Leadership Based on Asymmetric Information

Mana Komai and Mark Stegeman*

revision: April, 2009

Abstract

An organization makes collective decisions through neither markets nor contracts. Instead, rational agents voluntarily choose to follow a leader. The leader has no special talents but is distinguished by having exclusive access to information. If the leader satisfies a credibility condition, then incentive and coordination problems are solved and the unique nondegenerate equilibrium achieves the first best, even though every agent has incentives to shirk. A crucial feature is that the leader reveals part but not all of her information. If the credibility condition fails, as is more likely in a large organization, then the equilibrium changes discontinuously: the leader is ignored and the outcome is extremely inefficient. Appointing multiple leaders can help to restore credibility. If agents have heterogeneous costs, then appointing a leader with higher than average costs is another way to restore credibility, indicating that less leadership is sometimes better leadership. The leader's informational advantage can appear endogenously, in a model where players make private investments in information; in many cases the undesirable equilibrium, in which all players are informed, disappears. The equilibrium concept is an original extension of sequential equilibrium to a continuous state space.

1 Introduction

Observers of management and politics believe that leadership is important, but economists rarely study it.[1] In economists' typical models, managers simply fill gaps in authority. They variously act as owners' agents or exercise residual control when contracts are incomplete; politicians usually represent amalgamations of voters' preferences. These models have produced important insights but omit crucial aspects

*All of the theorems in this paper, except Theorems 5 and 6, have appeared previously in Komai and Stegeman (2004b). The suggestions of two anonymous referees, including the extension to the 3-stage model, greatly improved the paper. We are grateful to seminar participants at Arizona, VPI, and the NBER/NSF Decentralization Conference, and especially Martin Dufwenberg, Stephen Gilliland, Benjamin Hermalin, Mark Johnson, Roy Radner, and Mark Walker, for helpful comments and suggestions. The usual disclaimer applies.

[1]Economists have produced a few formal models of leadership. Examples include Hermalin (1993, 2007), Rotemberg and Saloner (1993, 2000), Kobayashi and Suihiro (2005), and Huck and Biel (2006).

of leadership. Leadership often includes motivating by example, separate from any formal authority.

This paper describes an "organization" which makes collective decisions through neither markets, nor contracts, nor voting, nor any grant of authority. There is no principal and no agent. Instead, rational actors choose to follow a better-informed leader absent any obligation to do so. We show that this behavior alleviates both of the classic problems of organizational design: conflicting interests and coordination failure.[2]

The first part of the paper, Sections 2-4, studies a 2-stage model under alternative information structures. If the incentive to shirk is merely moderate, then giving information to only one player, the "leader," leads to more efficient outcomes than giving the information to everyone (i.e., complete information). This happens because restricting access to information makes it harder to know when to shirk and, separately, radically reduces the number of equilibria.

The second part of the paper, Section 5, makes the information structure endogenous by adding an earlier stage in which the players individually and privately decide whether to pay a fixed research cost to become informed. To refine the equilibria of this more complicated extensive form, the paper proposes the concept of discrete sequential equilibrium, a straightforward extension of sequential equilibrium, which extends the sequential equilibrium in a straightforward way, by requiring strategies to be measurable in a finite partition of the continuous state space.

If the research cost is sufficiently small, in this 3-stage model, then complete information is an equilibrium outcome. If this cost falls in a higher but overlapping range, then the model supports the more efficient leader-follower equilibrium. For a wide range of parameter values (where the research cost exceeds the values in the region of overlap), the complete information equilibria disappear because research is too expensive for players who act autonomously, but the leader-follower equilibrium still exists. This shows that *the institution of leadership, which is more efficient than complete information even before considering the costs of information acquisition, can arise endogenously because costly acquisition by many players is unsupportable in equilibrium.*

These results have contrarian implications for the role of information in organizational design. Many researchers focus on how to reduce transactions costs by correcting

[2] See, for example, Milgrom and Roberts (1992).

information failures, but in our model efficiency *requires* an information failure.[3] A huge literature studies how to align agents' interests through well-designed contracts, but our model abandons any attempt to align interests. Instead, in the leader-follower equilibrium, each follower's ignorance deprives him of the information needed to protect his self-interest. Collectively, however, the followers benefit from their ignorance, and they understand this.[4]

The model also offers a contrarian view of leaders: they have no special talents and are distinguished merely by occupying the leadership position. Moreover, the third part of the paper, Section 6, establishes the surprising result that if players are differentiated, then the optimal leader has either average or unusually high costs of effort. Promoting high-cost persons to leadership has the double advantage of removing them from the ranks of followers, where they tend to shirk, and placing them into leadership, where their reluctance to act makes them more credible when they do.[5]

This apparently gloomy picture of unexceptional leaders and ignorant followers leads to happy results. In many cases coordination and shirking problems are completely solved: the unique productive equilibrium achieves the first best even though the underlying payoff structure makes coordination difficult and gives every agent incentives to shirk.

In reality, there may be reasons to seek low-cost leaders, reasons which do not appear in this simple model. This paper thus describes just one small piece of the puzzle of leadership, but this piece seems important. It shows that superficial notions of what makes a good leader – or a good follower – may not always be correct. In some cases, he who knows less works more, and he who leads least leads best.

One motivation for this work is the observation that the usual desiderata of organizational design – improving information and using contracts to correct incentives – impose costs which are often not modeled explicitly. Collecting and processing information is expensive, as are the design and enforcement of optimal contracts. Our model addresses these issues by proposing an environment which *minimizes* monitoring, contracting, and information exchange. Instead, the followers simply mimic the leader, and the simplicity of the information structure is exactly what mitigates the

---

[3] It is well known, at least since Hirshleifer (1971), that many situations exist in which information failures can increase efficiency.

[4] "They understand this" in the sense that a player who understands the model has no reason to change his equilibrium behavior.

[5] One reader has called this the Dilbert theory of leadership. This model may also suggest a theory of legitimacy: that it derives from superior information.

classic obstacles to efficiency.

Komai et al. (2007) show in a similar setting that giving a leader exclusive access to information can improve efficiency by reducing the incentive to shirk, but their simple example employs a linear payoff function. Linear payoffs eliminate most strategic interaction and with it the coordination problem which provides much of the motivation for the present analysis. Linear payoffs also preempt the problem of credibility failure, which here plays a central role. Finally, Komai et al. assume that all players are homogeneous, thereby excluding the question of the optimal leader, and they assume that the information structure is exogenous.

Sections 2 through 6 present formally the models and results just described. Section 5 includes the extension of sequential equilibrium to continuous state spaces, which may have broader application. Section 7 shows, by example, that if no single player would make a credible leader, then appointing multiple leaders can solve the credibility problem and support equilibria which are more efficient than any equilibrium under complete information. Section 8 discusses related papers, and Section 9 offers concluding remarks.

2 The Payoff Function

Each of $m > 1$ identical players, $I = \{1, 2, ..., m\}$, must decide whether to join a proposed project The project could, for example, be joining a group which will prepare a bid which if successful will help all m players, or supporting a political campaign which if successful will help all m players. Each player $i \in I$ chooses an action $a_i \in A = \{N, P\}$, where N denotes non-participation and P denotes participation in the project. If all players choose N then each earns a payoff of zero. If all choose P then each earns a random payoff x, which is common across players and distributed uniformly on the interval $[-1, 1]$.[6] The project is *good* if $x > 0$ or *bad* if $x \leq 0$.

If the players split, with a fraction $q \in \{0, \frac{1}{m}, \frac{2}{m}, ...1\}$ choosing P and the rest choosing N, then each participant earns

$$\pi(P, q; x) \equiv (x + \beta)q - \beta \ , \tag{1a}$$

and each non-participant earns

$$\pi(N, q; x) \equiv (x + \beta)q\alpha \ , \tag{1b}$$

[6] Assume that the state space $[-1, 1]$ is part of a probability space $([-1, 1], \mathfrak{B}, \mathfrak{U})$, where $\mathfrak{B}$ denotes the usual Borel field and $\mathfrak{U}$ denotes the uniform probability measure.

where $\alpha \in (0,1)$ and $\beta > 0$ are fixed parameters. Note that $\pi(P,1;x) = x$ and $\pi(N,0;x) = 0$, consistent with the assumptions in the previous paragraph.[7]

In the boundary case $\alpha = 1$ and $\beta = 0$, everyone gets xq, the product of project quality and the participation rate. The difference between participants' and non-participants' payoffs comes from $\beta > 0$, which can be interpreted (roughly) as a fixed cost of participation, and from $\alpha < 1$, the rate at which non-participants accrue project benefits, relative to participants. Given $\beta > 0$, $\alpha = 1$ represents the case of a pure public good, meaning that participants and nonparticipants benefit equally from the good, and $\alpha = 0$ represents the case of a pure network externality, meaning that only participants benefit from the "network."

To ensure that players never want to participate alone (i.e., to ensure that $\pi(P, \frac{1}{m}; 1) < \pi(N,0;1)$), assume:

$$\beta > \frac{1}{m-1} \ . \tag{2}$$

To ensure that players sometimes do want to participate (i.e., to ensure that $\pi(P,1;1) > \pi(N, 1-\frac{1}{m}; 1)$), assume:

$$\alpha < \frac{m}{(m-1)(1+\beta)} \ . \tag{3}$$

Two key properties of the payoff function π produce the two inefficiencies that motivate this study. First, for higher-quality projects ($x > -\beta$), non-participants benefit from others' participation.[8] This tends to cause inefficiency through *shirking* (or *free riding*). The payoff to shirking when all others participate is $\pi(N, \frac{m-1}{m}; x) - \pi(P,1;x) = (x+\beta)\alpha\frac{m-1}{m} - x$, which increases in α and β but decreases in x.

Second, higher-quality projects have the quite different property that any player i's return to participation is an increasing function of others' participation. Formally, $\pi(P, q_{-i} + \frac{1}{m}; x) - \pi(N, q_{-i}; x)$ increases in q_{-i} for x sufficiently large, where q_{-i} denotes the participation rate if player i chooses N.[9] This network effect can cause inefficiency through *coordination failure*. Section 3 uses a standard global games analysis to show that coordination problems can be severe in this model.

---

[7] Following the initial analysis in Sections 2-5, Section 6 amends the payoff function to incorporate heterogeneous participation costs.

[8] If $\beta > 1$, then $x + \beta > 0$, implying that non-participants always benefit from others' participation.

[9] For payoff function (1), both properties hold for $x > -\beta$. For the generalized payoff function studied in the appendix, the two properties are embodied in assumptions (10f) and (10k), which do not require that they hold for the same set of values of x.

For $x < -\beta$, $\pi(P, q_{-i} + \frac{1}{m}; x) - \pi(N, q_{-i}; x)$ is negative and decreasing in q_{-i}. The interpretation is that the private return to being the first participant may be bad, but the private return to being one of the last (e.g. decisive) participants may be worse.

The payoffs represented by π could be monetary, but they could also include non-transferable psychic, political, or status benefits. The incremental reward for participation (represented formally by $1 - \alpha > 0$) could come from earning more of these benefits and might or might not include transfers intended to reward participation. The model is agnostic about the constituents of π.

Ideally, the players could fix the shirking (i.e., moral hazard) problem by signing a contract, ex ante, which would transfer additional resources from nonparticipants to participants. In other words, players could agree to modify the given payoff structure π. A typical obstacle to such contractual solutions is players' inability to observe each others' behavior, but that is not the premise here.[10] Our assumption is that players do observe each others' participation, but for unspecified reasons they cannot execute a contract which improves upon whatever incentives to participate already exist in π. This could be because: behavior is not verifiable; drafting or enforcing performance clauses is costly; utility is only partially transferable; information is already incomplete when it is time to draft the contract; or simply because disagreements arise over the division of surplus. The experience of real organizations suggests that most employment contracts specify few specific contingencies.

Whatever the cause, the key premise is that the return to participation is insufficient to eliminate shirking, and it is hard to fix that problem contractually. To take the focus away from contracts and transfer schemes, this paper treats π as exogenous and instead considers the impact of alternative timing and information structures.

Projects which combine problems of shirking and coordination are common in reality. They may include preparing a bid, meeting a sales target, adopting a new procedure or technology, developing a product, supporting a database, implementing a restructuring plan, or establishing a working relationship with another unit. The model is agnostic about the process which generates these potential projects. (The projects could, for example, come from an unmodeled higher level in the organization.)

In political contexts, many "projects" are completely outside a contracted environment, and this can make the shirking problem severe. Legislators in the U.S. are relatively free agents, who are not bound to vote with their party's leadership. Citi-

[10] If the leader is chosen exogenously, as is assumed throughout the paper except in Section 5, then it is reasonable to suppose that players cannot normally observe each others' participation decisions but that the actions of the leader are a deliberately created exception. In Section 5, though, leaders effectively choose themselves by acquiring costly information, and in that case it may be hard to explain why their actions should be easier to observe than anyone else's. Therefore, to provide an interpretation that works throughout the entire paper, we here assume that all players' participation decisions are observable and that incentive (i.e. moral hazard) problems arise for other reasons.

zens who join advocacy groups, or endorse legislation or candidates, or work to support political causes, are often volunteers in every sense of the word. Such activities can be projects in the sense of our model.[11]

None of the results depend on the specific form of the payoff function. The appendix lists various properties of a general payoff function $\pi(\cdot, q; x)$, which are sufficient for all of the major results in this paper. These properties are, roughly: higher values of x represent better projects; it is never optimal to participate alone but there are increasing returns to participation; there exist projects for which full participation is individually rational but also a significant number of projects for which it is not;[12] and participating in good projects confers a positive externality on other players. Unlike (1), the more general payoff functions studied in the appendix allow $\frac{\partial \pi(N,q;x)}{\partial x} < 0$ for $x < 0$, which is plausible if political benefits arise from not participating in a bad project.

3 Complete Information

A player i who knows nothing about the value of x will never participate, regardless of other players' actions, because $Ex = 0$ implies that action P always gives lower expected utility than does action N (i.e., $-(1 - q_{-i} - \frac{1}{m})\beta < q_{-i}\alpha\beta$ for any $q_{-i} = 0, \frac{1}{m}, \frac{2}{m}, ... \frac{m-1}{m}$). To support participation, players need some information about x, which will help them to decide when to participate.

The simplest situation is that all players observe x and then decide simultaneously whether to participate. Call this the *complete information* model: each player i's strategy takes the form $s_i : [-1, 1] \rightarrow A$. It is easy to describe the symmetric Nash equilibria of this game, and it is natural to focus on equilibria that employ threshold strategies; a player adopts the *threshold strategy* t if she plays N if $x \leq t$ or P if $x > t$. All proofs are in the appendix.

Theorem 1 *In the complete information model, there exists $\tau^C \in (0, 1)$ such that full participation ($q = 1$) is a Nash equilibrium, given x, if and only if $x \geq \tau^C$; no participation ($q = 0$) is an equilibrium for all x. Therefore, symmetric adoption of the*

[11] The appendix (following assumption (10k)) describes a scenario in which a physically costless vote can be interpreted as the costly participation decision modeled here.

