# Confidence, Beliefs about Confidence, and Leadership in Teams*

Koji Abe[†] *Yokohama National University*
Hajime Kobayashi[‡] *Kansai University*
Hideo Suehiro[§] *Kobe University*

February 13, 2015

**Abstract**

We propose an economics of endogenous leadership by studying a two-agent team production game with endogenous timing. Each agent of the team must commit a level of effort in one of two periods that the agent prefers. At the end of each period, each agent observes the partner's move in this period. Both agents are rewarded by a team output determined by team productivity and total invested effort, while each agent must personally incur the cost of own effort.

We consider a mechanism of endogenous leadership driven by confidence. We assume that each agent is privately endowed with some level of confidence about team productivity by an information system that supplies agents with independent confidence, perfectly correlated confidence, or imperfectly correlated confidence as a mixture of the first two according to a parameter of the system. We then consider the following type of endogenous leadership. An agent with more confidence takes an initiative voluntarily and an agent with less confidence chooses to wait and responds to the leader's behavior with corresponding effort. We show that a degree of the correlation of agents' confidence determines whether this mechanism of endogenous leadership is realized in equilibrium; there exists a threshold of the parameter of the information system, and leadership is realized in equilibrium if and only if a degree of the correlation of agents' confidence is below the threshold.

JEL Classification: C72, D82
Keywords: Leadership, Team production, Endogenous Timing, Signaling

*Acknowledgement: To Be Added.
[†]Faculty of International Social Sciences, Yokohama National University, 79-4 Tokiwadai, Hodogaya, Yokohama 240-8501, Japan; abe@ynu.ac.jp.
[‡]Faculty of Economics, Kansai University, 3-3-35, Yamate, Suita, Osaka, 564-8680, Japan; khajime@kansai-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

Leadership is a dynamic interpersonal process that affects the performance of organizations. The seminal work of Hermalin (1998) shows that in spite of its complex appearance which seems to obstruct any theorization, leadership can be investigated by the economics approach. He considers a specific team environment for the possibility of leadership in which an agent of the team is assigned the leader role and the other agents are assigned the follower role by the authority. This environment is modeled as a team production game with exogenous sequence of moves in which the leader agent is requested to choose a level of effort first, the follower agents observe the leader's choice, and then they choose their levels of effort. He shows that if the team productivity is a random variable and the leader agent observes the realized value of the productivity while the follower agents do not, the leader's choice influences the followers' choices, that is, leadership is realized, because the leader's choice signals his private information about the realized value of the productivity so that the follower agents adjust their levels of effort according to their beliefs updated depending on the leader's choice.

Leadership is realized in diverse environments other than the one considered by Hermalin (1998). In particular, leadership often emerges without authority. Even if there is no directive from the authority about when to move and what to do for the team, a leadership process is triggered when an agent in the team takes an initiative voluntarily and the leadership is realized when this initiative influences the other agents to behave in a manner that benefits the team. We call this leadership *endogenous leadership*.

We propose an economics of endogenous leadership. We consider a two-agent team production game with endogenous timing. Each agent of the team chooses a level of effort in one of two periods that the agent prefers. At the end of each period, each agent observes the partner's move in this period. Both agents are rewarded by a team output determined by team productivity and total invested effort. Each agent must personally incur the cost of own effort. Each agent decides for themself both when to invest effort and what level of effort to invest.

We present a mechanism of endogenous leadership in this team in which leadership is driven by confidence about the team productivity. Each agent is endowed with imperfect private information about the team productivity. An agent who is confident that the team productivity is high takes the initiative, this initiative signals the agent's confidence to an agent who is not so confident and chooses to wait to see the other agent's behavior, and the less confident agent responds to the initiative with a level of effort corresponding to the agent's belief about the team productivity updated by the signal. We call this mechanism *endogenous signaling*, because the leader influences the follower through signaling by the initiative, and both who sends a signal and who receives the signal are determined endogenously by agents' choices of their timing for moves.

Our main result states that the mechanism of endogenous signaling may or may not function properly, depending on the nature of the information system by which each agent forms his confidence about the team productivity. Basically, the endogenous signaling is realized if the correlation in agents' confidence is limited so that each agent places at least a certain probability on the possibility that the other agent holds confidence differently from the agent, that is, an agent who is confident believes with some probability that the other agent is less confident, and an agent who is less confident believes with some probability that the other agent is confident.

In our attempt toward the economics of endogenous leadership, we address two theoretical issues. The first theoretical issue is to replicate a whole process of leadership in the endogenous timing game. The second theoretical issue is to examine how the incomplete information about common value affects agents in choosing both when to move and what to do in the endogenous timing game.

We consider the first issue. We need to explain simultaneously a pair of questions involved in endogenous leadership: initiative and influence. The question of initiative is why a member takes an initiative to work for the team in spite of no one forcing or ordering him to take the initiative or there being no direct incentive system for the member to do it. The question of influence is why a follower responds to an initiative by a leader with corresponding effort to work for the team in spite of the leader having no formal instrument to enforce the follower to take the desired behaviors. The question of influence arises both in endogenous leadership and in the environment of Hermalin (1998), while the question of initiative arises only in endogenous leadership. Hence, we need a theory of the mechanism by which agents are sorted by their own will to move at different points of time in the team production under the freedom to choose when to move.

By the types of sorting mechanisms, the existing studies of endogenous timing game are broadly classified into two groups. The first group mainly adopts private information and time preference as driving forces. This type of models explains sorting by the degree to which players face trade-off between the cost of waiting and the benefit of learning the rival's private information by waiting (or signaling the player's own private information by waiting). These models has been applied to investment decision (Hendricks and Kovenock (1989); Chamley and Gale (1994); Gul and Lundholm (1995)), bargaining (Kambe (1999); Abreu and Gul (2000)), war of attrition (Fudenberg and Tirole (1986)), English auction (Milgrom and Weber (1982)).

The second group explains sorting of timing of moves by the advantage or disadvantage of commitment determined by the strategic relations such as strategic complementarity or substitution. This type of models has been mainly applied to oligopolistic situation such as Cournot competition and Bertrand competition (Gal-Or (1987); Hamilton and Slutsky (1990); Mailath (1993); Normann (1997); van Damme and Hurkens (1999); Normann (2002); van Damme and Hurkens (2004); Amir and Stepanova (2006)).

Endogenous leadership in teams cannot be fully explained by those type of combination of the forces. Time preference is not necessarily present in team production. Rather, team output and contribution cost do not usually depend on timing of investing efforts. Then, team production problem is free from time discounting. Furthermore, as suggested by Holmström (1982), although there are possibly many types of team production function, the natural specification of the function is the product of total efforts and productivity parameter. This function is the prototype of team production function as appeared in many applications (e.g. Hermalin (1998)). Observe that the function does not imply neither strategic complementarity nor substitution. Therefore, the above two existing groups are not sufficient to explain endogenous leadership in teams.

We adopt a new combination of driving forces in order to explain endogenous leadership. The driving forces are *multi-sided private information* and *simple payoff externality*. Multi-sided private information means that every player has private information about team productivity. Simple payoff externality is that a player's payoff depends on the action choices of other players but a player's optimal action does not depend on the action choices by other players. The production function such as the product of total efforts and productivity parameter implies simple payoff externality.

The combination of multi-sided private information and simple payoff externality could make leadership emerge in teams as follows. Every agent holds a personal expectation about team productivity based on each agent's private information. The expectation pushes an agent to voluntarily move prior to others when the agent is more confident that the team productivity is high, and it induces an agent to wait for action at later timing when the agent is less confident. When an agent becomes a leader or a follower according to their private information, signaling from the leader to the follower is realized. Then, the sender obtains a benefit due to simple payoff externality from increased effort investment by the receiver. The receiver benefits from learning the sender's private information. When these matters are expected in a self-fulfilling manner, it sorts some agents into a leader and some other into a follower endogenously, and the leader's initiative influences the follower's choice upward. This is the mechanism of endogenous signaling.

We consider the second theoretical issue. We examine how the incomplete information about common value affects agents in choosing both when to move and what to do in the endogenous timing game. The multi-sided private information that we introduce to the team production is the information about the common value of the team productivity. In order to verify that the above-mentioned combination of multi-sided private information and simple payoff externality makes leadership emerge in teams, we need to investigate the effect of the incomplete information about the team productivity over the agents' behaviors.

