Experiments on the Emergence of Leadership in Teams*

Koji Abe† *Osaka University*
Hajime Kobayashi‡ *Osaka Prefecture University*
Hideo Suehiro§ *Kobe University*

August 9, 2010

Abstract

We study experimentally the emergence of leadership through endogenously formed leader–follower relationships in team production under asymmetric information. In a treatment that theoretically admits leadership, we observed that leadership emerged in a pattern of a particular sequential equilibrium: a subject with a higher expectation of team productivity leads, and one with a lower expectation follows. In a control treatment in which incompleteness of information is removed and one in which payoffs are changed to incorporate a prisoner's dilemma situation, leadership did not emerge. The results support the endogenous signaling theory of leadership in teams proposed by Kobayashi and Suehiro (2005, 2008).

JEL Classification: C72, C92, D82

Keywords: Leadership, Experiment, Endogenous signaling, Endogenous timing, Team production

---

*We are grateful to Masaki Aoyagi, Patricio Dalton, Eric van Damme, Eline van der Heijden, Toshiji Kawagoe, Naoko Nishimura, Jan Potters, and Tatsuyoshi Saijo for their helpful comments. We are also grateful to seminar participants at CentER of Tilburg University, ISER of Osaka University, KIER of Kyoto University, Nagoya University, VCASI, FESAMES 2009, and the JEA autumn meeting 2009. We gratefully acknowledge financial support from the Japan Society for the Promotion of Science under grant No.(C)18530138.

†Institute of Social and Economic Research, Osaka University, 6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan; `abe@iser.osaka-u.ac.jp`.

‡School of Economics, Osaka Prefecture University, 1-1, Gakuen, Sakai, Osaka, 599-8531, Japan; `khajime@eco.osakafu-u.ac.jp`.

§Graduate School of Business Administration, Kobe University, 2-1, Rokkodai, Nada, Kobe, Hyogo, 657-8501, Japan; `suehiro@kobe-u.ac.jp`.

1 Introduction

An organization achieves its economic performance through informal processes as well as formal ones. One of the most important informal processes is leadership in teams. In many actual teams, leadership enhances team performance. In such a case, a member of a team voluntarily moves ahead of the other members to commit to making a higher level of contribution to the team, and this leadership behavior influences the other members to make correspondingly higher levels of contributions. This mitigates the free-rider problem in teams.

In spite of its significant role in solving incentive problems, the mechanism of leadership in teams remains unexplored by economics in contrast with the development of theories of formalized incentive mechanisms for teams. The fundamental issue of the study is the emergence of leadership in teams. Leadership emerges in teams without a formalized enforcement mechanism because neither voluntary leadership behavior nor influenced followership behavior are observable or verifiable by a third party.

Kobayashi and Suehiro (2005, 2008) proposed an endogenous signaling theory of emergence of leadership. They studied a team production game in which each agent is endowed with partial private information about team productivity and selects a level of effort at a time chosen by the agent. They showed the existence of a sequential equilibrium in which a leader–follower relation emerges and the follower responds to the leader's choice. The theory implies that leadership emerges in teams through a process in which agents are endogenously sorted to be either a signal sender or a signal receiver, depending on the information they have.

Unfortunately, as is common to many signaling games, there are multiple sequential equilibria in the game of Kobayashi and Suehiro (2005, 2008), and one of them is a sequential equilibrium in which no agent takes the lead voluntarily. The endogenous signaling theory of leadership alone does not explain decisively the emergence of leadership in teams.

We experimentally study a simplified version of the game of Kobayashi and Suehiro (2005, 2008) to find whether leadership emerges in the experimental teams and whether subjects behave as the endogenous signaling theory of emergence of leadership predicts. To this end, we first develop a theoretical characterization of sequential equilibrium in the simplified game. We conclude that there is an interval defined by a set of team productivity parameters for the marginal cost of effort and if the marginal cost for a team falls within the interval, then there exist three kinds of sequential equilibria supporting emergence of leadership in the team in addition to the no-leadership equilibrium.

Based on this theoretical conclusion, we design the following experiment. Our experiment consists of one base treatment and two control treatments. For the base treatment, we choose a set of parameters for a team that satisfies the leadership condition. We design a game with these parameters and take it as the base treatment in which the endogenous signaling theory of leadership predicts emergence of leadership. For one control treatment, we remove the incompleteness of information about team productivity. If leadership emerges as a signaling process in the base treatment as the theory presumes, leadership should not emerge under the complete information treatment. For the other control treatment, we change a parameter of effort cost so as not to satisfy the leadership conditions. If leadership emerges as a sequential equilibrium as the theory postulates, leadership should also not emerge.

For each treatment, we statistically analyze whether the subjects behave as the endogenous signaling theory of leadership predicts. The subjects can be assumed to follow some behavior strategy independently, because we implemented our experiment by the anonymous random matching. We econometrically test whether this unknown behavior strategy corresponds to an equilibrium of the theory. For this purpose, we apply the error rate analysis introduced by Harless and Camerer (1995). According to this analysis, we postulate the null hypothesis that subjects behave according to a tested equilibrium strategy with error. The alternative hypothesis is that subjects follow some unknown behavior strategy. We estimate unknown parameters under each hypothesis by the maximum likelihood method. Based on the estimated parameters, we test the null hypothesis against the alternative hypothesis by the likelihood ratio test.

Some of the tested equilibria are asymmetric equilibria. For this test, we postulate the null hypothesis that a role is assigned to a subject, and he follows the strategy corresponding to the role. This hypothesis allows some correlation in role assignment. Therefore, in addition to the alternative hypothesis described above, we also try to test with another alternative hypothesis that a pair of subjects behave according to some unknown correlated strategy.

We also test a possibility that the data are generated by a mixture of multiple equilibria. For this test, we postulate the null hypothesis that a pair of subjects somehow coordinate on one of the equilibria and play according to the selected equilibrium. As for the tests of asymmetric equilibria, we also try to test with another alternative hypothesis that a pair of subjects behave according to some unknown correlated strategy. Although Rapoport and Amaldoss (2008) discussed difficulties in testing multiple equilibria including asymmetric equilibria, we statistically test this type of multiple equilibria.

The result is that leadership emerged in the base treatment in the pattern of a particular sequential equilibrium; that is, *leadership by confidence* as explored by Kobayashi and Suehiro (2005, 2008). Leadership by confidence is a symmetric strategy profile with the following properties. When an agent is endowed with private information that indicates high team productivity (H-type), that agent invests a high level of effort in period 1. When the private information indicates low team productivity (L-type), the agent suspends commitment in period 1. Then, the agent responds in period 2 to his partner's investment of a high level of effort in period 1 with the same high level of effort. He responds with a low level of effort to his partner's low level of effort or suspension of commitment. When one agent is of H-type and the other agent is of L-type, a leader–follower relationship endogenously emerges along the equilibrium path of leadership by confidence, and the H-type agent leads with a high level of effort while the L-type agent follows the leader with a corresponding high level of effort.

In the information-control treatment, leadership did not emerge as predicted by the theory. Combining the results of the base experiment with those of the information-control treatment, we can state that the existence of private information is necessary for leadership to emerge through an endogenously formed leader–follower relationship.

This supports the claim of Kobayashi and Suehiro (2005, 2008) that leadership emerges as a signaling process.

In the payoff-control treatment, leadership did not emerge so frequently but did not disappear completely. It is inferred that this disobedience to the sequential equilibrium stems from conditional cooperation by some subjects. Conditional cooperation is an attitude of individuals recently found in several experimental researches on those games in which players' rationality is in conflict with Pareto optimality (e.g. Fischbacher, Gächter, and Fehr (2001) and Herrmann and Thöni (2009)). This means that an agent is willing to cooperate if his opponents cooperate to realize Pareto optimality, and he is reluctant otherwise. Furthermore, if a subject believes that conditionally cooperative subjects exist, that subject may intend to exercise leadership based on the expectation of facing a conditionally cooperative follower. We call this type of leadership *leadership-for-conditional-cooperation*.

We estimate the proportions of these behavior modes by the error rate analysis.[1] When we postulate a model that combines with rational choices the behaviors related to conditional cooperation, the model fits to the experimental data significantly well. This result implies that a proportion of the subjects adopts conditional cooperation behavior, and another proportion adopts leadership-for-conditional-cooperation. However, the latter proportion is only 25%. Therefore, we can infer that the emergence of leadership in the base treatment is mainly driven by the endogenous signaling motives that Kobayashi and Suehiro (2005, 2008) advocate.

The rest of the paper is organized as follows. Section 2 reviews the related literature. Section 3 presents the endogenous signaling theory of leadership. Section 4 explains our experimental design. Section 5 reports the experimental results. Section 6 discusses the results. Proofs of propositions are relegated to Appendix.

2 Related Literature

Several works have been attempted to explain leadership by the theory of signaling. Our study of the emergence of leadership is related to the literature by decomposing the fundamental issue of leadership into two questions.[2] The first question is *how a leader leads*. The second question is *why a leader is there*.

The first question was addressed by Hermalin (1998). He studied team production under uncertainty using a signaling game with specific rules: (i) one member exclusively holds information about team productivity, (ii) the member must commit to a level of effort in the first period, and (iii) the other members choose their levels of effort in the second period after observing the leader's choice. He analyzed the separating equilibrium of this game in which the followers mimic the action of the leader. He showed that in this equilibrium, each member chooses a level of effort closer to the first best than the member would choose in a simultaneous effort choice game. His result is the first to show that a leader "leads by example".[3]

[1] See El-Gamal and Grether (1995), Costa-Gomes, Crawford, and Broseta (2001), and Costa-Gomes and Crawford (2006) for other experimental studies of mixed behaviors by the error rate analysis.

[2] Bolton, Brunnermeier, and Veldkamp (forthcoming) and Hermalin (forthcoming) give an overview of the economics of leadership from various perspectives.

[3] Vesterlund (2003) extended the Hermalin game to a case in which the leader must decide whether

The theoretical prediction of agents' behaviors suggested by Hermalin was tested by the experiment of Potters, Sefton, and Vesterlund (2007). Their experiment consists of four treatments: simultaneous moves with full information, simultaneous moves with asymmetric information, sequential moves with full information, and sequential moves with asymmetric information. They found that under asymmetric information, subjects' contribution was significantly larger for sequential moves than for simultaneous moves. They also found that under full information, subjects' contribution did not significantly differ between simultaneous moves and sequential moves. They concluded that leading by example was realized in the treatment of sequential moves with asymmetric information.

The second question was studied by two different approaches. In the first approach, a leader exists as the result of some form of selection process. Komai, Stegeman, and Hermalin (2007) theoretically showed that in a variant of Hermalin (1998), the welfare of a team is higher in the equilibrium of this game in which there is a leader than in a full information and simultaneous move game without a leader. They claimed that this implies that a leader must be there on the basis of Pareto optimality. Komai, Grossman, and Deters (2007) obtained an experimental result that supports this theory. Potters, Sefton, and Vesterlund (2005) theoretically and experimentally studied a contribution game with a voting stage in which agents select a leader. They found that subjects chose the Hermalin game at the voting stage in which subjects choose between the Hermalin game and the simultaneous move game, and subjects then played the Hermalin equilibrium.

In the second approach, a leader exists because he has volunteered to take that role. Andreoni (2006) studied a contribution game with an information gathering stage in which each individual may voluntarily choose to pay to investigate the value of public goods. After observing who purchased the information, each member decides on his contribution at a time chosen from a set of finite possible times. In this setting, he proved that if wealth varies between members, there exists a sequential equilibrium such that the richest member purchases the information and contributes first, and the other members contribute later depending on the level of contribution by the richest member.

Our preceding works, Kobayashi and Suehiro (2005, 2008) and the current paper, belong to the second approach. In the team production game with a continuum of effort levels, Kobayashi and Suehiro (2005) proved that if information is independent across agents, the three categories of leadership that we focus on in this paper are the only stable sequential equilibria, and at least one particular equilibrium other than leadership by confidence exists. Kobayashi and Suehiro (2008) emphasized the importance of leadership by confidence in the endogenous team production game. They derived a sufficient condition under which leadership by confidence exists as a stable sequential equilibrium.

In the second approach, our work and that of Andreoni (2006) differ in the forces driving the emergence of leadership. In Andreoni (2006), an agent who holds information becomes a leader, and one who does not becomes a follower. The driving force that determines who holds the information is heterogeneity of information acquisition

to acquire private information by paying a cost.

cost. In our work, possession of information is not the force driving the emergence of leadership. This difference appears prominently in the case of leadership by confidence. Remember that in our model, every agent holds private information. Heterogeneity in levels of confidence generated by private information is the force driving the emergence of leadership by confidence.

The present study implements experiments based on the model of Kobayashi and Suehiro (2005, 2008). This is the first attempt to test the endogenous signaling theory of leadership by carrying out an experiment.[4] Our experiment is also designed to capture the possibility of conditional cooperation, in contrast to Potters, Sefton, and Vesterlund (2007), who studied only a parameter case of rationality and Pareto optimality in line.

This paper is also related to the literature of experimental test of equilibrium refinements in signaling games. The existing studies including Brandts and Holt (1992, 1993) and Banks, Camerer, and Porter (1994) provided mixed evidence about the adequacy of equilibrium refinements. This paper shows that leadership by confidence predicted by more stringent refinement than sequential equilibrium is clearly selected.

3 The Theory of Emergence of Leadership in Teams

The theory that we test in our experiment is a simplified version of the endogenous signaling theory of leadership proposed by Kobayashi and Suehiro (2005, 2008).

3.1 The Model

We study team production with two symmetric agents $i = 1, 2$. Each agent i chooses a level e_i of effort from either a low level e_L or a high level e_H ($e_L < e_H$). A cost c_i of effort is c_L for $e_i = e_L$ and c_H for $e_i = e_H$ ($c_L < c_H$). The total effort determines a level of the team output by $\theta(e_i + e_j)$, where θ is a team productivity parameter. Each agent i is rewarded by the team output. The net payoff to agent i is given by:

$$U_i = \theta(e_i + e_j) - c_i.$$

Each agent is endowed with independent and partial information about the team productivity. Specifically, the parameter θ is a realization of a random variable on an independent product probability space as is shown in Table 1.

		ρ	$1-\rho$
	$1\backslash 2$	H	L
ρ	H	θ_H	θ_M
$1-\rho$	L	θ_M	θ_L

Table 1: Information Structure

[4] In the complete information setting, Arbak and Villeval (2007) and Rivas and Sutter (2009) experimentally studied on the emergence of voluntary leader. Huck and Rey-Biel (2006) theoretically studied voluntary leadership in the complete information setting by considering agents who dislike effort differentials.

There are four possible states: $(H, H), (H, L), (L, H), (L, L)$. In each state, agent 1 observes its first coordinate, called agent 1's type, and agent 2 observes its second coordinate, called agent 2's type. Each type is realized independently. Type H occurs with a probability ρ, and type L with $1 - \rho$. The team productivity is θ_H in the (H, H) state, θ_L in the (L, L) state, and θ_M in the (H, L) state and the (L, H) state. We assume that $\theta_L < \theta_M < \theta_H$.

Agents choose their effort levels according to the following time sequence. There are two periods, 1 and 2. In period 1, each agent i may exert an effort level e_i or may choose to do nothing (denoted as $\emptyset$). If he exerts an effort in period 1, then he cannot do anything in period 2. On the other hand, if he chooses to do nothing in period 1, then he must exert an effort in period 2.

In this sequence of moves, the two agents move (taking some e_i or $\emptyset$) independently and simultaneously in period 1. Each agent i immediately observes the behavior that the other agent j has taken. Agent i can then utilize this information for his choice in period 2 if he has chosen to do nothing in period 1. If both agents have chosen to do nothing in period 1, they invest some level of effort independently and simultaneously in period 2.

3.2 Strategy Profiles with Leadership

We study strategy profiles that support emergence of leadership in the model.

Agent i's strategy is a profile $(e^1_{i,t_i}, e^2_{i,t_i}(\cdot))_{t_i=H,L}$ of Bayesian strategies. e^1_{i,t_i} prescribes his behavior in period 1 for t_i-type and takes a value in $\{e_H, e_L, \emptyset\}$. $e^2_{i,t_i}(\cdot)$ prescribes his behavior in period 2 for t_i-type and assigns a value in $\{e_H, e_L\}$ for each possible value of agent j's choice in period 1.