[12] The appendix does not explicitly require a significant number of bad projects (i.e. projects with $x < 0$), but if almost all projects are good, then the leader's actions would convey very little information in a leader-follower equilibrium. Given the other assumptions of the model, this tends to cause violation of the credibility condition (4).

threshold strategy t constitutes a Nash equilibrium if and only if $t \geq \tau^C$.[13]

The threshold $\tau^C = \frac{(m-1)\alpha\beta}{m-(m-1)\alpha}$ (the solution to $\pi(P,1;\tau^C) = \pi(N, \frac{m-1}{m}; \tau^C)$) represents the lowest value of x at which each player is willing to participate if all others participate. If $x \geq \tau^C$, then the resulting subgame is a coordination game, where no participation and full participation are both Nash equilibria.[14]

Example: It is useful to illustrate Theorem 1 through the following simple example, which will be employed and extended throughout the paper. Suppose that $\alpha = \beta = \frac{1}{2}$ and $m = 9$. Then $\tau^C = 0.4$, implying that only 60% of good projects attract participants even in the most efficient Nash equilibrium. The remaining 40% of good projects are undone by shirking. In this equilibrium, each player's expected payoff is $\frac{1}{2}\int_{0.4}^{1} x dx = .21$.

Coordination failure may considerably increase these efficiency losses, because participation is weakly dominated at $x = \tau^C$ and "almost weakly dominated" when x exceeds τ^C only slightly. One way to model the coordination problem is to consider the perturbed model in which (in the style of Carlsson and van Damme (1993)) each player observes x with an arbitrarily small private error, which is distributed continuously and symmetrically about zero and which is statistically independent of other players' observation errors. Then threshold equilibria close to τ^C are ruled out.

Example (continued): In the perturbed version of the Example, consider the possibility of a threshold equilibrium with $t < 1$. Then a player who observes $x = t$ expects half of the remaining eight players to participate, implying that, on average, playing P gives him $\pi(P, \frac{5}{9}; t)$, while playing N gives him $\pi(N, \frac{4}{9}; t)$. A simple calculation shows that the latter exceeds the former for any $t < 1$, contradicting the

[13]For the payoff function (1), these are the only Nash equilibria of the complete information model, but this stronger statement does not extend to the more general class of payoff functions described in the appendix. Because we do not want to rely on the specific functional form (1), we emphasize results that extend directly to this more general class.

Specifying that the players move sequentially with perfect information, and refining the equilibrium by subgame perfection, changes this result by eliminating the $q = 0$ equilibrium for $x > \tau_C$, which in turn implies that the only equilibrium threshold strategy is $t = \tau_C$. For $x < \tau_C$, zero participation is the only equilibrium outcome under sequential moves because N is dominant for the last player and it follows from backward induction that every participant knows that he would be the last participant and that makes N optimal regardless of previous players' actions. For $x > \tau_C$, the last player to move plays P if all previous players have participated, and it follows from backward induction that the same is true for all previous players, implying that full participation is the only equilibrium outcome. The player who moves first effectively becomes the coordinator and selects away from any less efficient equilibrium. This differs from Varian's (1992) benchmark result that sequential contribution causes crowding out, because Varian assumes decreasing rather than increasing returns to contribution.

[14]This coordination game does not appear in Komai et al. (2007), because the linear payoffs in that model imply that a player's optimal action does not depend directly upon any other player's action. Therefore, none of the coordination problems discussed in the next several paragraphs arise in that paper.

indifference required by equilibrium. Therefore, the unique equilibrium of the global game produced by adding arbitrarily small noise to players' observations implies that no one ever participates!

These two sources of inefficiency are characteristic (to varying degrees) of all of the equilibria described by Theorem 1, for any values of the parameters $\alpha \in (0, 1)$, $\beta > 0$, and m. Shirking causes inefficient participation for $x \in (0, \tau^C)$, and introducing any uncertainty concerning others' play makes it hard to coordinate on efficient participation for $x > \tau^C$, if x is sufficiently close to τ^C.

Section 4 studies a 2-stage game with a different information structure, which can solve, at once, both sources of inefficiency: shirking and coordination failure. Each of these problems has generated a vast literature, but we are unaware of other studies which solve both problems in the same setting.

Section 5 then adds an earlier stage in which the information structure is determined endogenously. Depending on parameter values, either the complete information structure of this section or the incomplete information structure of Section 4 can appear in the equilibria of the 3-stage game.[15]

4 One Leader and Followers

Consider a different information structure, which allows just one specified player, $l \in I$, to observe x and to act before the other players. If this *leader* chooses P, then the remaining *followers,* $I \backslash \{l\}$, observe this and decide, independently and simultaneously, whether to choose N or P. If the leader chooses N, then the followers choose N by default; the interpretation is that they are unaware of the project. In this *leader-follower* game, the leader's (pure) strategy still has the form $s_l : [-1, 1] \to A$, and each follower i's (pure) strategy is $s_i \in A$, indicating what she will do if the leader participates.[16] Like the symmetric information game, the leader-follower game always has a trivial unproductive equilibrium in which no one ever participates. A *productive*

[15] It is sensible to study the 2-stage model (Section 4) before the 3-stage model (Section 5), because the 2-stage model corresponds roughly to the last two stages of the 3-stage model and so would be studied first anyway. Also, a leader can arise in many ways; the 3-stage model is only one account. Studying the two-stage model first allows us to remain agnostic about where the leader comes from.

[16] The formulation of the leader's strategy set does not allow her to reveal x directly, but in the productive equilibria of Theorem 4 she would anyway have no incentive to do so. (In other settings, such as that of Theorem 8, the possibility of revealing x could cause defections from the proposed leader-follower equilibrium.)

In many realistic settings verifiable disclosure of x would carry significant costs of assembling and interpreting evidence, and other costs. For example, the photographs famously released during the Cuban missile crisis also revealed secrets about U.S. intelligence-gathering.

The possibility of mixing is discussed below, in the context of stating the equilibrium concept.

equilibrium is one in which at least the leader participates with positive probability.

The practice of assigning information to just one player could arise endogenously from the research decisions of individual players, as in the expanded model of Section 5. Alternatively, the 2-stage model could reflect the decision of an outside agent (e.g., employer) who invests real resources, ex ante, into institutions which will discover x and establishes a leader who will get that information. A third interpretation is that the players choose the leader cooperatively, perhaps compensating her in advance for becoming an expert and acquiring unique knowledge about the value of x. In all cases, the payoffs represented by π should be interpreted as payoffs subsequent to any ex ante investments in the creation of the leader.[17]

This section shows that, in many cases, the leader-follower game solves both the coordination and shirking problems. It has a *unique* productive equilibrium, which achieves the *first best*.

4.1 The unique productive equilibrium.

An *equilibrium* of the leader-follower game is defined to be a Bayesian Nash equilibrium, a strategy profile, $S^* = (s_l^*; s_f^*, f \in I\backslash\{l\})$, such that: s_l^* maximizes leader l's expected payoff given the followers' strategies and any $x \in [-1, 1]$; and s_f^* maximizes follower f's expected payoff given the other players' strategies.[18] This formulation excludes mixed strategies. Mixed strategy equilibria are implausible because there always exists a pure strategy equilibrium and because increasing returns to participation imply that any mixed strategy equilibrium is unstable in the sense that is familiar from symmetric coordination games: myopic best replies to any small perturbation of the mixture lead to a pure strategy equilibrium.[19]

[17] Yet another interpretation is that the players learn about the value of x from an outside agent. who (for no additional compensation) agrees to share that information with only the leader. If the players can moreover persuade that agent to transmit the information in exactly the form that they would get it from the leader, then the leadership position has been "outsourced" and there is no need for a leader within the group. One general problem with outsourcing leadership is that the outside entity has no intrinsic incentive to report correctly. This issue is discussed after Theorem 3.

[18] For a "Bayesian" equilibrium, Harsanyi (1967) requires the leader to act optimally in a subset of $[0, 1]$ having probability one, which is equivalent to removing the reference to "any $x \in [0, 1]$." That would reduce the present equilibrium to a standard Nash equilibrium, consistent with Harsanyi's demonstration that the set of Bayesian equilibria corresponds exactly to the Nash equilibria of the strategic form such that strategies are expressed as a function of type. For a "Bayesian Nash" equilibrium, Crawford and Sobel (1982) and sequels require that the sender act optimally given *any* realization of his type, analogous to requiring "any $x \in [0, 1]$." The equilibrium concept employed here is thus a Bayesian Nash equilibrium in the sense of Crawford and Sobel rather than in the sense of Harsanyi.

[19] Of course there exist equilibria which incorporate trivial mixing: only a leader who observes his unique threshold state mixes. Any nontrivial mixing must be by the followers, and mixing of that kind is unstable, in the sense described in the text.

The simplest equilibrium is an unproductive one, in which the leader plays $s_f^*(x) = N$ for any x and every follower plays N. It is easy to see that this is not merely a Bayesian Nash equilibrium but can also be supported as a sequential equilibrium.[20] The more interesting equilibria are the productive ones, when they exist. Recall that, under full information, there always exists a *continuum* of productive equilibria (Theorem 1).

Lemma 2 shows that, in any productive equilibrium, all followers choose P. In other words, they follow the leader.

Lemma 2 *If* $(s_l^*; s_f^*, f \in I \backslash \{l\})$ *is a productive equilibrium, then* $s_f^* = P$ *for all* $f \in I \backslash \{l\}$.

The proof is somewhat complicated, but this paragraph sketches the argument. If the leader never participates in equilibrium, then the followers cannot participate and the equilibrium is unproductive.[21] Suppose that the leader participates with positive probability. She cannot gain by participating in bad projects. Therefore all followers should choose the same action, because $x \geq 0$ and the equilibrium participation of one follower imply, by increasing returns to participation, that other followers should also participate. If they all choose N, then the leader should not act alone and so should not choose P, but that is a contradiction. Therefore, every follower chooses P.

Lemma 2 implies that, in equilibrium, the leader neither gains nor loses by occupying that position. (In particular, the leader collects no informational rents.) Every player, including the leader, earns $\pi(P, 1; x) = x$ if the leader participates in a productive equilibrium, or $\pi(N, 0; x) = 0$ otherwise. In this sense, the players have common interests, and the leader consequently adopts the threshold strategy $t = 0$ in any productive equilibrium, participating in any good project and ignoring any bad project. That is efficient, but there is a caveat: *in some cases a productive equilibrium does not*

The linear payoff functions in Komai et al. (2007) eliminate this instability, and mixed strategy equilibria play an important role in the analysis of that example.

[20] The standard definition of a sequential equilibrium requires a finite state space, so imagine that the state space is any finite subset of $[0, 1]$. If the followers believe that the leader's unexpected participation conveys no information about the value of x, then, as noted in Section 3, each follower's unique optimal response must be N. Since the leader knows that he would participate alone, his optimal action is N regardless of the value of x.

[21] A previous version of the paper (Komai and Stegeman (2004b)) allows followers to participate even if the leader does not. The main impact of that change is to introduce the possibility of counterintuitive equilibria in which the followers always reject the leader by taking the opposite action. In such equilibria, the only reason for the leader to participate is to prevent the others from participating in very bad projects. The earlier paper rules out such equilibria by imposing a bound on how negative non-participants' payoffs can be, thereby bounding the leader's incentive to do this.

exist. The reason is that, in any productive equilibrium, the leader's participation causes the followers to infer that $x > 0$, but N may be dominant for the followers given that event. In that case, the leader is not *credible,* and all equilibria are unproductive. (It is common to use the word "credible" to describe a threat that is believable, but here "credible" describes a player who can send a signal which is informative in equilibrium.) Theorem 3 summarizes these observations and states the credibility condition.

Theorem 3 *There exists a productive equilibrium if and only if (4) holds. In that case, the unique productive equilibrium is: the leader adopts the threshold strategy* $t = 0$ *and every follower mimics the leader.*[22]

$$E_x[x - \pi(N, \frac{m-1}{m}; x) \mid x > 0] \geq 0 \tag{4}$$

Example (continued): For the payoff function of the Example in Section 3, $E_x[x - \pi(N, \frac{m-1}{m}; x) \mid x > 0] = \frac{1}{18}$. Therefore, (4) is satisfied and the leader-follower game has a unique productive equilibrium, in which everyone participates if $x > 0$ and no one participates if $x < 0$. In that equilibrium, each player's expected payoff is $\frac{1}{4}$. (This exceeds, as expected, the .21 payoff under complete information.)

A different way to achieve essentially the same result, in principle, would be to ask an outside disinterested person to act as the leader, without compensation. The distinterested person would get exclusive access to the value of x and then simply announce (without cost) whether or not everyone else should participate. This setup has the advantage of circumventing the credibility condition: if players are unwilling to participate given the information that $x > 0$, then a disinterested leader could elicit some participation by instead adopting a threshold $t > 0$, and she would have no incentive to deviate by participating for $x \in (0, t)$. On the other hand, an unpaid and disinterested leader introduces the significant practical problem that she might simply ignore the value of x and make an inaccurate announcement. For this reason, it is implausible that a real organization would rely upon such a leader, even if it were possible to find such a competent and truly (and verifiably) disinterested party.

It would be more natural for an organization to address a credibility problem by constructing special incentives for its leader, whether chosen from inside or outside the organization, but this solution goes beyond the intended scope of this paper. The

[22] The definition of a threshold strategy arbitrarily requires that an indifferent leader not participate. The statement of Theorem 3 thus ignores, for simplicity, the alternative but essentially equivalent equilibrium resulting from modifying the leader's strategy to require participation when $x = 0$. A similar caveat applies to Theorem 8.

purpose here is simply to show what can be done with an existing set of symmetrically interested players (with incentives described by π), abstracting from issues of contract design.

Inadequate information and communication are often viewed as sources of coordination failure. In contrast, the leader-follower equilibrium of Theorem 3 solves a coordination problem by *limiting* information, in two ways. First, it takes detailed information about the state *away* from the followers; this improves coordination by eliminating the situations in which the followers perceive participation to be barely sustainable, which in turn can lead to an extensive unraveling of cooperation (as seen in the analysis of the Example in Section 3). Taking information away from the followers is similar to taking away choices, and shrinking their strategy set makes it easier for them to coordinate.

Second, having removed the followers' detailed information about the state, the leader-follower model gives them just one binary datum about other players' actions: does the leader participate or not? This observation, which is simple and presumably inexpensive (if one were to model the costs of different information structures), provides just enough information to allow coordination. In summary, our solution to the *coordination problem* is unusual in requiring minimal information and communication.

Separately, the leader-follower game mitigates the *shirking problem* because the leader's exclusive access to x allows her to "blur" the followers' information and so "confuse" them into participating for values of x at which they would be unwilling to participate if they were fully informed. The rational followers are of course aware of their "confusion" but cannot fix it.[23]

The critical role played by the credibility condition (4) means that it deserves careful interpretation. The leader is more likely to be credible (i.e. (4) is more likely to be satisfied) when the benefits of shirking, as measured by $\pi(N, \frac{m-1}{m}; x)$, are relatively small. Smaller gains from shirking increase followers' willingness to participate, which helps the leader's credibility through no virtue of her own. The "credibility" of a leader thus depends not only upon the leader's qualities but also upon the followers' preferences and the circumstances in which they are all embedded.

Because $\pi(N, \frac{m-1}{m}; x) = (x + \beta)\frac{m-1}{m}\alpha$, credibility is associated with low values of β, α, and m. This is expected because all of these circumstances – small costs of

[23] The model of Section 5, with its endogenous information structure, allows the followers to pay a cost to learn x. The leader-follower equilibrium can survive this innovation, because that cost may exceed a follower's benefit from learning x.

participation, small payoffs for nonparticipants, and a small population – reduce the incentive to shirk.

The implication that leaders of large organizations (i.e., large m) are less likely to be credible may however be a previously unappreciated diseconomy of scale. A simple intuition is that the leader of a large organization acquires so much leverage that the discrepancy between her incentives and the incentives of the individual follower becomes too great to sustain credibility. The leader's increasing leverage, as the organization grows, compensates for the leader's otherwise growing incentive to shirk, but at some point the leader becomes so pivotal that credibility breaks down and the productive equilibrium disappears. The idea that large organizations are more likely to suffer from this particular kind of management failure seems to be new.

