3 was designed to capture different types of reasoning underlying strategic play in games with high cost Es, with Category 4 the corresponding category for games with low cost Es. There are four subcategories in each case. The most frequently coded subcategories are 3iii and 4iii, in both cases consisting of explicit reasoning from the point of view of Es’ potential responses to their choices as Ms. For example, to quote from a team deciding how to play as an MH:

⁴¹ These statements are also included in Category 1.

“Lets try 4 again” “OK” “I think the problem with 2 is that Bs [Es’] knew we were A1 [MH], so they chose X [IN] to maximize their payoff.” “I think you’re right.”

And to quote from a team in deciding how to play as an ML:

“If we enter 6 when we are A2 [ML] then everyone will know that we are an A2 [ML] and will guess accordingly, giving us a higher average....The 6 pays less but is very clear what we are.”

These subcategories are also among the highest of the 24 categories coded, with 3iii the third most frequent coding for games with high cost Es and 4iii the second most frequent for games with low cost Es. The second most common subcategories – 3ii and 4ii – are close variations on the same theme.⁴² From a game theorist’s point of view, this strategic empathy – reasoning from the other player’s point of view – goes to the heart of thinking and behaving strategically. The codings provide clear evidence that this type of reasoning underlies the development of strategic play.

The Fisher exact tests in Table 4 indicate that being coded for categories 3iii and 4iii are good predictors of whether or not, having played strategically once, teams will continue to play strategically. For games with high cost Es the only variable with any significant predictive power regarding continued strategic play is Category 3iii. This effect is quite strong; of the 27 teams coded for Category 3iii by either coder, *none* returned to non-strategic play. Likewise, Category 4iii is a strong predictor for whether strategic play will continue for games with low cost Es. Only one of the fifteen teams coded for 4iii by either coder returned to non-strategic

⁴² Category 3ii involves MHs choosing higher outputs in order to “fool” Es into thinking they were MLs. Category 4ii involves explicitly recognizing that the negative payoffs for MHs’ choice of 6 or 7 must make it obvious to Es that these choices are from MLs. In games with high cost Es, Categories 3ii and 3iii, while related, are not perfectly correlated. In games with low cost Es, virtually all teams coded for 4ii were also coded for 4iii, but not vice versa. The relatively high frequency of 3iii in games with low cost Es reflects play passing through a phase where MHs imitate MLs prior to the emergence of a separating equilibrium.

play.⁴³ Thus, thinking from the point of view of Es plays an important role in whether or not inexperienced Ms backslide into non-strategic play.

The dialogues provide a few additional insights. First, most experimenters suspect that switching roles in games speeds up the learning process. While it probably isn't worth the time and money to directly verify this methodological conjecture directly, the relatively frequent codings of Categories 5 and 5i (which code for recognizing separating choices by MLs while in the role of an E) in games with low cost Es provide indirect evidence to this effect.⁴⁴ Second, there are lessons to be learned from the less frequently coded categories that did not make it into Table 4. Category 8 codes directly for pure reinforcement learning. It is coded relatively infrequently, particularly for games with low cost Es (8.2% for games with high cost Es; 3.6% for games with low cost Es). Categories 3i or 4i were designed to capture Ms playing strategically purely on the basis of imitation. Neither is coded with any frequency (2.5% and 2.4% for 3i and 4i, respectively). The infrequent codings for reinforcement learning and pure imitation suggest that teams are not blindly choosing (avoiding) strategies that have done well (poorly) in the past. Rather, the substantially more frequent codings for Categories 3iii and 4iii suggest that they are trying to figure out *why* these strategies have done well or poorly in the past. Finally, in discussing why teams frequently fail to clear the TW threshold, psychologists often refer to "process loss." One extreme version of process loss is "truth loses" – when a subject who has failed to figure out the strategic aspects of the game convinces a more insightful teammate to play non-strategically. Category 14, designed to capture this, is not coded often -

⁴³ The other variables that have significant predictive power are M's choice of 5 or 6 when first playing strategically, and E's response. Teams reverting to non-strategic play are both more likely to have chosen 5 and to have been entered on compared with teams that continue to play strategically.

⁴⁴ However, subjects almost never explicitly draw on their own experiences as Es in deciding how to play as Ms – categories 3iv and 4iv are rarely coded (4.9% and 7.1 % respectively) prior to first strategic play.

less than 1% in games with high cost Es and 1.2% in games with low cost Es - indicating that this kind of process loss rarely occurs.

Conclusion 6: Use of feedback about the decisions of others is the most frequently coded category. Categories designed to capture Ms thinking from the point of view of Es' responses to Ms' choices are the most commonly coded justifications for strategic play, and the second and third most commonly coded categories overall for games with low and high cost Es respectively. This is consistent with the type of strategic anticipation that lies at the heart of game theory.

Conclusion 7: Reverting to non-strategic play after having played strategically occurs significantly less often for teams that have discussed their strategy as Ms from the point of view of Es. That is, teams recorded as thinking in game theoretic terms are more likely to continue to play strategically than those without such conversations.

B. Positive Learning Transfer for Teams in the Crossover Treatment: One of the most striking features of the teams data is the positive learning transfer in the crossover treatment. The codings provide some insight into the mechanism underlying this. First, most teams almost immediately recognized that the change in payoffs would increase entry rates substantially, particularly for output 4. Category 6 codes for this - recognizing that the change in payoff tables will change Es' choice (following the cross-over) - coded for 84.5% of all teams overall and coded (by at least one coder) for 76% of all teams in the *first play of the game* following the cross-over.⁴⁵ Second, in the first cycle following the crossover, category 4iii was coded for 35.7% of the teams, compared to 17.3% in the first cycle of inexperienced play for games with

⁴⁵ For example, from one M team, "Now they are making this interesting. ...Now it'd actually be worth it to pick X [IN] every time."

low cost Es. This represents a large growth in the frequency of strategic empathy, as well as an ability to transfer that strategic thinking, in the 2x2 treatment.⁴⁶

We can relate these observations, and the results of the cross-over treatments, to adaptive learning models offered in the literature. Standard adaptive learning models, for example stochastic fictitious play, predict *negative* learning transfer in the cross-over treatments as MLs with prior experience in games with high cost Es have to unlearn their expectations of high payoffs and low entry rates for choosing 4, beliefs which don't burden inexperienced subjects. One way to overcome this and generate the positive cross-game learning observed in the data is to introduce sophisticated learners into the model who, in this case, (i) anticipate that changing the payoff table for Es will change Es choices, and (ii) whose numbers increase as a result of prior experience with the game (see Cooper and Kagel, 2004).⁴⁷ From this perspective the negative transfer observed for 1x1 cross-over sessions suggests that sophisticated learning develops at a much slower rate in the 1x1 sessions than in the 2x2 sessions. More fundamental differences in the learning processes may exist as well.

Conclusion 8: The strong, positive learning transfer for teams is consistent with an adaptive learning model in which there are growing numbers of sophisticated learners who anticipate that the change in Es payoffs will promote increased entry and that play of 6 will deter entry. The codings provide supporting evidence for this conjecture. The negative learning transfer found in the 1x1 treatment suggests that sophisticated learning develops at a much slower

⁴⁶ There is also substantially less backsliding in the cross-over treatment than for inexperienced sessions in games with low cost Es (11.1% versus 29.0%; $Z = 1.91$, $p < .06$, 2-tailed test). This difference is consistent with the importance of Category 4iii in preventing backsliding, and the substantially higher frequency with which 4iii is coded in the cross-over treatment.

⁴⁷ For other examples of this sort of model, see Colin F. Camerer, Tech H. Ho, and Juin-Kuan Chong (2002); Milgrom and Roberts (1991); and Dale O. Stahl II (1996).

rate than in the 2x2 sessions, or some fundamental difference exists between the learning processes of teams and individuals.

VI. Conclusions and Discussion

This paper compares team versus individual play in a signaling game based on Milgrom and Roberts' (1982) entry limit pricing game. The focus of the paper is on differences in the learning/adjustment process between two person teams versus individual subjects. Strategic play develops more rapidly in teams in all three treatments. Further, the superiority of team play increases the more difficult it is for subjects to learn to play strategically, so that in the most challenging games teams meet or surpass the "truth wins" norm developed by psychologists for "eureka" type learning problems. Surpassing the truth wins norm is consistent with positive synergies between teammates, and is rarely reported in the psychology literature for similar types of problems. In addition, teams exhibit strong *positive* cross-game learning, whereas individuals exhibit *negative* cross-game learning. The positive learning transfer in the 2x2 treatment also contrasts with results typically reported in the psychology literature, where zero or even negative learning transfer is usually reported (Solomon and Perkins, 1989). It is consistent with adaptive learning models with growing numbers of sophisticated learners who anticipate their opponents' responses to changes in their opponents' payoffs. This interpretation is supported by the team dialogues.

Our results raise two questions. Why do teams perform relatively better, compared to individuals, the more difficult the learning/adjustment process is? How come teams meet, and even beat, the truth wins norm in games with low cost entrants and in the cross-over treatment given that psychology experiments usually report a failure to do so in similar problems? While greater incentives to play strategically can partially explain the strong performance of teams,

even after controlling for this factor teams meet the truth wins standard in games with low cost entrants and beat it in the cross-over treatment. Instead, the answer to both of these questions stems from several related sources.