We say that a strategy profile supports emergence of leadership if a play according to the strategy profile entails a leader–follower relation with a positive probability and the follower's choice depends on the leader's choice. More specifically, a strategy profile must induce two kinds of play with positive probabilities. The first kind of play is such that agent i chooses e_H in period 1 and agent j chooses $\emptyset$ in period 1 and, observing agent i's choice of e_H in period 1, agent j chooses e_H in period 2. The second kind is such that agent i chooses other than e_H in period 1 and agent j chooses $\emptyset$ in period 1 and, observing agent i's choice other than e_H in period 1, agent j chooses e_L in period 2.

Kobayashi and Suehiro (2005, 2008) identify three categories of strategy profiles as candidates for supporting emergence of leadership in sequential equilibrium: (i) leadership by confidence, (ii) leadership by identity, and (iii) leadership by identity with confidence.

A strategy profile is called leadership by confidence when it is a profile of symmetric strategies, $e^1_{i,H} = e_H$, $e^1_{i,L} = \emptyset$, and:

$$e^2_{i,L}(e^1_j) = \begin{cases} e_H & \text{if } e^1_j = e_H \\ e_L & \text{otherwise.} \end{cases}$$

To understand this strategy, note that an H-type agent holds a higher expectation about team productivity than an L-type agent; that is:

$$\rho\theta_H + (1-\rho)\theta_M > \rho\theta_M + (1-\rho)\theta_L.$$

In other words, an H-type agent is more confident about team productivity than an L-type agent. Then, leadership by confidence means that a more confident agent takes leadership by e_H and a less confident agent chooses to be a follower. The follower decides a level of effort depending on his partner's choice in period 1. He chooses e_H when observing that his partner has chosen e_H, and he chooses e_L otherwise.

Leadership by identity means that one agent (say, i) always takes leadership and the other agent ($j \neq i$) always chooses to be a follower. Formally, agent i's strategy is $e^1_{i,H} = e_H$, $e^1_{i,L} = e_L$ while agent j's strategy is $e^1_{j,t_j} = \emptyset$ for $t_j = H, L$ and:

$$e^2_{j,H}(e^1_i) = e_H \text{ for all } e^1_i,$$

$$e^2_{j,L}(e^1_i) = \begin{cases} e_H & \text{if } e^1_i = e_H \\ e_L & \text{otherwise.} \end{cases}$$

Finally, leadership by identity with confidence is a hybrid of the above two kinds of leadership. One agent takes leadership by e_H when he is more confident. He postpones his decision until period 2 when he is less confident. The other agent always chooses to wait and see in period 1. Formally, agent i's strategy is $e^1_{i,H} = e_H$, $e^1_{i,L} = \emptyset$, and $e^2_{i,L}(e^1_j) = e_L$ for all e^1_j, while agent j's strategy is the same as for leadership by identity.

3.3 Analysis of Sequential Equilibrium

We analyze necessary and sufficient conditions under which each of the above three categories of leadership is a sequential equilibrium.

First, a necessary and sufficient condition for leadership by confidence to be a sequential equilibrium is as follows.

Proposition 1 *Leadership by confidence is a sequential equilibrium if and only if:*

$$2\theta_L < \frac{c_H - c_L}{e_H - e_L} < \theta_M. \tag{1}$$

The intuition for condition (1) is as follows. The voluntary sorting into leader and follower according to agent's type is central for leadership by confidence to be a sequential equilibrium. Given that a partner follows leadership by confidence, every agent faces a basic trade-off in the choice of time to move. An agent choosing to follow enjoys valuable information in that his partner's type has been revealed through the partner's differentiated behavior in period 1. The team productivity can be inferred from the information about the partner's type together with the own private information. Based on this understanding of the team productivity, the follower can then choose an optimal level of effort in period 2. On the other hand, an agent choosing to lead with e_H enjoys a signaling benefit in that if the partner is of L-type, the partner changes a level of effort to e_H by observing the leader's choice of e_H in period 1.

Condition (1) resolves the trade-off properly in that the signaling benefit outweighs the value of information for the H-type while the opposite holds for the L-type. That is, the second inequality of (1) means that when an agent believes that the team productivity is θ_M, the marginal benefit of effort is larger than the marginal cost. This implies that the value of information from choosing to be a follower is null for the H-type, because he knows that the team productivity is either θ_M or even higher θ_H so that an optimal level of effort is e_H. Hence the signaling benefit induces an H-type agent to choose voluntarily to be a leader.

In contrast, an L-type agent is not sure whether the team productivity is θ_L or θ_M. The first inequality of (1) means that when the team productivity is θ_L, the marginal benefit of effort is smaller than the marginal cost, and even the marginal benefit of effort plus the signaling benefit of increased level of effort from a partner, that is $2\theta_L$, is smaller than the marginal cost. This implies first that the value of information from choosing to be a follower is strictly positive for the L-type because the follower would choose e_L if the team productivity is θ_L and would choose e_H if it is θ_M. Furthermore, the value of information is larger than the signaling benefit, because the increased level of effort from a partner does not compensate for the marginal cost when the team productivity is θ_L. Hence an L-type agent voluntarily chooses to be a follower and responds to a partner's behavior in period 2 as leadership by confidence prescribes.

Second, a necessary and sufficient condition for leadership by identity to be a sequential equilibrium is as follows.

Proposition 2 *Leadership by identity is a sequential equilibrium if and only if:*

$$\rho\theta_M + 2(1-\rho)\theta_L < \frac{c_H - c_L}{e_H - e_L} < \theta_M. \tag{2}$$

Condition (2) is similar to condition (1). Interpretation of the inequalities is the same. The only difference is the left-hand side of the first inequality. The marginal benefit from own effort increase is $\rho\theta_M + (1-\rho)\theta_L$, and the marginal benefit from inducing a partner's effort increase is $(1-\rho)\theta_L$. The difference comes from the fact that condition (2) is the incentive condition for an effort choice in period 1.

Third, the necessary and sufficient condition for leadership by identity with confidence is the same as for leadership by identity. To see this, note that the only difference between the two categories is the behavior of an L-type agent who is expected to be a leader. In leadership by identity, the L-type agent has to invest e_L in period 1. On the other hand, in leadership by identity with confidence, the agent has to wait and see in period 1 and invest e_L in period 2. Although the agent is able to observe the partner's behavior, his information is not updated because the partner always moves in period 2. Furthermore, the partner's behaviors when observing e_L and $\emptyset$ are e_L in both categories. Therefore, incentive conditions for an L-type leader are the same for the two categories.

Proposition 3 *Leadership by identity with confidence is a sequential equilibrium if and only if* (2)*.*

In propositions 1 to 3, we derive necessary and sufficient conditions (1) and (2) for each category of emergence of leadership to be a sequential equilibrium. Note that those conditions do not imply each other. Specifically, (1) implies (2) when $\theta_M \leq 2\theta_L$, and the reverse holds when $\theta_M \geq 2\theta_L$.

For the purpose of our experiments, we focus on the most suitable case in which both conditions are simultaneously satisfied; that is:

$$\max\{2\theta_L, \rho\theta_M + 2(1-\rho)\theta_L\} < \frac{c_H - c_L}{e_H - e_L} < \theta_M. \tag{3}$$

Under condition (3), there may exist a pathological sequential equilibrium. It is a profile of symmetric strategies, $e^1_{i,H} = e_L$, $e^1_{i,L} = \emptyset$, and:

$$e^2_{i,L}(e^1_j) = \begin{cases} e_H & \text{if } e^1_j = e_L \\ e_L & \text{otherwise.} \end{cases}$$

This strategy profile can be seen as a variant of leadership by confidence in which a leadership action is not e_H but e_L, and a follower responds to that e_L with e_H correspondingly. Since we are interested in emergence of leadership as defined in section 3.2, we exclude this pathological equilibrium from our experiment. By an argument similar to Proposition 1, we can show that this strategy profile is not a sequential equilibrium if either:

$$\rho\theta_H > \frac{c_H - c_L}{e_H - e_L}, \text{ or} \tag{4}$$

$$(1-\rho)\theta_L > \rho\Big[\theta_M - \frac{c_H - c_L}{e_H - e_L}\Big]. \tag{5}$$

Condition (4) means that an agent of H-type deviates to e_H in period 1. Condition (5) means that an agent of L-type mimics the behavior of H-type.

Finally, there always exists no-leadership equilibrium. A strategy profile is classified into no-leadership when $e^1_{i,t_i} = \emptyset$ for $i = 1, 2$ and $t_i = H, L$. Every agent chooses to wait and see in period 1 irrespective of his type. No-leadership is also a sequential equilibrium, because if every agent holds a pessimistic belief about the partner's type in period 2, no one has an incentive to invest some level of effort in period 1.

The above analysis is summarized as follows.

Proposition 4 *Suppose that condition* (3) *is satisfied. If either* (4) *or* (5) *holds, then there are exactly four sequential equilibria (in pure strategies). They are leadership by confidence, leadership by identity, leadership by identity with confidence, and no-leadership. Moreover, in no-leadership, the H-type chooses e_H and the L-type chooses e_L.*

Proposition 4 successfully identifies candidates for behaviors by agents in teams. However, the theory of sequential equilibrium alone does not pin down the emergence of leadership. The set of sequential equilibria in this game contains both emergence and non-emergence of leadership. Furthermore, in the sequential equilibria with emergence of leadership, there are three patterns of leadership. This commands further analysis with some refinements.

For this further analysis, we apply two alternative theories. The first theory is the intuitive criterion proposed by Cho and Kreps (1987). The criterion selects two sequential equilibria as follows.

Proposition 5 *Suppose that condition* (3) *is satisfied. Then there are two sequential equilibria that pass the intuitive criterion. They are leadership by confidence and leadership by identity.*

No-leadership and leadership by identity with confidence are eliminated by the intuitive criterion as follows. In no-leadership equilibrium, suppose that an H-type agent deviates to moving first with e_H and that this deviation induces his partner to believe that the first mover is of H-type. Then, that H-type agent will receive the same payoff as would an H-type leader under leadership by identity, because his partner chooses $\emptyset$ in both no-leadership and leadership by identity. Note that condition (2) guarantees that only the H-type has an incentive to deviate. The reason is the same for why only the H-type chooses e_H under leadership by identity. Hence, the deviation of that H-type agent successfully signals his type. Thus, no-leadership is eliminated.

In leadership by identity with confidence, suppose that an H-type follower deviates to moving first with e_H and that this deviation induces his partner to believe that the first mover is of H-type. That H-type follower will receive the same payoff as in leadership by confidence, because the H-type partner chooses e_H and the L-type partner chooses $\emptyset$ in both leadership by confidence and leadership by identity with confidence. This deviation successfully signals his type, because condition (1) guarantees that only the H-type has an incentive to deviate for the same reason that an H-type agent chooses e_H and an L-type agent chooses $\emptyset$ in leadership by confidence. Hence, leadership by identity with confidence is eliminated.

The second theory that we apply is the "mistaken theories" refinement. This theory was proposed by Kreps (1989, 1990) and developed in various forms by Suehiro (1992a, 1992b), Hillas (1994), Koçkesen and Ok (2004). The idea is to test the stability of a sequential equilibrium by taking into account a possibility of playing with an opponent who believes an alternative equilibrium holds. Here, we adopt the simplest version of the "mistaken theories" refinement. Each agent i is endowed with a theory of play that every agent almost certainly expects that a tested sequential equilibrium σ will be played. However, a prior over agents' theories admits a small probability of "mistaken theories"; that is, a small probability that agent j holds another theory that every agent almost certainly expects that an alternative sequential equilibrium σ' will be played. Each agent then follows his own theory in a play. When agent i sees a deviation by agent j from the tested equilibrium σ, agent i is forced to hold an out-of-equilibrium belief that the deviation has occurred because agent j followed σ'. The tested equilibrium σ is unstable against the alternative equilibrium σ' if σ prescribes a sequentially irrational behavior for some information set while σ' prescribes sequentially rational behaviors for all information sets. The tested equilibrium σ is unstable if it is unstable against some alternative sequential equilibrium. The "mistaken theories" refinement pins down a prediction of agents' behaviors in teams as follows (a formal definition of "mistaken theories" refinement is in the appendix).

Proposition 6 *Suppose that condition* (3) *is satisfied. Then leadership by confidence is the unique sequential equilibrium that is stable in "mistaken theories".*

Leadership by identity is eliminated by the "mistaken theories" refinement as follows. We test the stability of leadership by identity against leadership by confidence. Suppose that agent i of L-type is a leader in leadership by identity. He expects with a small probability that agent j believes leadership by confidence holds, and if agent j is of H-type, agent j moves first with e_H. Therefore, agent i's payoff will improve if he deviates to $\emptyset$ in period 1 and chooses e_H in period 2 when agent j moves first with e_H. Leadership by identity prescribes sequentially irrational behavior for agent i.

On the other hand, leadership by confidence prescribes sequentially rational behaviors for all information sets. Suppose that agent i believes leadership by confidence holds. Because leadership by confidence is a sequential equilibrium and the probability of "mistaken theories" is small, leadership by confidence prescribes sequentially rational behaviors for all the on-equilibrium information sets. A deviation from leadership by confidence is observed when agent j moves first with e_L. Agent i expects with a small probability that agent j believes leadership by identity holds, and if agent j is an L-type leader, he moves first with e_L. This is the only possible way that agent i could observe agent j move first with e_L. Therefore, when agent i is of L-type, it is sequentially rational for him to follow leadership by confidence by choosing $\emptyset$ and responding to agent j's e_L with e_L believing that agent j is of L-type. Hence, leadership by identity is unstable against leadership by confidence.

4 Experimental Design, Hypotheses, and Procedure

4.1 Experimental Design

The endogenous signaling theory of leadership hypothesizes that leadership emerges through endogenously formed signaling. When applied to the team production game of section 3.1, the theory predicts that this kind of leadership will emerge with a positive probability when the payoffs satisfy condition (3) (and (4) or (5)). To test this theory using the game of section 3.1, we designed an experiment with three treatments defined by two treatment variables: (i) information about team productivity and (ii) payoffs. Treatment 1 is a baseline treatment in which subjects are endowed with incomplete information about team productivity as described in section 3.1 and receive payoffs satisfying condition (3). According to the theory, leadership should emerge with a positive probability in this treatment. Treatment 2 is a treatment in which subjects are endowed with complete information about team productivity while the payoffs remain unchanged from Treatment 1. If leadership emerges as signaling as the theory presumes, leadership should not emerge in Treatment 2, because there is no private information to be signaled. Treatment 3 is a treatment in which information about team productivity remains incomplete as in Treatment 1 but the payoffs are changed so as not to satisfy condition (3). If leadership emerges in sequential equilibrium as the theory postulates, leadership should not emerge in Treatment 3.

For each of the three treatments, we prepared one game to be played by subjects. Game 1 for Treatment 1 is as follows. In the team production game of section 3.1, we set the parameters as $\theta_H = 210, \theta_M = 200, \theta_L = 80, \rho = \frac{1}{3}, e_H = 2, e_L = 1, c_H = 190$, and $c_L = 0$. This set of parameters obviously satisfies conditions (3) and (5). The payoffs of the game were described to subjects by using payoff matrices called points tables. In order to avoid transmitting semantic contents to subjects, e_L was called Alternative 1, and e_H was called Alternative 2. A payoff was called a point. We presented the following points tables X, Y, and Z for the payoff matrices calculated from the parameters, corresponding to the cases of team productivity being θ_L, θ_M, and θ_H respectively. The tables show payoffs for a row player.

Table Z

Your Choice / The Other's Choice	Alternative 2	Alternative 1
Alternative 2	650	440
Alternative 1	630	420

Table Y

Your Choice / The Other's Choice	Alternative 2	Alternative 1
Alternative 2	610	410
Alternative 1	600	400

Table X

Your Choice / The Other's Choice	Alternative 2	Alternative 1
Alternative 2	130	50
Alternative 1	240	160

Table 2: Points Tables of Game 1 and Game 2: X, Y, Z

We chose a payoff matrix applied to a pair of subjects by rolling two dice and using a rule shown in Table 3.[5] When both dice were 5 or 6, we selected points table Z. When both dice were from 1 to 4, we selected points table X. Otherwise, we selected points table Y.