A different model would allow the leader to send costless signals (talk) *in addition* to the single costly signal of participation. In general, such talk enriches the signal space and in the present context could include claims about the specific value of x. Crawford and Sobel (1982) show that such talk can be partially informative in equilibrium. Such equilibria tend however to be complicated, and it may be difficult to coordinate on them in realistic settings. Moreover, our payoff structure violates CS's assumptions in an important way. Here, the "sender" usually wants the "receiver" to make the maximum possible effort, which makes it hard to prevent the leader from defecting to whichever signal induces that maximum effort.[24]

4.2 First-best efficiency of the equilibrium.

It is intuitive that the equilibrium of Theorem 3 produces the first-best outcome, because it maximizes participation in good projects and minimizes participation in bad projects. Stating this claim formally requires a welfare measure. Let

$$W(q;x) = q\pi(P,q;x) + (1-q)\pi(N,q;x) \tag{5}$$

denote surplus per capita.

Theorem 4 *If (4) holds, then the unique productive equilibrium of Theorem 3 achieves the first best (i.e., maximizes $W(q;x)$ over q, given any x).*

Theorem 4 shows that leadership, created through access to superior information, not only mitigates shirking and coordination problems but (in this simple model) solves

[24] The possibility of costless talk differs from the possibility of directly revealing x, discussed in fn. 16.

them completely. Intuitively, the leader-follower equilibrium allows a representative leader to act on behalf of everyone. She participates only when participation maximizes total surplus and consequently achieves the unconstrained first best.

Using a single leader to attain the first-best outcome is subject to two important limitations. First, a small change in exogenous parameters can cause the credibility condition (4) to fail discontinuously, forcing an equilibrium with no participation at all.

Second, even if the efficient leader-follower equilibrium exists, the players must coordinate on that equilibrium rather than on the trivial no-participation equilibrium. One informal argument in favor of the leader-follower equilibrium invokes the idea of forward induction. Why would an organization create the institutions that select a leader, reveal x, and give the leader access to x, if its members intended to play to an equilibrium in which the leader never acts and the information is never used?[25]

5 Endogenous Information Acquisition

The game considered in Section 4 has two stages: the designated leader acts in stage 1 and the followers act in stage 2. This section expands this game to a 3-stage game, in which the information structure is endogenous. Both the symmetric equilibria and the leader-follower equilibrium can reappear in this model, and one goal of this section is to compare the circumstances that support these various equilibria.

One feature of the 3-stage model, which seems realistic but is unusual in similar formal models, is that no player ever observes directly which players are informed.

5.1 The three-stage model

In stage one of the new game, each player $i \in I = \{1, 2, .., m\}$ decides privately whether to invest in acquiring the information which the leader receives automatically in the 2-stage game. The cost of this *research* is $\kappa > 0$, where κ is an exogenous parameter. Each player who does research becomes *informed*, meaning that he learns the value of x and acquires the option to participate in the project; the interpretation is that he learns enough about the project to know what participation would require. In stage 2, each informed player simultaneously chooses whether to participate in the project, and everyone observes those participation decisions.

[25] If information acquisition is private and decentralized, as is assumed in the next section, then the forward induction argument is less convincing, and coordinating on a leader may be harder. Even in that case, though, any leader-follower equilibrium Pareto dominates the no-participation equilibrium, and this creates a reason to believe that such coordination may occur.

In stage 3, any informed player who did not participate in stage 2 has another chance to participate (but cannot undo a previous decision to participate). Also in stage 3, any uninformed player i can participate if anyone else participated in stage 2; the interpretation is that i has learned what participation requires by observing another's participation, though i still does not know the value of x.

If exactly one player becomes informed in stage 1, then the game in stages 2 and 3 is almost equivalent to the original 2-stage game. The main difference is that, unlike in the 2-stage game, where everyone knows which player is informed, here a player does not observe which other players are informed; he can infer this only from their subsequent participation decisions and (his beliefs about) their strategies. It will be convenient however to say that if a player i participates in stage 2, then other players "observe" in stage 3 that i became informed in stage 1, because the structure of the game implies that this must be true.

Each player's formal strategy must now specify his plan for action in each of the three stages. In stage 1, each player i chooses $s_i^1 \in \{R, U\}$, "R" for research or "U" to remain uninformed. For stage 2, we require for simplicity that player i adopt a threshold strategy $t_i^2 \in [-1, 1]$, meaning that if he played R in stage 1 then in stage 2 he plays N if $x \leq t_i^2$ or P if $x > t_i^2$. In stage 3, each player i observes a^2, the vector of actions in stage 2, $a^2 \in \{N, P\}^m$. If $a_i^2 = P$ then he has no more decisions to make; let A_i^2 denote the subset of $\{N, P\}^m$ such that $a_i^2 = N$. If $s_i^1 = R$, then we again require that player i adopt a threshold $t_i^3 \in [-1, 1]$, but now that threshold can depend on other players' stage 2 actions. Player i's stage 3 strategy thus has two parts, $t_i^3 : A_i^2 \to [-1, 1]$ for the case that $s_i^1 = R$ and $s_i^3 : A_i^2 \to \{N, P\}$ for the case that $s_i^1 = U$, subject to the constraint $s_i^3((N, N, ...N)) = N$.

Summarizing, a complete (pure) strategy for player i is a 4-tuple, $s_i = (s_i^1, t_i^2, s_i^3, t_i^3)$, with $s_i^1 \in \{R, U\}$, $t_i^2 \in [-1, 1]$, $s_i^3 : A_i^2 \to \{N, P\}$, and $t_i^3 : A_i^2 \to [-1, 1]$, subject to $s_i^3((N, N, ...N)) = N$. A mixed strategy generalizes these possibilities by allowing a player to map his information at each stage to a probability rather than to an action. It is unnecessary to develop a formal notation to represent mixed strategies, because mixing enters the analysis in a very simple way and never in the equilibrium strategies.

Each player's payoff is as before, the function (1) of all players' participation decisions and x, except that the number of persons participating now equals the number who choose P at any stage in the game, and a player who chooses R in stage 1 has an exogenous research cost $\kappa > 0$ subtracted from whatever payoff he would otherwise receive.

The next step is to define an appropriate refinement of the Bayesian Nash equilibrium, for the sequential environment. Then Theorems 5 and 6 will show that both the leader-follower equilibrium of Section 4 and the symmetric equilibrium of Section 3 can reappear in the 3-stage model, depending on the value of the research cost κ. Because the symmetric equilibrium requires everyone to do research, the efficiency advantage of the leader-follower equilibrium can be even greater in the 3-stage model than it was in the 2-stage model. Indeed, for a wide range of parameter values, the symmetric equilibrium is so costly that it (fortunately) disappears.

5.2 Definition of discrete sequential equilibrium

The Bayesian Nash equilibrium is no longer an adequate solution concept, because equilibria analogous to the equilibria studied in Sections 3 and 4 leave many information sets off the equilibrium path. For example, the Bayesian Nash equilibrium does not account for the incentives facing a player who observes another player unexpectedly defect to participation in stage 2. Given the absence of subgames, a natural way to refine the Bayesian Nash equilibrium is to require that it be sequential in the sense of Kreps and Wilson (1982), but standard definitions of a sequential equilibrium require a finite state space. A continuous state space typically introduces the problem that many information sets on the equilibrium path occur with zero probability ex ante, and it is not clear in general how to update beliefs given events of probability zero.

To extend the idea of sequential equilibrium to the current setting, define an *assessment* μ to be, as for Kreps and Wilson, a specification, at each information set, of a probability measure over the nodes in that information set.[26] To address the problem of the continuous state space, define an *interval partition* of the state space $[-1, 1]$ to be a partition comprising only intervals of positive length. For any such partition E, the pure strategy profile $s = ((s_i^1, t_i^2, s_i^3, t_i^3); i \in I)$ is said to be *E-discrete* if, for any $e \in E$ and $x, x' \in e$, the implied sequence of all players' actions is the same in states x and x'. The idea is that players' actions under s depend only upon the realization of E and not upon the realization of any finer partition of the state space. For any pure strategy profile s, it is clearly possible to construct a partition E such that s is E-discrete. (The boundaries separating the elements of E should include every value of x that any player uses as a decision threshold at any stage of the game.) The

[26] This is clearly possible, because we have already defined a field over the state space, and the set of possible actions is finite.

requirement that E comprise only intervals of states is natural and convenient but not essential.

A strategy profile is *fully mixed* if each player assigns positive probability to each of his possible actions, at each of his information sets. (Recall that, at any given information set, each player has only two possible actions.) A mixed strategy profile is E-discrete if, for any $e \in E$ and x,$x' \in e$, the implied probability of any given sequence of all players' actions is the same in states x and x'.

The crucial fact, which allows extension of the concept of sequential equilibrium to a continuous state space, is that every E-discrete and fully mixed strategy profile generates well-defined posterior beliefs at every information set. To see this, first consider information sets at which the acting player i has observed the true state x. In most cases, i has not observed all previous actions, but, given x, the fully mixed strategy profile implies a strictly positive prior probability density over all possible action sequences to that point, and player i can use his observations (if any) of past actions to compute the posterior probability distribution over whichever past actions he has not observed. At any other information set, the acting player i has observed nothing about x directly. Given each event $e' \in E$, such a player i can compute a posterior probability over previous unobserved actions, just as he would if he had observed a specific $x \in e'$, and his joint posterior over actions and states, given e', is the product of that posterior with the uniform distribution over the states in e'. Since each $e \in E$ occurs with positive probability, these calculations lead in the obvious way to a well-defined posterior distribution over all combinations of action profiles and states represented by nodes in the information set. (The appendix provides an explicit formula (21) for calculating this posterior distribution, at any information set.) By ensuring that the fully-mixed strategy profile s induces well-defined posterior beliefs at every information set, the requirement that s be E-discrete solves the problem introduced by the continuous state space.

Using the idea of E-discrete strategies to solve the problem of undefined posterior beliefs allows a straightforward extension of the concept of sequential equilibrium to the current problem. Specifically, an interval partition E of the state space, a strategy profile s^*, and an assessment (i.e. belief structure) μ^* form a *discrete sequential equilibrium* if there exists a sequence of fully mixed strategy profiles $s(k)$ and a sequence of assessments $\mu(k)$, $k = 1, 2, ...$, such that:

(i) $s(k)$ is E-discrete for all k;

(ii) $s(k)$ induces the posterior assessment $\mu(k)$, for all k;

(iii) $(s(k), \mu(k)) \rightarrow (s^*, \mu^*)$;[27]

(iv) s^* is sequentially rational for every player given the belief structure μ^* (in exactly the sense described by Kreps and Wilson).

This equilibrium concept extends in a natural way to many extensive games which are finite except for a continuous state space.[28]

5.3 The leader-follower equilibrium

This section describes an equilibrium of the 3-stage model, which replicates, in stages 2 and 3, the leader-follower equilibrium of the 2-stage model. This shows that the conclusions of Section 4 survive the extension to an endogenous information structure. The equilibrium requires that the cost of acquiring information be neither too high nor too low; these bounds are discussed below.

Theorem 5 states the equilibrium precisely, but this paragraph and the next describe the equilibrium strategies informally. Without loss of generality, assume that player $i = 1$ takes the role of leader. Her equilibrium strategy is to do research in stage 1 and then to participate in stage 2 according to whether $x > 0$. If she does not participate in stage 2, then in stage 3 she assumes that the only other participants will be whoever participated in stage 2 and optimizes accordingly. On the equilibrium path, the leader never participates in stage 3, because if $x > 0$ then she has already participated but if $x < 0$ then participation is suboptimal regardless of others' participation. If the leader forgets to do research in stage 1, then she has no decision in stage 2 and does not participate in stage 3 regardless of others' actions; that is optimal because the leader's equilibrium assessment draws no inferences about x from any other player's defection to participation in stage 2, and (as observed in Section 3) a player who knows nothing about the value of x never wants to participate.

For any follower $i > 1$, the equilibrium strategy is to do no research in stage 1, which leaves player i no decisions in stage 2, and in stage 3 to mimic the leader's stage 2 decision. If follower i accidentally does research in stage 1, then he does not participate in stage 2, because he expects to influence no one through his action and retains the option to participate in stage 3; in stage 3 he optimizes his participation

[27] The arrow indicates convergence in distribution.

[28] A limitation is that every information set in the game should be the cross product of a set of action profiles with a set of states (chosen by nature); if this is not true, then E-discreteness is insufficient to ensure that posterior beliefs are well-defined. The information sets of almost all commonly studied economic games have this product structure.

decision given x, assuming that everyone who has not yet participated will mimic the leader in stage 3.

The formal statement of the equilibrium strategies uses the following notation: for $a^2 \in \{N, P\}^m$, let $\#a^2$ denote the number of components of a^2 which are equal to P, and for $n \in \{0, 1, 2, ...m-1\}$ let $\chi(n) > 0$ denote the value of x such that, given x, a player is indifferent to participating if exactly n other players participate. It is immediate from (1) that

$$\chi(n) = \left(\frac{m}{n + 1 - \alpha n} - 1 \right) \beta$$

Note that $\chi(\cdot)$ is a strictly decreasing function, with $\chi(0) > 1$ and $\chi(m-1) = \tau^C > 0$.[29]

Theorem 5 *If* $\kappa \in (\underline{\kappa}, \overline{\kappa}) \equiv \left(\frac{\left[\left(\frac{m-1}{m}\right)\alpha\beta\right]^2}{4\left[1-\left(\frac{m-1}{m}\right)\alpha\right]}, \frac{1}{4} \right)$ *(an interval which is always nonempty), and the credibility condition (4) holds, then the following strategies, event partition* E*, and assessment* μ *constitute a discrete sequential equilibrium.*

(a) Strategies s *(† denotes strategy components which are never used on the equilibrium path):*

Leader ($i = 1$)	*Followers* ($i > 1$)
$s_1^1 = R$	$s_i^1 = U$
$t_1^2 = 0$ *(participate only in good projects)*	$t_i^2 = 1$ (†)
$s_1^3(\cdot) = N$ (†)	$s_i^3(\cdot) = a_1^2$ *(follow the leader)*
$t_1^3(a^2) = \chi(\#a^2)$ *(always implies N, on eq. path)*	$t_1^3(a^2) = \chi(\#a^2)$ *if* $a_1^2 = N$ (†)
	$t_1^3(a^2) = \tau^C$ *if* $a_1^2 = P$ (†)

(b) Event partition E*: the set of intervals of* $[-1, 1]$ *induced by the boundary points* $\{0, \tau^C, \chi(m-2), \chi(m-3), ..., \chi(1)\} \cap (-1, 1)$*, where each boundary point belongs to the lower interval.*

(c) Assessment (i.e. beliefs) μ*: In stages 2 and 3, every player* i*, who has not observed another player* j*'s stage 1 action, believes that* j *followed his stage 1 strategy, except: a player* $i > 1$*, who has observed* $x > 0$ *and player* 1*'s play of* N *in stage 2, believes in stage 3 that player* 1 *is informed with probability* $\frac{1}{2}$*.*[30] *A player* i *who has not (yet) observed* x *believes that* x *is distributed uniformly on* $[-1, 1]$*, except: a player* $i > 1$ *believes in stage 3 that* x *is distributed uniformly on* $[0, 1]$ *if player* $i = 1$ *chose* P *in stage 2, or uniformly on* $[-1, 0]$ *if player* $i = 1$ *chose* N *in stage 2.*

---

[29] Unlike the other theorems in this paper, Theorems 5 and 6 do not extend directly to the generalized state space and payoff function described in the appendix. The parts that do not extend are the calculations of the χ and κ thresholds. Theorems $5'$ and $6'$, in the appendix, generalize the calculation of those thresholds.