Some of the existing studies of endogenous timing games consider the incomplete information about the common value. They focus on a particular role that the incomplete information about the common value plays in endogenous timing games, that is, informational externality. Agents chooses their timing of moves to take advantage of the other agents' private information about the common value revealed through their actions in an environment in which an agent's payoff depends on the own decision and the common value, but not on the actions of others. For example, in the above-mentioned studies of strategic investment in the endogenous timing games, Chamley and Gale (1994) and Gul and Lundholm (1995) consider a case in which the value of investment is common to all the agents, each agent is endowed with partial information about the common value, and they can decide to invest or not after observing the other agents' investment decisions by delaying the own decisions. The equilibrium in these endogenous timing games depends on the joint distribution of the common value and the agents' type profile (a list of each agent's private information). However, what matters for the equilibrium is the conditional expectation of the common value given a profile of agents' types. The marginal distribution of a type profile is irrelevant. This is even more transparent in a more complex situation of English auction with interdependent values (which includes the case of common value as an extreme), in which agent's payoff also depends on the other agent's actions. Milgrom and Weber (1982) shows that the symmetric equilibrium of an agent's optimal stopping price given a history of the other agent's bidding behaviors is determined solely by the conditional expected value of the object given the bid history. The equilibrium is the same for any marginal distribution of a type profile as long as this conditional expected value is the same.

The incomplete information about the common value may be associated with marginal distributions of a type profile with different natures. In the investment models investigated by Chamley and Gale (1994) and Gul and Lundholm (1995), the common value is a strictly increasing function of the sum of agent's received signals and each signal is independently distributed across agents. The value of investment is common to all the agent while each agent's belief about the other agents' belief about the common value is constant irrespective of the own information. In contrast, in many English auctions with common values, agents' private information is correlated so that when an agent receives private information that indicates a higher common value, the agent believes that the other agents are likely to receive private information that also indicates higher common values.

In a team production with multi-sided private information about team productivity, an agent's payoff depends on the conditional expectation of the common value given a profile of agents' types as in the previous studies. The conditional expectation of the team productivity is higher if the agent is confident given the private information. The conditional expectation is also higher if the agent learns that the other agents are confident. A profile of agents' confidence is affiliated. This should be the driving force of influence of the leader over the follower.

Furthermore, an agent's belief about the other agent's confidence should be crucial for the agent to make up own mind to take the initiative. If the agent believes that the other agent is as confident as him, there is little point for the agent to signal the confidence to the other agent through the initiative. Hence, a new issue that was absent in the previous studies arises in our study of endogenous leadership: how a marginal distribution of a type profile affects the possibility of endogenous leadership and the agents' behaviors in the endogenous leadership.

To study the issue of an agent's belief about the other agent's confidence, we adopt a model of agent's private information about the team productivity that implies affiliation in a profile of agents' confidence and also accommodates a variety of marginal distribution of a type profile. For a fixed state space of team productivities, the model includes a parameter $\alpha$ in an interval $[0, 1]$ that determines a degree of correlation in types in the marginal distribution of a type profile. At an extreme value $\alpha = 1$ for the parameter, the model means that agents' types are independently distributed as in the models of Chamley and Gale (1994) and Gul and Lundholm (1995). At the other extreme value $\alpha = 0$ for the parameter, the model means that types are perfectly correlated. For a value of the parameter in between, agent's confidence is imperfectly correlated, and the lower the value $\alpha$, the more correlated agent's confidence is. On the other hand, the conditional expectation of the team productivity given a profile of agents' private information is kept constant over the entire space $[0, 1]$ of the parameter $\alpha$ while reflecting affiliation of agents' confidence. This enables us to separate the issue of an agent's belief about the other agent's confidence from the issue of affiliation of agents' confidence. We investigate how the equilibrium of endogenous leadership changes according to a degree of correlation in types described by the parameter $\alpha$.

We show that there exists a threshold $\alpha^*$ for the parameter that describes a degree of correlation in types, and there exists an equilibrium with endogenous leadership if and only if $\alpha \geq \alpha^*$, that is, the degree of correlation in agents' confidence is smaller than the threshold information system with the parameter value $\alpha^*$. The threshold $\alpha^*$ is $\alpha^* < 1$. Therefore, including the case of independently distributed types ($\alpha = 1$), endogenous leadership is realized in teams with information systems in a non-degenerate class. The threshold $\alpha^*$ may be $\alpha^* = 0$. This happens if the variation in possible values of the team productivity is large. Then, endogenous leadership is realized in teams irrespective of the nature of their information systems.

This result supports our practical observations about endogenous leadership. Leadership may or may not be realized endogenously. The result identifies in what kind of teams we observe endogenous leadership. No one would like to take the initiative voluntarily in those teams in which each agent believes with high probabilities that the other agents believe about the team productivity in the same way as theirs and choose exactly the same level of effort without the initiative ($\alpha < \alpha^*$). On the other hand, an agent with confidence about the team productivity would like to take the initiative if the agent believes with some probability that the other agents assess the team productivity differently from the agent and they choose lower levels of effort unless the agent takes the initiative ($\alpha \geq \alpha^*$). This is so even if the agent believes with high probabilities that the other agents confidence is similar and the agent believes that there is only a slight chance that the other agents are less confident ($\alpha^* = 0$) as long as each agent needs to adjust a level of effort to a large degree, because such agents with less confidence become followers and respond to the initiative with correspondingly high levels of effort, which would not be possible otherwise.

The rest of this paper is organized as follows. Section 2 presents a model of team production and a model of information system by which each agent is endowed with confidence about the team productivity. Section 3 defines the equilibrium strategy that we focus on for our study of endogenous leadership in the team described in Section 2. Section 4 presents our main result and prove it. Section 5 concludes with remarks.

## 2 The model

We study the team production game with two agents $i = 1, 2$. Each agent $i$ chooses a level of effort $e_i$ from the nonnegative real numbers with cost $c(e_i) = e_i^2/2$. The team performance is determined by $\theta \times (e_i + e_j)$, where $\theta$ is a random team productivity parameter. We assume that each agent $i$ is rewarded by the team performance itself. Hence, the net payoff to agent $i$ is given by:

$$U_i = \theta \times (e_i + e_j) - \frac{e_i^2}{2}.$$

The two notable features of the model are information structure and move structure, which we introduce in order.

*Information Structure*

We simply adopt a binary model of information. Team productivity $\theta$ takes a high value $\theta_H$ with probability $\rho$ and a low value $\theta_L < \theta_H$ with probability $1 - \rho$, where $\rho \in (0, 1)$ and $\theta_L, \theta_H \in (0, \infty)$ are arbitrarily fixed parameters. Agents do not directly observe the realized value of $\theta$. Instead, each agent $i$ receives a signal $s_i$ about $\theta$ before production. For $i = 1, 2$, signal $s_i$ takes a high value $s_H$ or a low value $s_L$ according to a conditional probability system $f$: given the realized value of $\theta$, random variables $(s_1, s_2)$ are jointly distributed according to $f(s_1, s_2|\theta)$.

In this paper, we specify $f$ as a mixture of two simple probability systems $f_I$ and $f_C$. Table 1 below describes probability system $f_I$. Using the prior probability of $\theta$, we can easily verify that random variables $(s_1, s_2)$ are independent from the ex-ante point of view, that is, by integrating out $\theta$, we have $f_I(s_1, s_2) = f_I(s_1)f_I(s_2)$.[1] Table 2 below describes probability system $f_C$. Obviously, this system informs both agents of the realized value of team productivity, that is, each agent obtains a perfectly correlated perfect information signal. We specify probability system $f$ by $f = \alpha f_I + (1 - \alpha)f_C$ using a parameter $\alpha \in [0, 1]$. This provides a tractable and the simplest parameterized family of information systems ranging from a situation ($\alpha = 1$) in which the agents obtain signals independently across them to a situation ($\alpha = 0$) in which

[1] To economize on notation, we let $f_I$ represent its unconditional form as well.

the agents obtain perfectly correlated perfect information signals.[2] That is to say, we can analyze how the intensity of correlation between agents' signals (or types) affects the consequences.[3]

$f_I(s_1, s_2|\theta_H)$

| $s_1 \setminus s_2$ | $s_H$ | $s_L$ |
|---|---|---|
| $s_H$ | $\rho$ | $\frac{1-\rho}{2}$ |
| $s_L$ | $\frac{1-\rho}{2}$ | 0 |