One possibility is that although the insight underlying the truth wins model is relevant to our results, the mathematical formula used to generate the truth wins standard is not appropriate here. This formula assumes that problem solving (playing strategically) involves a *single* “aha” insight. However, the dynamic leading to the separating equilibrium in games with low cost entrants consists of two distinct stages. Play first passes through a phase where MHs imitate MLs before MLs, under pressure from rising entry rates, begin to separate. Thus, this might be better likened to a multi-step problem in which getting to the right answer involves a series of smaller insights rather than a single “aha” type insight. In this case it might well be that the truth wins norm applies to each step of the learning process, with *both* team members being *equally* likely to provide the crucial insight for each step in the learning process. If this is the case, outcomes are likely to meet or beat the truth wins norm.⁴⁸

A second possibility is that playing as teams speeds up the development of strategic empathy, generating more “sophisticated” players who can think from the viewpoint of others. Increased numbers of sophisticated learners can account for teams meeting or beating the truth wins norm in games with low cost entrants, as well as the positive cross-game learning reported

⁴⁸ To be more formal, suppose that playing strategically involves two discrete insights. Assume that the likelihood of obtaining these insights is independent across tasks and across individuals. Let p_1 and p_2 be the probabilities of an individual obtaining each insight. The probability of an individual solving the problem is $p_1 p_2$. The TW norm is therefore $p_1 p_2 (2 - p_1 p_2)$. This however is not the probability that a team solves the problem in the absence of any synergies because it ignores the possibility that one team member obtains the first insight and the second team member obtains the second insight. The correct probability that a team solves the problem is $p_1 p_2 (4 - 2(p_1 + p_2) + p_1 p_2)$. Doing some algebra we can confirm that this probability is always greater than the TW norm. Of course, in such a setting it’s possible that teams perform no better, or even worse, than individuals in cases where team members who do not grasp the first insight are not able to work on the problem on any deeper level, and hence are of no use (or even a hindrance) to obtaining further insights.

for teams. Comparing team dialogues following the cross-over treatment with those of inexperienced teams provides direct evidence for large increases in the number of sophisticated learners as a result of experience in the teams treatment.⁴⁹

The question that remains is why do these factors play a role in our experiment but not in the typical psychology experiment? The answer is that our experimental procedures, typical of those employed in economics, differ in two significant ways from those commonly employed in psychology experiments. First, we are looking at a game, so that strategic interactions between teams with antagonistic goals plays a central role. Psychologists typically investigate individual decision problems (puzzles) where each team acts independently to solve a problem. If teams are better able to think from the point of view of others, this insight is relevant for games but not for individual decision problems. Second, our study consists of a number of replications of the same basic problem. In contrast, psychologists typically study one-shot learning problems. If teams differ not in making subjects more sophisticated initially, but instead in how fast subjects learn to become sophisticated, as our experimental results suggest, the ability of teams to meet or beat the truth wins norm will only be realized once they have gained sufficient experience. Both of these factors can also help explain the positive cross-game learning found with teams here, compared to the absence of positive cross-game learning in the typical psychology experiment.

Given the strong convergence of teams to equilibrium outcomes, particularly in the more difficult games, one might be tempted to conclude that increased reliance on team play is all that is needed to tidy up some of the embarrassing discrepancies between experimental data and

⁴⁹ Yet a third possibility is that there are significant strategic interactions within teams which spill over into enhanced strategic interactions between teams. However, the team dialogues provide very little evidence for this, as team members seem quite cooperative in their discussions. About the only clear strategic interplay we observe within teams is that at times one team member will give into their partner to play a strategy that they clearly think is wrong, in order to convince the partner that it is wrong (as opposed to continued discussions).

economic theory reported in the literature. However, the limited economics literature on team versus individual play indicates otherwise (see the references cited in Section II). Further, within the structure of the signaling games reported here, we have investigated the ability of teams to overcome violations of equilibrium refinements reported in earlier experiments. In particular, we have conducted treatments similar to the one reported in Cooper et al. (1997b) in which play by individuals violates one of the weakest equilibrium selection criteria - single round elimination of dominated strategies . Using experienced subjects, in two out of two sessions team play converges rapidly to an inefficient separating equilibrium that violates this selection criteria.⁵⁰ In short, the forward induction arguments underlying even very weak equilibrium selection criteria may be far too subtle for teams (as well as individuals) to adhere to.

Although the present paper answers a number of questions regarding team versus individual play in signaling games, many questions remain to be answered. Primary among these are (1) exploring why we observe negative learning transfer in the 1x1 game here unlike the positive learning transfer in the 1x1 game reported in Cooper and Kagel (2004), and (2) developing methods to directly compare the thought processes underlying individual subject play with those for teams. Work in progress on the first question suggests that the context used to frame the games plays a critical role in fostering positive transfer (see Cooper and Kagel, 2003). We are only now beginning work on the second question. From a broader perspective the team procedures employed here, and the instant messaging possibilities of modern laboratory software, open up a number of exciting possibilities for gaining direct insight into behavior in a wide variety of settings.

⁵⁰ These results may be found at <http://www.econ.sbs.ohio-state/kagel/violations.intuitive.pdf>.

**Table 1a:
Monopolist Payoffs**

High Cost Monopolist (MH)			Low Cost Monopolist (ML)		
Monopolist Output	Entrant Response		Monopolist Output	Entrant Response	
	IN	OUT		IN	OUT
1	150	426	1	250	542
2	168	444	2	276	568
3	150	426	3	330	606
4	132	408	4	352	628
5	56	182	5	334	610
6	-188	-38	6	316	592
7	-292	-126	7	213	486

**Table 1b:
Entrant Payoffs, High Cost Entrants**

Entrant's Strategy	Monopolist's Type	
	High Cost	Low Cost
IN	300	74
OUT	250	250

**Table 1c:
Entrant Payoffs, Low Cost Entrants**

Entrant's Strategy	Monopolist's Type	
	High Cost	Low Cost
IN	500	200
OUT	250	250

Table 2
Summary of Experimental Treatments

	1 x1 Treatment	2x2 Treatment
Inexperienced High Cost Entrants	5 sessions 70 subjects	6 sessions 128 subjects
Inexperienced Low Cost Entrants	4 sessions 64 subjects	4 sessions 104 subjects
Experienced Low Cost Entrants	3 sessions 42 subjects	3 sessions 67 subjects
Crossover Sessions ¹ High Cost → Low Cost Entrants	4 sessions 50 subjects	4 sessions 64 subjects

¹ Data for experienced subjects in games with high cost Es come from first cycle in this treatment.

Table 3: Categories for Coding Team Dialogues

Category ¹	Description ²
1	Myopic choice as an M
3i	Pooling as MH type: Pure imitation (No rationale given for why others should be imitated)
3ii	Pooling as an MH type: Idea is to “fool” E team
3iii	Pooling as an MH type: Discussion from point of view of a E team
3iv	Pooling as an MH type: Drawing on own actions as a E team
4i	Separating as an ML type: Pure imitation (No rationale given for why others should be imitated)
4ii	Separating as an ML type: Referring to negative numbers as making it obvious that a 6 or 7 couldn’t be an MH
4iii	Separating as an MH type: Discussion from point of view of Es
4iv	Separating as an ML type: Drawing on own actions as an E
5	Recognizing Separating as an E
5i	E team referring to negative numbers as making it obvious that a 6 or 7 couldn’t be an MH
6	Following Crossover: Recognizing that the change in payoff tables will change Es’ choices
7	MH types choose 2: Recognizing that the high frequency of IN at 3 & 4 make 2 the best choice (not myopic)
8	Reinforcement: Explicit reference to making choice due to the past success/failure of past actions
9	Other Regarding Behavior: Must be intent of decision, not an amusing side effect
10	Using the feedback provided about the decisions of others
11	Contamination: Talked with someone prior to the experiment about what choices to make
12	Expression of out of equilibrium beliefs
12i	Expression of out of equilibrium beliefs: Es will choose IN for anything other than 4
12ii	Expression of out of equilibrium beliefs: Es will choose OUT for anything other than 2
13	Correcting mistakes of partner (e.g. misreading the payoff table, misreading one’s own type, etc.)
14	Convincing a more knowledgeable/sophisticated teammate to deviate from a more profitable strategy
15	Level 1 reasoning by Es: MHs will choose 2, MLs will choose 4
16	Understanding pooling as an E in games with high cost Es

¹ Category 2 was “advocating myopic choice as a E player.” We dropped this category midway through the coding as being too ambiguous.

² For categories where the description starts with a particular action, we are looking for the justification given for taking this action.

Table 4
Subjects Playing Strategically for the First Time
Means of Independent Variables

High Cost Entrant Game, Inexperienced					Low Cost Entrant Game, Inexperienced				
Played Strategically Next Opportunity?	All	No	Yes	Fisher Exact	Played Strategically Next Opportunity?	All	No	Yes	Fisher Exact
M's Choice	3.687	3.692	3.698	.187	M's Choice	5.571	5.273	5.704	.028**
E's Choice (Entry = 1)	.180	.231	.116	.370	E's Choice (Entry = 1)	.286	.636	.148	.005***
Cycle 2	.246	.077	.233	.426	Cycle 2	.429	.364	.407	1.000
Ever Coded Category 1	.418	.423	.407	.829	Ever Coded Category 1	.298	.227	.352	.499
Ever Coded Category 3ii	.123	.115	.140	.291	Ever Coded Category 3iii	.286	.273	.315	.446
Ever Coded Category 3iii	.369	.000	.500	.000***	Ever Coded Category 4ii	.143	.000	.167	.276
Ever Coded Category 10	.557	.346	.628	.113	Ever Coded Category 4iii	.310	.091	.352	.030**
Ever Coded Category 13	.107	.038	.128	.258	Ever Coded Category 5 or 5i	.202	.182	.222	1.000
Ever Coded Category 15	.295	.077	.326	.126	Ever Coded Category 10	.595	.500	.667	.644
					Ever Coded Category 15	.107	.182	.093	.481
# Observations	61	13	43		# Observations	42	11	27	

* statistically significant at the 10% level

** statistically significant at the 5% level

*** statistically significant at the 1% level

Figure Captions

Figure 1: Pooled data from 1x1 sessions for games with high cost Es. Pure strategy pooling and separating equilibria exist. Bars indicate choice frequencies for MHs and MLs for each possible output level, with entry rates for those output levels shown in parentheses. Cycles consist of 12 (8) plays of the game for inexperienced (experienced) subjects.