[30] This probability does not affect any follower's decision but is included because a sequential equilibrium requires a complete set of beliefs. The probability is also arbitrary; it is straightforward to construct perturbed strategies to support any probability between zero and one.

For the 2-stage model, Theorem 3 also states that the leader-follower equilibrium is unique, but endogenous selection of the leader precludes uniqueness here, because any player can take that role.

To illustrate the 3-stage leader-follower equilbrium, reconsider the earlier Example.

Example (continued): Given the payoff function of the Example in Section 3, Theorem 5 shows that the leader-follower equilibrium exists if $\kappa \in \left(\frac{1}{45}, \frac{1}{4}\right)$. In that equilibrium, the leader's expected payoff is $\frac{1}{4} - \kappa$ and each follower's expected payoff is $\frac{1}{4}$.

If $\kappa > \frac{1}{4}$, then research is so expensive that the leader should (obviously) deviate away from research. In that case, the only apparent equilibria are those in which everyone remains uninformed and no one participates. If $\kappa < \frac{1}{45}$, then any follower should defect to research in stage 1, because this allows profitable shirking in stage 3 if he learns that $x \in (0, 0.4)$. (This is the kind of shirking that reduces cooperation in the complete information game of Section 3.) In summary, existence of the leader-follower equilibrium requires that research costs be small enough to justify the leader's research but large enough to discourage research by followers.

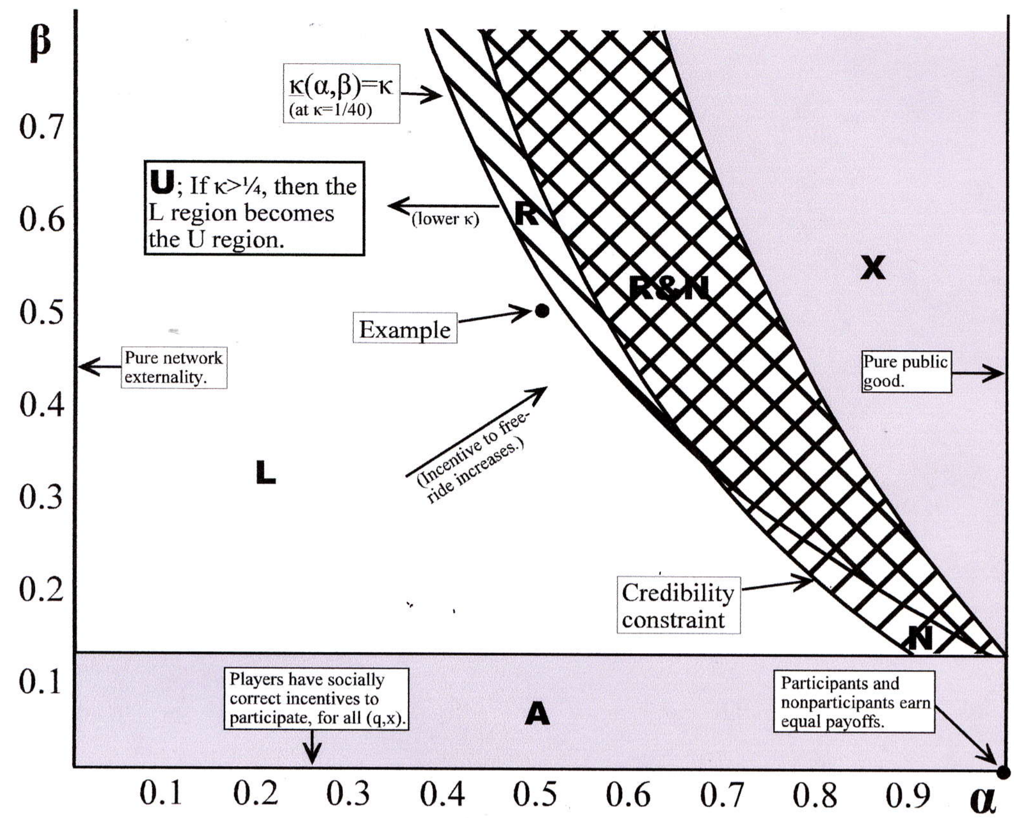
Instead of fixing (α, β), Figure 1 fixes $m = 9$ and $\kappa = \frac{1}{40}$ and shows that many values of (α, β) support the leader-follower equilibrium. Because $\frac{1}{40} \in \left(\frac{1}{45}, \frac{1}{4}\right)$, the parameter values $(\alpha, \beta) = (\frac{1}{2}, \frac{1}{2})$, from the Example, lie within the equilibrium region. As κ increases from $\kappa = \frac{1}{40}$, the equilibrium region becomes larger as the followers' incentive to defect to research declines (i.e., as region R shrinks), but only slightly larger, because the condition which prevents such defections gradually becomes weaker than the credibility constraint and no longer binds for low values of β (i.e., region R becomes a subset of region N).

5.4 The symmetric equilibrium

Another possible equilibrium is the symmetric equilibrium of Section 3, which can reappear in the 3-stage model if everyone does research in stage 1, takes no action in stage 2, and participates in stage 3 if x exceeds the common participation threshold. For the original model, Theorem 1 describes a continuum of symmetric equilibria, and the 3-stage model supports a similar continuum, but for simplicity this section focuses only on the most efficient symmetric equilibrium, which has a threshold equal to τ^C.

One reason for ignoring the less efficient symmetric equilibria, which have participation thresholds $t > \tau^C$, is that the possibility of sequential action destroys some of those equilibria. For instance, suppose that $m = 7$ and $\alpha = \beta = \frac{1}{4}$. Then $\tau^C = \frac{3}{44}$, but

Figure 1; Existence of the leader-follower equilibrium (m=9, κ=1/40).



Explanation of regions in (α,β)-space

A: This region was ruled out from the start (assumption (2)), because players sometimes have an incentive to act alone.

X: This region was ruled out from the start (assumption (3)), because players are never willing to participate.

N: The leader-follower equilibrium does not exist, because followers are unwilling to follow the leader (i.e., the credibility condition (4) fails).

R: The leader-follower equilibrium does not exist, because research is so cheap that followers will do research to learn when they should follow the leader. This region obviously expands as the research cost κ declines.

U: If κ>1/4, then the leader-follower equilibrium does not exist regardless of the value of (α,β), because the leader is unwilling to do research.

L: The leader-follower equilibrium exists.

Example: The point (α,β)=(½,½) represents the Example employed throughout the text. If κ=1/40, as shown, then the Example is within the equilibrium region.

the symmetric equilibria with thresholds $t > \frac{3}{4}$ are not discrete sequential in the 3-stage game, because if $x \in (\frac{3}{4}, t)$ then any player has an incentive to defect to participation in stage 2, because his visible commitment makes it worthwhile for any other player to become the second participant, in stage 3. If the game has even more stages, then the increasing possibilities for sequential commitment destroy even more of the inefficient (i.e. high threshold) symmetric equilibria.[31] For this reason, it seems especially appropriate in the multi-stage model to focus only on the symmetric equilibria which are most efficient.

The next theorem shows that the symmetric equilibrium, with all players informed and participation threshold $\tau^C \left(= \frac{(m-1)\alpha\beta}{m-(m-1)\alpha}\right)$, does reappear in the 3-stage model if κ is small enough to sustain each player's incentive to do research in stage 1.

Theorem 6 *If $\kappa < \kappa^* \equiv \left[x \frac{mx-(m-1)(x+2\beta)\alpha}{4m}\right]_{x=\tau^C}^{x=1}$, then the following strategies s, event partition E, and assessment μ constitute a discrete sequential equilibrium.*

Strategies s: Every player does research in stage 1, nothing in stage 2, and participates in stage 3 iff he has observed $x > \tau^C$. Formally: $s_1^1 = R$; $t_i^2 = 1$; $t_1^3(a^2) = \tau^C$ and $s_1^3(\cdot) = N$ for all a^2.

Event partition: $E = \{[-1, \tau^C], (\tau^C, 1]\}$.

Assessment μ: In stages 2 and 3, every player believes that every other player is informed. Every player who has not (yet) observed x believes that x is distributed uniformly on $[-1, 1]$.

Example (continued): Continuing the Example, Theorem 6 shows that the (most efficient) symmetric equilibrium, with participation threshold $\tau^C = 0.4$, reappears in the 3-stage model if $\kappa < .05$. In that equilibrium, each player's expected payoff is $.21 - \kappa$. If $\kappa \in (.05, .21)$, then each player earns a positive payoff in the symmetric equilibrium but could earn even more by defecting away from research and free-riding on others' participation.

Summarizing the results for the Example, the leader-follower equilibrium reappears in the 3-stage model if $\kappa \in \left(\frac{1}{45}, \frac{1}{4}\right)$, and the best symmetric equilibrium (i.e., with $t = \tau^C$) reappears if $\kappa < \frac{1}{20}$. Therefore, if $\kappa \in \left(\frac{1}{20}, \frac{1}{4}\right)$, then the 3-stage model supports the leader-follower equilibrium but not the symmetric equilibrium. In this range and in this sense, endogenous acquisition of information forces the emergence of a leader, if the players can coordinate on who plays that role. The intuition is that,

[31] This is similar to the point made in fn. 13, for a multi-stage game with complete information.

for $\kappa \in \left(\frac{1}{20}, \frac{1}{4}\right)$, research is cheap enough to support its acquisition by one player for the benefit of all but too expensive to support its acquisition by all players.

Summarizing the general results, Section 4 showed that the leader-follower equilibrium is more efficient than the symmetric equilibrium before accounting for research costs, and Theorem 6 shows that making research costly at the individual level reinforces that conclusion and can make the symmetric equilibrium so inefficient that it becomes unsustainable.

Figure 2 is similar to Figure 1, except that the higher research cost ($\kappa = \frac{1}{5}$ instead of $\kappa = \frac{1}{40}$) causes region R to disappear inside region N; the threat of follower research is no longer a binding constraint for the leader-follower equilibrium. Figure 2 also shows the region of the parameter space ($S\&L$) which supports the symmetric equilibrium of Theorem 6 as well as the leader-follower equilibrium of Theorem 5; this region is small because it is hard to support the symmetric equilibrium given such a large κ. In this sense, the 3-stage model "forces" the emergence of a leader for the many values of (α, β) which lie in the shaded L region.

5.5 First-best efficiency of the leader-follower equilibrium

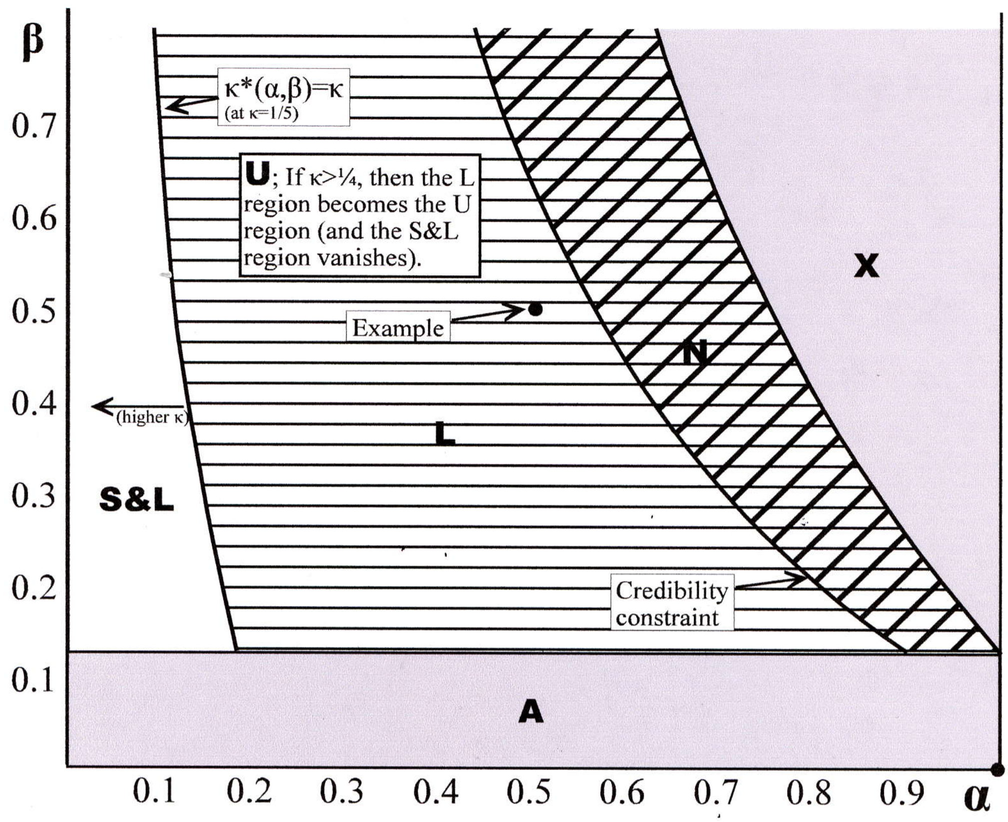
In the 2-stage model, the leader-follower equilibrium of Theorem 3 attains the first-best if it exists, meaning that it is impossible to guide players' actions to a better result. In the Example, the first-best surplus is $\frac{9}{4}$, comprising $\frac{1}{4}$ for each player (cf. Section 4). For the 3-stage version of the leader-follower equilibrium, the only difference in realized surplus is that now the leader must pay $\kappa \in (0, \frac{1}{4})$ to learn x; therefore, the expected surplus is $\frac{9}{4} - \kappa > 0$. If we understand the "first-best" to be subject to the constraint that no player can act contingently on information that no player has, then this is obviously the highest attainable surplus in the 3-stage model, because if no one pays to become informed then no one knows anything about the value of x and so no one is willing to participate, implying that surplus equals zero.

This reasoning generalizes beyond the Example: it is still true, in the 3-stage model, that the leader-follower equilibrium, if it exists, attains the first best.

If however κ is too small to support the leader-follower equilibrium, given fixed (α, β), then it is easy to show that the surplus attainable in any equilibrium falls short of the first-best by at least κ (the cost of informing a second player).[32] This implies that

[32] The first-best always requires exactly one player to learn x and every player to participate iff $x > 0$ (ignoring the ambiguity at $x = 0$). Theorem 3 implies that the leader-follower equilibrium is the only possible equilibrium of the 3-stage game in which exactly one player pays to learn x, so if κ falls outside of the range which supports that equilibrium, then the 3-stage game has no equilibrium in which exactly

Figure 2; Forced emergence of a leader (m=9, κ=1/5).



<u>Explanation of regions in (α,β)-space</u>

A,X: These regions were ruled out from the start (cf. Figure 1).

N: The leader-follower equilibrium does not exist, because followers are unwilling to follow the leader (i.e., the credibility condition (4) fails).

R: This region, which appeared in Figure 1, does not appear here, because at κ=1/5 research is so expensive that the requirement that followers do not deviate to research is not binding (i.e, R is a subset of region N).

U: If κ>1/4, then the leader-follower equilibrium does not exist regardless of the value of (α,β), because the leader is unwilling to do research.

L: The leader-follower equilibrium exists, but the symmetric equilibrium does not exist.

S&L: The symmetric and leader-follower equilibria both exist. The symmetric equilibrium exists because research is cheap enough to be worthwhile for everyone. This region obviously shrinks as the research cost κ rises, and it vanishes at κ=1/4.