$f_I(s_1, s_2|\theta_L)$

| $s_1 \setminus s_2$ | $s_H$ | $s_L$ |
|---|---|---|
| $s_H$ | 0 | $\frac{\rho}{2}$ |
| $s_L$ | $\frac{\rho}{2}$ | $1-\rho$ |

Table 1: Information structure $f_I$

$f_C(s_1, s_2|\theta_H)$

| $s_1 \setminus s_2$ | $s_H$ | $s_L$ |
|---|---|---|
| $s_H$ | 1 | 0 |
| $s_L$ | 0 | 0 |

$f_C(s_1, s_2|\theta_L)$

| $s_1 \setminus s_2$ | $s_H$ | $s_L$ |
|---|---|---|
| $s_H$ | 0 | 0 |
| $s_L$ | 0 | 1 |

Table 2: Information structure $f_C$

Abusing notations, we use $f$ to represent the conditional probability of one agent's signals given another agent's signals. Operator $\mathbb{E}$ is the expectation operator as usual. Then, the adopted information structure induces the following.[4]

- $f(s_L|s_H) = \alpha(1-\rho)$
- $f(s_H|s_H) = 1 - \alpha(1-\rho)$
- $f(s_L|s_L) = 1 - \alpha\rho$
- $f(s_H|s_L) = \alpha\rho$
- $\mathbb{E}[\theta|s_L, s_L] = \theta_L$
- $\mathbb{E}[\theta|s_L, s_H] = \mathbb{E}[\theta|s_H, s_L] = \frac{\theta_H+\theta_L}{2}$
- $\mathbb{E}[\theta|s_H, s_H] = \theta_H$

[2] Existing studies of endogenous leadership adopt $\alpha = 1$ if their models were written with the setup of this paper. See Kobayashi and Suehiro (2005, 2008); Abe et al. (2015).

[3] As an alternative family of information structures, we could consider conditional independence information structures. A conditional independence information structure describes a situation in which given the realized value of $\theta$, each agent obtain a noisy signal independently. Let $g(s|\theta)$ be such that $g(s_H|\theta_H) > g(s_H|\theta_L)$. Define conditional probability system $g(s_1, s_2|\theta)$ describing a information structure by $g(s_1, s_2|\theta) = g(s_1|\theta)g(s_2|\theta)$. By parameterizing it using parameters $g(s_H|\theta_H)$ and $g(s_H|\theta_L)$, we obtain a family of conditional independence information structures. Although adopting this family is indeed of interest, it is less tractable than the one adopted in this paper. Moreover, from the viewpoint of this paper, it is a drawback that the family cannot include an information structure satisfying the unconditional type independence: $g(s_1, s_2) = g(s_1)g(s_2)$ if and only if $g(s_H|\theta_H) = g(s_H|\theta_L)$, meaning any signal is uninformative.

[4] Note that the adopted information structure implies $\mathbb{E}[\theta|s_L, s_H] = \mathbb{E}[\theta|s_H, s_L] = \frac{\theta_H+\theta_L}{2}$. We could relax this limitation by selecting another probability system $h_I$ that satisfies $h_I(s_1, s_2) = h_I(s_1)h_I(s_2)$, instead of $f_I$. However, we adopt $f_I$ in this paper to reap the great benefit from its tractability.

- $\mathbb{E}[\theta|s_L] = \theta_L + \alpha\rho\frac{\theta_H - \theta_L}{2}$
- $\mathbb{E}[\theta|s_H] = \theta_H - \alpha(1-\rho)\frac{\theta_H - \theta_L}{2}$

*Move Structure*

Agents must choose levels of their efforts 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 an agent takes a level of effort in period 1, then the agent cannot do anything in period 2. On the other hand, if an agent chooses to do nothing in period 1, then the agent must take an effort level in period 2.

In this sequence of moves, the two agents must 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 a choice in period 2 if that agent has chosen to do nothing in period 1. If both agents have chosen to do nothing in period 1, both must invest some level of effort independently and simultaneously in period 2.

We describe the strategy in the team production game.

*Strategy*

Agent $i$'s strategy $\sigma_i$ is a profile $(\sigma^1_{i,s_i}, \sigma^2_{i,s_i})_{s_i=s_H,s_L}$ of Bayesian strategies. The part $\sigma^1_{i,s_i}$ prescribes agent $i$'s behavior in period 1 for $s_i$-type and it takes a value $\sigma^1_{i,s_i} = a^1_i \in \mathbb{R}_+ \cup \{\emptyset\}$. The part $\sigma^2_{i,s_i}$ prescribes agent $i$'s behavior in period 2 for $s_i$-type and it assigns a value $\sigma^2_{i,s_i}(a^1_j) = a^2_i \in \mathbb{R}_+$ for each possible value $a^1_j$ of agent $j$'s choice from $\mathbb{R}_+ \cup \{\emptyset\}$ in period 1.

# 3 Leadership by Confidence

We pay special attention on a particular type of symmetric strategy profile, which we call leadership by confidence.

**Definition 1.** A strategy profile $(\sigma_1, \sigma_2)$ is said to be a *leadership by confidence* if $\sigma^1_{1,s_H} = \sigma^1_{2,s_H} \in \mathbb{R}_+$ and $\sigma^1_{1,s_L} = \sigma^1_{2,s_L} = \emptyset$ for $i = 1, 2$.

Note that $\mathbb{E}[\theta|s_H] - \mathbb{E}[\theta|s_L] = (1-\alpha/2)(\theta_H - \theta_L) > 0$. Hence, a leadership by confidence means that each agent moves first if and only if that agent is more confident about team productivity.

We say that a strategy profile supports emergence of leadership if a play according to the strategy profile entails a leader–follower (i.e. a first- and second-mover) relationship with a positive probability and the follower's choice depends on the leader's choice. A leadership by confidence is such a candidate because it satisfies the first half requirement of the emergence of leadership by definition: an agent with a higher confidence of team productivity leads, and one with a lower confidence follows. If this happens in equilibrium, then a leadership by confidence also satisfies the second half requirement of the emergence of leadership: the emergence of leader transmits a leader's good signal to the follower, it makes the follower's confidence upward revision, and it changes the follower's effort level. Therefore, if a leadership by confidence is an equilibrium, then it supports the emergence of leadership.

A leadership by confidence describes a particular type of leadership based on endogenous signaling behaviors. That is, 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 their information. Among strategy profiles with the endogenous signaling structure, we consider leadership by confidence as a natural behavioral pattern and pay special attention on it.[5]

We characterize in the next section a necessary and sufficient condition for a leadership by confidence to be a sequential equilibrium using intensity parameter $\alpha$ of correlation between agents' types. Since agents' types correspond to their confidence of team productivity, this means that our equilibrium condition is written in terms of intensity of correlation between agents' confidence.

In what follows, we will introduce the notion of the first-move incentive. We will study this notion in order to examine the existence of leadership by confidence equilibrium in the next section.

Suppose a leadership by confidence prevails. Since it is a symmetric strategy profile, we abbreviate agent index $i$ and denote the strategy by $\sigma = (\sigma^1, \sigma^2)$, where $\sigma^1_{s_H} = e$ for some $e \in \mathbb{R}_+$ and $\sigma^1_{s_L} = \emptyset$. We identify $\sigma^2$ from the sequential equilibrium requirement. Consider an information set of period 2 in which an agent with $s_H$-signal observes the other agent's $\sigma^1_{s_H} = e$. At this information set, the agent updates the confidence of team productivity to $\mathbb{E}[\theta|s_H, s_H]$. As a result, the agent's sequentially rational choice identifies $\sigma^2_{s_H}(e) = \mathbb{E}[\theta|s_H, s_H]$ as the unique maximizer of $\mathbb{E}[\theta|s_H, s_H]\{e+e'\} - (e')^2/2$ with choice variable $e'$. We likewise identify $\sigma^2_{s_H}(\emptyset) = \mathbb{E}[\theta|s_H, s_L]$, $\sigma^2_{s_L}(e) = \mathbb{E}[\theta|s_L, s_H]$, and $\sigma^2_{s_L}(\emptyset) = \mathbb{E}[\theta|s_L, s_L]$.