Figure 2: Pooled data from 2x2 sessions for games with high cost Es. Pure strategy pooling and separating equilibria exist. Bars indicate choice frequencies for MHs and MLs for each possible output level, with entry rates for those output levels shown in parentheses. Cycles consist of 12 (8) plays of the game for inexperienced (experienced) subjects.

Figure 3: Comparing the development of strategic play for MHs in 2x2 with 1x1 sessions in games with high cost Es. Vertical axis shows frequency of MHs choosing outputs 3-5.

Horizontal axis shows cycle of play. Bars give the 90% confidence interval for the truth wins standard.

Figure 4: Pooled data from 1x1 sessions for games with low cost Es. Only pure strategy separating equilibria exist. Bars indicate choice frequencies for MHs and MLs for each possible output level, with entry rates for those output levels shown in parentheses. Cycles consist of 12 (8) plays of the game for inexperienced (experienced) subjects.

Figure 5: Pooled data from 2x2 sessions for games with low cost Es. Only pure strategy separating equilibria exist. Bars indicate choice frequencies for MHs and MLs for each possible output level, with entry rates for those output levels shown in parentheses. Cycles consist of 12 (8) plays of the game for inexperienced (experienced) subjects.

Figure 6: Comparing the development of strategic play for MLs in 2x2 with 1x1 sessions in games with low cost Es. Vertical axis shows frequency of MLs choosing outputs 5-7.

Horizontal axis shows cycle of play. Bars give the 90% confidence interval for the truth wins standard.

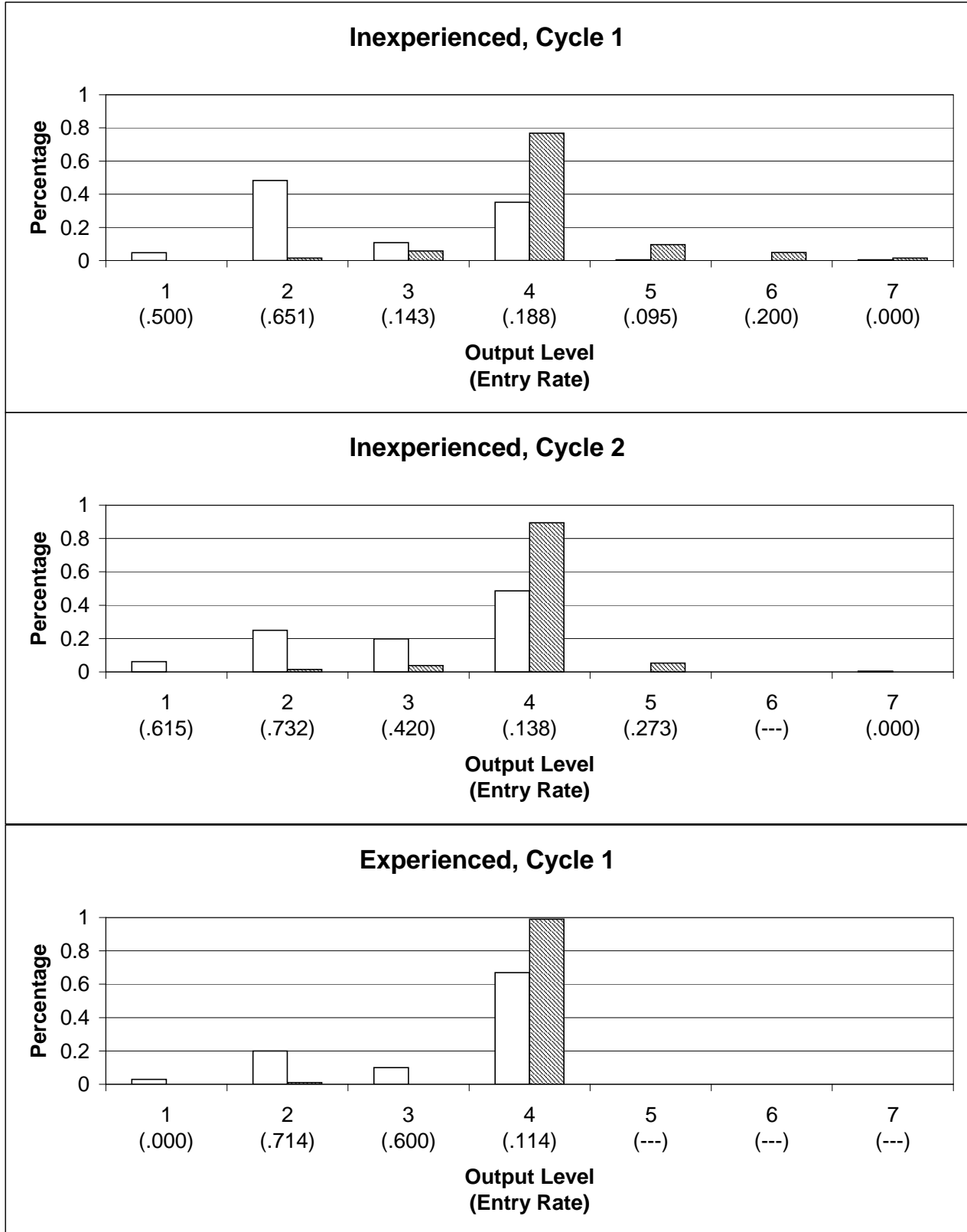
Figure 7: Pooled data from 1x1 and 2x2 sessions for the cross-over treatment. Data are for play following the cross-over where only pure strategy separating equilibria exist. Bars indicate choice frequencies for MHs and MLs for each possible output level, with entry rates for those output levels shown in parentheses. Cycles consist of 8 plays of the game.

Figure 8: Comparing the development of strategic play for MLs in 2x2 with 1x1 sessions *following* the cross-over to games with low cost Es. Vertical axis shows frequency of MLs choosing outputs 5-7. Horizontal axis shows cycle of play. Bars give the 90% confidence interval for the truth wins standard.

Figure 9: Cross-game learning: Comparing the development of strategic play for MLs in the cross-over treatment (data from Figure 7) with play in games with only low cost Es (data from Figures 4 and 5). Vertical axis shows frequency of MLs choosing outputs 5-7. Horizontal axis shows cycle of play.