Dice # 1 \ Dice # 2	5 or 6	1 to 4
5 or 6	Z	Y
1 to 4	Y	X

Table 3: Selection of Points Tables

At the beginning of a play, one subject was informed whether the result of one dice was "1 to 4" or "5 or 6". "1 to 4" corresponds to the L-type, and "5 or 6" corresponds to the H-type. The other subject was informed of the result of the other dice in the same way. Then, each subject of the pair was requested simultaneously to choose either Alternative 1, Alternative 2, or $\times$. A choice of $\times$ means doing nothing at this time and corresponds to $\emptyset$ in our model. The choice made by each subject was reported to the other subject. The subjects were then requested simultaneously to choose either Alternative 1 or Alternative 2 if they had chosen $\times$ at the preceding

[5]We prepared a pair of results of rolling dice in advance by using a table of random numbers. This procedure was explained to subjects, and it was understood that points tables were selected properly.

time. If a subject had chosen either Alternative 1 or Alternative 2, he was requested to choose ××, which means that he had to do nothing at this time.[6,7] Finally, payoffs for each subject of the pair were determined by the subjects' choices of alternatives according to the points table selected for the pair at the beginning of the game.

Game 2 for Treatment 2 is a game obtained by changing part of the rules of Game 1 in such a way that a pair of subjects are given complete information about the points table to be used by the pair. Specifically, after rolling two dice to select a points table for a pair, we reported the results of the two dice to both subjects of the pair.[8] Each subject was then requested to find the points table that applied to his pair. After this, subjects proceeded to make their first choices. All the other rules of the game were kept unchanged.

Game 3 for Treatment 3 is a game obtained from Game 1 by changing points tables X, Y, and Z in Table 2 into those in Table 4. This change in payoffs is obtained when the value of parameter c_H is changed from $c_H = 190$ to $c_H = 220$ while the remaining parameters are kept unchanged. The set of parameters satisfies neither condition (1) nor condition (2). All the other rules of the game were kept unchanged.

Table Z

Your Choice / The Other's Choice	Alternative 2	Alternative 1
Alternative 2	620	410
Alternative 1	630	420

Table Y

Your Choice / The Other's Choice	Alternative 2	Alternative 1
Alternative 2	580	380
Alternative 1	600	400

Table X

Your Choice / The Other's Choice	Alternative 2	Alternative 1
Alternative 2	100	20
Alternative 1	240	160

Table 4: Points Tables of Game 3: X, Y, Z

4.2 Hypotheses

A set of hypotheses to be tested in our experiment is summarized as follows.

Treatment 1 is the case that we studied in section 3.3. According to Propositions 4, 5, and 6, we conclude for Game 1 that leadership will emerge with a positive probability, but the predicted behaviors depend on the refinement criterion that we adopt. Which

[6] Subjects were requested to choose ×× so that no subject was able to detect, by observing others in the laboratory, who had chosen × at the preceding time.

[7] We use the personal pronoun "he" throughout our description of the experiment for simple exposition. However, as we explain later, the subjects were randomly selected so that there were both male and female subjects in our experiment.

[8] In an explanation of the rules of Game 2 to subjects, we clearly stated that the same information is delivered to both subjects in a pair.

refinement criterion best explains actual human behaviors in Game 1 is not determined a priori and can be verified experimentally. Therefore, we postulate the following hypothesis for Treatment 1.

Hypothesis 1 *In Game 1, subjects will behave according to one of the following patterns.*

[SE] *Leadership by confidence, leadership by identity, leadership by identity with confidence, or no-leadership will be adopted.*

[CK] *Leadership by confidence or leadership by identity will be adopted.*

[MT] *Leadership by confidence will be adopted.*

The alternatives **[SE]**, **[CK]**, and **[MT]** of predicted patterns correspond to the sequential equilibrium, the Cho–Kreps stable sequential equilibrium, and the "Mistaken Theories" stable sequential equilibrium respectively adopted as the refinement criterion.

Game 2 for Treatment 2 is no longer the team production game that we studied in section 3.1. A pair of subjects play with complete information about payoffs of either X, Y, or Z depending on the result of rolled dice. If the team production game with payoffs of X were played simultaneously, it would be a dominant equilibrium for subjects to choose Alternative 1 (e_L). Similarly, it would be a dominant equilibrium for subjects to choose Alternative 2 (e_H) for Y and for Z. From these facts, it is straightforward to see that the sequential equilibrium of Game 2 is the following.

Proposition 7 *In the sequential equilibria of Game 2, an agent chooses Alternative 1 (e_L) if the points table is X, and Alternative 2 (e_H) if the points table is Y or Z. He chooses the alternatives in either period 1 or 2. His choice of alternatives in period 2 is independent of partner's choice in period 1.*

Choice of alternative is determined by the points table and is independent of both the time of move and the partner's choice. Hence the sequential equilibria do not support emergence of leadership in Game 2.

In Proposition 7, there is no a priori reason to assume the same probability that subjects choose prescribed efforts in period 1 for X, Y, and Z. Therefore, we postulate the following hypothesis for Treatment 2.

Hypothesis 2 *In Game 2, when payoffs are X, a subject will choose Alternative 1 (e_L) in period 1 with probability α_X and will choose $\emptyset$ with probability $1-\alpha_X$ to respond with Alternative 1 (e_L) in period 2 to any choice of his partner. When payoffs are Y, he will choose Alternative 2 (e_H) in period 1 with probability α_Y and will choose $\emptyset$ with probability $1-\alpha_Y$ to respond with Alternative 2 (e_H) in period 2 to any choice of his partner. When payoffs are Z, he will behave as for Y with probability α_Z instead of α_Y.*

In Game 3 for Treatment 3, if the team production game with payoffs X were played simultaneously, it would be a dominant equilibrium for the subjects to choose

Alternative 1 (e_L). This would also be the case if payoffs were Y and Z. From these facts, it is straightforward to see that the sequential equilibrium of Game 3 is the following.

Proposition 8 *In the sequential equilibria of Game 3, an agent chooses Alternative 1(e_L) in any points table. He chooses it in either period 1 or 2. His choice in period 2 is independent of the partner's choice in period 1.*

The chosen alternative is Alternative 1(e_L) and is independent of both time of move and partner's choice. Hence the sequential equilibria do not support emergence of leadership in Game 3.

In Proposition 8, there is no a priori reason to assume the same probability with which subjects choose Alternative 1(e_L) in period 1 for the L-type and the H-type. Therefore, we postulate the following hypothesis for Treatment 3.

Hypothesis 3 *In Game 3, a subject of H-type will choose Alternative 1 (e_L) in period 1 with probability α_H and will choose $\emptyset$ with probability $1 - \alpha_H$ to respond with Alternative 1 (e_L) in period 2 to any choice of his partner. A subject of L-type will choose Alternative 1 (e_L) in period 1 with probability α_L and will choose $\emptyset$ with probability $1 - \alpha_L$ to respond with Alternative 1 (e_L) in period 2 to any choice of his partner.*

4.3 Procedure

For each of the three treatments, we recruited 42 subjects from undergraduate students in their first and second years at the departments of business and economics of Kobe University. We randomly selected 42 students from applicants and randomly divided them into three groups of 14 students. Each group was assigned to one session of the experiment. Thus, three sessions were implemented for each treatment. No student attended more than one session.

Each session was implemented as follows. The experiment took place in a large classroom. We explained our experiment by reading an instruction out loud. We then conducted a quiz to test how correctly each subject understood the rules of the game. We identified 10 students according to their grades in the quiz. We randomly assigned subject numbers 1 through 10 to these students.

Subjects 1 through 10 then played the game for the treatment for 10 rounds. They were randomly and anonymously matched into pairs in each round by the method of randomized block design. Specifically, subjects 1 to 5 were randomly matched with subjects 6 to 10 in round 1 through round 5 under the stipulation that no one played with another subject twice. These matchings were repeated for round 6 through round 10.[9]

At the end of each round, each subject was informed of a complete history of play by his pair in the round. On the other hand, no subject had any information about plays made by any pair other than his pair throughout 10 rounds of play.

[9] The remaining 4 subjects who performed less well in the quiz were assigned subject numbers 11 through 14 according to their grades. They also played the game for the treatment for 10 rounds. Subject 11 was matched with subject 12 and subject 13 was matched with subject 14 throughout rounds 1 to 10. This asymmetry in dealing with the subjects according to their subject numbers was not revealed to any subject. All subjects were told that 14 subjects were randomly and anonymously matched into pairs for each round.

After 10 rounds of play, every subject answered a free-format questionnaire that asked him to explain how he played in the experiment. Each session took about two and a half hours to complete.

Every subject received a reward in Japanese Yen (100 Yen is approximately 1 US Dollar) according to the following formula.

$$2000 + 0.25 \times [\text{Total Points That You Earned}]$$

This formula for calculating rewards was clearly explained to all subjects in advance of play. The average amount of reward in our experiment was 3041 Japanese Yen.[10]

5 Results

We first examine pooling the data in our experiment. Then, we report the results of Treatment 1, 2, and 3 respectively.

5.1 Pooling the Data

A sample of a round in a treatment is a set of 15 outcomes played by randomly formed pairs of 30 subjects. In order to get enough sample sizes for our hypothesis tests, we pool samples from different rounds.

Strictly speaking, a subject's behavior may be influenced by his experience of plays in the previous rounds except round 1. Therefore, for pooling samples properly, we should check the independence of samples across different experiences of plays. In our games, however, a number of possible histories of plays is large (64 possible outcomes for a play) and quickly becomes enormous as rounds progress so that it is impractical to test the independence of samples across different experiences of plays. Therefore, we follow the approach of Costa-Gomes, Crawford, and Broseta (2001) (p.1215) and test the independence of samples across different rounds by the Fisher test. We examine the following three null hypotheses; the samples in rounds 1 through 5 are statistically independent, the samples in rounds 6 through 10 are statistically independent, and the samples in rounds 1 through 10 are statistically independent.

Table 5 shows the simulated p-values.[11] None of the null hypotheses is rejected at the 5 % significance level.[12] Therefore, we will use the data of pooled samples from rounds 1 through 5, from rounds 6 through 10, and from rounds 1 through 10 in our hypothesis tests.

[10] A standard wage for an undergraduate student from a part-time job in this area is 800 Japanese Yen (8 US Dollars) per hour. A reward of 2000 Japanese Yen for attending the experiment corresponds to the cash a subject might earn for the duration of the session.

[11] The simulated values are computed by using randomly generated 2×10^6 possible joint distributions of outcomes that are compatible with the observed marginal distributions.

[12] The p-value for the independence of the distributions of samples in rounds 1 through 5 in Treatment 1 is only slightly above the significance level. We will return to an implication of this result in section 6.1.

	Rounds 1–5	Rounds 6–10	Rounds 1–10
Treatment 1	0.06	0.96	0.48
Treatment 2	0.21	0.37	0.20
Treatment 3	0.84	0.91	0.87

Table 5: Independence across Rounds

5.2 Result of Treatment 1

Table 6 shows the data of Treatment 1. The panel of rounds 1–5 shows the distribution of plays by 75 different pairs randomly formed from 30 subjects in rounds 1 through 5. The far left column classifies four possible pairs of types of subjects 1 to 5 (left) and subjects 6 to 10 (right). The top two rows classify 16 possible plays. The first row shows a pair of choices made in period 1. The item on the left is a choice made by a subject 1 to 5, and the item on the right is a choice made by a subject 6 to 10. The second row shows a corresponding pair of choices made in period 2. The symbols 2, 1, and × mean choosing Alternative 2 (e_H), choosing Alternative 1 (e_L), and doing nothing ($\emptyset$). For example, the number 3 is entered in the "(H, H)" row and "(2, 2) in period 1 and (×, ×) in period 2" column. This means that in rounds 1 through 5, there were exactly 3 pairs of subjects both of whom were assigned H types and chose Alternative 2 in period 1. The panel of rounds 6–10 shows the corresponding distribution of plays in rounds 6 through 10.

Period 1	2, 2	2, 1	1, 2	1, 1	2, ×	2, ×	1, ×	1, ×	×, 2	×, 2	×, 1	×, 1	×, ×	×, ×	×, ×	×, ×
Period 2	×, ×	×, ×	×, ×	×, ×	×, 2	×, 1	×, 2	×, 1	2, ×	1, ×	2, ×	1, ×	2, 2	2, 1	1, 2	1, 1
H, H	3	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0
H, L	1	1	0	0	13	3	0	1	0	0	0	0	0	1	0	1
L, H	0	0	0	0	0	0	0	1	11	4	0	0	0	0	1	0
L, L	0	0	0	0	2	0	0	4	1	1	0	3	0	1	0	18

Panel: Rounds 1–5

Period 1	2, 2	2, 1	1, 2	1, 1	2, ×	2, ×	1, ×	1, ×	×, 2	×, 2	×, 1	×, 1	×, ×	×, ×	×, ×	×, ×
Period 2	×, ×	×, ×	×, ×	×, ×	×, 2	×, 1	×, 2	×, 1	2, ×	1, ×	2, ×	1, ×	2, 2	2, 1	1, 2	1, 1
H, H	9	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0
H, L	0	1	0	0	11	1	0	2	0	0	0	0	0	1	0	0
L, H	2	0	1	0	0	0	0	0	12	3	0	0	0	0	0	0
L, L	0	0	0	0	2	0	0	2	0	0	0	2	0	0	0	23

Panel: Rounds 6–10

Table 6: The Data of Treatment 1 (distributions of plays by pairs)

We test Hypothesis 1 using the data of rounds 1–5, 6–10, and 1–10. In our hypothesis testing, we apply the error rate analysis introduced by Harless and Camerer (1994, 1995) and El-Gamal and Grether (1995) and applied to various issues by Costa-Gomes, Crawford, and Broseta (2001), Costa-Gomes and Crawford (2006). In this analysis, we postulate a null hypothesis that subjects behave according to a tested pattern in Hypothesis 1 with error. We estimate unknown parameters under the hypothesis by the maximum likelihood method. Based on the estimated parameters, we test the null hypothesis against some appropriate alternative hypothesis by the likelihood ratio test.

The most natural alternative hypothesis is that subjects follow some behavior strategy independently, because we implemented our experiment by the anonymous random matching. If this alternative hypothesis is true, leadership by confidence ([MT]) and no-leadership (as a special case of [SE]) are possible among those patterns described in Hypothesis 1, because they are symmetric equilibria. Therefore, we test these possibilities against the alternative hypothesis that subjects follow some unknown behavior strategy independently.

Hypothesis 1 includes the possibilities of asymmetric equilibria (leadership by identity and leadership by identity with confidence as special cases of [SE]) and the possibilities of multiple equilibria being played in a coordinated way ([CK] and [SE]). Although these possibilities seem limited under our random matching environment, we will verify these possibilities statistically.[13] For these tests, the natural alternative hypothesis is that subjects follow some unknown correlated strategy.

5.2.1 The Test of Symmetric Equilibrium

Leadership by confidence and no-leadership in Hypothesis 1 are symmetric equilibria. We first test leadership by confidence, and the test is run as follows. The null hypothesis is that subjects almost certainly play according to leadership by confidence while choosing an alternative out of the equilibrium with probability ε.

For the data of rounds 1–5, we construct the following log likelihood function of parameter ε under the null hypothesis.[14]

$$L_0(\varepsilon) = 3 \times \ln\left[(1-2\varepsilon)^2\right] + 2 \times \ln\left[(1-2\varepsilon)\varepsilon \cdot (1-\varepsilon)\right] + \cdots \tag{6}$$

The brackets in the first term show the probability under the null hypothesis for the event that both subjects in a pair choose Alternative 2 in period 1 when they are of H types. The remaining terms correspond to probabilities for observed plays similarly computed under the null hypothesis.

We maximize the likelihood function (6) with respect to parameter ε. The maximum likelihood estimate is $\varepsilon^* = 0.06$ and the maximized log likelihood is $L_0(\varepsilon^*) = -114.67$. These values are shown in MT row of Table 7.

[13] A test for multiple equilibria being played by subjects was discussed by Rapoport and Amaldos (2008). However, they did not provide a statistical analysis because they judged it implausible that subjects would coordinate an equilibrium, round by round, under their anonymous random matching procedure. Here we formulate such a statistical test.

[14] We drop probabilities of type realizations from the likelihood function, because they are cancelled out by corresponding probabilities in a log likelihood function under an alternative hypothesis in the hypothesis test below.