We derive the first-move incentive for each type agent from the implied incentive compatibility constraint. The constraint that an agent with $s_H$-signal does not mimic an agent with $s_L$-signal is as follows.

$$\begin{aligned}
&f(s_L|s_H)\mathbb{E}[\theta|s_H, s_L]\big\{e + \mathbb{E}[\theta|s_L, s_H]\big\} + f(s_H|s_H)\mathbb{E}[\theta|s_H, s_H]\big\{e + e\big\} - \frac{e^2}{2} \\
&\geq f(s_L|s_H)\left[\mathbb{E}[\theta|s_H, s_L]\big\{\mathbb{E}[\theta|s_H, s_L] + \mathbb{E}[\theta|s_L, s_L]\big\} - \frac{\mathbb{E}[\theta|s_H, s_L]^2}{2}\right] \\
&\quad + f(s_H|s_H)\left[\mathbb{E}[\theta|s_H, s_H]\big\{\mathbb{E}[\theta|s_H, s_H] + e\big\} - \frac{\mathbb{E}[\theta|s_H, s_H]^2}{2}\right].
\end{aligned}$$

The left hand side means the expected payoff obtained by choosing $e$ at period 1, and the right hand side means the expected payoff obtained by choosing $\emptyset$ at period 1 and taking $\sigma^2$ at period 2.

Observe that both the expected payoffs in the incentive compatibility constraint directly depend on the belief about the other agent's signal given own signal, that is, $f(s_L|s_H)$. Since a signal is directly related to the own confidence about team productivity, the dependence of $f(s_L|s_H)$ on the constraint means that the belief about the other agent's confidence plays a crucial role in leadership by confidence. This implies that the intensity $\alpha$ of correlation between agents' confidence strongly influences the incentive for the leadership behavior.

Rearranging and simplifying that constraint, we have:

$$\begin{aligned}
&f(s_L|s_H)\mathbb{E}[\theta|s_H, s_L]\big\{\mathbb{E}[\theta|s_L, s_H] - \mathbb{E}[\theta|s_L, s_L]\big\} \\
&\geq f(s_L|s_H)\frac{\mathbb{E}[\theta|s_H, s_L]^2}{2} + f(s_H|s_H)\frac{\mathbb{E}[\theta|s_H, s_H]^2}{2} - \left\{\mathbb{E}[\theta|s_H]e - \frac{e^2}{2}\right\}.
\end{aligned}$$

[5] Kobayashi and Suehiro (2005) study several kinds of strategy profiles with the endogenous signaling structure including leadership by confidence under type independence. Kobayashi and Suehiro (2008) further study leadership by confidence under type independence. Abe et al. (2015) experimentally study endogenous leadership under type independence and indeed report that the observed outcomes are consistent to a leadership by confidence.

The left hand side measures the value of signaling for an agent with $s_H$-signal because this measures the expected influence of the leader behavior on the other agent's period 2 choice. The right hand side measures the value of learning for an agent with $s_H$-signal because this measures the expected net benefit that the agent enjoy by knowing the other agent's signal and by making adjustments to period 2 choice. Note that the value of learning depends on $e$. When we emphasize the dependence, we use the value of adjustment from $e$ as an expression of the value of learning.

We define the first-move incentive with leader choice $e$ for an agent with $s_H$-signal by the difference between the value of signaling and the value of adjustment from $e$ for the agent. Since this depends on $\alpha$ through $f$ and $\mathbb{E}[\theta|s_H]$, denote it by $\Delta_H(e,\alpha)$. Let $\Delta_L(e,\alpha)$ be the corresponding one for an agent with $s_L$-signal. That is, they are as follows.

$$\begin{aligned}\Delta_H(e,\alpha) &= f(s_L|s_H)\mathbb{E}[\theta|s_H,s_L]\{\mathbb{E}[\theta|s_L,s_H]-\mathbb{E}[\theta|s_L,s_L]\}\\&\quad - f(s_L|s_H)\frac{\mathbb{E}[\theta|s_H,s_L]^2}{2} - f(s_H|s_H)\frac{\mathbb{E}[\theta|s_H,s_H]^2}{2} + \left\{\mathbb{E}[\theta|s_H]e-\frac{e^2}{2}\right\},\\ \Delta_L(e,\alpha) &= f(s_L|s_L)\mathbb{E}[\theta|s_L,s_L]\{\mathbb{E}[\theta|s_L,s_H]-\mathbb{E}[\theta|s_L,s_L]\}\\&\quad - f(s_L|s_L)\frac{\mathbb{E}[\theta|s_L,s_L]^2}{2} - f(s_H|s_L)\frac{\mathbb{E}[\theta|s_L,s_H]^2}{2} + \left\{\mathbb{E}[\theta|s_L]e-\frac{e^2}{2}\right\}.\end{aligned}$$

Our analysis in the next section is based on the next lemma. This states that it is determined by no-mimic conditions on the path of play whether a leadership by confidence constitutes a sequential equilibrium (Kreps and Wilson (1982)).

**Lemma 1.** Let $\sigma = (\sigma^1, \sigma^2)$ be a strategy such that $\sigma^1_{s_H} = e$ for some $e \in \mathbb{R}_+$ and $\sigma^1_{s_L} = \emptyset$. Leadership by confidence $(\sigma,\sigma)$ constitutes a sequential equilibrium (by specifying appropriate beliefs and behavior off the path of play explicitly) if and only if $\sigma^2_{s_H}(e) = \mathbb{E}[\theta|s_H,s_H]$, $\sigma^2_{s_H}(\emptyset) = \mathbb{E}[\theta|s_H,s_L]$, $\sigma^2_{s_L}(e) = \mathbb{E}[\theta|s_L,s_H]$, $\sigma^2_{s_L}(\emptyset) = \mathbb{E}[\theta|s_L,s_L]$, and $\Delta_H(e,\alpha) \geq 0 \geq \Delta_L(e,\alpha)$.

*Proof.* The "only if" part is obvious.

We consider the "if" part. Consider a situation in which the other agent plays $a^1$ at period 1. Denote by $\mu(a^1)$ the symmetric belief that the other agent has $s_H$-signal. Assume $\mu(a^1) = 1$ if $a^1 = e$ and $\mu(a^1) = 0$ otherwise. Then, this is a consistent belief for $\sigma$, and every prescribed 2nd period choice is sequentially rational under this belief.

Since condition $\Delta_H(e,\alpha) \geq 0 \geq \Delta_L(e,\alpha)$ assures that agents have no incentive to mimic the period 1 behavior of the agent with a signal different from theirs, we consider the other kinds of deviation at period 1.

We show that any agent with $s_H$-signal has no incentive to choose $e' \neq e$ in period 1. Given $\mu$ above and the sequentially rational second period choices, we have:

$$\begin{aligned}&f(s_L|s_H)\mathbb{E}[\theta|s_H,s_L]\{e+\mathbb{E}[\theta|s_L,s_H]\} + f(s_H|s_H)\mathbb{E}[\theta|s_H,s_H]\{e+e\} - \frac{e^2}{2}\\&\geq f(s_L|s_H)\left[\mathbb{E}[\theta|s_H,s_L]\{\mathbb{E}[\theta|s_H,s_L]+\mathbb{E}[\theta|s_L,s_L]\} - \frac{\mathbb{E}[\theta|s_H,s_L]^2}{2}\right]\\&\quad + f(s_H|s_H)\left[\mathbb{E}[\theta|s_H,s_H]\{\mathbb{E}[\theta|s_H,s_H]+e\} - \frac{\mathbb{E}[\theta|s_H,s_H]^2}{2}\right]\\&\geq f(s_L|s_H)\left[\mathbb{E}[\theta|s_H,s_L]\{e'+\mathbb{E}[\theta|s_L,s_L]\} - \frac{(e')^2}{2}\right]\end{aligned}$$

$$+ f(s_H|s_H)\left[\mathbb{E}[\theta|s_H,s_H]\{e'+e\} - \frac{(e')^2}{2}\right]$$

$$= f(s_L|s_H)\mathbb{E}[\theta|s_H,s_L]\{e' + \mathbb{E}[\theta|s_L,s_L]\} + f(s_H|s_H)\mathbb{E}[\theta|s_H,s_H]\{e'+e\} - \frac{(e')^2}{2},$$

where the first expression corresponds to the expected payoff of an agent with $s_H$-signal obtained by choosing the supposed $e$ in period 1, the second expression corresponds to the one obtained by choosing $\emptyset$ in period 1 and then taking $\sigma^2$, the third expression corresponds to the one obtained by choosing $\emptyset$ in period 1 and then choosing $e'$ irrespective of the other agent's period 1 choice, and the last expression corresponds to the one obtained by choosing $e'$ in period 1. Hence, any agent with $s_H$-signal has no incentive to deviate from taking the supposed $e$ in period 1.