Figure 1

1 x 1 Sessions, Games with High Cost Entrants



MH

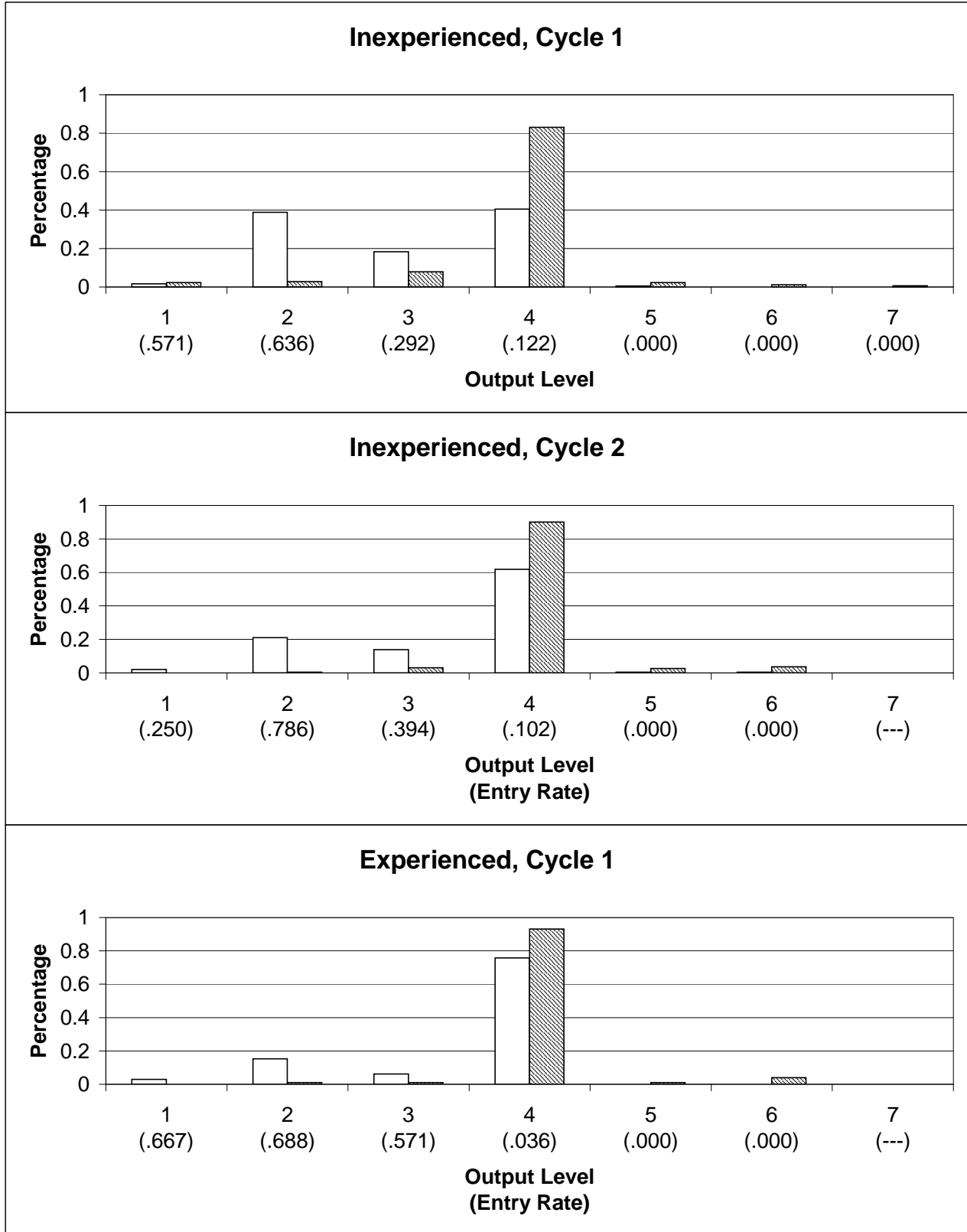


ML



Figure 2

2 x 2 Sessions, Games with High Cost Entrants



MH



ML



Figure 3

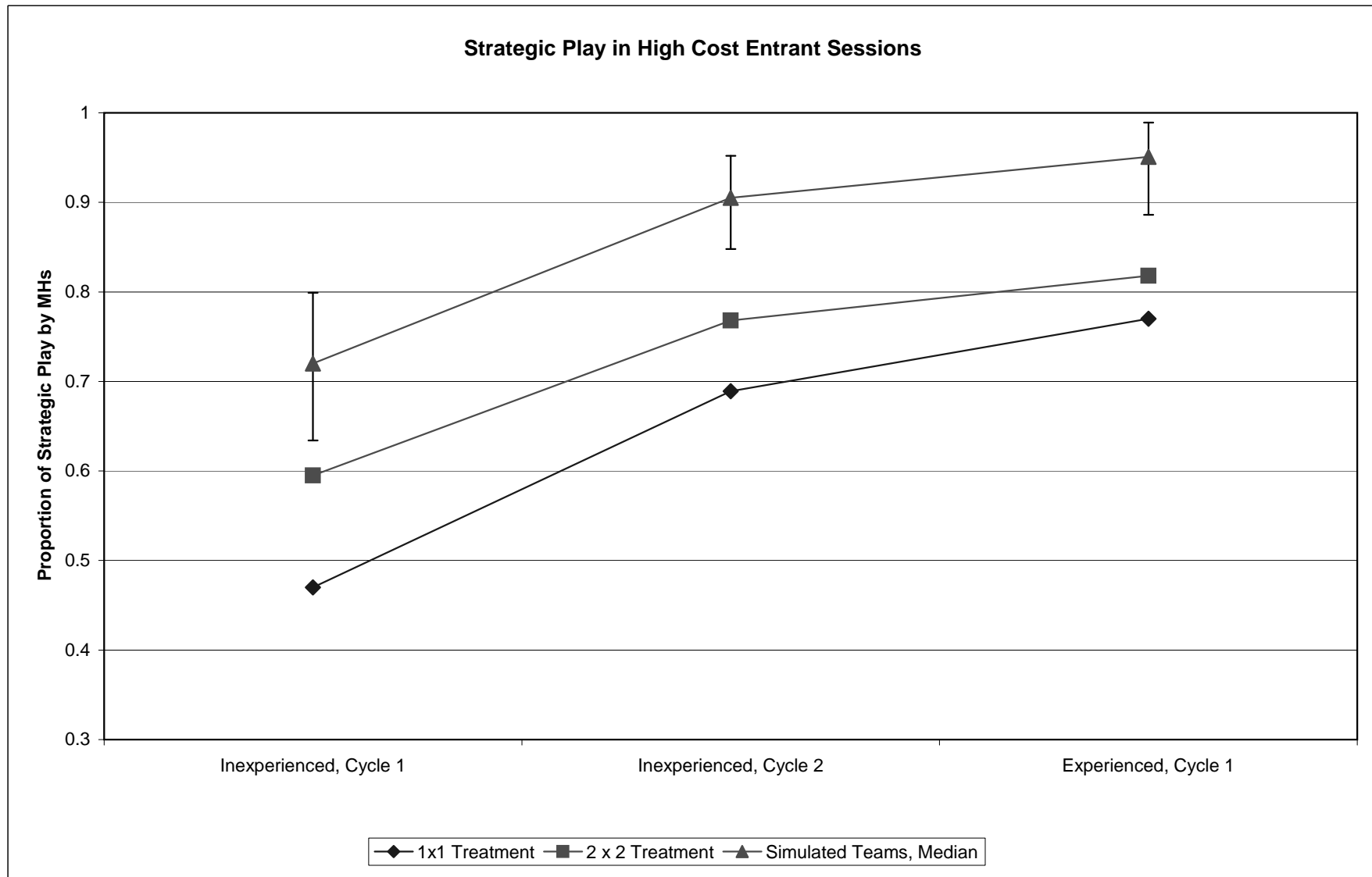
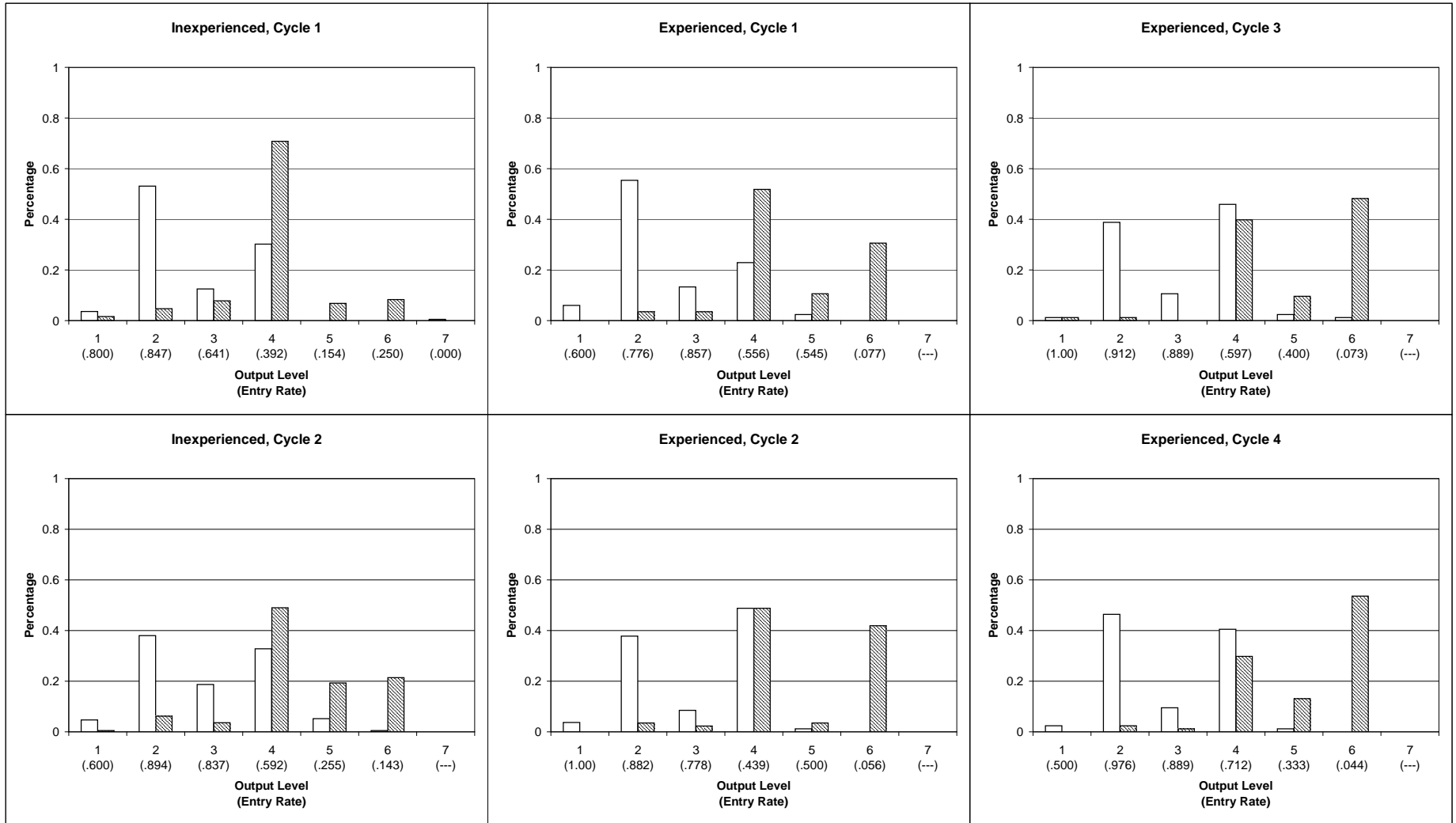