Our alternative hypothesis is that every subject plays according to some behavior strategy $\beta = (\beta_H, \beta_L)$. When a subject is of H type, he follows the behavior strategy $\beta_H = (\beta_H^1, \beta_H^2)$. β_H^1 is the probability distribution $\beta_H^1 = (\beta_H^1(2), \beta_H^1(1), \beta_H^1(\times))$ that prescribes the probabilities that Alternative 2, Alternative 1, and $\times$ will be chosen in period 1. β_H^2 is the profile of three probability distributions $\beta_H^2 = ((\beta_H^2(2|2), \beta_H^2(1|2)), (\beta_H^2(2|1), \beta_H^2(1|1)), (\beta_H^2(2|\times), \beta_H^2(1|\times)))$. The probability distribution $(\beta_H^2(2|2), \beta_H^2(1|2))$ prescribes the probabilities that Alternative 2, Alternative 1 will be chosen in period 2 when he sees his partner choose Alternative 2 in period 1. The probability distributions $(\beta_H^2(2|1), \beta_H^2(1|1))$ and $(\beta_H^2(2|\times), \beta_H^2(1|\times))$ prescribe corresponding probabilities when he sees his partner choose Alternative 1 in period 1 and when he sees his partner choose $\times$ in period 1 respectively. Similarly, $\beta_L = (\beta_L^1, \beta_L^2)$ is a behavior strategy for the L-type.

We construct the following log likelihood function of parameters $\beta_H^1(2)$, $\beta_H^1(1)$, $\beta_H^2(2|2)$, $\beta_H^2(2|1)$, $\beta_H^2(2|\times)$, $\beta_L^1(2)$, $\beta_L^1(1)$, $\beta_L^2(2|2)$, $\beta_L^2(2|1)$, $\beta_L^2(2|\times)$ for the data of rounds 1–5 under the alternative hypothesis.

$$\begin{aligned}&L_1(\beta_H^1(2), \beta_H^1(1), \beta_H^2(2|2), \beta_H^2(2|1), \beta_H^2(2|\times), \beta_L^1(2), \beta_L^1(1), \beta_L^2(2|2), \beta_L^2(2|1), \beta_L^2(2|\times))\\&= 3 \times \ln\left[(\beta_H^1(2))^2\right] + 2 \times \ln\left[\beta_H^1(2)(1 - \beta_H^1(2) - \beta_H^1(1)) \cdot \beta_H^2(2|2)\right] + \cdots \qquad (7)\end{aligned}$$

The brackets in the first term show the probability under the alternative hypothesis for the event that both subjects in a pair choose Alternative 2 in period 1 when they are of H types. The remaining terms correspond to probabilities for observed plays similarly computed under the alternative hypothesis.

We maximize the likelihood function (7) with respect to parameters $\beta_H^1(2)$, $\beta_H^1(1)$, $\beta_H^2(2|2)$, $\beta_H^2(2|1)$, $\beta_H^2(2|\times)$, $\beta_L^1(2)$, $\beta_L^1(1)$, $\beta_L^2(2|2)$, $\beta_L^2(2|1)$, $\beta_L^2(2|\times)$. The maximum likelihood estimates are observed frequencies of the corresponding choices, and the maximized log likelihood is $L_1^* = -101.84$. The maximized log likelihood value is shown in Table 7.

We then test leadership by confidence ([MT] of Hypothesis 1) with the likelihood ratio test. The log likelihood ratio is:

$$2(L_1^* - L_0(\varepsilon^*)) = 25.66.$$

This value is shown in Table 7, together with the degree of freedom and the p-value for the χ^2 test.

The tests of leadership by confidence for the data of rounds 6–10 and 1–10 are conducted in a parallel way, and the results are shown in Table 7. The tests of no-leadership are conducted in a parallel way, and the results are also shown in NL rows of Table 7.

5.2.2 The Test of Asymmetric Equilibrium and Multiple Equilibria

The test of multiple equilibria in [SE] being played in a coordinated way is run as follows. The null hypothesis is that a pair of subjects first adopts an equilibrium in [SE] with an unknown probability and then that pair of subjects almost certainly play the adopted equilibrium while choosing an alternative out of the equilibrium with probability ε. In the null hypothesis, we denote the unknown probabilities of leadership by confidence, leadership by identity, leadership by identity with confidence, and no-leadership by ρ_{LC}, ρ_{LI}, ρ_{LIC}, and $1-\rho_{LC}-\rho_{LI}-\rho_{LIC}$. In leadership by identity and leadership by identity with confidence, we assume that each subject in a pair is assigned as leader with probability $\frac{1}{2}$.[15]

For the data of rounds 1–5, we construct the following log likelihood function of parameters ρ_{LC}, ρ_{LI}, ρ_{LIC}, and ε under the null hypothesis.

$$
\begin{aligned}
L_0(\rho_{LC},\rho_{LI},\rho_{LIC},\varepsilon) \quad = \quad & 3\times\ln\Big[\rho_{LC}\cdot(1-2\varepsilon)^2 \\
& \quad +\rho_{LI}\times\frac{1}{2}\cdot(1-2\varepsilon)\varepsilon+\rho_{LI}\times\frac{1}{2}\cdot\varepsilon(1-2\varepsilon) \\
& \quad +\rho_{LIC}\times\frac{1}{2}\cdot(1-2\varepsilon)\varepsilon+\rho_{LIC}\times\frac{1}{2}\cdot\varepsilon(1-2\varepsilon) \\
& \quad +(1-\rho_{LC}-\rho_{LI}-\rho_{LIC})\times\varepsilon^2\Big] \\
& +2\times\ln\Big[\rho_{LC}\cdot(1-2\varepsilon)\varepsilon\cdot(1-\varepsilon) \\
& \quad +\rho_{LI}\times\frac{1}{2}\cdot(1-2\varepsilon)^2\cdot(1-\varepsilon)+\rho_{LI}\times\frac{1}{2}\cdot\varepsilon^2\cdot(1-\varepsilon) \\
& \quad +\rho_{LIC}\times\frac{1}{2}\cdot(1-2\varepsilon)^2\cdot(1-\varepsilon)+\rho_{LIC}\times\frac{1}{2}\cdot\varepsilon^2\cdot(1-\varepsilon) \\
& \quad +(1-\rho_{LC}-\rho_{LI}-\rho_{LIC})\times\varepsilon(1-2\varepsilon)\cdot(1-\varepsilon)\Big] \\
& +\cdots
\end{aligned}
\tag{8}
$$

The brackets in the first term show the probability under the null hypothesis for the event that both subjects in a pair choose Alternative 2 in period 1 when they are of H types. The probability is the sum of probabilities with which this event occurs under alternative equilibria. For example, the first probability $\rho_{LC}\cdot(1-2\varepsilon)^2$ corresponds to the case in which leadership by confidence is adopted for an equilibrium to be played with the probability ρ_{LC} and both subjects 1 to 5 and subjects 6 to 10 choose Alternative 2 prescribed by leadership by confidence in period 1 with probabilities $1-2\varepsilon$. The remaining terms correspond to probabilities for observed plays similarly computed under the null hypothesis.

We maximize the likelihood function (8) with respect to parameters ρ_{LC}, ρ_{LI}, ρ_{LIC}, and ε. The maximum likelihood estimates are $\rho^*_{LC}=0.70$, $\rho^*_{LI}=0.10$, $\rho^*_{LIC}=0.19$, and $\varepsilon^*=0.06$ and the maximized log likelihood is $L_0(\rho^*_{LC},\rho^*_{LI},\rho^*_{LIC},\varepsilon^*)=-111.03$. These values are shown in SE row of Table 7.

Our alternative hypothesis is that subjects play according to some correlated strategy with a distribution p over paths. To each path of choices and each pair of types, the distribution p assigns the probability that subjects play the path of choices when the pair of types is realized. For example, when both subjects are of H types, they play path $\{(2,2),(\times,\times)\}$ (both subjects choose 2 in period 1) with probability $p\big(\{(2,2),(\times,\times)\}\ |HH\big)$. Following the standard formulation of correlated strategy, we model the probability $p\big(\{(2,2),(\times,\times)\}\ |HH\big)$ as a probability conditional on the

[15] In our formulation, we hypothesized that when a pair of subjects coordinate on leadership by identity or leadership by identity with confidence, one of the subjects takes the role of leader with probability $\frac{1}{2}$ because there is no a priori reason to assume particular asymmetric probabilities.

realized pair of types.

We construct the following log likelihood function of parameters p for the data of rounds 1–5 under the alternative hypothesis.

$$L_1(p) = 3 \times \ln \Big[p\big(\{(2,2),(\times,\times)\} \;|HH\big)\Big] + 4 \times \ln \Big[p\big(\{(2,\times),(\times,2)\} \;|HH\big)\Big] + \cdots \tag{9}$$

The brackets in the first term show the probability under the alternative hypothesis for the event that both subjects in a pair choose Alternative 2 in period 1 when they are of H types. The remaining terms correspond to probabilities for observed plays similarly computed under the alternative hypothesis.

We assume that the unknown correlated strategy treats subjects symmetrically. Under this assumption, any path and its mirror image must be considered identical paths. For example, $\{(2,\times),(\times,2)\}$ and $\{(\times,2),(2,\times)\}$ when a pair of types are HH are identical paths. Therefore, the number 4 in the second term in (9) is obtained by summing the numbers of observations $\{(2,\times),(\times,2)\}$ and $\{(\times,2),(2,\times)\}$ when a pair of types are HH.

We maximize the likelihood function (9) with respect to parameters p. The maximum likelihood estimates are observed frequencies of the corresponding paths, and the maximized log likelihood is $L_1^* = -84.82$. The maximized log likelihood value is shown in Table 7.

We then test [SE] of Hypothesis 1 with the likelihood ratio test. The log likelihood ratio is:

$$2(L_1^* - L_0(\rho_{LC}^*, \rho_{LI}^*, \rho_{LIC}^*, \varepsilon^*)) = 52.42.$$

This value is shown in SE row of Table 7, together with the degree of freedom and the p-value for the χ^2 test.

The test of whether the data of rounds 1–5 support the pattern [CK] of Hypothesis 1 is conducted in a parallel way by setting $\rho_{LI} = 1 - \rho_{LC}$, $\rho_{LIC} = 0$ for the null hypothesis. Similarly, the test of whether the data of rounds 1–5 support leadership by identity (LI) and leadership by identity with confidence (LIC) of Hypothesis 1 is conducted by setting $\rho_{LI} = 1$ and $\rho_{LIC} = 1$ for the null hypothesis respectively. The results of these tests are shown in Table 7. Table 7 also shows tests of whether leadership by confidence and no-leadership respectively fit the data of rounds 1–5 against this alternative hypothesis. The tests of Hypothesis 1 for the data of rounds 6–10 and 1–10 are conducted in a parallel way, and the results are shown in Table 7.

	Estimates				Alternative Hypothesis: Behavior Strategy					Alternative Hypothesis: Correlated Strategy				
	ε	ρ_{LC}	ρ_{LI}	ρ_{LIC}	L_0^*	L_1^*	$2(L_1^* - L_0^*)$	χ^2-d.f.	p-value	L_0^*	L_1^*	$2(L_1^* - L_0^*)$	χ^2-d.f.	p-value
Rounds 1–5														
SE	0.06	0.70	0.10	0.19	—	—	—	—	—	-111.03	-84.82	52.42	29	4.89E-03
CK	0.07	0.87	0.13	—	—	—	—	—	—	-112.56	-84.82	55.48	31	4.43E-03
MT	0.06	1.00	—	—	-114.67	-101.84	25.66	9	2.31E-03	-114.67	-84.82	59.70	32	2.11E-03
LI	0.12	—	1.00	—	—	—	—	—	—	-165.16	-84.82	160.68	32	4.51E-19
LIC	0.07	—	—	1.00	—	—	—	—	—	-125.39	-84.82	81. 14	32	3.79E-06
NL	0.22	—	—	—	-201.98	-101.84	200.29	9	2.89E-38	-201.98	-84.82	234.32	32	1.24E-32
Rounds 6–10														
SE	0.05	0.94	0.02	0.04	—	—	—	—	—	-78.02	-63.43	29.18	29	0.46
CK	0.05	0.97	0.03	—	—	—	—	—	—	-78.12	-63.43	29.38	31	0.55
MT	0.05	1.00	—	—	-78.22	-73.10	10.24	9	0.33	-78.22	-63.43	29.58	32	0.59
LI	0.12	—	1.00	—	—	—	—	—	—	-161.95	-63.43	197.04	32	1.18E-25
LIC	0.07	—	—	1.00	—	—	—	—	—	-113.89	-63.43	100.92	32	4.59E-09
NL	0.23	—	—	—	-199.71	-73.10	253.23	9	2.07E-49	-199.71	-63.43	272.56	32	5.81E-40
Rounds 1–10														
SE	0.05	0.82	0.07	0.10	—	—	—	—	—	-191.73	-153.80	75.86	29	4.60E-06
CK	0.06	0.92	0.08	—	—	—	—	—	—	-192.63	-153.80	77.66	31	7.05E-06
MT	0.07	1.00	—	—	-194.79	-177.87	33.83	9	9.57E-05	-194.79	-153.80	81.98	32	2.89E-06
LI	0.12	—	1.00	—	—	—	—	—	—	-327.18	-153.80	346.76	32	1.62E-54
LIC	0.07	—	—	1.00	—	—	—	—	—	-239.35	-153.80	171.10	32	6.24E-21
NL	0.22	—	—	—	-401.76	-177.87	447.77	9	8.60E-91	-401.76	-153.80	495.92	32	1.38E-84

Table 7: The Test Results of Hypothesis 1

Table 7 shows the following result for Hypothesis 1.

Result 1 *(i) All the patterns in Hypothesis 1 are rejected for the data of rounds 1–5.*
(ii) All the patterns in Hypothesis 1 are accepted for the data of rounds 6–10. For patterns [SE] and [CK], the estimated probabilities of adopting an equilibrium are highly concentrated on leadership by confidence; $\rho^*_{LC} = 0.94$ *in [SE] and* $\rho^*_{LC} = 0.97$ *in [CK].*
(iii) All the patterns in Hypothesis 1 are rejected for the pooled data of rounds 1 through 10.
(iv) The pattern [MT] (leadership by confidence) is accepted for the data of rounds 6–10 against both the alternative hypothesis that subjects follow some behavior strategy and the alternative hypothesis that subjects follow some correlated strategy.

Result 1 is understood intuitively in another light by converting the original data of plays by pairs in Table 6 into data of behaviors by subjects. The panel of rounds 1–5 in Table 8 shows the distributions of 150 behaviors chosen by 30 subjects in rounds 1–5. Subpanel H-1 shows a distribution of choices made by the H-type in period 1. Subpanel H-2 shows a joint distribution of choices made by the H-type in period 2 and choices made by his partner in period 1. The number of observations of $\times$ in panel H-1 is equal to the total number of observations in panel H-2, because the H-type's choices in period 2 occurs if and only if he has chosen $\times$ in period 1. Subpanels L-1 and L-2 show the corresponding distributions for the L-type. The panel of rounds 6–10 shows the corresponding distributions of behaviors in rounds 6–10.

Period 1

	Alternative 2 (e_H)	Alternative 1 (e_L)	$\times$ ($\emptyset$)	Total
H	**43**	1	8	52

Panel H-1

Period 2

	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
H	2 (e_H)	**4**	0	4
	1 (e_L)	**0**	1	1
	$\times$ ($\emptyset$)	**2**	1	3
	Total	6	2	8

Panel H-2

	Alternative 2 (e_H)	Alternative 1 (e_L)	$\times$ ($\emptyset$)	Total
L	5	9	**84**	98

Panel L-1

	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
L	2 (e_H)	**27**	8	35
	1 (e_L)	0	**8**	8
	$\times$ ($\emptyset$)	1	**40**	41
	Total	28	56	84

Panel L-2

Panel: Rounds 1–5

Period 1				
	Alternative 2 (e_H)	Alternative 1 (e_L)	× ($\emptyset$)	Total
H	**52**	2	4	58

Panel H-1

Period 2				
	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
H	2 (e_H)	**3**	0	3
	1 (e_L)	**0**	0	0
	× ($\emptyset$)	**1**	0	1
	Total	4	0	4

Panel H-2

	Alternative 2 (e_H)	Alternative 1 (e_L)	× ($\emptyset$)	Total
L	4	6	**82**	92

Panel L-1

	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
L	2 (e_H)	**25**	4	29
	1 (e_L)	0	**6**	6
	× ($\emptyset$)	0	**47**	47
	Total	25	57	82

Panel L-2

Panel: Rounds 6–10

Table 8: The Distributions of Behaviors of Subjects in Treatment 1

In the panel of rounds 6–10 in Table 8, the boldface numbers are the numbers of observations of behaviors prescribed by leadership by confidence. The observations clearly concentrate on the prescribed behaviors. Subpanel H-1 shows that subjects of H-type choose the prescribed Alternative 2 in period 1 at a rate of $\frac{52}{58} = 0.90$. Subpanel H-2 shows that when subjects of H-type mistakenly choose × in period 1, they respond perfectly with the prescribed Alternative 2. Subpanel L-1 shows that subjects of L-type choose the prescribed × in period 1 at a rate of $\frac{82}{92} = 0.89$. Subpanel L-2 shows that when subjects of L-type choose the prescribed × in period 1, they respond to the partner's choice of Alternative 2 with the prescribed Alternative 2 at a rate of $\frac{25}{29} = 0.86$ and to Alternative 1 and × with the prescribed Alternative 1 perfectly. The rates of obedience to leadership by confidence are high for all the information sets. There is no particular tendency to deviate from leadership by confidence. These features in observed subjects' behaviors fit [MT] of Hypothesis 1 well. The panel of rounds 1–5 shows similar fitness less clearly.