Similarly, we can show that any agent with $s_L$-signal has no incentive to choose $e' \neq e$ in period 1. □

# 4 Existence of leadership by confidence equilibrium

We characterize a necessary and sufficient condition for leadership by confidence to be a sequential equilibrium using intensity parameter $\alpha$ of correlation between agents' confidence. Consider a leadership by confidence strategy $\sigma = (\sigma^1, \sigma^2)$ such that $\sigma^1_{s_H} = e$ for some $e \in \mathbb{R}_+$, $\sigma^1_{s_L} = \emptyset$, $\sigma^2_{s_H}(e) = \mathbb{E}[\theta|s_H,s_H]$, $\sigma^2_{s_H}(\emptyset) = \mathbb{E}[\theta|s_H,s_L]$, $\sigma^2_{s_L}(e) = \mathbb{E}[\theta|s_L,s_H]$, and $\sigma^2_{s_L}(\emptyset) = \mathbb{E}[\theta|s_L,s_L]$. Then, Lemma 1 states that $\sigma$ constitutes a sequential equilibrium if and only if $\Delta_H(e,\alpha) \geq 0 \geq \Delta_L(e,\alpha)$. In this section, we examine this inequalities in terms of $\alpha$. Specifically, we consider the set of $\alpha$ such that $\Delta_H(e,\alpha) \geq 0 \geq \Delta_L(e,\alpha)$ for some $e$ and show below that the set is written as $[\alpha^*, 1]$ for some $\alpha^* \in [0,1)$.

We provide the basic properties of $\Delta_H$ and $\Delta_L$.

**Lemma 2.** The basic properties of $\Delta_H$ and $\Delta_L$ are listed:

(a) We can write $\Delta_H(e,\alpha)$ and $\Delta_L(e,\alpha)$ as follows.

$$\Delta_H(e,\alpha) = \Delta_H(\mathbb{E}[\theta|s_H],\alpha) - \frac{\{e - \mathbb{E}[\theta|s_H]\}^2}{2},$$

$$\Delta_L(e,\alpha) = \Delta_L(\mathbb{E}[\theta|s_L],\alpha) - \frac{\{e - \mathbb{E}[\theta|s_L]\}^2}{2}.$$

(b) $\Delta_H(\mathbb{E}[\theta|s_H],\alpha)$ is a strictly increasing function of $\alpha$ with $\Delta_H(\mathbb{E}[\theta|s_H],\alpha)|_{\alpha=0} = 0$.

*Proof.* (a) They are just rewriting and immediately follow.
(b) We can calculate $\Delta_H(\mathbb{E}[\theta|s_H],\alpha)$ as follows.

$$\begin{aligned}
&\Delta_H(\mathbb{E}[\theta|s_H],\alpha) \\
&= f(s_L|s_H)\{\mathbb{E}[\theta|s_H,s_H] - \mathbb{E}[\theta|s_H,s_L]\} \\
&\quad \times \left\{ f(s_L|s_H)\left(\frac{\mathbb{E}[\theta|s_H,s_H] - \mathbb{E}[\theta|s_H,s_L]}{2}\right) - \frac{\mathbb{E}[\theta|s_H,s_H] - 3\mathbb{E}[\theta|s_H,s_L]}{2} \right\}.
\end{aligned}$$

Since $f(s_L|s_H) = \alpha(1-\rho)$, $\Delta_H(\mathbb{E}[\theta|s_H],\alpha)$ is a convex function of $\alpha$, corresponding to a parabola opening upward. Moreover, we can easily verify that its larger root is $\alpha = 0$. This completes the proof. □

Note that if an agent with $s_H$-signal maximizes the expected payoff using just own signal, the optimal level of effort is $\mathbb{E}[\theta|s_H]$. Lemma 2-(a) states that the first-move incentive is also maximized by choosing $\mathbb{E}[\theta|s_H]$, and in general, $\Delta_H(e, \alpha)$ is computed by subtracting a cost of effort adjustment from the maximized first-move incentive. The same applies to $\Delta_L$. Lemma 2-(b) states that the maximized value of $\Delta_H(\mathbb{E}[\theta|s_H], \alpha)$ is always nonnegative, which means that an agent with $s_H$-signal basically has an incentive to lead.

From Lemma 2-(a), it immediately follows that for any $e$, we have:

$$\frac{\partial}{\partial e}\Delta_H(e,\alpha) - \frac{\partial}{\partial e}\Delta_L(e,\alpha) = -\big\{e - \mathbb{E}[\theta|s_H]\big\} + \big\{e - \mathbb{E}[\theta|s_L]\big\} = \mathbb{E}[\theta|s_H] - \mathbb{E}[\theta|s_L] > 0.$$

Hence, $\Delta_H$ and $\Delta_L$ have a strict form of single-crossing property.

**Lemma 3.** *Suppose $\Delta_H(e, \alpha) = \Delta_L(e, \alpha)$. Then, $\Delta_H(e', \alpha) > \Delta_L(e', \alpha)$ for any $e' > e$ and $\Delta_H(e'', \alpha) < \Delta_L(e'', \alpha)$ for any $e'' < e$.*

To make use of the single-crossing property of $\Delta_H$ and $\Delta_L$, we pay attention on the intersections, that is, levels of effort for which the first-move incentives of both types are the same. Define $\hat{e}(\alpha)$ implicitly by the equation $\Delta_H(\hat{e}(\alpha), \alpha) = \Delta_L(\hat{e}(\alpha), \alpha)$. Since $\Delta_H$ and $\Delta_L$ have continuous partial derivatives and satisfy $\frac{\partial}{\partial e}\{\Delta_H(e, \alpha) - \Delta_L(e, \alpha)\} \neq 0$ as demonstrated above, by a standard implicit function theorem, $\hat{e}(\alpha)$ is indeed a continuous function.[6]

We examine the existence of leadership by confidence equilibrium in terms of $\alpha$ through the analysis of the first-move incentive with $\hat{e}(\alpha)$.

**Lemma 4.**

(a) Suppose $\Delta_H(\hat{e}(\alpha), \alpha) = \Delta_L(\hat{e}(\alpha), \alpha) \geq 0$. Then, there exists $e^* \geq \mathbb{E}[\theta|s_H]$ such that $\Delta_H(e^*, \alpha) \geq 0 \geq \Delta_L(e^*, \alpha)$.

(b) Suppose $\Delta_H(\hat{e}(\alpha), \alpha) = \Delta_L(\hat{e}(\alpha), \alpha) < 0$. Then, a necessary and sufficient condition for the existence of $e^*$ such that $\Delta_H(e^*, \alpha) \geq 0 \geq \Delta_L(e^*, \alpha)$ is $\hat{e}(\alpha) \leq \mathbb{E}[\theta|s_H]$. Moreover, in the case that such $e^*$ exists, we can take $e^* \geq \mathbb{E}[\theta|s_H]$ without loss of generality.

*Proof.* First, we prove that if we have $\Delta_H(e', \alpha) \geq 0 \geq \Delta_L(e', \alpha)$ for some $e'$, then we also have $\Delta_H(e'', \alpha) \geq 0 \geq \Delta_L(e'', \alpha)$ for some $e'' \geq \mathbb{E}[\theta|s_H]$. Take $e'$ arbitrarily such that $\Delta_H(e', \alpha) \geq 0 \geq \Delta_L(e', \alpha)$. If $e' \geq \mathbb{E}[\theta|s_H]$, then we have nothing to prove. Hence, assume $e' < \mathbb{E}[\theta|s_H]$. Note from Lemma 3 that $\Delta_H(e, \alpha) > \Delta_L(e, \alpha)$ for all $e > e'$. Hence, it follows from $e' < \mathbb{E}[\theta|s_H]$ together with Lemma 2-(b) that $\Delta_H(\mathbb{E}[\theta|s_H], \alpha) > \max\{0, \Delta_L(\mathbb{E}[\theta|s_H], \alpha)\}$. In case of $\Delta_L(\mathbb{E}[\theta|s_H], \alpha) \leq 0$, set $e'' = \mathbb{E}[\theta|s_H]$. In case of $\Delta_L(\mathbb{E}[\theta|s_H], \alpha) > 0$, Lemma 2-(a) implies that $\Delta_L(e, \alpha)$ reaches zero as $e$ goes up from $\mathbb{E}[\theta|s_H]$ in advance of $\Delta_H$. Hence, we have $e''$ such that $\Delta_H(e'', \alpha) > 0 \geq \Delta_L(e'', \alpha)$.