Figure 4

1 x 1 Sessions, Games with Low Cost Entrants



MH

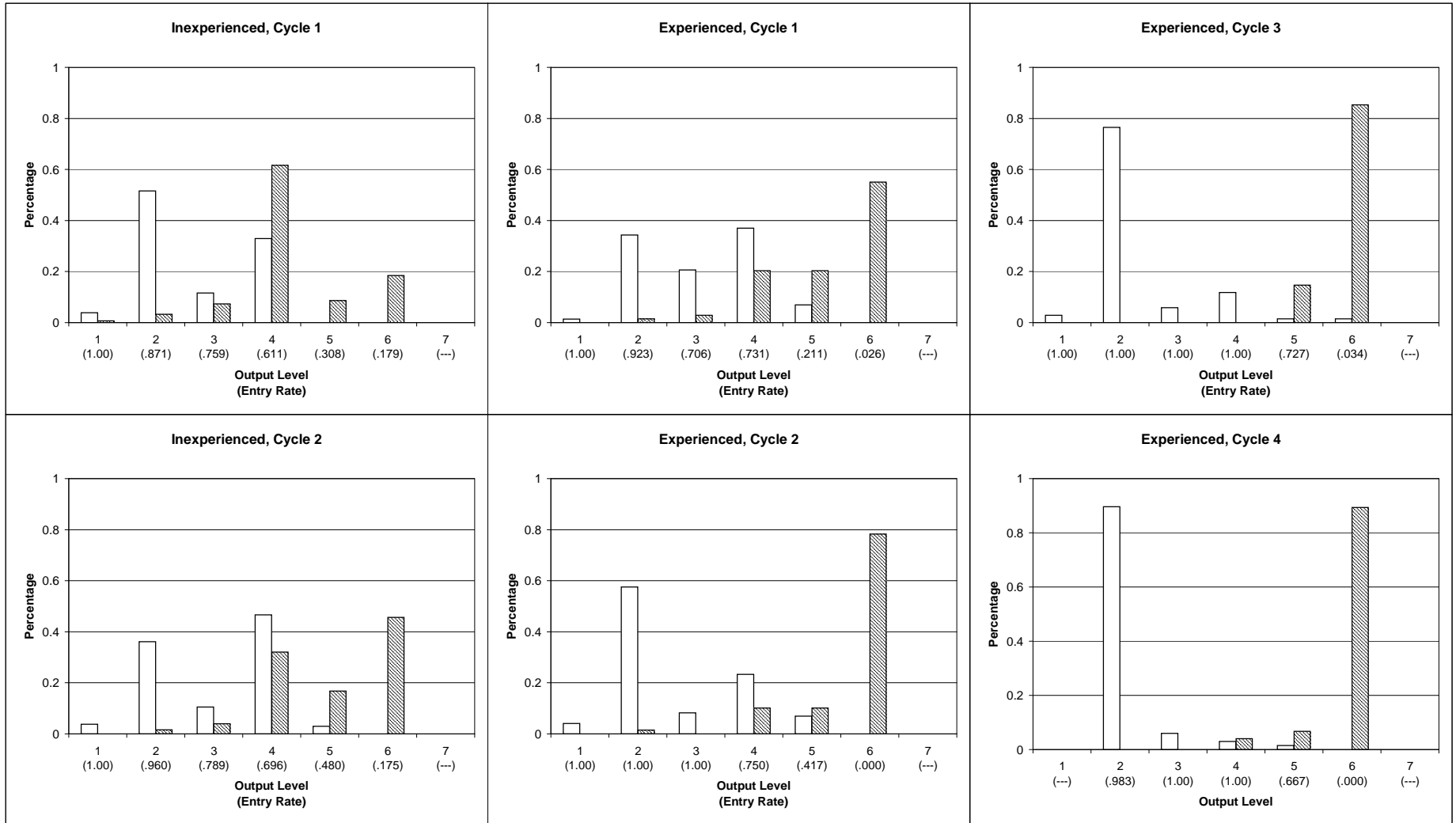


ML



Figure 5

2 x 2 Sessions, Games with Low Cost Entrants



MH



ML



Figure 6

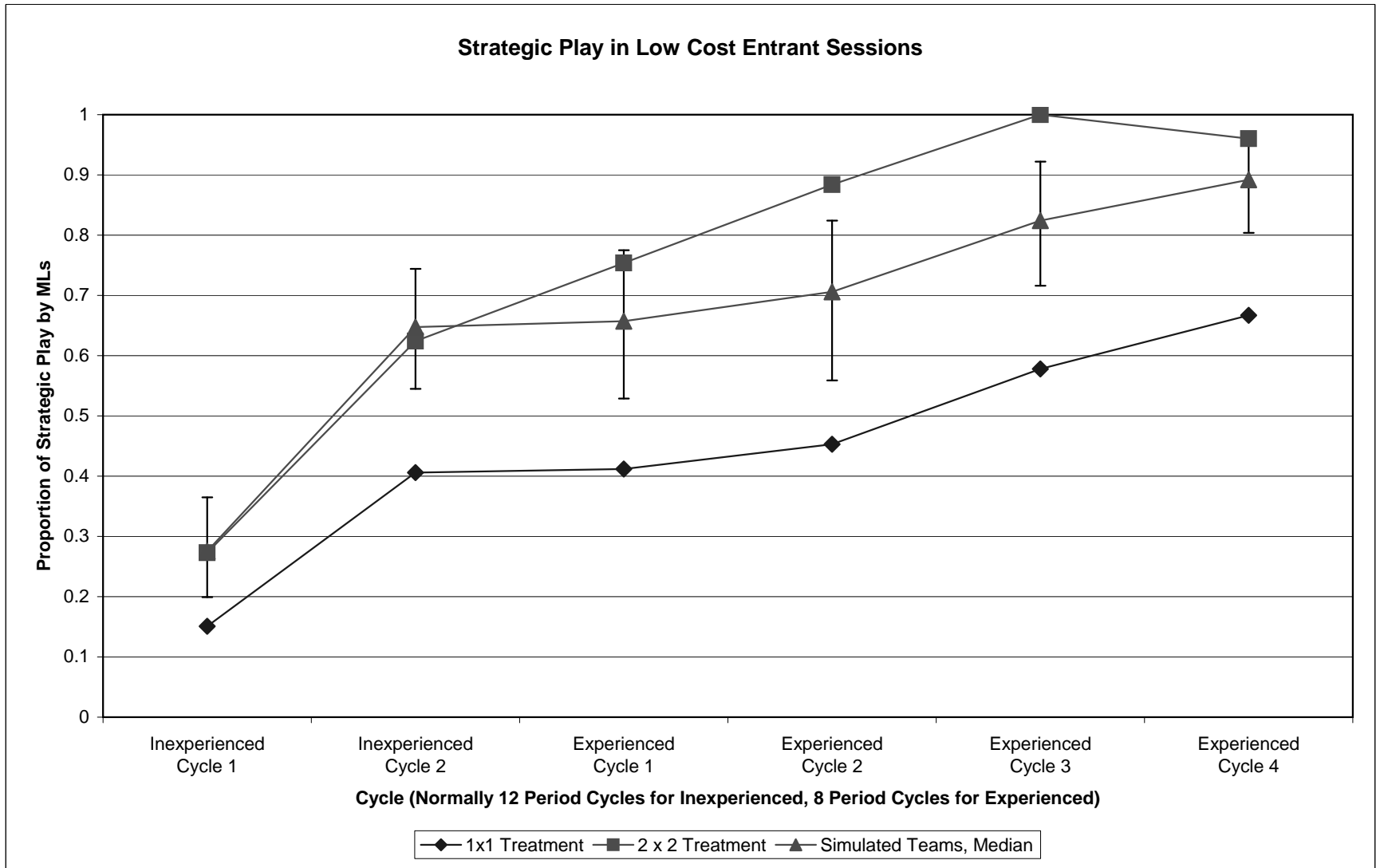


Figure 7

1 x 1 vs. 2 x 2 Sessions, Crossover from High Cost Entrant Game to Low Cost Entrant Game