Example: The point (α,β)=(½,½) represents the Example employed throughout the text. If κ=1/5, as shown, then the Example is well within the equilibrium region.

increasing the real cost of information can increase welfare, because a small increase in κ may allow a discrete efficiency gain through the emergence of the leader-follower equilibrium.

6 One Leader of a Heterogeneous Population

If the credibility condition fails, then the leader-follower equilibrium disappears in dramatic fashion: a leader who is not credible cannot persuade anyone to participate and never participates herself. (In Figures 1 and 2, credibility fails in region N.) If however players are heterogeneous, a possibility not yet considered, then the richer set of potential leaders improves the chance of finding one who is credible, though she might not achieve the first best. The choice of leader then becomes a substantial question.

Suppose, specifically, that players vary according to their costs of participation. Instead of a common participation cost β, assume that each player i's participation cost is $\beta + \Delta_i$, where

$$\Delta_i \equiv \left(i - \frac{m+1}{2} \right) \delta \tag{6}$$

for some fixed increment $\delta > 0$. Then player $i = 1$ has the lowest participation cost, $\beta + \Delta_1 = \beta - \frac{m-1}{2}\delta$, and player m has the highest participation cost, $\beta + \Delta_m = \beta + \frac{m-1}{2}\delta$. The mean participation cost is still β, but individuals' costs are dispersed symmetrically around β. Heterogeneous costs mean that the player-specific payoff functions π_i replace the original payoff function π:

$$\begin{aligned} \pi_i(P, q; x) &\equiv \pi(P, q; x) - \Delta_i = (x + \beta)q - \beta - \Delta_i \\ \pi_i(N, q; x) &\equiv \pi(N, q; x) = (x + \beta)q\alpha \end{aligned} \tag{7}$$

To ensure that heterogeneous participation costs do not cause radical changes in the equilibria, all of the results in this section assume that $\delta > 0$ is sufficiently small. For example, the results assume that *each* player's payoff function satisfies conditions analogous to (2) and (3). Assumption B in the Appendix describes a specific upper bound on δ, which is sufficient to ensure that all of the following results obtain.

---

one player learns x. Any equilibrium in which more than one player learns x cannot improve upon the first-best participation decisions and so must fall short of first-best surplus by at least κ. Any equilibrium in which no player learns x has no participation and thus zero payoffs, but the 3-stage leader-follower equilibrium must generate payoffs of at least κ for every player except the leader, whose utility must be non-negative; therefore, for any κ that supports the 3-stage leader-follower equilibrium or any smaller κ, players' actions can be guided to attain surplus of at least $(m-1)\kappa$, implying that the surplus from no participation falls short of first-best surplus by at least κ.

Participation costs have various interpretations. Player m could be the least productive or the busiest player, according to whether $\beta + \Delta_i$ represents a direct cost or an opportunity cost. The latter possibility may be more interesting, because the results will favor high-cost leaders, which in that case means leaders whose time is valuable and who are disinclined to waste it chasing low-return projects. If high costs instead reflect incompetence, then a recommendation to pick high-cost leaders conflicts with obvious offsetting considerations, which are outside the present model.

To simplify matters, this section returns to the 2-stage game of Section 4. This means that the leader is specified exogenously, and the main question is which player should fill that role. The addition of heterogeneous participation costs is the only difference between the present model and that of Section 4.

6.1 The unique productive equilibrium, for a given leader

Players' varying costs of participation imply that the credibility condition (4) now varies across players. Because it is awkward to work with a sequence of inequalities like (4), Lemma 7 describes a vector of numbers, $(\tau_i^F; i \in I)$, which will simplify the statement of the players' credibility conditions.

Lemma 7 *If $\delta > 0$ is sufficiently small given m, α, and β, then the following is true. For each $f \in I$, there exists a unique $\tau_f^F \in (-1, 1)$ such that $E_x[x - \Delta_f - \pi(N, \frac{m-1}{m}; x) \mid x > \tau_f^F] = 0$, and τ_f^F is strictly increasing in f.*

If player f learns that $x > t$ for some t, and he believes that every other player will participate, then the construction of τ_f^F implies that f should participate too if and only if $t \geq \tau_f^F$. Because different leaders will adopt different decision thresholds t, the credibility condition for follower f depends also upon who is the leader. If leader $l \in I$ assumes that every follower will mimic her, then action P gives her $x - \Delta_l$ and N gives her zero. Therefore, she will participate iff $x > \Delta_l$. The implied new credibility condition, which indicates whether follower f is willing to mimic leader l in a leader-follower equilibrium, is:

$$\Delta_l \geq \tau_f^F$$

(This is equivalent to (4) if $\Delta_f = \Delta_l = 0$.) If $\Delta_l < \tau_f^F$, then leader l participates for so many mediocre projects that follower f elects not to mimic l, even if she expects everyone else to do so.

Heterogeneous participation costs raise the possibility of equilibria in which some followers mimic the leader and others do not, but choosing $\delta > 0$ sufficently small suppresses such complicated outcomes by ensuring that all followers act identically. Then the followers' credibility conditions can be summarized in one condition:

$$\Delta_l \geq \tau^F_{\max(I \backslash \{l\})} \tag{8}$$

If this inequality holds for a given leader l, then the signal conveyed by leader l's participation is strong enough to support participation by everyone else, and l is said to be credible. In that case, any player $l' > l$ would also make a credible leader, because $\tau^F_{\max(I \backslash \{l\})}$ is weakly decreasing in l while Δ_l is strictly increasing. Leaders with higher participation costs Δ_l are more likely to be credible because followers know that, in equilibrium, they are more selective when deciding which projects are worthwhile.

Theorem 8 describes equilibria similar to those of Theorem 3. For any given leader, Theorem 8 shows that the uniqueness of the productive equilibrium survives small differences in participation costs.

Theorem 8 *If $\delta > 0$ is sufficiently small given m, α, and β, then the following is true. For any $l \in I$, there exists a productive equilibrium, with player l as leader, if and only if (8) holds. In that case, the unique such productive equilibrium is: l adopts the threshold strategy $t = \Delta_l$ and every follower mimics l.*[33]

6.2 The efficient choice of leader

To make efficiency claims, it is first necessary to revise the welfare measure, to account for players' heterogeneous participation costs. Given q, $W_\delta(q;x)$ computes the highest surplus per capita attainable if qm players participate, assuming that the participants are the low-cost players $1, 2, ...qm$.

$$\begin{aligned} W_\delta(q;x) &\equiv \frac{1}{m}\sum\nolimits_{i=1}^{qm} \pi_i(P,q;x) + (1-q)\pi(N,q;x) \\ &= W(q;x) + \frac{\delta q(1-q)m}{2} \end{aligned} \tag{9}$$

The efficiency results for heterogeneous players are straightforward and intuitive and stated only informally.[34] Assume throughout that $\delta > 0$ is sufficiently small, given m, α, and β. Also assume that the credibility condition (8) holds for some $l \in I$, implying that l would be a credible leader, and let $\widetilde{L} \equiv \{\widetilde{l}, \widetilde{l}+1, ...m\}$ denote the set of all such players.

[33] The definition of a threshold strategy arbitrarily requires that an indifferent leader not participate (cf. fn. 22).

[34] An earlier version of the paper provided precise statements and proofs of these results. These are available from the authors.

If $\widetilde{L}$ owns more than half of the players, then the optimal leader is the median player $i = \frac{m+1}{2}$ if m is odd or one of the two "median" players, $i \in \left\{\frac{m}{2}, \frac{m}{2}+1\right\}$, if m is even. If m is odd, then the median player is exactly representative and the associated leader-follower equilibrium achieves the first-best. (The proof is similar to the proof of Theorem 4.)

If $\widetilde{L}$ owns at most half of the players, then the welfare-maximizing leader is $\widetilde{l}$, the lowest-cost player among all of the potential credible leaders. Such a leader will not achieve the first best, because her participation cost is higher than average, implying that she will skip some projects which would increase total surplus if everyone participated. Nevertheless, she is the optimal leader subject to the constraint that the leader must be credible.

These results imply that *it is never optimal to appoint a leader who has unusually low costs of participation*, because it is better to appoint a leader with higher costs, who is both more representative and at least as credible. This conclusion however ignores (because the present model ignores) the possibility that a leader with unusually low costs has associated special talents. In a real situation, the choice of an optimal leader has many aspects, and the present results merely warn against the natural presumption that higher productivity is always good. A leader who is too productive cannot commit to showing restraint in project selection.

Example (continued). Let $\delta = .02$, so that the players' participation costs, from $i = 1$ to $i = m$, are $(.42, .44, .46, .48, .50, .52, .54, .56, .58)$. Assumptions (2) and (3) are still satisfied for each player.[35] Now $\Delta_l = .02l - .1$ for each potential leader $l \in I$, and $\tau_f^F = .072f - .56$ for each potential follower $f \in I$. For the median leader $l = 5$, $\Delta_5 = 0 < .088 = \tau^F_{\max(I \setminus \{l\})}$, violating the credibility condition (8). Only followers $f \in \{1, 2, 3, 4, 6, 7\}$ are willing to follow leader $l = 5$ (and then only if $f \in \{8, 9\}$ were also willing to follow her). Leaders $l \in \{6, 7, 8\}$ are similarly not credible. If however the highest-cost player $l = 9$ is appointed to lead, then $\Delta_l = .08 > .016 = \tau^F_{\max(I \setminus \{l\})}$, and she is credible. Promoting the highest-cost player to leadership has the double advantage of moving her from a position in which her high cost harms efficiency to

[35] For general payoff functions π, Assumption B in the Appendix describes conditions (more than) sufficient to ensure that δ is "sufficiently small" in the sense assumed by Theorem 8. Setting $x_0 = -.05$ and $\omega = .2$, the present example satisfies parts (i) and (ii) of Assumption B but not part (iii), because here $\frac{\omega}{2m} < \delta < \frac{\omega}{m}$ instead of $\delta < \frac{\omega}{2m}$. The marginal contribution of the latter inequality, for general payoff functions, is to rule out productive equilibria in which all followers $f < l$ play P and all followers $f > l$ play N, given some leader $l \in \{2, 3, ...m - 1\}$. It is easy to confirm that the present example has no such equilibria. Therefore, $\delta = .02$ is small enough to ensure that the example satisfies all the claims of Theorem 8. (Choosing $\delta = .01$ would satisfy all parts of Assumption B, but that choice is less interesting because in this particular example it implies that no leader would be credible.)

a position in which it can enhance efficiency; in other words, it helps both sides of inequality (8).

In equilibrium, the leader $l = 9$ participates in 92% of the good projects, which, with everyone following her, capture 98% of the potential surplus. This is much better performance than accrues under full information, which can support full participation only for $x > .544$ (even if one ignores the coordination problem); this represents only about 46% of the good projects.

7 Multiple leaders

If none of the players would make a credible leader, then it may be possible to establish credibility by appointing multiple leaders. This section demonstrates this possibility by example.

Consider the 2-stage model of Section 4, with homogeneous players. Assume that $\alpha = \frac{1}{2}$, $\beta = \frac{4}{5}$, and $m = 5$. The high participation cost β creates a strong incentive to shirk and take a free ride on others' participation. If everyone is informed, then the propensity to shirk appears in the high implied value for τ^C, $\tau^C = \frac{8}{15}$, which implies that most good projects are ignored. In the 2-stage leader-follower model, the credibility condition fails, implying that all projects are ignored.

One way to improve on these outcomes is to designate several leaders, who each observe x and act simultaneously, after which the remaining, uninformed, players decide simultaneously whether to participate. Adding leaders reduces each leader's leverage, because she no longer influences the decision of the other leader(s). Therefore, each leader becomes less willing to participate and more selective in her choice of projects, and this can make multiple leaders credible where a single leader would not be.

For the parameter values specified, assume that there are $n = 2$ leaders. Then there exists a Bayesian Nash equilibrium in which each leader finds it optimal to participate for $x > 4/45$, rather than for $x > 0$ as would be required in a one-leader equilibrium, and each follower participates if and only if both leaders participate. The slightly higher participation threshold is enough to make the leaders credible.[36]

[36] If both leaders participate, then each follower would rather participate than be the lone defector; the follower would be indifferent if the leaders used a threshold of $\frac{1}{15}$ instead of $\frac{4}{45}$. If either leader does not participate, then it is easy to confirm that no follower wants to be only the first or second participant, regardless of the value of x, and the same argument shows that neither leader wants to participate if $x \leq \frac{4}{45}$. If $x = \frac{4}{45}$, then each leader would be indifferent between full participation (if that were an option) and free-riding on the other leader, if the other leader participated alone; therefore each leader strictly prefers to participate given $x > \frac{4}{45}$. The coordination problem between the leaders means that there exists a continuum of leader-follower equilibria, similar to the continuum of equilibria

The leader-follower model and the complete information model may be viewed as the extreme cases of a range of models, which are distinguished by the number of informed players, $n \in \{1, 2, 3, 4, m\}$, who act before uninformed players. In the present example, one ($n = 1$) leader is not credible, but two ($n = 2$) leaders are credible, though the equilibrium is inefficient because projects $x \in (0, \frac{4}{45})$ are ignored. Under complete information, $n = m$, credibility is assured (i.e., not an issue) because there are no followers, but efficiency losses are much greater because projects $x \in (0, \frac{24}{45})$ are ignored. We prove no theorem, but these observations suggest that the most efficient number of leaders (n) is in general the smallest number that is credible. Only if $n = 1$ can such a leader-follower equilibrium achieve full efficiency, but equilibria with $n = 2, 3, ...m - 1$ credible leaders can still improve on the best complete information equilibrium at $n = m$.[37]

8 Related Literature

Hermalin (1998) emphasizes the importance of informal authority, which stems from superior information rather than from a formal position. He studies a team leader who, like ours, has private information about the return to effort and increases observable effort when the return is high. The leader's effort fully reveals his information, but the leader-follower equilibrium produces more efficient outcomes than does the equilibrium under full information, mainly because it improves the leader's incentives to work.

Komai, Stegeman, and Hermalin (2007, henceforth KSH) show that a leader-follower equilibrium can be more efficient if the leader's action is not fully revealing, through the same kind of pooling of states that relieves incentive problems in the present paper. KSH study a very simple game in which utility is linear in actions, which implies that, in any given state, players' optimal actions are independent of others' actions. By studying non-linear utility functions, this paper introduces a substantial issue of coordination failure and the separate and plausible problem that productive equilibria can disappear due to credibility failure.[38] Solving the shirking and coordination

---

in Section 3. If the two leaders act sequentially, as suggested by a referee, then that coordination problem is solved, much as in fn. 13, and the equilibrium described in the text is the essentially unique equilibrium with two leaders and followers. The referee's suggestion led to the example in this section.

[37] In Komai and Stegeman (2004), we study a somewhat similar implementation of multiple leaders. In that model, each follower observes only one leader, so that each leader has her own band of followers. This reduces the leader's leverage much more dramatically than the model described here, making it easier to establish credibility (but with commensurate efficiency losses).

[38] In KSH's model and notation, $Nc < 1$ is sufficient for the existence of a productive equilibrium. This condition is analogous to assumption (3) in the present model. KSH have no condition analogous to the credibility condition (4)

problems simultaneously, and all issues connected to credibility failure, are thus new with this paper. Relative to KSH, this paper also: establishes the uniqueness of the productive equilibrium; extends the analysis to heterogeneous players and the problem of choosing the optimal leader; shows that appointing several leaders can sometimes solve a credibility problem; and makes the information structure endogenous.