Thanks to the above-mentioned fact, we just examine below whether $\Delta_H(e^*, \alpha) \geq 0 \geq \Delta_L(e^*, \alpha)$ holds for some $e^*$.

(a) Suppose $\Delta_H(\hat{e}(\alpha), \alpha) = \Delta_L(\hat{e}(\alpha), \alpha) \geq 0$. Note from Lemma 2-(a) that, given $\alpha$, $\Delta_H$ and $\Delta_L$ are both parabolas opening down. It hence follows from Lemma 3 that there is $e^* > \hat{e}(\alpha)$ such that $\Delta_H(e^*, \alpha) \geq 0 \geq \Delta_L(e^*, \alpha)$.

[6] Although $\hat{e}(\alpha)$ may take a negative value for some $\alpha$ under some parameters $\rho$, $\theta_H$, and $\theta_L$, it is not a problem because we use $\hat{e}(\alpha)$ just for analysis. Indeed, together with Lemma 3, a negative $\hat{e}(\alpha)$ means that $\Delta_H(e, \alpha) > \Delta_L(e, \alpha)$ for all $e \geq 0$. Then, it immediately follows from Lemma 2-(a) that there exists $e^* \geq \mathbb{E}[\theta|s_H]$ such that $\Delta_H(e^*, \alpha) \geq 0 \geq \Delta_L(e^*, \alpha)$. In Lemma 4 below, we provide a general result about the existence of such $e^*$ independently of the sign of $\hat{e}(\alpha)$.

(b) Suppose $\Delta_H(\hat{e}(\alpha), \alpha) = \Delta_L(\hat{e}(\alpha), \alpha) < 0$. Consider first the case of $\hat{e}(\alpha) \leq \mathbb{E}[\theta|s_H]$. Then, from Lemma 2-(b), it must be $\hat{e}(\alpha) < \mathbb{E}[\theta|s_H]$. Note from Lemma 2-(a) that $\Delta_H(e, \alpha)$ is increasing in $e \in [\hat{e}(\alpha), \mathbb{E}[\theta|s_H])$. By Lemma 2-(b) and continuity of $\Delta_H(e, \alpha)$, there is a $e^* \in [\hat{e}(\alpha), \mathbb{E}[\theta|s_H]]$ such that $\Delta_H(e^*, \alpha) = 0$. Lemma 3 then implies $\Delta_H(e^*, \alpha) = 0 > \Delta_L(e^*, \alpha)$. Consider second the case of $\hat{e}(\alpha) > \mathbb{E}[\theta|s_H]$. Lemma 2-(a) implies $\Delta_H(e, \alpha) < 0$ for any $e > \hat{e}(\alpha)$. Then, it from Lemma 3 follows that we never have $\Delta_H(e, \alpha) \geq 0 \geq \Delta_L(e, \alpha)$. □

Combining with Lemma 1, Lemma 4 states that what value of $\alpha$ guarantees the existence of leadership by confidence equilibrium. Lemma 4-(a) states that for $\alpha$ such that $\Delta_H(\hat{e}(\alpha), \alpha) = \Delta_L(\hat{e}(\alpha), \alpha) \geq 0$, we can find an $e^*$ with which the leadership by confidence constitutes a sequential equilibrium. Lemma 4-(b) states that in case $\Delta_H(\hat{e}(\alpha), \alpha) = \Delta_L(\hat{e}(\alpha), \alpha) < 0$, for $\alpha$ such that $\hat{e}(\alpha) \leq \mathbb{E}[\theta|s_H]$, we can find an $e^*$ with which the leadership by confidence constitutes a sequential equilibrium.

We are now ready to state the main result.

**Proposition 1.**

(a) Suppose $\theta_H \geq 2\theta_L$. Then, there exists a leadership by confidence equilibrium.

(b) Suppose $\theta_H < 2\theta_L$. Then, there exists a unique $\alpha^* \in (0, 1)$ such that there exists a leadership by confidence equilibrium if and only if $\alpha \geq \alpha^*$.

Figure 1 demonstrates Proposition 1-(b) and displays the set of $(\alpha, \rho)$ under which a leadership by confidence equilibrium exists, where we set $\theta_H = 3/2$ and $\theta_L = 1$. The left side line of the union of orange and blue regions corresponds to $\alpha^*$ in Proposition 1-(b).[7] Hence, Proposition 1-(b) states that a leadership by confidence equilibrium exists if and only if $(\alpha, \rho)$ is in the the right hand side of the $\alpha^*$ line.

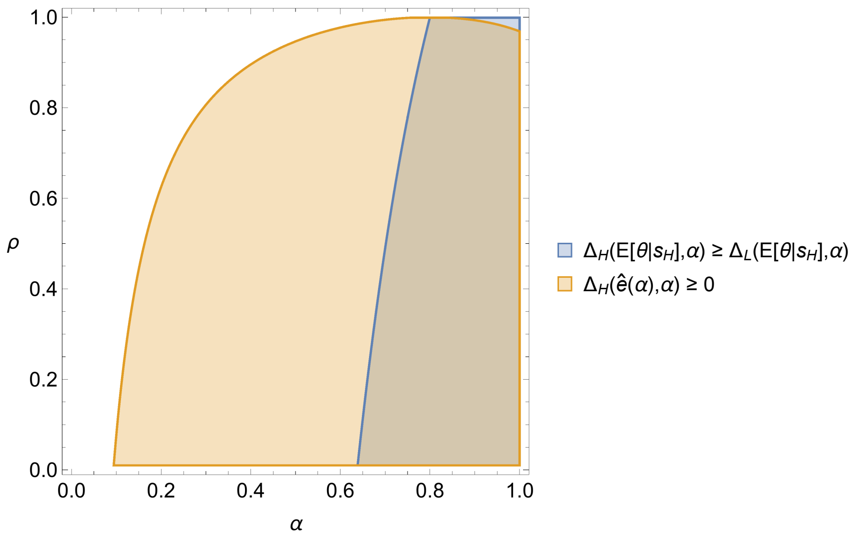


Figure 1: The set of $(\alpha, \rho)$ under which a leadership by confidence equilibrium exists.

We prove Proposition 1.

[7]The blue region represents the set of $(\alpha, \rho)$ under which we have $\Delta_H(\mathbb{E}[\theta|s_H], \alpha) \geq \Delta_L(\mathbb{E}[\theta|s_H], \alpha)$, and the left side line of the blue region corresponds to $\beta$, which we consider the proof in Proposition 1.

*Proof.* We below show that (i) when $\theta_H \geq 2\theta_L$, $\Delta_H(\mathbb{E}[\theta|s_H], \alpha) - \Delta_L(\mathbb{E}[\theta|s_H], \alpha) \geq 0$ for all $\alpha \in [0, 1]$ and (ii) when $\theta_H < 2\theta_L$, there exists a unique $\beta \in (0, 1)$ such that $\Delta_H(\mathbb{E}[\theta|s_H], \alpha) \geq \Delta_L(\mathbb{E}[\theta|s_H], \alpha)$ if and only if $\alpha \geq \beta$. If it is the case of (i), then Lemma 3 implies $\hat{e}(\alpha) \leq \mathbb{E}[\theta|s_H]$ for every $\alpha \in [0, 1]$. Therefore, we can immediately conclude the existence of a leadership by confidence equilibrium from Lemma 4. If it is the case of (ii), then we show that there exists a unique $\alpha^* \in (0, \beta)$ such that (ii-a) $\Delta_H(\hat{e}(\alpha), \alpha) = \Delta_L(\hat{e}(\alpha), \alpha) \geq 0$ for all $\alpha \in [\alpha^*, \beta]$ and (ii-b) $\Delta_H(\hat{e}(\alpha), \alpha) = \Delta_L(\hat{e}(\alpha), \alpha) < 0$ and $\hat{e}(\alpha) > \mathbb{E}[\theta|s_H]$ for all $\alpha \in [0, \alpha^*)$. Therefore, we can conclude that there exists a leadership by confidence equilibrium if and only if $\alpha \geq \alpha^*$ from Lemma 4.