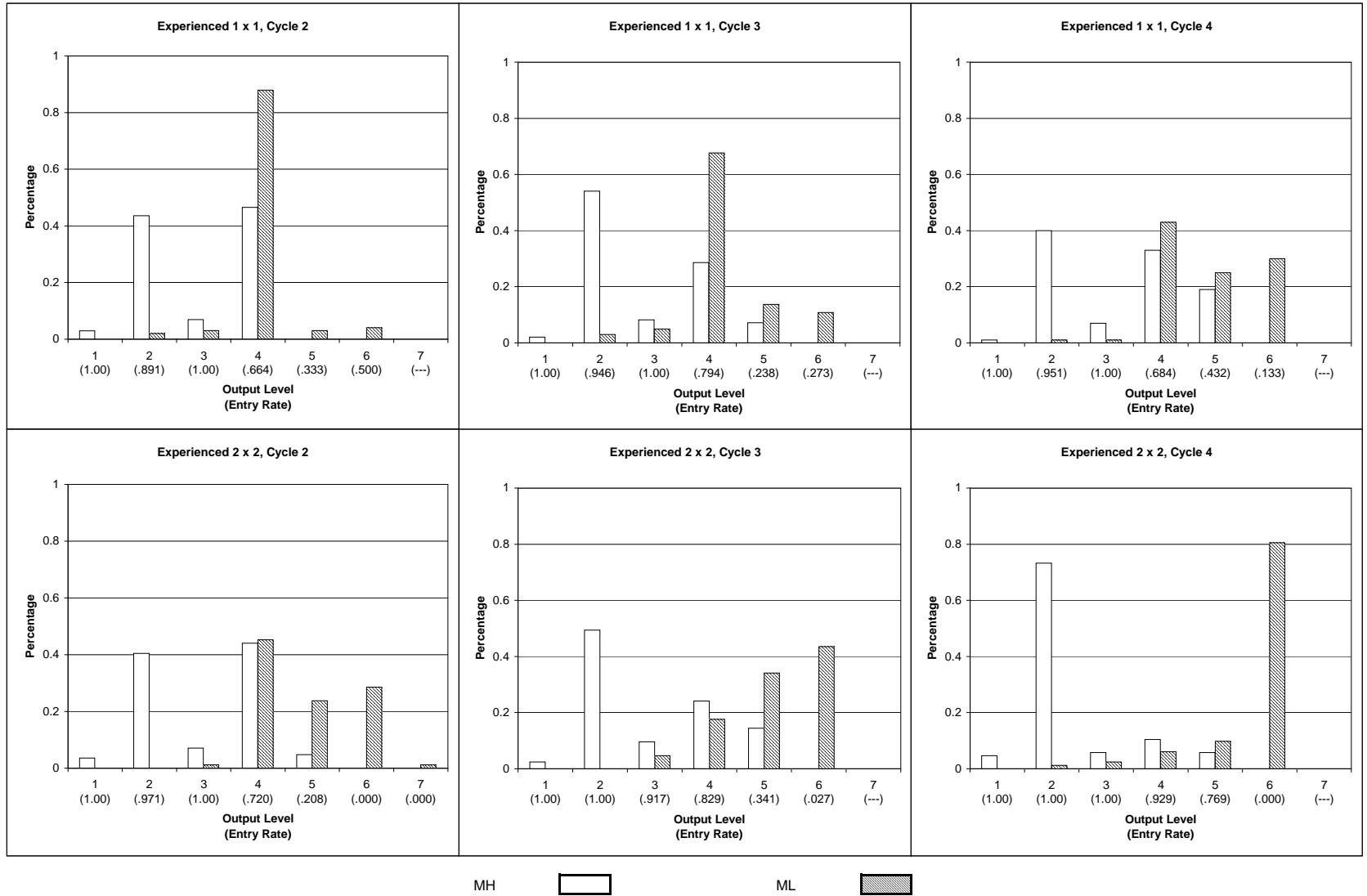


Figure 8

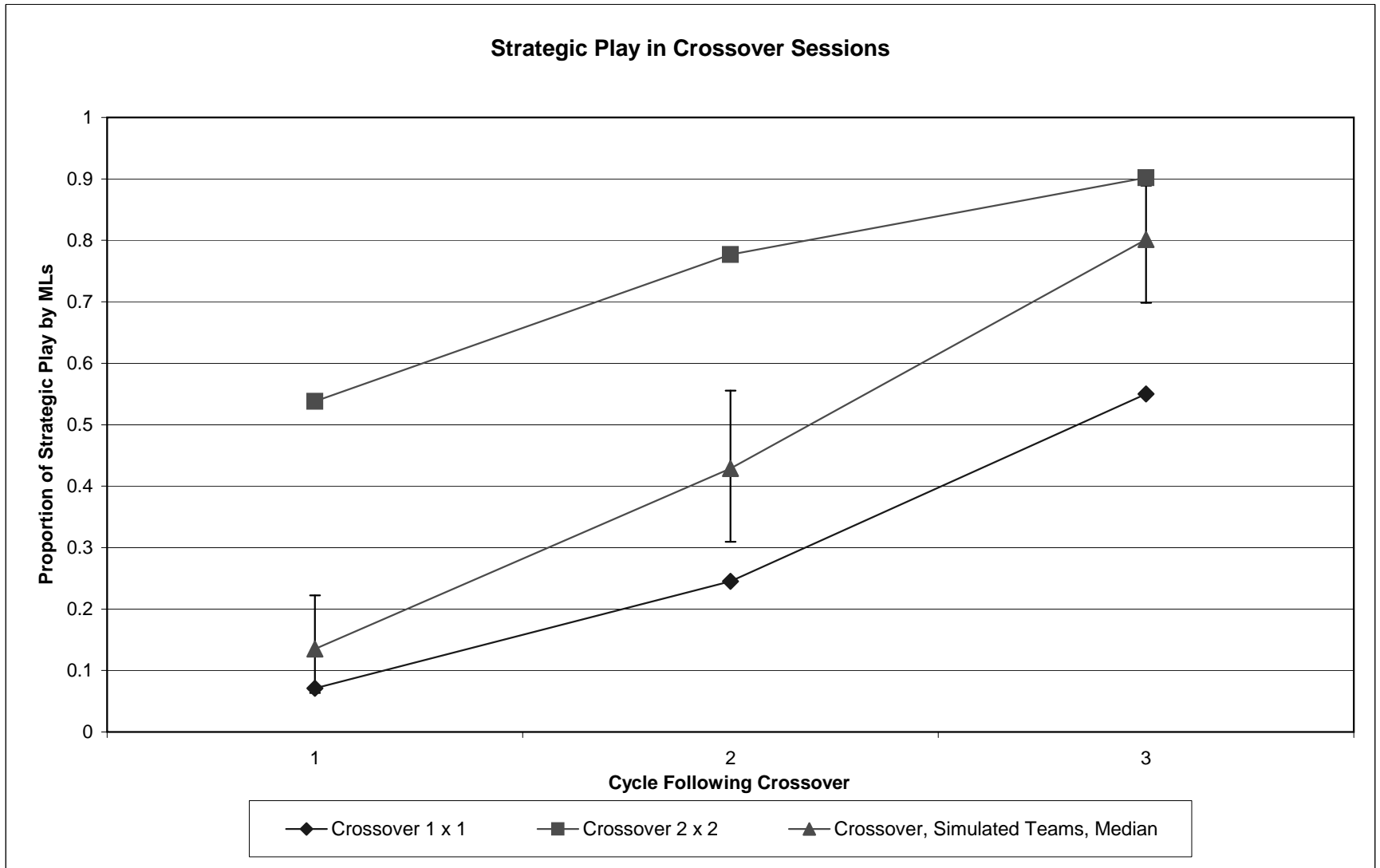
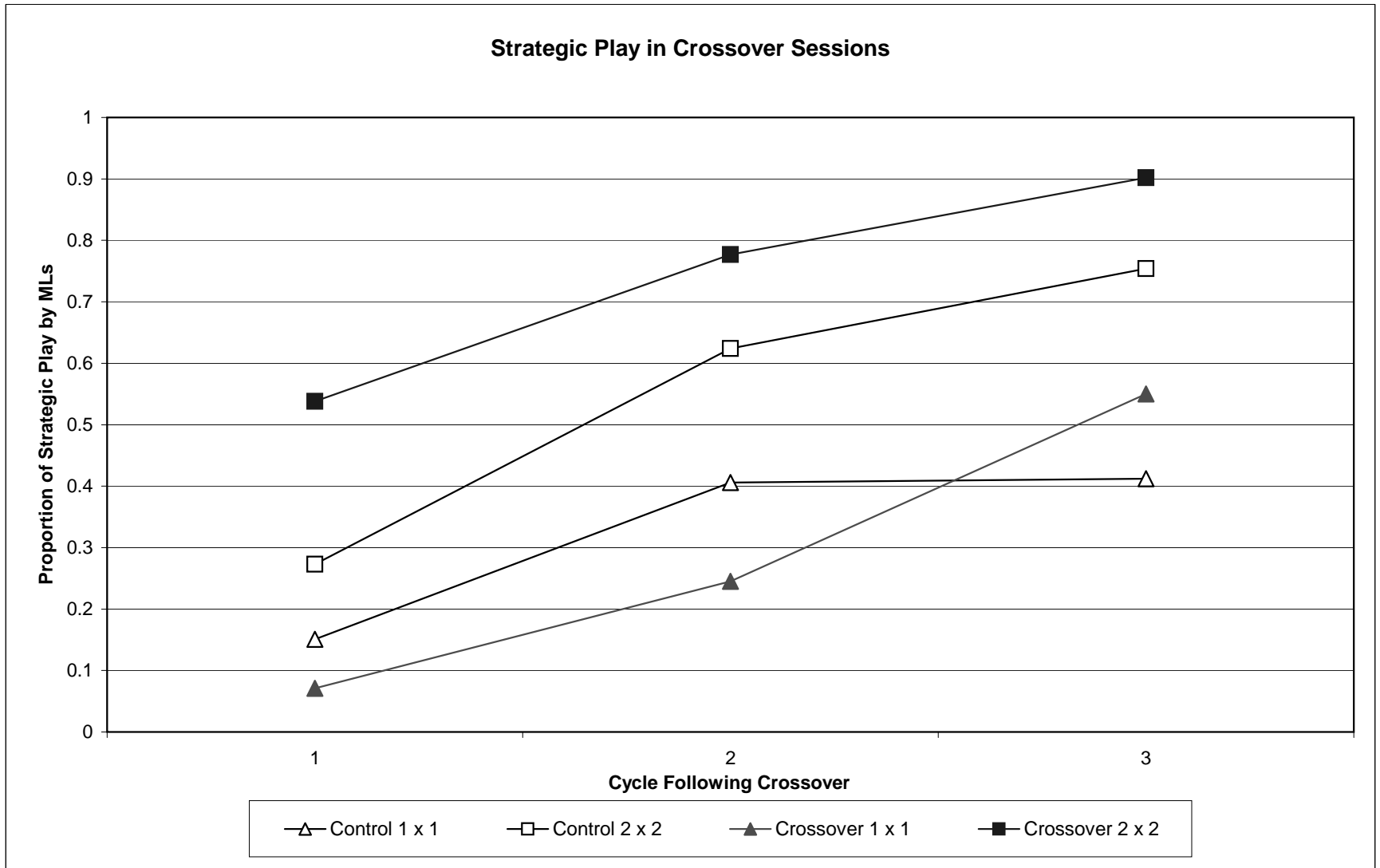


Figure 9



Appendix: This appendix contains the details of the regressions reported in the text. Additional analysis examining the robustness of the regression results to alternative specifications and additional details about the codings of the dialogues are posted at

<http://www.econ.ohio-state.edu/kagel/onlineappendix.twoheads.pdf> as well as on the AER web site.