Andreoni (2004), building on a model of Vesterlund (2003), assumes that the leader is chosen through a war of attrition and decides whether to acquire costly information about the quality of a public good. Andreoni shows that the lowest-cost player (in his model, the richest person) becomes the leader, because he is the one who gains most from provision of the public good. This is quite different from our conclusion in favor of high-cost leaders. Andreoni's result differs mainly because his leader self-selects and pays a penalty for being the leader, assumptions motivated by the example of charitable fund-raising. In our model of heterogeneous players, the leader is selected exogenously to maximize efficiency, and he does not bear, or is implicitly compensated for, the cost of acquiring information.

Levy and Razin (2004) describe a game between two countries deciding how much to cooperate. The key feature is that a democracy must share with its rival all information possessed by its decision-maker, the public. This constraint allows democracies to cooperate where autocracies cannot. (In contrast, our leader-follower model enhances cooperation because an informed player uses information that she *cannot* share with her rivals.) Levy and Razin show that in some cases democracies cooperate better if decision-makers in both countries are completely uninformed.[39]

Daughety and Reinganum (2006, henceforth DR) describe a model of joint production, in which workers who act simultaneously use their observable but noncontractible effort to signal (to buyers) their private information about the quality of their work. As in Hermalin (1998), the effort invested in signaling improves efficiency, and DR show that the resulting equilibria often Pareto dominate equilibria under complete information. DR note that this is an instance of adverse selection mitigating moral hazard, and the same could be said of our model, except that here it is mainly players' uncertainty about the return to effort (rather than their desire to send signals to others) which generates the Pareto improvement, and the selection effect in our equilibrium is

[39] Levy and Razin's model creates an explicit role for a leader. Each of their democracies has a leader who must decide whether to reveal information, while the public makes the decision. The leader and the public have common interests; the leader essentially decides, through cheap talk, whether the public should make the decision in a (completely) informed or (completely) uninformed way, mindful that the rival will see the information also.

favorable rather than "adverse."

Crawford and Sobel (1982, CS) study a game in which an informed player sends costless messages to an uninformed player who then acts. If their interests differ, then CS show that equilibrium messages must be coarse enough to give the sender no incentive to defect, which implies that they are not fully informative, which implies that equilibrium actions are inefficient. In this paper, the participation message is coarse (specifically, binary) by assumption, which as in CS makes it less susceptible to manipulation, but in our case the loss of information, combined with the useful incentive effects of a *costly* signal, helps efficiency.

Austen-Smith (1994) adds to CS's model the possibility that the sender is uninformed, and only the sender observes whether the sender is informed. Austen-Smith also assumes, as in the present model, that signals are costly. In equilibrium, senders who receive no information or sufficiently bad information send no signal. Adding noise through the possibility of an uninformed sender means that sending no signal implicitly sends a signal which is less negative (holding the sender's strategy fixed), implying that the receiver's response to no signal is better for the sender than it would otherwise be, implying that the sender more often sends no signal, implying that there is more likely to exist an equilibrium in which the sender's decision to send a signal is informative. The possibility of an uninformed sender can thus improve equilibrium information and efficiency. In contrast, adding noise in the present model, by transforming a complete information setting to one in which only one player is informed and can send only a discrete signal, improves efficiency by strictly degrading the information available to the followers. Somewhat in the style of Austen-Smith, Blume, Board, and Kawamura (2007, BBK) add the possibility of noisy message transmission to CS's model and show that noise can improve efficiency. As in Austen-Smith's model, BBK's noise can improve efficiency by making the most negative message less negative.

Our model is also related to the idea of information cascades, but unlike the many studies emphasizing the inefficiencies caused by cascades, we use the leader-follower relationship to improve efficiency.

The management literature includes myriad studies of leadership, many of them empirical and most omitting the formal modeling familiar to economists. Such studies address what leaders do, how they do it, how they can do it better, how to adjust their environment so that they can do it better, and which personal attributes are important for leadership, in various settings.[40] Economists have made relatively few contributions to this literature, but Rotemberg and Saloner (1993), for example, present a model that compares the effectiveness of selfish and empathetic managers in different situations. The recent management literature (e.g., Case (1995)) tends to promote sharing information with employees, but this paper provides a reason to doubt that such policies are always wise.[41] Prendergast (1993) presents a quite different model with a related message. He shows that if managers rely on information provided by workers, then workers' incentive to conform means that it may be best to insulate them from managers' other sources of information.

Komai (2002) extends the present analysis to a continuous action set. Komai and Stegeman (2004a) study the possibility of solving credibility problems by dividing authority among several leaders (cf. fn. 36).

9 Summary and Concluding Remarks

Our model exhibits the familiar problems of moral hazard and coordination failure, but our remedy reverses the usual method of mechanism design. Instead of using contracts to align agents' incentives, given an exogenous information structure, we leave contracts in the background and redesign agents' information. Instead of trying to improve information, subject to monitoring and processing costs, we keep critical data away from decision-makers. By depriving agents of the fine information required for profitable defections, our low-cost information structure promotes cooperation as well as coordination.

This analysis seems realistic for some settings. For example, a knowledgeable person who makes a personal commitment to a political campaign rarely mentions the candidate's weaknesses; he typically expresses similar enthusiasm for all of his endorsements. The support that his stance attracts from other activists may be a public good that benefits all, though many might shirk – withhold their active support from a particular candidate Smith – if they knew that Smith was only marginally better than his opponent. The followers benefit from their own ignorance about Smith, because it inhibits such shirking.

Instead of considering where to assign *authority*, our model assigns it to no one.

[40] Dansereau and Yammarino (1998) survey some of these studies.

[41] Milgrom and Roberts (1992) survey general themes from both economists' formal models of organizations and the less formal management literature. They note the consensus view that the "key problem in achieving effective coordination and adaptation is that the information needed to determine the best use of resources... is not freely available to everyone."

Authority arises informally (as Hermalin (1998) proposes) from superior information. Informal authority can provide an unorthodox answer to the old question of the optimal degree of decentralization. Our leader-follower model is completely decentralized in the sense that each agent acts noncooperatively, given the symmetric payoff function π – yet authority is completely centralized in the sense that, in equilibrium, the leader effectively makes every decision.

Our simple model shows that it is possible to model organizations without invoking contracts, prices, residual authority, or bargaining. This minimalist theory of organizations can help to explain why leaders exist, but it yields a minimalist theory of leadership. Our leaders have no special skills or authority beyond receiving better information.

Our model is unusual in correcting a leader's incentives by magnifying her impact rather than her compensation. When credibility fails, it is because the leader has too much impact – she becomes too pivotal – relative to the other players. This is more likely in a large organization, and the problem of credibility failure may be an important and unappreciated diseconomy of scale. To restore cooperation in such cases, a leader must find a way to convince her followers that she will not (literally) mislead them.

We have shown that one way to restore the leader's credibility is to appoint a high-cost leader, because her desire to ration her effort leads to fewer initiatives that burden others, and this can increase the cooperation that she receives when she does promote a project. The discrete gain from restoring the leader's credibility dominates the loss from ignoring a few good projects. A high-cost leader might also shirk in unobservable aspects of project execution, a bad outcome outside of our model, but delegation of those responsibilities could mitigate this negative impact of a high-cost leader. Optimal design of the leadership role may (as in this model) emphasize simple demonstration effects rather than complex private decisions.

Many CEOs seem to fail because they initiate too many unsuccessful projects – new products, new acquisitions, new reorganizations. In some cases this may reflect inability to distinguish good from bad projects, a possibility which our model assumes away. In other cases the CEO may fail because of excess enthusiasm for change, reflecting insensitivity to the costs which change imposes on others in the organization. This is the credibility issue captured, in very simple form, in our model. The willingness not to promote change, or not to (too often) put one's own stamp on the organization, may be one attribute of successful leaders, but it rarely appears in the formal modeling of management.

We have shown by example that another way to establish credibility is to appoint multiple leaders. By making each leader less pivotal, the leaders can be jointly credible where a single leader would not be.

Another unusual feature of our model is that we allow (in Section 5) leaders to step forward and appoint themselves, by taking the initiative to acquire better information. In many cases, a distinction between leader(s) and followers *must* emerge, because it is not individually rational for every player to become informed.

To keep the focus on information (and reflecting our belief that utility may be less transferable than is often assumed), we have suppressed transfers and any differences among the payoff functions of leaders and followers. A more complete analysis would consider how various transfer schemes might affect our conclusions and, conversely, how the issues identified in this paper may affect the design of optimal contracts.

10 Appendix

10.1 The general model

All results are proved for a generalization of the model, which incorporates two changes: the distribution of states is more general than $U[-1,1]$, and the payoff function is more general than (1). The purpose of the generalization is to indicate which properties of the original model, and especially the payoff function, drive the results.

Assume that x is distributed on $X \equiv [\underline{x}, \overline{x}] \subset \Re$, with $0 \in (\underline{x}, \overline{x})$, according to a continuous and strictly positive density ϕ, which induces a probability measure $\mathfrak{P}$ for the probability space $(X, \mathfrak{B}, \mathfrak{P})$. Note that in general $Ex \neq 0$. The strategy of any player who observes x is now a function of X instead of $[-1,1]$, and the definition of equilibrium is revised to incorporate this generalization.

The new payoff function $\pi : A \times [0,1] \times [\underline{x}, +\infty) \rightarrow \Re$ generalizes (1) and for convenience extends its domain beyond $q = 0, \frac{1}{m}, \frac{2}{m}, ...$ to $q \in [0,1]$ and beyond X to $[\underline{x}, +\infty)$. The payoff function π can take any functional form but must be twice-continuously differentiable and satisfy assumptions (10), parts (a) through (l); except as noted, each assumption applies throughout the domain of π. It is trivial to confirm that the original payoff function (1) satisfies all twelve assumptions.

The extension to player heterogeneity proceeds exactly as in Section 6. If players are heterogeneous, then $\pi_i(N,q;x) = \pi(N,q;x)$ and $\pi_i(P,q;x) = \pi(P,q;x) - \Delta_i$, where

$\Delta_i = \left(i - \frac{m+1}{2}\right)\delta$. The homogeneous case appears when $\Delta_i = \delta = 0$.

Except for the results in Section 5, the generalizations of the state space and payoff function require no changes in the statements of the lemmas and theorems in the text. This appendix proves the results, exactly as originally stated, in the more general setting. The exceptions are Theorems 5′ and 6′, which take a slightly different form in the present more general model but still imply the original statements of Theorems 5 and 6 as special cases.

10.2 Restrictions on the general payoff function π.

Assumptions (10) restrict the form of the generalized payoff function, for the case of homogeneous players. Assumptions (10a) and (10b) are, in part, normalizations. They say that full participation gives every player a payoff of x, while zero participation gives everyone a zero payoff.

$$\pi(P,1;x) = x \tag{10a}$$

$$\pi(N,0;x) = 0 \tag{10b}$$

The next three assumptions describe the ways in which higher values of x represent better projects. They imply that, for any given rate of participation: higher values of x increase total surplus, especially participants' surplus, and increase (or make less negative) participants' marginal return from others' participation.

$$\frac{\partial\left[q\pi(P,q;x)+(1-q)\pi(N,q;x)\right]}{\partial x} > 0 \qquad for\ q>0 \tag{10c}$$

$$\frac{\partial \pi}{\partial x}(P,q;x) > \max\left(0, \frac{\partial \pi}{\partial x}(N,q;x)\right) \qquad for\ q>0 \tag{10d}$$

$$\frac{\partial^2 \pi}{\partial q \partial x}(P,q;x) \geq 0 \tag{10e}$$

Unlike (1), the generalized payoff function allows nonparticipants to prefer lower values of x ($\frac{\partial \pi}{\partial x}(N,q;x) < 0$). This could happen, for example, if participation means voting for an unpopular bill. Then the political advantage of opposing the bill might increase as the quality of the bill decreases.

The next two assumptions help to ensure that, in equilibrium, either everyone follows the leader or no one does. Assumption (10f) states that, for good projects, higher participation increases (or makes less negative) the individual return to participation. This could reflect technical returns to scale or network effects, or social (e.g., "safety in numbers") effects. For bad projects ($x < 0$), the derivative condition in (10f) could plausibly fail, because the first few participants might have only a modest negative impact, but the final few participants might allow the project to reach full destructive fruition, contradicting increasing returns to participation. Nonetheless, assumption (10g), which implies merely that the best outcome for a participant is either full participation or zero participation by others, is (as Lemma 2′ shows) sufficient to rule out intermediate levels of participation for $x < 0$.

$$\frac{\partial \left[\pi(P, q + \frac{1}{m}; x) - \pi(N, q; x)\right]}{\partial q} > 0 \quad for\ x \geq 0 \tag{10f}$$

$$\pi(P, q; x) \ is\ quasi\text{-}convex\ in\ q \tag{10g}$$

Assumptions (10h)-(10j) bound the circumstances that support participation. Given the previous assumptions, (10h) and (10i) imply that a player is willing to participate only if she learns something favorable about the realization of x *and* believes that others may also participate (cf. (14)). Assumption (10j) states that a player does prefer to participate if she believes that x takes its maximal value and that everyone else will participate. For the particular payoff function (1), assumptions (10i) and (10j) are equivalent to the parameter restrictions (2) and (3).

$$E_x \left[\pi(P, 1; x) - \pi(N, 1 - \frac{1}{m}; x)\right] < 0 \tag{10h}$$

$$\pi(P, \frac{1}{m}; x) < 0 \quad \text{for } x \in X \tag{10i}$$

$$\pi(P, 1; \overline{x}) - \pi(N, 1 - \frac{1}{m}; \overline{x}) > 0 \tag{10j}$$

Assumption (10k) introduces the free riding externality. It says that if one player defects from full participation in a good project, then she earns a positive payoff from others' participation.

$$\pi(N, 1 - \frac{1}{m}; x) > 0 \qquad for\ x \geq 0 \tag{10k}$$

The incentive to free ride can arise in unconventional ways. If P represents a legislator's vote for a bill and $x = 0$ the political payoff if the vote is unanimous, then $\pi(N, 1 - \frac{1}{m}; 0) > 0$ may describe the payoff gained from defecting to oppose the bill, a payoff which arises only because the defector's opposition calls attention to flaws in the bill and makes it less popular. (The realism of this scenario rests in the observation that majorities often place great value on getting a unanimous vote.)

Finally, (10l) ensures that, for good projects, surplus maximization requires either zero participation or full participation.

$$\frac{\partial^2 W(q;x)}{\partial q^2} > 0 \quad for\ x \geq 0 \tag{10l}$$

10.3 Elementary properties of π.

It is useful to define summary notation for player i's gain from participation, as a function of q_{-i}, the fraction of the players who participate if i does not participate. If player i participates, then $q = q_{-i} + \frac{1}{m}$; if not, then $q = q_{-i}$. Let $a \equiv \frac{m+1}{2}$ denote the "average player" and for $i \in I \cup \{a\}$ let:

$$\gamma_i(q_{-i};x) \equiv \pi\left(P, q_{-i} + \frac{1}{m}; x\right) - \pi\left(N, q_{-i}; x\right) - \Delta_i. \tag{11}$$

Assumption (10f), and assumptions (10d) and (10e), imply respectively:

$$\partial\gamma_i/\partial q_{-i} \quad > \quad 0 \quad for\ x \geq 0 \tag{12}$$

$$\partial\gamma_i/\partial x \quad > \quad 0 \tag{13}$$

Given (10h), (12) implies that nonparticipation is a strictly dominant action for any player who learns nothing about the value of x:

$$E_x[\gamma(q_{-i};x)] < 0 \quad \text{for all } q_{-i}. \tag{14}$$

10.4 Bounds on δ, for the heterogeneous case.

In the heterogeneous case ($\delta > 0$), some players' payoff functions π_i could violate assumptions (10h)-(10j), even as the "average" payoff function π satisfies them, but Assumption B (below) requires that δ be small enough to ensure that π_i satisfies (10h)-(10j), for all $i \in I$ and that W_δ satisfies (10l). That immediately implies that π_i satisfies (10b)-(10k). Assumption (10a) cannot be similarly extended to the heterogeneous case, because heterogeneous players inevitably disagree about which projects are "good." Instead, Definition A establishes a lower bound $x_0 < 0$, such that certain assumptions about good projects extend to projects $x \geq x_0$, and the second part of Assumption B then requires δ to be small enough to ensure that all projects worse than x_0 look "bad" to every player. Definition A and Assumption B place a third bound on δ, which allows an extension of (12) to cross-player comparisons.