Since $\Delta_H(\mathbb{E}[\theta|s_H], \alpha)$ and $\Delta_L(\mathbb{E}[\theta|s_H], \alpha)$ are both quadratic functions of $\alpha$, the difference is also a quadratic function. Reducing it with a bit of calculation, we have:

$$
\begin{aligned}
&\Delta_H(\mathbb{E}[\theta|s_H], \alpha) - \Delta_L(\mathbb{E}[\theta|s_H], \alpha) \\
&= \left(\frac{\theta_H - \theta_L}{2}\right)\left\{\left(\frac{\theta_H - \theta_L}{2}\right)(1-\rho)\alpha^2 + \left(\frac{7\theta_L - 3\theta_H}{4}\right)\alpha + (\theta_H - 2\theta_L)\right\}.
\end{aligned}
$$

To emphasize that $\Delta_H(\mathbb{E}[\theta|s_H], \alpha) - \Delta_L(\mathbb{E}[\theta|s_H], \alpha)$ is a function of $(\alpha, \rho)$, define a function $G$ by $G(\alpha, \rho) := \Delta_H(\mathbb{E}[\theta|s_H], \alpha) - \Delta_L(\mathbb{E}[\theta|s_H], \alpha)$ for all $\alpha \in [0, 1]$ and for all $\rho \in (0, 1)$. Observe that $G(\alpha, \rho) \geq \lim_{\rho \to 1} G(\alpha, \rho)$ for all $\alpha \in [0, 1]$ and for all $\rho \in (0, 1)$. Observe also that for all $\rho \in (0, 1)$,

$$
\begin{aligned}
G(1, \rho) &\geq \lim_{\rho \to 1} G(1, \rho) \\
&= \left(\frac{\theta_H - \theta_L}{2}\right)\left\{\left(\frac{7\theta_L - 3\theta_H}{4}\right) + (\theta_H - 2\theta_L)\right\} \\
&= \left(\frac{\theta_H - \theta_L}{2}\right)\left(\frac{\theta_H - \theta_L}{4}\right) \\
&> 0.
\end{aligned}
$$

We prove (i). Suppose $\theta_H \geq 2\theta_L$. This implies $\lim_{\rho \to 1} G(0, \rho) \geq 0$. Then, $\lim_{\rho \to 1} G(\alpha, \rho)$ is a linear function of $\alpha$ that has nonnegative values at both the end points of domain. Hence, it must be globally nonnegative. Therefore, $G(\alpha, \rho) \geq \lim_{\rho \to 1} G(\alpha, \rho) \geq 0$ for all $\alpha \in [0, 1]$ and for all $\rho \in (0, 1)$.

We prove (ii). Suppose $\theta_H < 2\theta_L$. This implies $G(0, \rho) < 0$ and $7\theta_L - 3\theta_H > 0$. Note from the latter that for all $\alpha \in [0, 1]$ and for all $\rho \in (0, 1)$,

$$
\frac{\partial}{\partial \alpha} G(\alpha, \rho) = \left(\frac{\theta_H - \theta_L}{2}\right)\left\{(\theta_H - \theta_L)(1-\rho)\alpha + \left(\frac{7\theta_L - 3\theta_H}{4}\right)\right\} > 0.
$$

Hence, $G(\alpha, \rho)$ is a strictly increasing continuous function of $\alpha$ that changes from negative to positive. This means that there exists a unique number $\beta \in (0, 1)$, depending on $\rho$, such that $G(\alpha, \rho) \geq 0$ if and only if $\alpha \geq \beta$. By the construction of $\beta$ and Lemma 3, we have $\hat{e}(\beta) = \mathbb{E}[\theta|s_H]|_{\alpha=\beta}$ and $\hat{e}(\alpha') > \mathbb{E}[\theta|s_H]|_{\alpha=\alpha'}$ for any $\alpha' < \beta$. The latter proves a part of (ii-b). Moreover, we prove in the appendix that $\frac{\partial}{\partial \alpha}\{\Delta_H(\mathbb{E}[\theta|s_H], \alpha) - \Delta_H(\hat{e}(\alpha), \alpha)\} < 0$ for all $\alpha < \beta$. This, together with Lemma 2-(b), implies that $\Delta_H(\hat{e}(\alpha), \alpha)$ is a strictly increasing continuous function of $\alpha$ that switches from negative $\Delta_H(\hat{e}(0), 0)$ to positive $\Delta_H(\hat{e}(\beta), \beta) = \Delta_H(\mathbb{E}[\theta|s_H], \alpha)|_{\alpha=\beta}$. Therefore, there exists a unique number $\alpha^* \in (0, \beta)$ such that (ii-a) $\Delta_H(\hat{e}(\alpha), \alpha) = \Delta_L(\hat{e}(\alpha), \alpha) \geq 0$ for all $\alpha \in [\alpha^*, \beta]$ and (ii-b) $\Delta_H(\hat{e}(\alpha), \alpha) = \Delta_L(\hat{e}(\alpha), \alpha) < 0$ and $\hat{e}(\alpha) > \mathbb{E}[\theta|s_H]$ for all $\alpha \in [0, \alpha^*)$. □

Proposition 1-(a) states that, if the variation in possible values of the team productivity is large, then a leadership by confidence equilibrium exists. Proposition 1-(b) states that, even if the variation is limited, a leadership by confidence equilibrium exists as long as the degree of correlation between agents' confidence is relatively small.

To obtain the intuition about how the intensity of correlation affects the existence, recall Lemma 4. This lemma states that we never have a leadership by confidence equilibrium if and only if $\Delta_H(\hat{e}(\alpha),\alpha) = \Delta_L(\hat{e}(\alpha),\alpha) < 0$ and $\hat{e}(\alpha) > \mathbb{E}[\theta|s_H]$. From the single-crossing property of the first-move incentives stated in Lemma 3, it is indeed the case that $\Delta_H(e,\alpha) \geq 0$ implies $\Delta_L(e,\alpha) > \Delta_H(e,\alpha)$.

Consider the limit case of $\alpha = 0$, that is, signals are now perfectly correlated perfect information signals. Hence, we have $f(s_L|s_H) = f(s_H|s_L) = 0$. In that case, the value of signaling and the value of learning (or the value of adjustment from $e$) for each type are as follows.

$$\begin{aligned}
&\text{Value of signaling for } s_H = 0,\\
&\text{Value of signaling for } s_L = \mathbb{E}[\theta|s_L,s_L]\big\{\mathbb{E}[\theta|s_L,s_H] - \mathbb{E}[\theta|s_L,s_L]\big\}\\
&\text{Value of adjustment from } e \text{ for } s_H = \frac{\big\{e - \mathbb{E}[\theta|s_H,s_H]\big\}^2}{2}\\
&\text{Value of adjustment from } e \text{ for } s_L = \frac{\big\{e - \mathbb{E}[\theta|s_L,s_L]\big\}^2}{2}
\end{aligned}$$

The value of of signaling for an agent with $s_H$-signal is zero because the agent knows that the other agent receives $s_H$-signal and hence is never influenced by leading by example. On the other hand, the value of signaling for an agent with $s_L$-signal is positive. The values of adjustment from $e$ are just the adjustment costs from $e$ to the "observed" team productivities in this limit case. Then, from this, we find that $\Delta_H(e,\alpha) \geq 0$ is achieved if and only if $e = \mathbb{E}[\theta|s_H,s_H]$, and in that case, we have $\Delta_H(\mathbb{E}[\theta|s_H,s_H],\alpha) = 0$. In other words, if an agent with $s_H$-signal knows that the other agent also knows the team productivity is high, that agent has basically no incentive to lead. On the other hand, if the variation in possible values of the team productivity is small ($\theta_H < 2\theta_L$), then we have:

$$\begin{aligned}
&\text{Value of signaling for } s_L - \text{Value of adjustment from } \mathbb{E}[\theta|s_H,s_H] \text{ for } s_L\\
&= \mathbb{E}[\theta|s_L,s_L]\big\{\mathbb{E}[\theta|s_L,s_H] - \mathbb{E}[\theta|s_L,s_L]\big\} - \frac{\big\{\mathbb{E}[\theta|s_H,s_H] - \mathbb{E}[\theta|s_L,s_L]\big\}^2}{2}\\
&= \mathbb{E}[\theta|s_L,s_L]\frac{\big\{\mathbb{E}[\theta|s_H,s_H] - \mathbb{E}[\theta|s_L,s_L]\big\}}{2} - \frac{\big\{\mathbb{E}[\theta|s_H,s_H] - \mathbb{E}[\theta|s_L,s_L]\big\}^2}{2}\\
&= \frac{\big\{2\mathbb{E}[\theta|s_L,s_L] - \mathbb{E}[\theta|s_H,s_H]\big\}\big\{\mathbb{E}[\theta|s_H,s_H] - \mathbb{E}[\theta|s_L,s_L]\big\}}{2}\\
&= \frac{(2\theta_L - \theta_H)\big\{\mathbb{E}[\theta|s_H,s_H] - \mathbb{E}[\theta|s_L,s_L]\big\}}{2}\\
&> 0.
\end{aligned}$$