5.3 Result of Treatment 2

Table 9 shows the data of Treatment 2 in the same format as that for Treatment 1.

Period 1	2, 2	2, 1	1, 2	1, 1	2, ×	2, ×	1, ×	1, ×	×, 2	×, 2	×, 1	×, 1	×, ×	×, ×	×, ×	×, ×
Period 2	×, ×	×, ×	×, ×	×, ×	×, 2	×, 1	×, 2	×, 1	2, ×	1, ×	2, ×	1, ×	2, 2	2, 1	1, 2	1, 1
H, H	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
H, L	10	0	0	0	3	1	0	0	0	0	0	0	3	0	0	0
L, H	4	0	0	0	5	2	0	0	3	0	0	0	0	0	0	0
L, L	0	0	0	13	0	0	0	11	0	0	0	7	0	0	0	9

Panel: Rounds 1–5

Period 1	2, 2	2, 1	1, 2	1, 1	2, ×	2, ×	1, ×	1, ×	×, 2	×, 2	×, 1	×, 1	×, ×	×, ×	×, ×	×, ×
Period 2	×, ×	×, ×	×, ×	×, ×	×, 2	×, 1	×, 2	×, 1	2, ×	1, ×	2, ×	1, ×	2, 2	2, 1	1, 2	1, 1
H, H	4	0	0	0	2	1	0	0	1	1	0	0	1	0	0	0
H, L	6	0	0	0	4	0	0	0	1	1	0	0	1	0	1	0
L, H	3	0	0	0	2	1	0	0	2	0	0	0	0	1	0	0
L, L	0	0	0	15	0	0	0	14	0	0	0	11	0	0	0	2

Panel: Rounds 6–10

Table 9: The Data of Treatment 2 (distributions of plays by pairs)

We test Hypothesis 2 for the data of rounds 1–5, 6–10, and 1–10 as we did for Hypothesis 1 by applying the error rate analysis. The test of whether the data of rounds 1–5 supports Hypothesis 2 is run as follows. The null hypothesis is Hypothesis 2. We construct the following log likelihood function of parameters α_X, α_Y, α_Z, and ε for the data of rounds 1–5 under the null hypothesis.

$$\begin{aligned} L_0(\alpha_X, \alpha_Y, \alpha_Z, \varepsilon) &= 3 \times \ln\left[((1-\varepsilon)\alpha_Z)^2\right] \\ &\quad +1 \times \ln\left[(1-\varepsilon)\alpha_Z(1-\varepsilon)(1-\alpha_Z)\cdot(1-\varepsilon)\right] \\ &\quad +\cdots \end{aligned} \tag{10}$$

The brackets in the first term show the probability under the null hypothesis for the event that when table Z is selected, a pair of subjects both choose Alternative 2 in period 1 with probabilities $(1-\varepsilon)\alpha_Z$ prescribed by the sequential equilibrium. The remaining terms correspond to probabilities for observed plays similarly computed under the null hypothesis.

We maximize the likelihood function (10) with respect to parameters α_X, α_Y, α_Z, and ε. The maximum likelihood estimates are $\alpha_X^* = 0.55$, $\alpha_Y^* = 0.68$, $\alpha_Z^* = 0.87$, and $\varepsilon^* = 0.02$, and the maximized log likelihood is $L_0(\alpha_X^*, \alpha_Y^*, \alpha_Z^*, \varepsilon^*) = -112.91$. These values are shown in Table 10.

The alternative hypothesis that we adopt is the same as for the test of symmetric equilibrium in Treatment 1. Every subject plays according to some behavior strategy $\beta = (\beta_X, \beta_Y, \beta_Z)$. A subject follows β_X, β_Y, and β_Z when tables X, Y, and Z are selected respectively. A corresponding log likelihood function under the alternative hypothesis is constructed and maximized with respect to the parameters in the behavior strategy β. The value L_1^* of the maximized log likelihood is shown in Table 10.

We then conduct a χ^2 test with the likelihood ratio:

$$2(L_1^* - L_0(\alpha_X^*, \alpha_Y^*, \alpha_Z^*, \varepsilon^*)) = 17.17.$$

This value is shown in Table 10 together with the degree of freedom and the p-value for the χ^2 test. The tests of Hypothesis 2 for the data of rounds 6–10 and 1–10 are similar and shown in Table 10.

	Estimates				L_0^*	L_1^*	$2(L_1^* - L_0^*)$	χ^2-d.f.	p-value
	ε	α_X	α_Y	α_Z					
Rounds 1–5									
SE	0.02	0.55	0.68	0.87	-112.91	-104.33	17.17	11	0.10
Rounds 6–10									
SE	0.03	0.65	0.63	0.65	-124.43	-109.78	29.29	11	2.04E-03
Rounds 1–10									
SE	0.02	0.60	0.66	0.71	-239.61	-218.03	43.16	11	1.02E-05

Table 10: The Test Results of Hypothesis 2

Table 10 shows the following result for Hypothesis 2.

Result 2 *Hypothesis 2 is accepted for the data of rounds 1–5. It is rejected for the data of rounds 6–10 and 1–10.*

The failure of Hypothesis 2 to fit the data of rounds 6–10 and 1–10 is understood intuitively in another light by converting the original data of plays by pairs in Table 9 into data of behaviors by subjects. Table 11 shows the data as in Table 8 for Treatment 1. The boldface numbers present the numbers of observations of behaviors prescribed by the sequential equilibrium. In spite of the fact that obedience to the sequential equilibrium is perfect in period 1 for both rounds 1–5 and 6–10, obedience in period 2 is perfect only for table X, and the fitness is poor for tables Y and Z for the data of rounds 6–10.

Period 1

	Alternative 2 (e_H)	Alternative 1 (e_L)	× ($\emptyset$)	Total
Z	**7**	0	**1**	8

Panel Z-1

Period 2

	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
Z	2 (e_H)	**1**	0	1
	1 (e_L)	**0**	0	0
	× ($\emptyset$)	**0**	0	0
	Total	1	0	1

Panel Z-2

	Alternative 2 (e_H)	Alternative 1 (e_L)	× ($\emptyset$)	Total
Y	**42**	0	**20**	62

Panel Y-1

	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
Y	2 (e_H)	**11**	3	14
	1 (e_L)	**0**	0	0
	× ($\emptyset$)	**6**	0	6
	Total	17	3	20

Panel Y-2

	Alternative 2 (e_H)	Alternative 1 (e_L)	× ($\emptyset$)	Total
X	0	**44**	**36**	80

Panel X-1

	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
X	2 (e_H)	0	**0**	0
	1 (e_L)	0	**18**	18
	× ($\emptyset$)	0	**18**	18
	Total	0	36	36

Panel X-2

Panel: Rounds 1–5

Period 1

	Alternative 2 (e_H)	Alternative 1 (e_L)	× ($\emptyset$)	Total
Z	**13**	0	**7**	20

Panel Z-1

Period 2

	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
Z	2 (e_H)	**3**	2	5
	1 (e_L)	**0**	0	0
	× ($\emptyset$)	**2**	0	2
	Total	5	2	7

Panel Z-2

	Alternative 2 (e_H)	Alternative 1 (e_L)	× ($\emptyset$)	Total
Y	**29**	0	**17**	46

Panel Y-1

	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
Y	2 (e_H)	**9**	2	11
	1 (e_L)	**0**	0	0
	× ($\emptyset$)	**4**	2	6
	Total	13	4	17

Panel Y-2

	Alternative 2 (e_H)	Alternative 1 (e_L)	× ($\emptyset$)	Total
X	0	**55**	**29**	84

Panel X-1

	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
X	2 (e_H)	0	**0**	0
	1 (e_L)	0	**25**	25
	× ($\emptyset$)	0	**4**	4
	Total	0	29	29

Panel X-2

Panel: Rounds 6–10

Table 11: The Distributions of Behaviors of Subjects in Treatment 2

5.4 Result of Treatment 3

Table 12 shows the data of Treatment 3 in the same format as Table 6 for the data of Treatment 1.

Period 1	2, 2	2, 1	1, 2	1, 1	2, ×	2, ×	1, ×	1, ×	×, 2	×, 2	×, 1	×, 1	×, ×	×, ×	×, ×	×, ×
Period 2	×, ×	×, ×	×, ×	×, ×	×, 2	×, 1	×, 2	×, 1	2, ×	1, ×	2, ×	1, ×	2, 2	2, 1	1, 2	1, 1
H, H	2	0	1	0	1	0	0	1	1	2	0	0	1	2	0	1
H, L	0	0	0	1	1	1	0	2	0	0	0	4	1	1	0	4
L, H	0	0	3	1	0	0	0	6	1	0	0	2	0	0	2	3
L, L	0	0	1	4	0	0	0	8	0	0	0	3	0	0	1	13

Panel: Rounds 1–5

Period 1	2, 2	2, 1	1, 2	1, 1	2, ×	2, ×	1, ×	1, ×	×, 2	×, 2	×, 1	×, 1	×, ×	×, ×	×, ×	×, ×
Period 2	×, ×	×, ×	×, ×	×, ×	×, 2	×, 1	×, 2	×, 1	2, ×	1, ×	2, ×	1, ×	2, 2	2, 1	1, 2	1, 1
H, H	0	0	0	0	1	3	0	1	1	1	0	0	0	0	0	2
H, L	0	1	0	0	1	1	0	0	0	0	0	1	0	4	0	5
L, H	0	0	2	1	1	1	0	1	0	0	0	0	0	1	1	3
L, L	0	0	0	2	1	1	0	9	0	0	0	4	0	2	0	23

Panel: Rounds 6–10

Table 12: The Data of Treatment 3 (distributions of plays by pairs)

We test Hypothesis 3 for the data of rounds 1–5, 6–10, and 1–10 as we did for Hypothesis 1 by applying the error rate analysis. The test of whether the data of rounds 1–5 supports Hypothesis 3 is run as follows. The null hypothesis is Hypothesis 3. We construct the following log likelihood function of parameters α_H, α_L, and ε for the data of rounds 1–5 under the null hypothesis.

$$L_0(\alpha_H, \alpha_L, \varepsilon) = 2 \times \ln \left[\varepsilon^2\right] + 1 \times \ln \left[(1-\varepsilon)\alpha_H \varepsilon\right] + \cdots \tag{11}$$

The brackets in the first term show the probability under the null hypothesis for the event that both subjects in a pair mistakenly choose Alternative 2 out of the sequential equilibrium in period 1 when they are of H types. The remaining terms correspond to probabilities for observed plays similarly computed under the null hypothesis.

We maximize the likelihood function (11) with respect to parameters α_H, α_L, and ε. The maximum likelihood estimates are $\alpha_H^* = 0.38$, $\alpha_L^* = 0.19$, and $\varepsilon^* = 0.12$ and the maximized log likelihood is $L_0(\alpha_H^*, \alpha_L^*, \varepsilon^*) = -172.12$. These values are shown in Panel SE of Table 13.

The alternative hypothesis that we adopt is the same as for the test of symmetric equilibrium in Treatment 1. Every subject plays according to some behavior strategy $\beta = (\beta_H, \beta_L)$. A subject follows β_H when his type is H and β_L when his type is L. A corresponding log likelihood function under the alternative hypothesis is constructed and maximized with respect to the parameters in the behavior strategy β. The value L_1^* of maximized log likelihood is shown in Table 13.

We then conduct a χ^2 test with the likelihood ratio:

$$2(L_1^* - L_0(\alpha_H^*, \alpha_L^*, \varepsilon^*)) = 53.35.$$

This value is shown in Table 13 together with the degree of freedom and the p-value for the χ^2 test. The tests of Hypothesis 3 for the data of rounds 6–10 and 1–10 are similar and shown in Panel SE of Table 13.

	Estimates					L_0^*	L_1^*	$2(L_1^* - L_0^*)$	χ^2-d.f.	p-value
	ε	ρ_D	ρ_{LCC}	α_H	α_L					
Rounds 1–5										
SE	0.12	1.00	—	0.38	0.19	-172.12	-145.45	53.35	7	3.16E-09
Rounds 6–10										
SE	0.11	1.00	—	0.22	0.07	-151.20	-134.44	33.53	7	2.11E-05
Rounds 1–10										
SE	0.12	1.00	—	0.30	0.14	-327.77	-285.53	84.48	7	1.67E-15

Panel: SE

	Estimates					L_0^*	L_1^*	$2(L_1^* - L_0^*)$	χ^2-d.f.	p-value
	ε	ρ_D	ρ_{LCC}	α_H	α_L					
Rounds 1–5										
MIX	0.02	0.50	0.25	0.28	0.73	-147.07	-145.45	3.24	5	0.66
Rounds 6–10										
MIX	0.04	0.55	0.25	0.06	0.37	-135.27	-134.44	1.67	5	0.89
Rounds 1–10										
MIX	0.03	0.52	0.25	0.18	0.54	-287.61	-285.53	4.16	5	0.53

Panel: MIX

Table 13: The Test Results of Hypothesis 3

Panel SE of Table 13 shows the following result for Hypothesis 3.

Result 3 *Hypothesis 3 is rejected for the data of rounds 1–5, 6–10, and 1–10.*

The failure of Hypothesis 3 to fit the data is understood intuitively in another light by converting the original data of plays by pairs in Table 12 into data of behaviors by subjects. Table 14 shows the data as in Table 8 for Treatment 1. The boldface numbers show the numbers of observations of behaviors prescribed by the sequential equilibrium. Obedience to the sequential equilibrium is perfect for the information sets for the H-type and the L-type in period 2 when the subject sees his partner choose 1 in period 1. The rates of obedience to the sequential equilibrium are also fairly high for particular information sets for the L-type: a choice in period 1 and a choice in period 2 when the subject sees his opponent choose $\times$ in period 1. The rates are $\frac{92}{93} = 0.99$ and $\frac{37}{39} = 0.95$ for the data of rounds 1–5 and $\frac{104}{108} = 0.96$ and $\frac{61}{64} = 0.95$ for the data of rounds 6–10. However, the rates of obedience to the sequential equilibrium are not as high for the other information sets.

Period 1

	Alternative 2 (e_H)	Alternative 1 (e_L)	$\times$ ($\emptyset$)	Total
H	15	**8**	**34**	57

Panel H-1

Period 2

	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
H	2 (e_H)	2	**2**	4
	1 (e_L)	0	**11**	11
	$\times$ ($\emptyset$)	8	**11**	19
	Total	10	24	34

Panel H-2

	Alternative 2 (e_H)	Alternative 1 (e_L)	$\times$ ($\emptyset$)	Total
L	1	**35**	**57**	93

Panel L-1

	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
L	2 (e_H)	2	**1**	3
	1 (e_L)	0	**15**	15
	$\times$ ($\emptyset$)	2	**37**	39
	Total	4	53	57

Panel L-2

Panel: Rounds 1–5

Period 1				
	Alternative 2 (e_H)	Alternative 1 (e_L)	× (∅)	Total
H	11	**2**	**29**	42

Panel H-1

Period 2				
	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
H	2 (e_H)	3	**5**	8
	1 (e_L)	0	**3**	3
	× (∅)	5	**13**	18
	Total	8	21	29

Panel H-2

	Alternative 2 (e_H)	Alternative 1 (e_L)	× (∅)	Total
L	4	**23**	**81**	108

Panel L-1

	Choice / Partner's Choice	Alternative 2 (e_H)	Alternative 1 (e_L)	Total
L	2 (e_H)	2	**2**	4
	1 (e_L)	0	**13**	13
	× (∅)	3	**61**	64
	Total	5	76	81

Panel L-2

Panel: Rounds 6–10

Table 14: The Distributions of Behaviors of Subjects in Treatment 3

6 Discussions

6.1 Discussion of Result 1

In Treatment 1, all the patterns [SE], [CK], and [MT] of Hypothesis 1 are accepted in rounds 6–10. Note that the maximum likelihoods of [SE] and [CK] are better than that of [MT] only by 0.20 and 0.10 respectively by sacrificing the degrees of freedom by 3 and 1. Note also that the estimated proportion of leadership by confidence being played in [SE] and [CK] are 0.94 and 0.97. These results seem to suggest that leadership by confidence prevails in rounds 6–10 as the endogenous signaling theory of leadership predicts.