The overall effect of Assumption B is to ensure that $\delta > 0$ changes players' participation thresholds without introducing other, more radical, changes in equilibrium behavior. Where the results refer to δ "sufficiently small," they mean small enough to satisfy Assumption B.

Definition A. Because $\pi \in C^2$, there exists $x_0 \in (\underline{x}, 0)$ such that the inequalities of (10f), (10k), (10l), and (12) hold for all $x \geq x_0$. If players are heterogeneous ($\delta > 0$), then fix such a $x_0 < 0$; if not, then fix $x_0 = 0$. Then (12) implies that $\partial\gamma_i(q_{-i}; x)/\partial q_{-i} > \omega$ for all $x \geq x_0$ and q_{-i}, for some fixed $\omega > 0$. (Note that the value of $\partial\gamma_i(q_{-i}; x)/\partial q_{-i}$ does not depend on i and δ.) Fix such an ω.

Assumption B. In the model with heterogeneous players, $\delta > 0$ is small enough, given m and π and x_0, to ensure: (i) assumptions (10b) through (10l) hold with π_i replacing π for all $i \in I$, the constraint $x \geq x_0$ replacing $x \geq 0$, and W_δ replacing W; (ii) $x_0 < \Delta_1 (< 0)$; (iii) $\delta < \frac{\omega}{2m}$.

10.5 Elementary properties of π and (for the heterogeneous case) π_i, given small δ.

Given Assumption B(i), (10h) and (10j) imply:

$$E_x\left[\gamma_i(1 - \frac{1}{m}; x)\right] < 0 \quad \text{for all } i \tag{11h$'$}$$

$$\gamma_i(1 - \frac{1}{m}; \overline{x}) > 0 \quad \text{for all } i \tag{11j$'$}$$

It is useful to define notation for leaders' and followers' gain from participation, in the specific context of a leader-follower equilibrium. For any $l, f \in I \cup \{a\}$ and $t \in X$, let

$$h_l(x) \equiv \pi_l(P, 1; x) = x - \Delta_l \tag{15}$$

$$k_f(t) \equiv E_x[\gamma_f(1 - \frac{1}{m}; x) \mid x > t] \tag{16}$$

Lemma C shows that for three kinds of players – players with complete information, leaders who will be followed, and uninformed followers of the latter – the return to participation increases monotonically in x. For each kind of player, Lemma C also characterizes the participation threshold.

Lemma C. If $\delta > 0$ is sufficiently small given m and π, or if $\delta = 0$, then the following is true. For any $i, l, f \in I \cup \{a\}$:

(i) There exists unique $\tau_i^C \in (x_0, \overline{x})$ such that $\gamma_i(1 - \frac{1}{m}; \tau_i^C) = 0$.

(ii) The unique solution to $h_l(x) = 0$ is $x = \Delta_l \in (x_0, \tau_l^C)$.

(iii) There exists unique $\tau_f^F \in (\underline{x}, \tau_f^C)$ such that $k_f(\tau_f^F) = 0$; $k_f' > 0$; and if $\delta > 0$ then τ_f^F is strictly increasing in f.

Proof. If $\delta > 0$, then fix δ small enough to satisfy Assumption B. Fix any $i, l, f \in I \cup \{a\}$. The definitions of h_i and Δ_i, and Assumption B(ii), imply $h_i(x_0) < 0$. Likewise $h_l(x_0) < 0$. The definitions of π_i, γ_i, and h_i imply $\gamma_i(1 - \frac{1}{m}; x_0) = h_i(x_0) -$

$\pi\left(N, 1-\frac{1}{m}; x_0\right)$, and (10k) and Definition A (the choice of x_0) imply $\pi\left(N, 1-\frac{1}{m}; x_0\right) > 0$, so $\gamma_i(1-\frac{1}{m}; x_0) < 0$. Since (13) implies $\partial\gamma_i(1-\frac{1}{m}; x)/\partial x > 0$, claim (i) follows from (10j′).

By definition (as above) $\gamma_l(1-\frac{1}{m}; \tau_l^C) = h_l(\tau_l^C) - \pi\left(N, 1-\frac{1}{m}; \tau_l^C\right) = 0$. Part (i) implies that $\tau_l^C > x_0$, which with (10k) and Definition A implies $\pi\left(N, 1-\frac{1}{m}; \tau_l^C\right) > 0$. Therefore, $h_l(\tau_l^C) > 0$, the previous paragraph shows that $h_l(x_0) < 0$, and clearly $h_l' > 0$, so $h_l(x) = 0$ has a unique solution $x \in (x_0, \tau_l^C)$. Definition (15) shows that $x = \Delta_l$. That establishes (ii).

Let $\Phi(x) = \int_0^x \phi(t)dt$. Using $\partial\gamma_f/\partial x > 0$ (13), we have, for $t < \overline{x}$:

$$\begin{aligned} k_f'(t) &= \frac{\partial E_x[\gamma_f(1-\frac{1}{m}; x) \mid x > t]}{\partial t} \\ &= \frac{\phi(t)}{1-\Phi(t)} \times \left[\int_t^{\overline{x}} \frac{\gamma_f(1-\frac{1}{m}; x)\phi(x)dx}{1-\Phi(t)} - \gamma_f(1-\frac{1}{m}; t)\right] \\ &> \frac{\phi(t)}{1-\Phi(t)} \times \left[\int_t^{\overline{x}} \frac{\gamma_f(1-\frac{1}{m}; t)\phi(x)dx}{1-\Phi(t)} - \gamma_f(1-\frac{1}{m}; t)\right] = 0 \end{aligned}$$

Therefore, $k_f'(t) > 0$ for $t < \overline{x}$. From (10h′): $k_f(\underline{x}) = E_x[\gamma_f(1-\frac{1}{m}; x)] < 0$. By definition, $\gamma_f(1-\frac{1}{m}; \tau_f^C) = 0$, which with (13) and $\tau_f^C < \overline{x}$ implies $k_f(\tau_f^C) = E_x[\gamma_f(1-\frac{1}{m}; x) \mid x > \tau_f^C] > 0$. These three facts show that $k_f(\tau_f^F) = 0$ has a unique solution $\tau_f^F \in (\underline{x}, \tau_f^C)$. If $\delta > 0$, then (11) and (16) show that $k_f(t)$ is strictly decreasing in f, which with $k_f'(t) > 0$ implies that the solution to $k_f(t) = 0$ must be strictly increasing in f. That establishes (iii). □

10.6 Proofs of results (Sections 3, 4, and 6)

Proof of Theorem 1. For any $i, j \in I$, $\delta = 0$ implies $\gamma_i = \gamma_j$ and $\tau_i^C = \tau_j^C$ (from Lemma C(i)) and $x_0 = 0$. Let $\tau^C \in (0, \overline{x})$ and γ denote these common values. Assumptions (10b) and (10i) imply: (a) given any x, it is a Nash equilibrium for no one to participate. If every player participates, then one player's gain from participation is $\gamma(1-\frac{1}{m}, x)$. Lemma C(i) and (13) imply that $\gamma(1-\frac{1}{m}, x) < 0$ for $x < \tau^C$ and $\gamma(1-\frac{1}{m}, x) > 0$ for $x > \tau^C$. Therefore: (b) full participation is a Nash equilibrium iff $x \geq \tau^C$. Statements (a) and (b) imply the result. □

It is convenient to prove the remaining results in a sequence different from the sequence of appearance. Lemma 2′ generalizes Lemma 2 to the case of heterogeneous players. After proving results 6-8 for the heterogeneous case, results 2-5 for the homogeneous case are little more than special cases.

Lemma 2′. If $\delta > 0$ is sufficiently small given m and π, or if $\delta = 0$, then the following is true. Any productive equilibrium $(s_l^*; s_f^*, f \in I\backslash\{l\})$ has (a) $s_l^*(x) = N$ for all $x < x_0$ and (b) $s_f^* = P$ for all $f \in I\backslash\{l\}$.

Proof. If $\delta > 0$, then fix δ small enough to satisfy Assumption B.

(a) Assumption (10b) implies that the leader's payoff from nonparticipation is zero, while her payoff from participation is $\pi_l(P, \frac{n+1}{m}; x)$, where n denotes the number of followers who participate. Fix any $x < x_0$. Lemma C(ii) shows that $h_l(x) = \pi_l(P, 1; x) < 0$, Assumptions (10i) and B(i) imply that $\pi_l(P, \frac{1}{m}; x) < 0$, and (10g) then implies that $\pi_l(P, \frac{n+1}{m}; x) < 0$ for all n, implying that $s_l^*(x) = N$ is a dominant action for the leader.

(b) The choice of ω (Definition A) implies $\gamma_f(\frac{i}{m}; x) - \gamma_f(\frac{i-1}{m}; x) > \frac{\omega}{m}$, implying $\gamma_{f+1}(\frac{i}{m}; x) - \gamma_f(\frac{i-1}{m}; x) > \frac{\omega}{m} - \delta$ and $\gamma_{f+2}(\frac{i}{m}; x) - \gamma_f(\frac{i-1}{m}; x) > \frac{\omega}{m} - 2\delta$ for all $x \geq x_0$ and $f, i \in I$ such that these expressions are well-defined. Therefore, $\delta < \frac{\omega}{2m}$ (Assumption B(iii)) implies that for all $f, i \in I$ such that the expressions are well-defined:

$$\gamma_{f+1}\left(\frac{i}{m}; x\right) > \gamma_f\left(\frac{i-1}{m}; x\right) \text{ for all } x \geq x_0 \tag{17a}$$

$$\gamma_{f+2}\left(\frac{i}{m}; x\right) > \gamma_f\left(\frac{i-1}{m}; x\right) \text{ for all } x \geq x_0 \tag{17b}$$

Consider a follower f's optimization problem in an equilibrium $(s_l^*; s_f^*, f \in I\backslash\{l\})$. Let $X^* \equiv \{x \in X \mid s_l^*(x) = P\}$ denote the states in which the leader participates; the productivity of the equilibrium requires that X^* occur with positive probability. If the leader participates but some follower f does not, then let q_{-f}^* denote the fraction of players who participate, as implied by $(s_{f'}^*, f' \in I\backslash\{f, l\})$. Let κ_f denote f's gain from participation, given the leader's action P:

$$\kappa_f \equiv E_x[\gamma_f(q_{-f}^*; x) \mid x \in X^*] \tag{18}$$

Since X^* occurs with positive probability, the equilibrium condition for any follower f requires:

$$\kappa_f > 0 \text{ implies } s_f^* = P; \quad \kappa_f < 0 \text{ implies } s_f^* = N \tag{19}$$

The next step is to show that all followers choose the same strategy. Suppose that $s_f^* = P$ for some follower $f \in I\backslash\{l\}$. Let f denote the largest such follower. Then $q_{-i}^* \geq q_{-f}^*$ for all $i \in I\backslash\{l\}$, which with (12) and Assumption B(i) implies that $\gamma_f\left(q_{-i}^*; x\right) \geq \gamma_f(q_{-f}^*; x)$ for all $i \in I\backslash\{l\}$ and $x \geq x_0$, which with (11) implies that $\gamma_i\left(q_{-i}^*; x\right) > \gamma_f(q_{-f}^*; x)$ for all followers $i < f$ and $x \geq x_0$. Since Lemma 2′(a) implies that $x \geq x_0$ for all $x \in X$, it follows from (18) that $\kappa_i > \kappa_f$ for all followers $i < f$. The equilibrium condition (19) then implies that $s_i^* = P$ for all followers $i < f$. Suppose that $s_{f'}^* = N$ for some $f' \in I\backslash\{l\}$; it follows that $f' > f$. The argument now proceeds by contradiction. By definition (18):

$$\kappa_f \equiv E_x[\gamma_f(\frac{f-1}{m};x) \mid x \in X^*] \qquad if\ l < f \qquad (20a)$$

$$\kappa_f \equiv E_x[\gamma_f(\frac{f}{m};x) \mid x \in X^*] \qquad if\ l > f \qquad (20b)$$

$$\kappa_{f+1} \equiv E_x[\gamma_{f+1}(\frac{f}{m};x) \mid x \in X^*] \qquad if\ l < f \qquad (20c)$$

$$\kappa_{f+1} \equiv E_x[\gamma_{f+1}(\frac{f+1}{m};x) \mid x \in X^*] \qquad if\ l > f+1 \qquad (20d)$$

$$\kappa_{f+2} \equiv E_x[\gamma_{f+2}(\frac{f+1}{m};x) \mid x \in X^*] \qquad if\ l = f+1 \qquad (20e)$$

If $l < f$, then (17a), (20a), and (20c) imply that $\kappa_{f+1} > \kappa_f$. If $l > f+1$, then (17a), (20b), and (20d) imply that $\kappa_{f+1} > \kappa_f$. If $l = f+1$, then the existence of f' implies that $f+2 \leq m$; (17b), (20b), and (20e) imply that $\kappa_{f+2} > \kappa_f$. In every case, $\kappa_i > \kappa_f$ for some follower $i > f$, and the equilibrium condition (19) implies that $s_i^* = P$, contradicting the choice of f. Therefore, f' does not exist and all followers choose the same strategy.

If that strategy is N then (10b) and (10i) imply that the leader never participates, implying that X^* occurs with zero probability, a contradiction. Therefore, all followers choose P. □

Proof of Lemma 7. Lemma C(iii) immediately implies Lemma 7.

Proof of Theorem 8 (including the case $\delta = 0$)**.** If $\delta > 0$, then fix δ small enough to satisfy Assumption B. Fix $l \in I$. Lemma 2′ shows that a productive equilibrium requires all followers $f \in I\backslash\{l\}$ to play P. Given this, it is sufficient to show:

(i) If all followers play P, then the unique optimal strategy for leader l is the threshold strategy Δ_l.[42]

(ii) If the leader plays the threshold strategy Δ_l, and all other followers play P, then P is optimal for follower f iff $\Delta_l \geq \tau_f^F$.

Proof of (i): Leader l earns payoff $\pi(P, 1; x) - \Delta_l = x - \Delta_l$ from playing P or zero (given (10b)) from playing N. Therefore, optimization for leader l requires: if $x > \Delta_l$ then $s_l^*(x) = P$; if $x < \Delta_l$ then $s_l^*(x) = N$.

Proof of (ii): Follower f's equilibrium condition (19) shows that P is optimal for f iff $\kappa_f \geq 0$. Other players' equilibrium strategies imply that $\kappa_f = k_f(\Delta_l)$. Therefore, P is optimal iff $k_f(\Delta_l) \geq 0$, which by Lemma C(iii) is true iff $\Delta_l \geq \tau_f^F$. □

[42] This claim disregards the inconsequentially differentiated strategies such that the leader participates at the threshold $x = \tau_l^L$ (cf. fn. 13).

The remaining results assume that $\delta = 0$, which implies that $\Delta_i = 0$ for all $i \in I$.