That is to say, in the case that the variation in possible values of the team productivity is small, the adjustment cost from $\mathbb{E}[\theta|s_H,s_H]$ to $\mathbb{E}[\theta|s_L,s_L]$ is also small, and hence the positive value of signaling dominates the adjustment cost. This means that if an agent with $s_L$-signal knows that the other agent also knows the team productivity is low, that agent rather has an incentive to lead. In general, if the intensity of correlation between agents' confidence is large, the first-move incentive for an agent with $s_H$-signal is small, while the first-move incentive for an agent with $s_L$-signal is not so small because of the positive value of signaling and the small value of adjustment from the supposed initiative effort. Therefore, we never have a leadership by confidence equilibrium in such cases.

# 5 Concluding Remarks

We presented a mechanism of endogenous leadership and identified a set of necessary and sufficient condition on the information system about the team productivity under which this mechanism of endogenous leadership succeeds in realizing leadership in equilibrium.

We focused on the possibility of endogenous leadership and left several important issues out of the analysis. First, we did not study the welfare effect of endogenous leadership in team production. Although leadership including endogenous leadership is believed in practice to advance the efficiency of team production, we need to verify it theoretically.

Second, we assumed that every agent receives the same payoff from the team output. Once we consider the issue of team design, this assumption must be replaced by the question of what is the optimal rewarding schedule for agents. These issues among others are left for future research.

# Appendix

*Supplement for the proof of Proposition 1.*

Suppose $\theta_H < 2\theta_L$. Consider the following:

$$\begin{aligned}\frac{\partial}{\partial\alpha}\left\{\Delta_H(\mathbb{E}[\theta|s_H],\alpha) - \Delta_H(\hat{e}(\alpha),\alpha)\right\} &= \frac{\partial}{\partial\alpha}\left(\frac{\left\{\hat{e}(\alpha) - \mathbb{E}[\theta|s_H]\right\}^2}{2}\right)\\ &= \left\{\hat{e}(\alpha) - \mathbb{E}[\theta|s_H]\right\}\frac{\partial}{\partial\alpha}\left\{\hat{e}(\alpha) - \mathbb{E}[\theta|s_H]\right\},\end{aligned}$$

where the first equality follows from Lemma 2-(a). As stated in the proof of Proposition 1, we have $\hat{e}(\alpha) - \mathbb{E}[\theta|s_H] > 0$ for $\alpha < \beta$. We below show that $\frac{\partial}{\partial\alpha}\left\{\hat{e}(\alpha) - \mathbb{E}[\theta|s_H]\right\} < 0$ for all $\alpha \in [0,1]$. Therefore, we can conclude that $\frac{\partial}{\partial\alpha}\left\{\Delta_H(\mathbb{E}[\theta|s_H],\alpha) - \Delta_H(\hat{e}(\alpha),\alpha)\right\} < 0$ for all $\alpha \in [0,\beta)$.

Reaping the benefit from tractability of the adopted information structure such as $\mathbb{E}[\theta|s_H,s_H] - \mathbb{E}[\theta|s_H,s_L] = \mathbb{E}[\theta|s_H,s_L] - \mathbb{E}[\theta|s_L,s_L]$, solve the implicit equation $\Delta_H(e,\alpha) = \Delta_L(e,\alpha)$ explicitly and derive the function $e = \hat{e}(\alpha)$ as follows.

$$\hat{e}(\alpha) = \frac{4\theta_H + 8\theta_L - \alpha\left\{(5-4\rho)\theta_H + (3+4\rho)\theta_L\right\}}{4(2-\alpha)}.$$

Differentiate this with respect to $\alpha$ and obtain:

$$\frac{\partial}{\partial\alpha}\hat{e}(\alpha) = \frac{\theta_L - 3\theta_H + 4\rho(\theta_H - \theta_L)}{2(2-\alpha)^2}.$$

We can easily verify that, when $\theta_H < 3\theta_L$ (and hence when $\theta_H < 2\theta_L$), this is a strictly decreasing function of $\alpha \in [0,1]$ for any fixed $\rho \in (0,1)$. Hence, we have:

$$\begin{aligned}\frac{\partial}{\partial\alpha}\big\{\hat{e}(\alpha) - \mathbb{E}[\theta|s_H]\big\} &\leq \frac{\partial}{\partial\alpha}\hat{e}(\alpha)\bigg|_{\alpha=0} - \frac{\partial}{\partial\alpha}\mathbb{E}[\theta|s_H] \\ &= \frac{\theta_L - 3\theta_H + 4\rho(\theta_H - \theta_L)}{8} + \frac{(1-\rho)(\theta_H - \theta_L)}{2} \\ &= \frac{\theta_H - 2\theta_L}{8} \\ &< 0,\end{aligned}$$

where the last inequality directly follows from the supposition $\theta_H < 2\theta_L$. □

## References


**Abe, Koji, Hajime Kobayashi, and Hideo Suehiro.** 2015. "Experiments on the Emergence of Leadership in Teams," unpublished manuscript.

**Abreu, Dilip, and Faruk Gul.** 2000. "Bargaining and Reputation," *Econometrica*, 68(1): 85–117.

**Amir, Rabah, and Anna Stepanova.** 2006. "Second-Mover Advantage and Price Leadership in Bertrand Duopoly," *Games and Economic Behavior*, 55(1): 1–20.

**Chamley, Christophe, and Douglas Gale.** 1994. "Information Revelation and Strategic Delay in a Model of Investment,," *Econometrica*, 62(5): 1065–1085.

**Fudenberg, Drew, and Jean Tirole.** 1986. "A Theory of Exit in Duopoly," *Econometrica*, 54(4): 943–960.

**Gal-Or, Esther.** 1987. "First Mover Disadvantage with Private Information," *Review of Economic Studies*, 54(2): 279–292.

**Gul, Faruk, and Russell Lundholm.** 1995. "Endogenous Timing and the Clustering of Agents' Decisions," *Journal of Political Economy*, 103(5): 1039–1066.

**Hamilton, Jonathan H., and Steven M. Slutsky.** 1990. "Endogenous Timing in Duopoly Games: Stackelberg or Cournot Equilibria," *Games and Economic Behavior*, 2(1): 29–46.

**Hendricks, Kenneth, and Dan Kovenock.** 1989. "Asymmetric Information, Information Externalities, and Efficiency: The Case of Oil Exploration," *RAND Journal of Economics*, 20(2): 164–182.

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

**Holmström, Bengt.** 1982. "Moral Hazard in Teams," *Bell Journal of Economics*, 13(2): 324–340.

**Kambe, Shinsuke.** 1999. "Bargaining with Imperfect Commitment," *Games and Economic Behavior*, 28(2): 217–237.

**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.

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

**Mailath, George J.** 1993. "Endogenous Sequencing of Firm Decisions," *Journal of Economic Theory*, 59(1): 169–182.

**Milgrom, Paul R., and Robert J. Weber.** 1982. "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50(5): 1089–1022.

**Normann, Hans-Theo.** 1997. "Endogenous Stackelberg Equilibria with Incomplete Information," *Journal of Economics*, 66(2): 177–187.

**Normann, Hans-Theo.** 2002. "Endogenous Timing with Incomplete Information and with Observable Delay," *Games and Economic Behavior*, 39(2): 282–291.

**van Damme, Eric, and Sjaak Hurkens.** 1999. "Endogenous Stackelberg Leadership," *Games and Economic Behavior*, 28(1): 105–129.

**van Damme, Eric, and Sjaak Hurkens.** 2004. "Endogenous Price Leadership," *Games and Economic Behavior*, 47(2): 404–420.