Although the theory points out that there are four sequential equilibria, the experiment discovered that a particular equilibrium occurs. This observation corresponds to the prediction of the "mistaken theories" refinement.

In rounds 1–5, although statistically insignificant, the estimated parameters and the maximized log likelihoods have tendencies parallel to those in rounds 6–10. It seems that subjects followed leadership by confidence to a considerable extent from the beginning and an equilibration to leadership by confidence had occurred through plays in earlier rounds.[16]

6.2 Discussion of Result 2

According to Result 2, the behaviors in rounds 1–5 were consistent with the theory of sequential equilibrium. In rounds 6–10, the fitness of the theory is not as good as for

[16]The rejection of Hypothesis 1 for rounds 1–10 can be understood if one takes the data of rounds 1–5 as showing an equilibration process and the data of rounds 6–10 as showing an equilibrium. Remember from Table 5 that the p-value of the Fisher test for rounds 1–5 is 0.06 and that for rounds 6–10 is 0.96.

rounds 1–5. The reason is that there are 6 observations of responses to Alternative 2 with Alternative 1 and to $\times$ with Alternative 1 when tables Y and Z were selected.

This result for rounds 6–10 can be interpreted using the spite dilemma proposed by Saijo and Nakamura (1995) and Saijo (2008). They define a spite behavior as a choice motivated by the difference between own payoff and opponent's payoff. The spite dilemma occurs when the subject's mind wavers between choosing a spite behavior and choosing a payoff-maximizing behavior. They found that there were a certain number of subjects who chose spite behaviors when faced with a spite dilemma.[17]

In our Game 2, the spite behavior is choosing Alternative 1 for all tables X, Y, and Z. Therefore, the spite dilemma occurs in tables Y and Z. Result 2 can be interpreted as saying that in rounds 6–10, the behaviors of sequential equilibrium were observed in table X that has no spite dilemma, while 6 spite behaviors were observed in tables Y and Z that have the dilemma. The 6 spite behaviors were taken by 4 subjects, and the questionnaire responses of three of them suggest that they chose the spite behaviors intentionally.

If we follow this interpretation, we can state that our subjects, with the exception of a few spiteful ones, followed the sequential equilibrium in Treatment 2. Results 1 and 2 together suggest that the existence of private information is necessary for leadership to stably emerge through an endogenously formed leader–follower relationship. It supports the claim of the endogenous signaling theory of leadership that leadership emerges as a signaling process.

6.3 Discussion of Result 3

Hypothesis 3 was rejected. The subjects were not shown to follow the sequential equilibrium of Game 3. More specifically, as noted after Result 3, the data fit well to some part of the prediction of the sequential equilibrium and do not fit as well to some other part.

To understand the failure of Hypothesis 3, we conduct a further exploration of the data of Treatment 3 to search for a set of meaningful behavior patterns if not a single behavior in equilibrium. There are three patterns of behavior that may be candidates for such a set. The first is dominance behavior. It is the behavior strategy in the sequential equilibrium hypothesized in Hypothesis 3. A subject chooses Alternative 1 in either period 1 or period 2.

The second is conditional cooperation behavior. Several experiments on games in which players' rationality is in conflict with Pareto optimality have revealed conditional cooperation. Conditional cooperation means that an agent is willing to cooperate if opponents cooperate to realize Pareto optimality, and that agent is reluctant otherwise. In Game 3, a subject is in a situation of the same kind. It is rational for a subject to choose Alternative 1 for any payoffs of X, Y, and Z, while it is Pareto optimal for a pair of agents to choose Alternative 1 if payoffs are X and to choose Alternative 2 if payoffs are either Y or Z. Rationality is in conflict with Pareto optimality in payoffs Y and Z. When a subject is of H-type, he knows that the payoffs are either Y or

[17] Spite behaviors are observed in various environments. See Cason, Saijo, and Yamato (2002), Brandts, Saijo, and Schram (2004), Cason, Saijo, Yamato, and Yokotani (2004), and Fehr, Hoff, and Kshetramade (2008).

Z and that it is Pareto optimal for a pair of agents to choose Alternative 2. The conditional cooperation behavior in Game 3 means that a subject of H-type behaves in a way compatible with the conditional cooperation. Namely, he chooses $\times$ in period 1 and responds to his partner's choice of Alternative 2 or $\times$ with Alternative 2 and to Alternative 1 with Alternative 1.

Furthermore, even when a subject is of L-type, he may conceive himself as being in the same situation as the H-type. Suppose he expects that his partner will choose Alternative 2 in period 1 only when he is of H-type. When the partner does so in fact, it signals to the L-type subject that the payoffs are Y, and he knows that his partner has committed to the Pareto optimal choice. The conditional cooperation behavior in Game 3 also means that the L-type subject chooses $\times$ in period 1 and responds to a partner's choice of Alternative 2 with Alternative 2 and to Alternative 1 or $\times$ with Alternative 1.

The third type of behavior is leadership-for-conditional-cooperation (LCC). LCC behavior means that a subject of H-type chooses Alternative 2 in period 1 while a subject of L-type chooses $\times$ in period 1 and responds to his partner's choice of Alternative 2 with Alternative 2 and to Alternative 1 or $\times$ with Alternative 1. When there is a chance for a partner to follow LCC behavior, it makes sense for a subject to choose the conditional cooperation behavior stated above.

We postulate the following hypothesis.

Hypothesis 4 *In Game 3, a subject will play dominance behavior with probability ρ_D, will play conditional cooperation behavior with probability ρ_{CC}, and will play LCC behavior with probability $\rho_{LCC} = 1 - \rho_D - \rho_{CC}$.*

We test Hypothesis 4 for the data of rounds 1–5, 6–10, and 1–10 as for Hypothesis 1 by applying the error rate analysis. The test of whether the data of rounds 1–5 support Hypothesis 4 is as follows. The null hypothesis is Hypothesis 4. When dominance behavior is followed, a subject of H-type chooses Alternative 1 in period 1 with probability α_H and chooses $\emptyset$ with probability $1 - \alpha_H$ to respond to any choice of his partner with Alternative 1 while a subject of L-type chooses Alternative 1 in period 1 with probability α_L and chooses $\emptyset$ with probability $1 - \alpha_L$ to respond to any choice of his partner with Alternative 1.

We construct the following log likelihood function of parameters ρ_D, ρ_{LCC}, α_H, α_L, and ε for the data of rounds 1–5 under the null hypothesis.

$$\begin{aligned}
&L_0(\rho_D, \rho_{LCC}, \alpha_H, \alpha_L, \varepsilon) \\
&= 2 \times \ln\Big[\{\rho_D \times \varepsilon + \rho_{LCC} \times (1 - 2\varepsilon) + (1 - \rho_D - \rho_{LCC}) \times \varepsilon\}^2\Big] \\
&+1 \times \ln\Big[\{\rho_D \times (1 - \varepsilon) \times \alpha_H + \rho_{LCC} \times \varepsilon + (1 - \rho_D - \rho_{LCC}) \times \varepsilon\} \\
&\qquad\qquad \times\{\rho_D \times \varepsilon + \rho_{LCC} \times (1 - 2\varepsilon) + (1 - \rho_D - \rho_{LCC}) \times \varepsilon\}\Big] \\
&+\cdots \qquad\qquad (12)
\end{aligned}$$

The brackets in the first term show the probability under the null hypothesis for the event that both subjects choose Alternative 2 in period 1 when they are of H types. The probability is the product of probabilities that a subject chooses Alternative 2 in period

1 when he is of H-type. The probability $\rho_D \times \varepsilon$ in the braces in the first term means that when a subject is of H-type, he follows dominance behavior with probability ρ_D and chooses Alternative 2 in period 1 with probability ε. The remaining probabilities in the braces correspond to those of LCC behavior and conditional cooperation behavior. The remaining terms correspond to probabilities for observed plays similarly computed under the null hypothesis.

We maximize the likelihood function (12) with respect to parameters ρ_D, ρ_{CC}, α_H, α_L, and ε. The maximum likelihood estimates are $\rho_D^* = 0.50$, $\rho_{LCC}^* = 0.25$, $\alpha_H^* = 0.28$, $\alpha_L^* = 0.73$, and $\varepsilon^* = 0.02$ and the maximized log likelihood is $L_0(\rho_D^*, \rho_{LCC}^*, \alpha_H^*, \alpha_L^*, \varepsilon^*) = -147.07$. These values are shown in Panel MIX of Table 13.

The alternative hypothesis, the log likelihood function under the alternative hypothesis and the value L_1^* of maximized log likelihood are the same as in the analysis of Hypothesis 3. The value L_1^* is replicated in Panel MIX of Table 13.

We then conduct a χ^2 test with the likelihood ratio:

$$2(L_1^* - L_0(\rho_D^*, \rho_{LCC}^*, \alpha_H^*, \alpha_L^*, \varepsilon^*)) = 3.24.$$

This value is shown in Panel MIX of Table 13 together with the degree of freedom and the p-value for the χ^2 test. The tests of Hypothesis 4 for the data of rounds 6–10 and the pooled data of rounds 1 through 10 are similar and shown in Panel MIX of Table 13.

Panel MIX of Table 13 shows the following result for Hypothesis 4.

Result 4 *Hypothesis 4 is accepted for the data of rounds 1–5, 6–10, and 1–10.*

Result 4 means that the data of Treatment 3 is explained by the combination of the three behavior patterns. The fractions of these patterns are estimated to be $\rho_D^* = 0.5$, $\rho_{CC}^* = 0.25$, and $\rho_{LCC}^* = 0.25$ in rounds 1–5 and $\rho_D^* = 0.55$, $\rho_{CC}^* = 0.20$, and $\rho_{LCC}^* = 0.25$ in rounds 6–10.

An implication of this result for the endogenous signaling theory of leadership is as follows. In Game 3, subjects follow dominance behavior with a probability of more than one-half. This fact partially supports the prediction of the theory for Game 3. Furthermore, when subjects are of H-types, only 25% corresponding to LCC behavior adopt leadership behavior in Game 3 while $100 \times (1 - 2\varepsilon) = 90\%$ are expected to lead in Game 1 ([MT] in rounds 6–10 in Table 7). This supports the prediction of the theory that leadership does not emerge in Game 3 while it does in Game 1. Hence, Result 4 suggests that, in spite of Result 3, there is no reason to abandon the endogenous signaling theory of leadership.

Finally, we propose a direction for future research. Result 4 implies that a successful leader–follower relation is realized with a nonnegligible probability in Game 3 for which the endogenous signaling theory predicts no emergence of leadership. Specifically, when both subjects are of H-type and one of them follows LCC behavior while the other follows conditional cooperation behavior, the former leads with Alternative 2 and the latter responds with Alternative 2. The probability of this emergence of leadership is $2\rho_{LCC}^* \rho_{CC}^* = 0.125$ in rounds 1–5 and 0.1 in rounds 6–10. The same successful leader–follower relation is also realized when one subject is of H-type and follows LCC behavior while the other is of L-type and follows either LCC or conditional cooperation behavior. The probability is $\rho^*_{LCC}(\rho^*_{LCC} + \rho^*_{CC}) = 0.125$ in rounds 1–5 and 0.1125 in rounds 6–10. This suggests that there is a leadership mechanism in teams other than endogenous signaling by rational agents. This leadership mechanism needs to be theorized. For this purpose, first, we need to define the preferences of individuals that lead to conditional cooperation and LCC behavior. Then, we need to establish a theory that explains the combination of the three behavior patterns as an equilibrium under those preferences. These issues should be explored in future research.

Appendix

Proof of Proposition 1

First, examine the sequential rationality of the strategy for L-types. Consider an L-type's behaviors in period 2. Let $\mu(e_j^1)$ be a follower's belief that the opponent j is of H-type when observing that he chooses e_j^1 in period 1. When $e_j^1 = e_H$, a consistent belief is $\mu(e_H) = 1$. This means that the follower believes that team productivity is $\mu(e_H)\theta_M + (1 - \mu(e_H))\theta_L = \theta_M$. Then, choosing e_H is sequentially rational for an L-type agent if and only if $\theta_M(e_H + e_H) - c_H > \theta_M(e_L + e_H) - c_L$; that is:

$$\theta_M > \frac{c_H - c_L}{e_H - e_L}. \tag{13}$$

Similarly, when $e_j^1 = \emptyset$, a consistent belief is $\mu(\emptyset) = 0$. This means that team productivity is $\mu(\emptyset)\theta_M + (1 - \mu(\emptyset))\theta_L = \theta_L$. Then, choosing e_L is sequentially rational for an L-type agent if and only if $\theta_L(e_L + e_L) - c_L > \theta_L(e_H + e_L) - c_H$; that is:

$$\theta_L < \frac{c_H - c_L}{e_H - e_L}. \tag{14}$$

Let us consider an L-type's behavior in period 1. An L-type believes with probability ρ that the opponent is of H-type and team productivity is θ_M and believes with probability $1 - \rho$ that the opponent is of L-type and team productivity is θ_L. Then, an L-type agent does not mimic an H-type if and only if:

$$\begin{aligned}
&\rho\{\theta_M(e_H + e_H) - c_H\} + (1 - \rho)\{\theta_L(e_L + e_L) - c_L\} \\
&> \rho\{\theta_M(e_H + e_H) - c_H\} + (1 - \rho)\{\theta_L(e_H + e_H) - c_H\},
\end{aligned}$$

that is:

$$2\theta_L < \frac{c_H - c_L}{e_H - e_L}. \tag{15}$$

Second, examine the sequential rationality of the strategy for H-types. An H-type believes with probability ρ that the opponent is of H-type and team productivity is θ_H and believes with probability $1 - \rho$ that the opponent is of L-type and team productivity is θ_M. Then, the expected payoff from the equilibrium is:

$$\rho\{\theta_H(e_H + e_H) - c_H\} + (1 - \rho)\{\theta_M(e_H + e_H) - c_H\}.$$

If an H-type deviates to $\emptyset$, a sequentially rational choice in period 2 is e_H irrespective of his belief about the opponent's type under condition (13). The expected payoff from the deviation is:

$$\rho\{\theta_H(e_H + e_H) - c_H\} + (1-\rho)\{\theta_M(e_H + e_L) - c_H\}.$$

The payoff from the equilibrium is larger than that from the deviation.

Because (15) implies (14), a necessary and sufficient condition for leadership by confidence to be a sequential equilibrium is the pair of conditions (13) and (15); that is, (1).

Proof of Proposition 2

Let agent i be the leader and j be the follower. First, examine the sequential rationality of the behaviors of the follower j. The sequential rationality condition of the behaviors of an L-type is the same as for leadership by confidence; that is, (13) and (14) hold. Condition (13) also implies that the behaviors of an H-type are sequentially rational, because:

$$\begin{aligned}&\mu\{\theta_H(e_H + e_i^1) - c_H\} + (1-\mu)\{\theta_M(e_H + e_i^1) - c_H\} \\ &> \mu\{\theta_H(e_L + e_i^1) - c_L\} + (1-\mu)\{\theta_M(e_L + e_i^1) - c_L\}\end{aligned}$$

for any $e_i^1 = e_H, e_L$ and any belief μ that his partner i is of H-type.

Second, examine the sequential rationality of the behaviors of the leader i. The sequential rationality condition of the behaviors of the H-type is the same as for leadership by confidence; that is, the condition (13) because the H-type follower's choice is e_H irrespective of the leader's choice.