All of the regressions are probits where the dependent variable is whether Ms have played strategically or not. In games with high cost (low cost) Es the data set only includes choices of MHs (MLs). For MHs (MLs) output levels 3- 5 (5-7) are coded as strategic play.

Because of repeated play by the same subjects the probits must account for the presence of individual effects. This task is complicated by the structure of the 2x2 data, since in the experienced sessions subjects were matched with a new partner (as was the case when the software malfunctioned and had to be restarted). This requires accounting not just for potential correlation between observations from the same team, but also for potential correlation between observations from different teams that shared a common member. In dealing with this problem we employ the relatively conservative approach suggested by Moulton (1986) and Liang and Zeger (1986) to correct the standard errors for clustering.¹ A cluster is defined in terms of “chunks”: any two observations that share a common team member must be included in the same chunk.² For example, suppose that subjects A and B were a team in an inexperienced subject session. In the following experienced subject session, A is teamed with C and B is teamed with D. Any observations that include subjects A, B, C, or D are included in a single chunk, even though observations involving C and D as inexperienced subjects have only a tenuous connection. By taking this conservative approach we bias the results against finding statistical significance.

Games with High Cost Es: Results for games with high cost Es are reported in Table A1. Explanatory

¹ This is the correction used by Stata for clustering. The qualitative results reported are not sensitive to how the individual and team effects are handled – see <http://www.econ.ohio-state.edu/kagel/onlineappendix.twoheads.pdf>.

² For the 1x1 data, all observations from a single subject constitute a chunk.

variables include dummies for the cycle,³ interactions between dummies for the cycle and a dummy for the 2x2 treatment, and a measure of Es' choices. The first inexperienced cycle of the 1 x 1 treatment serves as a base. The 2x2-cycle interaction terms capture differences between behavior by teams in that cycle and behavior by individuals from 1 x 1 sessions in the same cycle. The measure of Es' behavior consists of the difference in the current cycle between the entry rates for output levels 2 and 4, which serves as a proxy for the incentives for strategic play. This is calculated across all observations from the same session, since subjects got to see all entry decisions in their session.

Model 1 tests for 2x2 effects without controlling for entry rates. A marginally significant 2x2 effect is found in the first inexperienced cycle, with no significant effects thereafter. Model 2 adds the control for Es' behavior. Ms are only weakly responsive to the entry rate differential between 2 and 4, as the parameter value is far from statistically significant at conventional levels. The impact on the estimated 2x2 effects is minimal. Model 3 checks whether the impact of entry rate differentials differs between the 2x2 and 1x1 treatments. While the difference is impressive, the parameters are estimated imprecisely so that we cannot reject a null hypothesis of no difference in responsiveness between treatments.

Games with Low Cost Es: The probits for this treatment are reported in Table A2. The regression specification is similar to Table A1 except that the entry rate differential used is between output levels 4 and 6, reflecting the incentives to play strategically for MLs. Note that the regressions include a dummy for the third cycle of inexperienced play, but no interaction term between this dummy and the 2x2 dummy since none of the inexperienced 2x2 sessions included a third cycle.

Model 1 tests for differences in the frequency of strategic play without any controls for entry rates. The results strongly support Hypothesis 1, as the 2x2 interactions are strongly significant in all cycles.⁴ Model 2 adds the entry rate differential as a control for incentives to play strategically as an ML. This is

³ In regressions not reported here, we added controls for the varying length of a cycle. These additions are statistically significant but do not affect the qualitative results we report.

⁴ The interaction term for the third cycle of the experienced sessions had to be dropped because there was no variation in the data – all MLs in the relevant cell played strategically. This can be interpreted as yielding an arbitrarily large parameter estimate that would be statistically significant at any desired level.

statistically significant at the 1% level. With the inclusion of the entry rate differential, the 2x2 interaction terms are weakened across the board with several of them failing to achieve statistical significance. There is still a clear 2x2 effect, but at least part of this effect must be attributed to changes in the incentives to play strategically as an ML. Model 3 permits differing sensitivity to entry rate differentials between the 2x2 and 1x1 treatments. The results indicate that subjects are roughly twice as sensitive to the entry rate differential in the 1x1 treatment, but the standard errors of the estimates are sufficiently large that the null hypothesis of no difference cannot be rejected at the 10% level. Thus, we regard the estimates as suggestive. With this in mind, the strength of the 2x2 effect is greater overall in Model 3 than in Model 2.

Cross-Over Sessions: The first half of Table A3 reports probits comparing the 2x2 treatment with the 1x1 treatment for the crossover sessions. The data set includes all plays by MLs following the crossover. The regression specification is the same as for games with low cost Es. However, since teams are never rematched here, clustering at the chunk level is equivalent to clustering at the team level.

Model 1 confirms the obvious – MLs play strategically significantly more often following the crossover in 2x2 sessions than in 1x1 sessions. Model 2 indicates that this difference is robust to controls for differing entry rates.

The second half of Table A3 contains probits for the cross-game learning effects. This includes sessions where subjects only have experience with the low cost E game, which serve as controls here, as well as the cross-over sessions following the crossover. The variables of interest here are the dummies for the three cycles following the crossover. The structure of the dummies is such that these three parameter estimates capture, respectively, the differences between the first, second and third cycle following the crossover and the first, second and third of inexperienced play in the controls.

For the 1x1 data, Model 1 shows a small but statistically significant *negative* crossover effect. Model 2 indicates that this negative effect is robust to controls for the entry rate differential. Thus, previous experience games with high cost Es inhibits strategic play in games with low cost Es.

For the 2x2 data, Model 1 shows a statistically significant *positive* crossover effect. As shown in Model 2, the size of this effect is substantially reduced by adding controls for entry rate differentials, only retaining statistical significance for the first cycle following the crossover. It is difficult to sort out causality here – previous experience with the high cost entrant game leads both to immediately higher levels of strategic play and to immediately higher entry rate differentials. Based on the dialogues, we are disinclined to attribute the higher rates of strategic play solely to differences in feedback.

Comparing the cross-game learning probits, it is worth noting that once again the responsiveness to entry rates is substantially larger in the 1x1 treatment. Even though this result never quite achieves statistical significance, its pervasiveness in games with low cost Es suggests that teams rely less on the feedback than do individuals. This implies differences in the basic process used by teams and individuals to reason about the limit pricing game.

Table A1
 Probit Regressions, High Cost Entrant Sessions
 Standard Errors Corrected for Clustering at the “Chunk” Level

Dependent Variable: Strategic Choice by MHs
 (1003 obs, 174 teams)

Variable	Model 1	Model 2	Model 3
Constant	-.077 (.129)	-.241 (.240)	-.154 (.263)
Inexperienced Cycle 2	.568 ^{***} (.133)	.525 ^{***} (.140)	.548 ^{***} (.140)
Experienced	.815 ^{***} (.189)	.748 ^{***} (.204)	.784 ^{***} (.208)
2x2 * Inexperienced Cycle 1	.316 [*] (.187)	.299 (.187)	.000 (.529)
2x2 * Inexperienced Cycle 2	.240 (.211)	.197 (.223)	-.206 (.722)
2x2 * Experienced	.170 (.274)	.164 (.269)	-.232 (.809)
Entry Rate Differential		.359 (.425)	
1 x 1 * Entry Rate Differential			.168 (.472)
2x2 * Entry Rate Differential			.777 (.852)
Log Likelihood	-609.56	-608.64	-608.07

- * statistically significant at the 10% level
- ** statistically significant at the 5% level
- *** statistically significant at the 1% level

Table A2
 Probit Regressions, Low Cost Entrant Sessions
 Standard Errors Corrected for Clustering at the “Chunk” Level

Dependent Variable: Strategic Choice by MLs
 (1375 obs, 176 teams)