Let us turn to considering L-type behavior. The L-type believes with probability ρ that the opponent is of H-type and team productivity is θ_M and believes with probability $1-\rho$ that the opponent is of L-type and team productivity is θ_L. The L-type agent does not mimic an H-type if and only if:

$$\begin{aligned}&\rho\{\theta_M(e_L + e_H) - c_L\} + (1-\rho)\{\theta_L(e_L + e_L) - c_L\} \\ &> \rho\{\theta_M(e_H + e_H) - c_H\} + (1-\rho)\{\theta_L(e_H + e_H) - c_H\},\end{aligned}$$

that is:

$$\rho\theta_M + 2(1-\rho)\theta_L < \frac{c_H - c_L}{e_H - e_L}. \tag{16}$$

Because (16) implies (14), a necessary and sufficient condition for leadership by identity to be a sequential equilibrium is the pair of conditions (13) and (16); that is, (2).

Proof of Proposition 4

We classify agent i's strategy $(e_{i,t_i}^1, e_{i,t_i}^2(\cdot))_{t_i=H,L}$ into four groups according to $e_{i,H}^1 = \emptyset$ or $e_{i,H}^1 \neq \emptyset$ and $e_{i,L}^1 = \emptyset$ or $e_{i,L}^1 \neq \emptyset$. Taking symmetry into account, this defines ten groups of strategy profiles that we must study for equilibrium.

1. *The group of $e^1_{i,H} \neq \emptyset$ and $e^1_{i,L} = \emptyset$ for $i = 1,2$:* For each agent $i = 1,2$, there are two candidates of an H-type's behavior in period 1 for equilibrium: $e^1_{i,H} = e_H$ and $e^1_{i,H} = e_L$. The case of $e^1_{i,H} = e_H$ for both $i = 1,2$ corresponds to leadership by confidence. As is proved in Proposition 1, leadership by confidence is a sequential equilibrium when condition (3) is satisfied.

Consider the case of $e^1_{i,H} = e_L$, which may constitute the pathological sequential equilibrium. The pathological sequential equilibrium requires that $e^2_{j,L}(e_H) = e_L$, because otherwise agent i of H-type would deviate to choosing e_H in period 1 under condition (3). Suppose (4) in addition. Then, agent i of H-type has an incentive to deviate to choosing e_H in period 1. His expected payoff from the prescribed behavior $e^1_{i,H} = e_L$ is $\rho[\theta_H(e_L + e^1_{j,H}) - c_L] + (1-\rho)[\theta_M(e_L + e_H) - c_L]$. The expected payoff from the deviation to e_H in period 1 is $\rho[\theta_H(e_H + e^1_{j,H}) - c_H] + (1-\rho)[\theta_M(e_H + e_L) - c_H]$. The latter is higher than the former under (4).

Suppose (5) instead of (4). Then, agent i of L-type has an incentive to mimic an H-type and deviate to choosing e_L in period 1. His expected payoff from the prescribed behaviors $e^2_{i,L}(e^1_{j,H}) = e_H$ and $e^2_{i,L}(\emptyset) = e_L$ is $\rho[\theta_M(e_H + e^1_{j,H}) - c_H] + (1-\rho)[\theta_L(e_L + e_L) - c_L]$. The expected payoff from the devation to e_L in period 1 is $\rho[\theta_M(e_L + e^1_{j,H}) - c_L] + (1-\rho)[\theta_L(e_L + e_H) - c_L]$. The latter is higher than the former under (5).

Hence, leadership by confidence is the unique sequential equilibrium in this group of strategy profiles when condition (3) is satisfied and in addition (4) or (5) is satisfied.

2. *The group of $e^1_{i,H} \neq \emptyset$ and $e^1_{i,L} = \emptyset$ and $e^1_{j,H} = e^1_{j,L} = \emptyset$:* There are two candidates of an H-type agent i's behavior in period 1 for equilibrium: $e^1_{i,H} = e_H$ and $e^1_{i,H} = e_L$. The case of $e^1_{i,H} = e_H$ corresponds to leadership by identity with confidence. As is stated in Proposition 3, leadership by identity with confidence is a sequential equilibrium when condition (3) is satisfied.

Consider the case of $e^1_{i,H} = e_L$. Under condition (3), agent i of L-type chooses e_L in period 2. Then, if he deviates to choosing e_L in period 1, agent j of L-type will respond with e_H and this improves the expected payoff for the agent i of L-type. Therefore, $e^1_{i,H} = e_L$ does not constitute a sequential equilibrium. Hence, leadership by identity with confidence is the unique sequential equilibrium in this group of strategy profiles.

3. *The group of $e^1_{i,H} \neq \emptyset$ and $e^1_{i,L} \neq \emptyset$ and $e^1_{j,H} = e^1_{j,L} = \emptyset$:* There are four candidates of an H-type agent i's behavior in period 1 for equilibrium: (i) $e^1_{i,H} = e^1_{i,L} = e_H$, (ii) $e^1_{i,H} = e_H$ and $e^1_{i,L} = e_L$, (iii) $e^1_{i,H} = e_L$ and $e^1_{i,L} = e_H$, and (iv) $e^1_{i,H} = e^1_{i,L} = e_L$. The case of (ii) corresponds to leadership by identity.

The case of (i) is not a sequential equilibrium. Under condition (3), it must be the case that $e^2_{j,L}(e_H) = e_L$. Then, agent i of L-type improves his expected payoff under condition (3) by choosing e_L instead of e_H.

For case (iii), if agent i of L-type deviates to choosing e_L in period 1, agent j of L-type responds to the deviation with e_H. Therefore, agent i of L-type has an incentive for the deviation.

Finally, the case of (iv) is not a sequential equilibrium. Under condition (3), it must be the case that $e^2_{j,L}(e_L) = e_L$. Then, agent i of H-type improves his expected payoff under condition (3) by choosing e_H instead of e_L.

Hence, leadership by identity is the unique sequential equilibrium in this group of strategy profiles.

4. *The group of* $e^1_{i,H} = e^1_{i,L} = \emptyset$ *for* $i = 1, 2$: This group is no-leadership. Under condition (3), it must be the case that $e^2_{i,H}(\emptyset) = e_H$ and $e^2_{i,L}(\emptyset) = e_L$. Suppose that if a partner deviates to choosing a level of effort in period 1, an agent of L-type responds with e_L, believing that the partner is of L-type. This prevents any agent from deviating from choosing $\emptyset$ in period 1. Therefore, no-leadership with $e^2_{i,H}(\emptyset) = e_H$ and $e^2_{i,L}(\emptyset) = e_L$ is a sequential equilibrium.

5. *The group of* $e^1_{i,H} \neq \emptyset$ *and* $e^1_{i,L} \neq \emptyset$ *for* $i = 1, 2$: Under condition (3), it must be the case that $e^1_{i,H} = e_H$ and $e^1_{i,L} = e_L$. Then, an agent of L-type improves his expected payoff by choosing $\emptyset$ in period 1 and responding to his partner's choice of e_H (resp., e_L) with e_H (resp., e_L) in period 2. Therefore, there is no sequential equilibrium in this group.

6. *The group of* $e^1_{i,H} \neq \emptyset$ *and* $e^1_{i,L} \neq \emptyset$, *and* $e^1_{j,H} \neq \emptyset$ *and* $e^1_{j,L} = \emptyset$: For the same reason as for group 5, agent i of L-type improves his expected payoff by choosing $\emptyset$ in period 1 and resopnding to his partner's choice of $e^1_{j,H}$ (resp., $\emptyset$) with e_H (resp., e_L) in period 2.

7. *The group of* $e^1_{i,H} \neq \emptyset$ *and* $e^1_{i,L} \neq \emptyset$, *and* $e^1_{j,H} = \emptyset$ *and* $e^1_{j,L} \neq \emptyset$: Agent j's behavior is separating. Furthermore, under condition (3), agent j of H-type chooses e_H in period 2 irrespective of agent i's choice in period 1. Therefore, agent i of L-type improves his expected payoff by choosing $\emptyset$ in period 1 and responding to his partner's choice of $\emptyset$ (resp., $e^1_{j,L}$) with e_H (resp., e_L) in period 2. Hence, there is no sequential equilibrium in this group.

8. *The group of* $e^1_{i,H} = \emptyset$ *and* $e^1_{i,L} \neq \emptyset$ *for* $i = 1, 2$: For the same reason as for group 7, agent i of L-type deviates to choosing $\emptyset$ in period 1 and responding to his partner's choice of $\emptyset$ (resp., $e^1_{j,L}$) with e_H (resp., e_L) in period 2. Therefore, there is no sequential equilibrium in this group.

9. *The group of* $e^1_{i,H} = \emptyset$ *and* $e^1_{i,L} \neq \emptyset$, *and* $e^1_{j,H} \neq \emptyset$ *and* $e^1_{j,L} = \emptyset$: Sequential equilibrium requires that $e^2_{j,L}(\emptyset) = e_H$ and $e^2_{j,L}(e^1_{i,L}) = e_L$. Then, if agent i of L-type deviates to choosing $\emptyset$ in period 1, it induces agent j of L-type to respond with e_H. This improves the expected payoff for the agent i of L-type. He improves his expected payoff even further by responding to his partner's choice of $\emptyset$ (resp., $e^1_{j,L}$) with e_H (resp., e_L) in period 2. Therefore, there is no sequential equilibrium in this group.

10. *The group of* $e^1_{i,H} = \emptyset$ *and* $e^1_{i,L} \neq \emptyset$, *and* $e^1_{j,H} = e^1_{j,L} = \emptyset$: Agent i of L-type improves his expected payoff by deviating to choosing $\emptyset$ in period 1 for the same reason as for group 9. Therefore, there is no sequential equilibrium in this group.

Proof of Proposition 5

No-leadership does not pass the intuitive criterion. Consider a deviation to moving first with e_H. Suppose that a partner's belief given the deviation is that the deviator is of H-type. Then, an H-type agent prefers the deviation to the equilibrium behavior, because the payoff from the deviation:

$$\rho[\theta_H(e_H + e_H) - c_H] + (1 - \rho)[\theta_M(e_H + e_H) - c_H]$$

is larger than the payoff from the equilibrium behavior:

$$\rho[\theta_H(e_H + e_H) - c_H] + (1 - \rho)[\theta_M(e_H + e_L) - c_H]$$

by $(1 - \rho)\theta_M(e_H - e_L) > 0$. On the contrary, an L-type agent prefers the equilibrium behavior to the deviation because the payoff from the equilibrium behavior:

$$\rho[\theta_M(e_L + e_H) - c_L] + (1 - \rho)[\theta_L(e_L + e_L) - c_L]$$

is larger than the payoff from the deviation:

$$\rho[\theta_M(e_H + e_H) - c_H] + (1 - \rho)[\theta_L(e_H + e_H) - c_H]$$

by $(e_H - e_L)[\frac{c_H - c_L}{e_H - e_L} - (\rho\theta_M + 2(1-\rho)\theta_L)] > 0$ under the condition (2). The presumed belief given the deviation is self-fulfilling. Therefore, the deviation upsets no-leadership.

Leadership by identity with confidence does not pass the intuitive criterion either. Consider a deviation by a follower to moving first with e_H. Suppose that a partner's belief given the deviation is that the deviator is of H-type. Then, an H-type follower prefers the deviation to the equilibrium behavior for the same reason as in the above argument of no-leadership. An L-type follower prefers the equilibrium behavior to this deviation because the payoff from the equilibrium behavior:

$$\rho[\theta_M(e_H + e_H) - c_H] + (1 - \rho)[\theta_L(e_L + e_L) - c_L]$$

is larger than the payoff from the deviation:

$$\rho[\theta_M(e_H + e_H) - c_H] + (1 - \rho)[\theta_L(e_H + e_H) - c_H]$$

by $(1-\rho)(e_H - e_L)[\frac{c_H - c_L}{e_H - e_L} - 2\theta_L] > 0$ under the condition (1). The presumed belief given the deviation is self-fulfilling. Therefore, the deviation upsets leadership by identity with confidence.

Leadership by confidence passes the intuitive criterion. The only out-of-equilibrium information set is the one reached by a deviation to moving first with e_L. Suppose that a partner's belief given the deviation is that the deviator is of H-type. An H-type agent prefers the equilibrium behavior to that deviation because the payoff from the equilibrium behavior:

$$\rho[\theta_H(e_H + e_H) - c_H] + (1 - \rho)[\theta_M(e_H + e_H) - c_H]$$

is larger than the payoff from the deviation:

$$\rho[\theta_H(e_L + e_H) - c_L] + (1 - \rho)[\theta_M(e_L + e_H) - c_L]$$

by $\rho(e_H - e_L)[\theta_H - \frac{c_H - c_L}{e_H - e_L}] + (1-\rho)(e_H - e_L)[\theta_M - \frac{c_H - c_L}{e_H - e_L}] > 0$ under the condition (1).

Hence, the out-of-equilibrium belief in the sequential equilibrium that the deviator is of L-type is reasonable.

Finally, leadership by identity also passes the intuitive criterion. There are three kinds of out-of-equilibrium information set. The first is a leader's information set reached when a leader deviates to $\emptyset$ and a follower deviates to moving first with e_H. Suppose that the leader's belief given the follower's deviation is that the deviator is of H-type. Given a follower's expectation that a leader will choose $\emptyset$ with probability zero, the follower expects that the partner's behavior would not be changed if he deviated to moving first with e_H. This means that the payoff from that deviation for an H-type follower is:

$$\rho[\theta_H(e_H + e_H) - c_H] + (1 - \rho)[\theta_M(e_H + e_L) - c_H],$$

which is the same as the payoff from the equilibrium behavior. The H-type follower is indifferent between that deviation and the equilibrium behavior. Hence, the out-of-equilibrium belief in the sequential equilibrium that the deviator is of L-type is reasonable.

The second kind of out-of-equilibrium information set is similar to the first and is reached when a leader deviates to $\emptyset$ and a follower deviates to moving first with e_L. Now, the payoff for an H-type follower from the deviation to moving first with e_L is:

$$\rho[\theta_H(e_L + e_H) - c_L] + (1 - \rho)[\theta_M(e_L + e_L) - c_L],$$

which is smaller than the payoff from the equilibrium behavior by $\rho(e_H - e_L)[\theta_H - \frac{c_H - c_L}{e_H - e_L}] + (1 - \rho)(e_H - e_L)[\theta_M - \frac{c_H - c_L}{e_H - e_L}] > 0$ under the condition (1). Hence, the out-of-equilibrium belief in the sequential equilibrium that the deviator is of L-type is reasonable.

The third kind of out-o-equilibrium information set is a follower's information set reached when a leader chooses $\emptyset$. Suppose that the follower's belief given this deviation is that the deviator is of H-type. Then, the payoff from the deviation for an H-type leader is:

$$\rho[\theta_H(e_H + e_H) - c_H] + (1 - \rho)[\theta_M(e_H + e_H) - c_H],$$

which is the same as the payoff from the equilibrium behavior. The H-type deviator is indifferent between that deviation and the equilibrium behavior. Hence, the out-of-equilibrium belief in the sequential equilibrium that the deviator is of L-type is reasonable.

Remark. As we noted in the discussion leading to Proposition 4, there may exist the pathological sequential equilibrium under condition (3). However, this equilibrium does not pass the intuitive criterion. Consider a deviation to moving first with e_H. Suppose that a partner's belief given the deviation is that the deviator is of H-type. Then, an H-type agent prefers the deviation to the equilibrium behavior under condition (3). On the other hand, an L-type agent has no incentive to deviate from the equilibrium behavior for the same reason that an L-type agent has no incentive to mimic the H-type behavior in leadership by confidence. Therefore, the deviation upsets the pathological equilibrium.

Proof of Proposition 6

A Formal Definition of a "Mistaken Theories" Refinement

We adopt a simple version of the "mistaken theories" refinement. We test the stability of a sequential equilibrium σ by each sequential equilibrium σ' other than σ. We say that a sequential equilibrium (strategy profile) σ is a "theory" of play for agent i when he believes with almost probability one that σ is a sequential equilibrium to be played by himself and agent j. We say that the theory σ is a "mistaken theory" when it is actually the case that agent j's theory is σ'; that is, agent j believes with almost probability one that σ' is a sequential equilibrium to be played by himself and agent i. Let $(\sigma^1, \sigma^2) \in \{\sigma, \sigma'\} \times \{\sigma, \sigma'\}$ denote a pair of agent 1's theory σ^1 and agent 2's theory σ^2. Let $p \in \Delta(\{\sigma, \sigma'\} \times \{\sigma, \sigma'\})$ denote a probability measure over the space of possible pairs of agents' theories. We imagine an extended incomplete information game in which a pair (σ^1, σ^2) of agents' theories is realized with probability $p(\sigma^1, \sigma^2)$ and agents play our team production game, each knowing their own theory privately. We study a play in which agent i follows prescription σ_i^i for agent i in the theory σ^i in which he believes.