Variable	Model 1	Model 2	Model 3
Constant	-1.032 ^{***} (.150)	-.451 ^{***} (.169)	-.385 ^{**} (.180)
Inexperienced Cycle 2	.795 ^{***} (.157)	.270 (.177)	.212 (.188)
Inexperienced Cycle 3	1.006 ^{***} (.244)	.149 (.311)	.040 (.348)
Experienced Cycle 1	.809 ^{***} (.221)	.190 (.238)	.119 (.251)
Experienced Cycle 2	.915 ^{***} (.235)	.605 ^{**} (.236)	.585 ^{**} (.238)
Experienced Cycle 3	1.230 ^{***} (.227)	.638 ^{**} (.248)	.576 ^{**} (.258)
Experienced Cycle 4	1.463 ^{***} (.236)	.628 ^{**} (.275)	.526 [*] (.296)
2x2 * Inexperienced Cycle 1	.424 ^{**} (.189)	-.046 (.195)	-.149 (.207)
2x2 * Inexperienced Cycle 2	.553 ^{***} (.197)	.463 ^{**} (.205)	.469 ^{**} (.200)
2x2 * Experienced Cycle 1	.909 ^{***} (.260)	.709 ^{***} (.225)	.810 ^{***} (.232)
2x2 * Experienced Cycle 2	1.312 ^{***} (.336)	.480 (.309)	.669 ^{**} (.330)
2x2 * Experienced Cycle 3	Dropped (No variation)	Dropped (No variation)	Dropped (No variation)
2x2 * Experienced Cycle 4	1.320 ^{***} (.307)	.934 ^{***} (.399)	1.228 ^{***} (.393)
Entry Rate Differential		1.791 ^{***} (.443)	
1 x 1 * Entry Rate Differential			2.076 ^{***} (.622)
2x2 * Entry Rate Differential			1.036 ^{***} (.276)
Log Likelihood	-751.45	-703.53	-700.84

* statistically significant at the 10% level
 ** statistically significant at the 5% level
 *** statistically significant at the 1% level

Table A3
 Probit Regressions, Crossover Sessions, Low Cost Entrants
 Standard Errors Corrected for Clustering at the “Chunk” Level

Dependent Variable: Strategic Choice by MLs

Test of Truth Wins Norm			Tests of Cross-Game Learning				
	All Crossover Data (552 obs, 92 teams)			1 x 1 Treatment (1119 obs, 116 teams)		Team Treatment (739 obs, 152 teams)	
Variable	Model 1	Model 2	Variable	Model 1	Model 2	Model 1T	Model 2T
Constant	-1.471*** (.218)	-1.068*** (.287)	Constant	-1.032*** (.150)	-.626*** (.164)	-.603*** (.120)	-.327*** (.123)
Crossover Cycle 2	.781*** (.240)	.413 (.298)	Cycle 2 ¹	.795*** (.157)	.281 (.178)	.919*** (.139)	.827*** (.129)
Crossover Cycle 3	1.596*** (.238)	1.244*** (.265)	Cycle 3 ²	.809*** (.220)	.163 (.246)	1.289*** (.151)	1.079*** (.135)
Team * Crossover Cycle 1	1.560*** (.281)	1.023*** (.375)	Control, Inexperienced Cycle 3	1.006*** (.244)	.107 (.319)		
Team * Crossover Cycle 2	1.450*** (.299)	1.162*** (.343)	Control, Experienced Cycle 2	.915*** (.234)	.597** (.237)	1.798*** (.176)	1.409*** (.232)
Team * Crossover Cycle 3	1.170*** (.308)	.738* (.404)	Control, Experienced Cycle 3	1.230*** (.227)	.614** (.253)	Dropped (No Variation)	
Entry Rate Differential		1.139* (.623)	Control, Experienced Cycle 4	1.463*** (.236)	.589** (.285)	2.353*** (.190)	1.863*** (.284)
1 x 1 * Entry Rate Differential			Crossover, Cycle 1	-.439* (.264)	-.648* (.370)	.692*** (.215)	.415* (.221)
Team * Entry Rate Differential			Crossover, Cycle 2	-.453* (.258)	-.627** (.255)	.444* (.250)	.153 (.289)
			Crossover, Cycle 3	.349 (.274)	.322 (.265)	.610** (.296)	.323 (.305)
			Entry Rate Differential		1.901*** (.490)		1.004*** (.339)
Log Likelihood	-280.30	-274.94	Log Likelihood	-655.43	-601.54	-376.00	-371.91

* statistically significant at the 10% level

** statistically significant at the 5% level

*** statistically significant at the 1% level

¹ Equals 1 for observations in the second inexperienced cycle of the control sessions and the second cycle following the crossover in the crossover sessions.

² Equals 1 for observations in the first experienced cycle of the control sessions and the third cycle following the crossover in the crossover sessions.

References

- Alchian, Armen A. and Demsetz, Harold.** “Production, Information Costs, and Economic Organization.” *American Economic Review*, December 1972, 62(5), pp. 777-95.
- Battalio, Raymond C., Samuelson, Larry, and Van Huyck, John.** “Optimization Incentives and Coordination Failure in Laboratory Stag Hunt Games.” *Econometrica*, May 2001, 69(3), pp. 749-64.
- Brandts, Jordi and Holt, Charles A.** “An Experimental Test of Equilibrium Dominance in Signaling Games.” *American Economic Review*, December 1992, 82(5), pp. 1350-65.
- Bornstein, Gary and Yaniv, Ilan.** “Individual and Group Behavior in the Ultimatum Game: Are Groups More “Rational” Players?” *Experimental Economics*, June 1998, 1(1), pp.101-08.
- Camerer, Colin F.; Ho, Tech-Hua, and Chong, Juin-Kuan.** “Sophisticated Experience-Weighted Attraction Learning and Strategic Teaching in Repeated Games,” *Journal of Economic Theory*, May 2002, 104(1), pp. 137-88.
- Cason, Timothy N. and Mui, Vai-Lam.** “A Laboratory Study of Group Polarization in the Team Dictator Game,” *Economic Journal*, September 1997, 107(444), 1465-83.
- Cho, In-Koo and Kreps, David M.** “Signaling Games and Stable Equilibria.” *Quarterly Journal of Economics*, May 1987, 102(2), pp.179-221.
- Cooper, David J.; Garvin, Susan and Kagel, John H.** “Signaling and Adaptive Learning in an Entry Limit Pricing Game.” *RAND Journal of Economics*, Winter 1997a, 28(4), pp. 662-83.
- _____. “Adaptive Learning vs. Equilibrium-Refinements in an Entry Limit Pricing Game.” *The Economic Journal*, May 1997b, 107(442), pp. 553-75.
- Cooper, David J. and Kagel, John H.** “Lessons Learned: Generalizing Learning Across Games.” *American Economic Review*, May 2003, 93(2), pp. 202-07.

_____. "Learning and Transfer in Signaling Games." Working paper, Case Western Reserve University, 2004.

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

Cox, James C. "Trust, Reciprocity, and Other Regarding Preferences: Group vs. Individuals and Males vs. Females," in Rami Zwick and Amnon Rapoport, eds., *Advances in Experimental Business Research*, Dordrecht: Kluwer Academic Publishers, 2002, pp 331-50.

Cox, James C. and Hayne, Stephen C. "Barking Up the Wrong Tree: Are Small Groups Rational Agents?" Mimeo, University of Arizona, 2002.

Davis, James, H. "Some Compelling Intuitions About Group Consensus Decisions, Theoretical and Empirical Research, and Interpersonal Aggregation Phenomena: Selected Examples, 1950-1990." *Organizational Behavior and Human Decision Processes*, June 1992, 52(1), pp. 3-38.

Hill, Gayle W. "Group Versus Individual Performance: Are $N + 1$ Heads Better Than One?" *Psychological Bulletin*, May 1982, 91(3), pp. 517-39.

Kagel, John H. And Levin, Dan. *Common Value Auctions and the Winner's Curse*. Princeton, NJ: Princeton University Press, 2002.

Kerr, Norbert L.; MacCoun, Robert J. and Kramer, Geoffrey P. "Bias in Judgment: Comparing Individuals and Groups." *Psychological Review*, October 1996, 103(4), pp. 687-719.

Kocher, Martin G and Sutter, Matthias. "When the 'Decision Maker' Matters: Individual Versus Team Behavior in Experimental 'Beauty-Contest' Games" *Economic Journal* (forthcoming).

Liang, Kung-Yee, and Scott L. Zeger. "Longitudinal Data Analysis Using Generalized Linear

Models,” *Biometrika*, April 1986, 73(1), pp.13-22.

Lorge, Irving and Solomon, Herbert. “Two models of group behavior in the solution of eureka-type problems.” *Psychometrika*, June 1955, 20(2), pp. 139-48.

Milgrom, Paul. and Roberts, John. “Limit Pricing and Entry Under Incomplete Information: An Equilibrium Analysis.” *Econometrica*, March 1982, 50(2), pp. 443-59.

_____. “Adaptive and Sophisticated Learning in Normal Form Games,” *Games and Economic Behavior*, February 1991, 3(1), pp. 82-100.

Moulton, Brent. "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, August 1986, 32(3), pp. 385-97.

Roth, Alvin E. and Erev, Ido, “Learning in extensive form games: Experimental data and simple dynamic models in the intermediate term,” *Games and Economic Behavior*, January 1995, 8 (1), pp. 164-212.

Shaw, Majorie. “A comparison of individuals and small groups in the rational solution of complex problems.” *American Journal of Psychology*, July 1932, 44(3), pp. 491-504.

Solomon, Gavriel and Perkins, David N. “Rocky Roads to Transfer: Rethinking Mechanisms of a Neglected Phenomenon.” *Educational Psychologist*, Spring 1989, 24(2), pp. 113-42.

Stahl, Dale O. II “Boundedly Rational Rule Learning in a Guessing Game,” *Games and Economic Behavior*, October 1996, 16(2), pp. 303-30.

Sutter, Matthais. “Are Four Heads Better than Two? An Experimental Beauty-Contest Game with Teams of Different Sizes,” Max Planck Institute for Research into Economic Systems, Discussion Paper No. 15-2004, 2004.