The stability test of σ by σ' is defined as follows.

Definition. *A sequential equilibrium σ is not stable against a sequential equilibrium σ' in "mistaken theories" if there exists a sequence $\{(p^n, \mu^n)\}_{n=1}^{\infty}$ of probability measures p^n over the space of possible pairs of agents' theories and belief systems μ^n in the extended incomplete information game of team production with priors p^n that satisfies the following: (1-a) p^n has full support over $\{\sigma, \sigma'\} \times \{\sigma, \sigma'\}$ and (1-b) $\lim_{n\to\infty} p^n(\sigma^1, \sigma^2) > 0$ if $\sigma^1 = \sigma^2$ and $\lim_{n\to\infty} p^n(\sigma^1, \sigma^2) = 0$ otherwise; (2) μ^n is consistent with agents' conforming to prescriptions of their theories; and (3) in the extended incomplete information game of team production with priors p^n, the tested sequential equilibrium σ prescribes a sequentially irrational behavior given the beliefs μ^n for some information set while the other sequential equilibrium σ' prescribes sequentially rational behaviors given the beliefs μ^n for all information sets.*

Condition (1-b) requires that when σ^i is an agent i's theory, he believes with almost probability one that the theory is shared with agent j. Conditions (1-a) and (2) force a particular out-of-equilibrium belief that a deviation has occurred because a deviator follows the other sequential equilibrium. Condition (3) requires that the tested equilibrium σ has a problem in sequential rationality and the alternative equilibrium does not.

For an asymmetric sequential equilibrium such as leadership by identity, we extend a notion of "theory" as follows. In leadership by identity, let σ_ℓ denote the strategy for the leader and σ_f for the follower. Then, a theory of leadership by identity includes a strategy profile $\sigma_{\ell,f} = (\sigma_\ell, \sigma_f)$ and a strategy profile $\sigma_{f,\ell} = (\sigma_f, \sigma_\ell)$. When we test the stability of leadership by identity against another sequential equilibrium σ', we consider a probability measure p over a space $\{\sigma_{\ell,f}, \sigma_{f,\ell}, \sigma'\} \times \{\sigma_{\ell,f}, \sigma_{f,\ell}, \sigma'\}$.

Finally, we say that a sequential equilibrium σ is not stable in "mistaken theories" if it is not stable against some other sequential equilibrium in "mistaken theories".

Proof of Proposition 6

We show that leadership by identity is unstable using the "mistaken theories" refinement. Let us test leadership by identity against leadership by confidence σ'. As explained, we consider $\sigma_{\ell,f} = (\sigma_\ell, \sigma_f)$ and $\sigma_{f,\ell} = (\sigma_f, \sigma_\ell)$ to be a theory of leadership by identity. Consider a sequence of probability measures $\{p^n\}_{n=1}^{\infty}$ over a space $\{\sigma_{\ell,f}, \sigma_{f,\ell}, \sigma'\} \times \{\sigma_{\ell,f}, \sigma_{f,\ell}, \sigma'\}$ with properties (1-a) and (1-b). An example is $p^n(\sigma^1, \sigma^2) = \frac{1}{3} - \varepsilon_n$ if $\sigma^1 = \sigma^2$ and $\frac{\varepsilon_n}{2}$ otherwise where $\lim_{n\to\infty} \varepsilon_n = 0$. Let μ^n be a belief system consistent with a play under p^n.

Then, it is not sequentially rational for agent 1 with a theory $\sigma_{\ell,f}$ to move first with e_L when he is of L-type. The expected payoff from moving first with e_L is:

$$\begin{aligned} &q^n(\sigma_{\ell,f}|\sigma_{\ell,f})\Big\{\rho[\theta_M(e_L+e_H)-c_L]+(1-\rho)[\theta_L(e_L+e_L)-c_L]\Big\} \\ &+q^n(\sigma_{f,\ell}|\sigma_{\ell,f})\Big\{\rho[\theta_M(e_L+e_H)-c_L]+(1-\rho)[\theta_L(e_L+e_L)-c_L]\Big\} \\ &+q^n(\sigma'|\sigma_{\ell,f})\Big\{\rho[\theta_M(e_L+e_H)-c_L]+(1-\rho)[\theta_L(e_L+e_L)-c_L]\Big\}, \end{aligned} \tag{17}$$

where we calculate the conditional probabilities by:

$$q^n(\sigma_{\ell,f}|\sigma_{\ell,f}) = \frac{p^n(\sigma_{\ell,f}, \sigma_{\ell,f})}{p^n(\sigma_{\ell,f}, \sigma_{\ell,f}) + p^n(\sigma_{\ell,f}, \sigma_{f,\ell}) + p^n(\sigma_{\ell,f}, \sigma')}$$

and so forth. If he deviates to choosing $\emptyset$ in period 1 and responding to e_H with e_H and otherwise with e_L, the expected payoff is:

$$\begin{aligned} &q^n(\sigma_{\ell,f}|\sigma_{\ell,f})\Big\{\rho[\theta_M(e_L+e_H)-c_L]+(1-\rho)[\theta_L(e_L+e_L)-c_L]\Big\} \\ &+q^n(\sigma_{f,\ell}|\sigma_{\ell,f})\Big\{\rho[\theta_M(e_H+e_H)-c_H]+(1-\rho)[\theta_L(e_L+e_L)-c_L]\Big\} \\ &+q^n(\sigma'|\sigma_{\ell,f})\Big\{\rho[\theta_M(e_H+e_H)-c_H]+(1-\rho)[\theta_L(e_L+e_L)-c_L]\Big\}. \end{aligned} \tag{18}$$

The payoff (18) is larger than (17) by:

$$\big(q^n(\sigma_{f,\ell}|\sigma_{\ell,f}) + q^n(\sigma'|\sigma_{\ell,f})\big)\rho(e_H - e_L)\left(\theta_M - \frac{c_H - c_L}{e_H - e_L}\right) > 0$$

under condition (1).

On the other hand, it is sequentially rational for an agent with a theory σ' to follow it. This fact is straightforward for an information set on the equilibrium path of leadership by confidence, because $\lim_{n\to\infty} q^n(\sigma'|\sigma') = 1$ and it is strictly optimal at the information set to follow leadership by confidence. In leadership by confidence, when an agent is of H-type, an out-of-equilibrium information set is reached by mistakenly choosing $\emptyset$ in period 1. Because his choice in period 2 does not affect his partner's choice, it is optimal for the H type agent to choose e_H; that is, the prescription of leadership by confidence. For an agent (say, agent 1) of L-type, an out-of-equilibrium information set is reached when he chooses $\emptyset$ in period 1 and sees that his partner (agent 2) has chosen e_L in period 1. If agent 1 with a theory σ' is of L-type and is reached at this information set, then a consistent belief assigns probability one for the partner agent 2 of L-type with a theory $\sigma_{f,\ell}$, and probability zero otherwise. Therefore, the conditional expected payoff from choosing e_L is $\theta_L(e_L + e_L) - c_L$. The conditional expected payoff from deviating to e_H is $\theta_L(e_H + e_L) - c_H$. The former is larger than the latter by:

$$(e_H - e_L)\left(\frac{c_H - c_L}{e_H - e_L} - \theta_L\right) > 0$$

under condition (1). Hence, leadership by identity is unstable against leadership by confidence.

It can be verified by similar arguments that leadership by identity with confidence and no-leadership are also unstable against leadership by confidence. It is also straightforward to see that leadership by confidence is stable by the "mistaken theories" refinement because in the above proof of instability of leadership by identity, $\{p^n\}_{n=1}^{\infty}$ was taken arbitrarily up to the properties (1-a) and (1-b), and leadership by confidence was proved to be sequentially rational at all the information sets.

Remark. As we noted in the discussion leading to Proposition 4, there may exist the pathological sequential equilibrium under condition (3). However, this equilibrium is unstable by the "mistaken theories" refinement. Let us test the pathological equilibrium σ' against leadership by identity. As explained, we consider $\sigma_{\ell,f} = (\sigma_\ell, \sigma_f)$ and $\sigma_{f,\ell} = (\sigma_f, \sigma_\ell)$ to be a theory of leadership by identity. Consider a sequence of probability measures $\{p^n\}_{n=1}^{\infty}$ over a space $\{\sigma_{\ell,f}, \sigma_{f,\ell}, \sigma'\} \times \{\sigma_{\ell,f}, \sigma_{f,\ell}, \sigma'\}$ with properties (1-a) and (1-b). An example is $p^n(\sigma^1, \sigma^2) = \frac{1}{3} - \varepsilon_n$ if $\sigma^1 = \sigma^2$ and $\frac{\varepsilon_n}{2}$ otherwise where $\lim_{n\to\infty} \varepsilon_n = 0$. Let μ^n be a belief system consistent with a play under p^n. Then, an L-type agent who believes in σ' should choose e_H in period 2 when observing the partner's choice of e_H in period 1, because the partner's behavior is a sign that he is of H type and he behaves as a leader believing in leadership by identity. Therefore, σ' prescribes a sequentially irrational behavior. On the other hand, leadership by identity prescribes sequentially rational behaviors for all information sets. Particularly, an H-type follower does not have an incentive to deviate to moving first with e_L under the expectation of playing with an L-type agent who believes in σ', because the probability of being matched with this agent vanishes as $\varepsilon_n \to 0$ and the benefit of conforming to the choice e_H by leadership by identity dominates.

Even when the pathological sequential equilibrium exits, leadership by confidence is stable against the equilibrium by the "mistaken theories" refinement. An L-type agent who believes the pathological sequential equilibrium should choose e_H in period 2 when observing the partner's choice of e_H in period 1. The pathological sequential equilibrium prescribes a sequentially irrational choice of e_L for this case.

References

Andoreoni, James. 2006. "Leadership Giving in Charitable Fund-Raising," *Journal of Public Economic Theory*, 8(1): 1–22.

Arbak, Emrah, and Marie-Claire Villeval. 2007. "Endogenous Leadership: Selection and Influence," Institute for the Study of Labor, IZA Discussion Papers: No. 2732.

Banks, Jeffrey, Colin Camerer, and David Porter. 1994. "An Experimental Analysis of Nash Refinements in Signaling Games," *Games and Economic Behavior*, 6:(1) 1–31.

Bolton, Patrick, Markus K. Brunnermeier, and Laura L. Veldkamp. Forthcoming. "Economists' Perspectives of Leadership," in Nitin Nohria and Raksesh Khurana, eds., *Handbook of Leadership Theory and Practice*. Boston, MA: Harvard Business School Press.

Brandts, Jordi, and Charles A. Holt. 1992. "An Experimental Test of Equilibrium Dominance in Signaling Games," *American Economic Review*, 82:(5) 1350–1365.

Brandts, Jordi, and Charles A. Holt. 1993. "Adjustment Patterns and Equilibrium Selection in Experimental Signaling Games," *International Journal of Game Theory*, 22:(3) 279–302.

Brandts, Jordi, Tatsuyoshi Saijo, and Arthur Schram. 2004. "How Universal is Behavior? A Four Country Comparison of Spite and Cooperation in Voluntary Contribution Mechanisms," *Public Choice*, 119:(3–4) 381–424.

Cason, Timothy N., Tatsuyoshi Saijo, and Takehiko Yamato. 2002. "Voluntary Participation and Spite in Public Good Provision Experiments: An International Comparison," *Experimental Economics*, 5:(2) 133–153.

Cason, Timothy N., Tatsuyoshi Saijo, Takehiko Yamato, and Konomu Yokotani. 2004. "Non-excludable Public Good Experiments," *Games and Economic Behavior*, 49:(1) 81–102.

Cho, In-Koo, and David M. Kreps. 1987. "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102(2): 179–221.

Costa-Gomes, Miguel A., and Vincent P. Crawford. 2006. "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *American Economic Review*, 96(5): 1737–1768.

Costa-Gomes, Miguel A., Vincent P. Crawford, and Bruno Broseta. 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, 69(5): 1193–1235.

El-Gamal, Mahmoud A., and David M. Grether. 1995. "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*, 90(432): 1137–1145.

Fehr, Ernst, Karla Hoff, and Mayuresh Kshetramade. 2008. "Spite and Development," *American Economic Review*, 98(2): 494–499.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment," *Economics Letters*, 71(3): 397–404.

Harless, David W., and Colin F. Camerer. 1994. "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica*, 62(6): 1251–1289.

Harless, David W., and Colin F. Camerer. 1995. "An Error Rate Analysis of Experimental Data Testing Nash Refinements," *European Economic Review*, 39(3–4): 649–660.

Hermalin, Benjamin E. 1998. "Toward an Economic Theory of Leadership: Leading by Example," *American Economic Review*, 88(5): 1188–1206.

Hermalin, Benjamin E. Forthcoming. "Leadership and Corporate Culture," in Robert Gibbons and John Roberts, eds., *Handbook of Organizational Economics*. Princeton, NJ: Princeton University Press.

Herrmann, Benedikt, and Christian Thöni. 2009. "Measuring Conditional Cooperation: A Replication Study in Russia," *Experimental Economics*, 12(1): 87–92.

Hillas, John. 1994. "Sequential Equilibria and Stable Sets of Beliefs," *Journal of Economic Theory*, 64(1): 78–102.

Huck, Steffen, and Pedro Rey-Biel. 2006. "Endogenous Leadership in Teams," *Journal of Institutional and Theoretical Economics*, 162(2): 253–261.

Kobayashi, Hajime, and Hideo Suehiro. 2005. "Emergence of Leadership in Teams," *Japanese Economic Review*, 56(3): 295–316.

Kobayashi, Hajime, and Hideo Suehiro. 2008. "Leadership by Confidence in Teams," Graduate School of Business Administration, Kobe University, Discussion Paper Series: No. 2008-35.

Koçkesen, Levent, and Efe A. Ok. 2004. "Strategic Delegation by Unobservable Incentive Contracts," *Review of Economic Studies*, 71(2): 397–424.

Komai, Mana, Philip Grossman, and Travis Deters. 2007. "Leadership and Information: An Experimental Study," Department of Economics, St. Cloud State University, Working Papers: No. 2006-04.

Komai, Mana, Mark Stegeman, and Benjamin E. Hermalin. 2007. "Leadership and Information," *American Economic Review*, 97(3): 944–947.

Kreps, David M. 1989. "Out-of-Equilibrium Beliefs and Out-of-Equilibrium Behavior," in Frank H. Hahn, ed., *The Economics of Missing Markets, Information, and Games*. Oxford: Claredon Press.

Kreps, David M. 1990. *A Course in Microeconomic Theory.* Princeton, NJ: Princeton University Press.

Kreps, David M., and Robert Wilson. 1982. "Sequential Equilibria," *Econometrica*, 50(4): 863–894.

Potters, Jan, Martin Sefton, and Lise Vesterlund. 2005. "After You – Endogenous Sequencing in Voluntary Contribution Games," *Journal of Public Economics*, 89(8): 1399–1419.

Potters, Jan, Martin Sefton, and Lise Vesterlund. 2007. "Leading-by-Example and Signaling in Voluntary Contribution Games: An Experimental Study," *Economic Theory*, 33(1): 169–182.

Rapoport, Amnon, and Wilfred Amaldoss. 2008. "In Search of Experimental Support for an Asymmetric Equilibria Solution in Symmetric Investment Games," *Journal of Economic Behavior and Organizations*, 67(1): 200–203.

Rivas, M. Fernanda, and Matthias Sutter. 2009. "Leadership in Public Goods Experiments – On the Role of Reward, Punishment and Endogenous Leadership," Unpublished Paper.

Saijo, Tatsuyoshi. 2008. "Spiteful Behavior in Voluntary Contribution Mechanism Experiments," in Charles R. Plott and Vernon L. Smith, eds., *Handbook of Experimental Economics Results*. Vol. 1. Amsterdam: Elsevier/North-Holland.

Saijo, Tatsuyoshi, and Hideki Nakamura. 1995. "The 'Spite' Dilemma in Voluntary Contribution Mechanism Experiments," *Journal of Conflict Resolution*, 39(3): 535–560.

Suehiro, Hideo. 1992a. "A 'Mistaken Theories' Refinement," Tilburg University, CentER Discussion Paper Series: No. 9224.

Suehiro, Hideo. 1992b. "Robust Selection of Equilibria," Tilburg University, CentER Discussion Paper Series: No. 9225.

Vesterlund, Lise. 2003. "The Informational Value of Sequential Fundraising ," *Journal of Public Economics*, 87(3–4): 627–657.