Proof of Lemma 2. Lemma 2′ immediately implies Lemma 2. $\square$

Proof of Theorem 3. Consider any $f \in I\backslash\{l\}$. Equations (10a), and (11) imply that $\gamma_f(\frac{m-1}{m}; x) = x - \pi(N, \frac{m-1}{m}; x)$. Therefore, using (16), (4) is equivalent to $k_f(0) \geq 0$, which Lemma C(iii) shows is equivalent to $\tau_f^F \leq 0$, which is equivalent to (8). The proof of Theorem 6 encompasses the case $\delta = 0$; given the equivalence of (4) and (8), this case implies Theorem 3. $\square$

Proof of Theorem 4 Fix any $x \geq 0$. Then (10l) implies that $W(q; x)$ is maximized at (only) $q = 0$ or $q = 1$. Because $W(0; x) = \pi(N, 0; x) = 0$ and $W(1; x) = \pi(P, 1; x) = x$, $q = 1$ is the unique maximizer for $x > 0$, and $q = 0$ and $q = 1$ both maximize $W(q; 0)$. Therefore, $W(q; 0) \leq 0$ for all $q > 0$, but (10c) implies that $\partial W(q; x)/\partial x > 0$ for all $x \leq 0$ and $q > 0$, so $W(q; x) < 0$ for all $x < 0$ and $q > 0$. Since $W(0; x) = 0$ for all $x < 0$, it follows that $q = 0$ is the unique maximizer of $W(q; x)$ for all $x < 0$. Summarizing, $q = 1$ maximizes $W(q; x)$ for all $x \geq 0$, and $q = 0$ maximizes $W(q; x)$ for all $x \leq 0$. The claim follows immediately from Theorem 3. $\square$

10.7 Generalizing the definition of γ (Section 5).

The 3-stage equilibria described in Section 5 employ the function $\chi(\cdot)$, which is calculated from the original payoff function (1). To extend those equilibria to the more general payoff function π, it is necessary to define $\chi(\cdot)$ more generally. Since the 3-stage model assumes that players are homogeneous ($\delta = 0$), the player subscripts on the function γ_i (the return to participation) can be omitted for the purpose of defining $\chi(\cdot)$.

Assume that $\dfrac{\partial \gamma}{\partial x}(q_{-i}; x)$, which must be strictly positive by (13), has a strictly positive lower bound given any fixed q_{-i}. This assumption is innocuous, because the continuity of $\dfrac{\partial \pi}{\partial x}$ already ensures that such a bound exists for $x \in X$, the economically relevant region. Let $\chi(m-1)$ denote the unique value of x satisfying $\gamma(1 - \frac{1}{m}; x) = 0$; Lemma C(i) shows that this equation has a unique solution $x = \tau^C \in (x_0, \overline{x}) = (0, \overline{x})$ Therefore, $\chi(m-1) = \tau^C > 0$. It has just been shown that $\gamma(1 - \frac{1}{m}; \chi(m-1)) = 0$, which from (12) implies that $\gamma(0; \chi(m-1)) < 0$, and the strictly positive lower bound on $\dfrac{\partial \gamma}{\partial x}$ then implies that $\gamma(0; x') = 0$ for some unique $x' > \chi(m-1)$; let $\chi(0)$ denote x'. Summarizing: $0 < \chi(m-1) < \chi(0)$ and $\gamma(\frac{m-1}{m}; \chi(m-1)) = \gamma(0; \chi(1)) = 0$. For integer $n \in \{0, 1, ...m-1\}$, (12) and (13) imply that $\gamma(\frac{n}{m}; x) = 0$ for some unique $x \in [\chi(m-1), \chi(0)]$ – denote this value $\chi(n)$ – and $\chi(n)$ is strictly positive and strictly decreasing in n. That completes the reconstruction of $\chi(\cdot)$.

10.8 The explicit updating rule (Section 5).

It is useful to state explicitly the rule for updating beliefs, in the 3-stage model. This rule generates well-defined posterior distributions at every stage 3 information set, given any fully-mixed and E-discrete strategy profile. (Updating is trivial in stage 2, because players observe only the results of their own research.) Let $z \subset X \times A^{12}$ denote player i's information set in stage 3, where A^{12} denotes all possible combinations of players' actions in stages 1 and 2; note that z comes from the structure of the game and does not include any information inferred from equilibrium strategies. Let $A(z) \subset A^{12}$ denote the action profiles which are consistent with i's observations of actions at z; then either $z = X \times A(z)$ or $z = \{x\} \times A(z)$ for some $x \in X$. Given any fixed (pure or mixed) strategy profile s, and for any Borel-measurable $f \subset X$ and $a \in A^{12}$, let $pr(f, a, z; s)$ denote the joint prior probability that (i) $x \in f$, (ii) a is the profile of actions taken through stage 2, and (iii) player i reaches z. Assuming that the fixed strategy profile s is E-discrete for some E, then for any $e \in E$ and $a \in A^{12}$ let $\psi(a|e; s)$ denote the prior probability that players choose a in event e. Lemma D states the Bayesian posterior distribution at any information set where this posterior is well-defined.

Lemma D: Fix an interval partition E and a (pure or mixed) E-discrete strategy profile s and player i's stage 3 information set z, such that some action profile in $A(z)$ occurs with positive probability under s (i.e. $\psi(a'|e'; s) > 0$ for some $a' \in A(z)$ and $e' \in E$). Also, fix an arbitrary Borel-measurable $f \subset X$ and action profile $a \in A^{12}$.

If player i has not observed x then:

$$\begin{aligned} pr(f, a|z; s) &= \sum\nolimits_{e \in E} \left[\frac{pr(a, f \cap e; s) pr(z|a, f \cap e; s)}{pr(z; s)} \right] \qquad (21a) \\ & \sum\nolimits_{e \in E} \left[\frac{pr(a, f \cap e; s)}{pr(z; s)} \right] \quad \text{if } pr(f, a|z; s) > 0 \\ & \sum\nolimits_{e \in E} \left[\frac{\phi(f \cap e) pr(a|f \cap e; s)}{\sum_{e' \in E} \phi(e') pr(z|e'; s)} \right] \\ & \sum\nolimits_{e \in E} \left[\frac{\phi(f \cap e) \psi(a|e; s)}{\sum_{e' \in E} \phi(e') \sum_{a' \in A(z)} \psi(a'|e'; s)} \right] \end{aligned}$$

If player i has observed x, then

$$pr(f, a|z; s) = \frac{\phi(f|\{x(z)\}) \psi(a|e(z); s)}{\sum_{a' \in A(z)} \psi(a'|e(z); s)} \qquad (21b)$$

wherever this is well-defined, where $x(z)$ denotes the state observed at z and $e(z)$ denotes the element of E implied by that state.

Proof: The assumptions for (21a) immediately imply $pr(z; s) > 0$, and the first and third steps, in (21a), follow from the routine axioms of probability. To justify the other steps suppress the "s" argument and suppose that $pr(f, a|z) > 0$. Then obviously $a \in A(z)$, but because player i cannot distinguish two nodes that imply the same action profile, it follows that $a \notin A(z')$ for all $z' \neq z$, implying $pr(z|a) = 1$, implying that $pr(z|a, f \cap e) = 1$ for $f \cap e$ such that $pr(a, f \cap e) > 0$; that establishes step 2. Substituting arbitrary $e' \in E$ for $f \cap e$ shows that $pr(z|a, e') = 1$ for e' such that $pr(a, e') > 0$, but the latter condition is equivalent to $pr(a|e') > 0$, implying altogether that $pr(z|a, e') = 1$ for e' such that $pr(a|e') > 0$, which implies that $pr(z, a|e') = pr(a|e')(= \psi(a|e'))$. Summing both sides yields $\sum_{a' \in A(z)} pr(z, a'|e') = \sum_{a' \in A(z)} \psi(a'|e')$, but the definition of $A(z)$ implies that the left-hand side equals $pr(z|e')$, so $pr(z|e') = \sum_{a' \in A(z)} \psi(a'|e')$, as needed for step 4. Step 4 also uses the fact that the strategy profile is $E-$discrete, which implies that $pr(a|f \cap e) = \psi(a|e)$.

Similar but simpler arguments establish (21b). □

The formulas (21) show that *player i's posterior beliefs are well-defined* at stage 3 information set z if and only if $\sum_{a' \in A(z)} \psi(a'|e; s) > 0$ for some event $e \in E$, or in the true event if player i has observed the state. It is intuitive that this is the condition which is necessary for well-defined posterior probabilities; it says essentially that the z is on the equilibrium path (i.e. with positive probability). It must hold at every information set if players' strategies are fully mixed.

10.9 Generalized statements of Theorems 5 and 6 (Section 5).

Unlike most of the results in the paper, the generalization of the model requires restatements of Theorems 5 and 6. The reason is that the thresholds described in Theorems 5 and 6 must now be described more generally. Theorem 5 is a special case of Theorem 5′, and Theorem 6 is a special case of Theorem 6′.

Theorem 5′: If $\int_0^{\tau^C} \phi(x) \left[\pi(N, 1 - \frac{1}{m}; x) - \pi(P, 1; x)\right] dx < \kappa < \int_0^{\overline{x}} \phi(x)\pi(P, 1; x)dx$, and the credibility condition (4) holds, then the following strategies s, event partition E, and assessment μ constitute a discrete sequential equilibrium.

(a) Strategies s († denotes strategy components which are never used on the equilibrium path):

Leader ($i = 1$)	Followers ($i > 1$)
(a) $s_1^1 = R$	(e) $s_i^1 = U$
(b) $t_1^2 = 0$	(f) $t_i^2 = 1$ (†)
(c) $s_1^3(\cdot) = N$ (†)	(g) $s_i^3(\cdot) = a_1^2$
(d) $t_1^3(a^2) = \chi(\#a^2)$ (> 0)	(h) $t_1^3(a^2) = \chi(\#a^2)$ if $a_1^2 = N$ (†)
	(i) $t_1^3(a^2) = \tau^C$ if $a_1^2 = P$ (†)

(b) Event partition E: the set of intervals of X induced by the boundary points $\{0, \tau^C, \chi(m-2), \chi(m-3), ..., \chi(1)\} \cap int(X)$, where each boundary point belongs to the lower interval.

(c) Assessment μ: In stages 2 and 3, every player i, who has not observed another player j's stage 1 action, believes that j has followed his stage 1 strategy, except: a player $i > 1$, who has observed $x > 0$ and player 1's play of N in stage 2, believes in stage 3 that player 1 is informed with probability $\frac{1}{2}$. A player i who has not (yet) observed x believes that x is distributed according to its prior ϕ, except: a player $i > 1$ believes in stage 3 that x distributed according to the conditional density $\phi(\cdot|x > 0)$ if player $i = 1$ chose P in stage 2, or according to $\phi(\cdot|x < 0)$ if player $i = 1$ chose N in stage 2.

Proof: Because $\tau^C = \chi(m-1)$ and χ is strictly monotonic, each element of E has positive measure, implying that it also has positive probability under ϕ, as required by the definition of a discrete sequential equilibrium. Let s^* and μ^* denote the equilibrium strategy profile and assessment.

Because each decision is binary, the sequence of fully mixed strategies can be defined in a simple way. For $k = 2, 3, 4, ...$ let $s(k)$ denote the strategy profile such that each player at each information set, follows his equilibrium strategy with probability $1 - \frac{1}{k}$ and takes the alternative action with probability $\frac{1}{k}$, and these decisions are statistically independent of x (if the player is informed) and across information sets. Clearly $s(k) \to s^*$ as k diverges, and $s(k)$ is E-discrete for all k, because by construction players adopt the same mixed actions in any two states x and x' in the same element of E. Let $\mu(k)$ denote the assessment implied by $s(k)$ (i.e., as computed by formula (21))

The next step is to show that $\mu(k) \to \mu^*$ (at each information set) as k diverges. For each k, the construction of $s(k)$ implies that every action in stages 1 and 2, except for player 1's stage 2 participation decision a_1^2, is statistically independent of x. Therefore, the posterior distribution of x, at any information set where x has not been observed, is the prior ϕ except where this is updated by whatever information has been revealed (to a player $i > 1$) by a_1^2. The construction of $s(k)$ implies: the probability that $x \in X'$, for any fixed $X' \subset X \cap (-\infty, 0]$, is $\frac{\Phi(X')\frac{1}{k}}{\Phi(0)\frac{1}{k}+[1-\Phi(0)]\left(1-\frac{1}{k}\right)} \to 0$ given $a_1^2 = P$, or $\frac{\Phi(X')\left[\left(1-\frac{1}{k}\right)^2+\frac{1}{k}\right]}{\Phi(0)(1-\frac{1}{k})^2+[1-\Phi(0)](1-\frac{1}{k})\frac{1}{k}+\frac{1}{k}} \to \frac{\Phi(X')}{\Phi(0)}$ given $a_1^2 = N$, as k diverges; the probability that $x \in X'$, for any fixed $X' \subset X \cap (0, +\infty)$, is $\frac{\Phi(X')\left(1-\frac{1}{k}\right)}{\Phi(0)\frac{1}{k}+[1-\Phi(0)](1-\frac{1}{k})} \to \frac{\Phi(X')}{1-\Phi(0)}$ given $a_1^2 = P$, or $\frac{\Phi(X')\left[\left(1-\frac{1}{k}\right)\frac{1}{k}+\frac{1}{k}\right]}{\Phi(0)\left(1-\frac{1}{k}\right)^2+[1-\Phi(0)]\left(1-\frac{1}{k}\right)\frac{1}{k}+\frac{1}{k}} \to 0$ given $a_1^2 = N$, as k diverges. These limiting beliefs match the equilibrium assessment.

Now consider players' beliefs about other players' information. The construction of $s(k)$ implies trivially that in stage 2 each player i believes that each other player followed his equilibrium strategy, since i has observed nothing that could be affected by any other players' actions. At any stage 3 information set: player i's posterior belief that another player $j > 1$ is informed, given $a_j^2 = N$, is $\frac{1}{1+k} \to 0$ as k diverges; for $i > 1$, player i's posterior belief that player 1 is informed, given $a_1^2 = N$, is $\frac{\Phi(0)\left(1-\frac{1}{k}\right)}{\Phi(0)\left(1-\frac{1}{k}\right)+\frac{1}{k}} \to 1$ if i is uninformed $\frac{\left(1-\frac{1}{k}\right)^2}{\left(1-\frac{1}{k}\right)^2+\frac{1}{k}} \to 1$ if i has observed $x \leq 0$, or $\frac{1-\frac{1}{k}}{2-\frac{1}{k}} \to \frac{1}{2}$ if i has observed $x > 0$, as k diverges. These limiting beliefs again match the equilibrium assessment. That completes the proof that $\mu(k) \to \mu^*$.

It remains to show that the equilibrium strategies s^* are optimal, given the equilibrium assessment μ^*. Because each player has a finite number of opportunities to act, it is sufficient to establish the optimality of each player's behavior strategy at each of his information sets, assuming equilibrium play at any subsequent information sets.

We proceed by backward induction, from stage 3. Consider a player $i > 1$. Given her current assessment of players' information, s^* implies that any player $i' > 1$ who has not yet participated will imitate a_1^2. Suppose that i is uninformed. If no other player $i' > 1$ has participated, then i is in the same position as in the participation equilibrium of Theorem 3, and she should imitate a_1^2. If the situation is the same except that some other player $i' > 1$ has already participated, then the choices are effectively unchanged if $a_1^2 = P$ (because those players would participate anyway) and i should still play P; if $a_1^2 = N$ then i believes that $x < 0$ and so prefers to play N regardless of how many others participate. Therefore, strategy component (g) is optimal.

Suppose that i has observed x. If $a_1^2 = P$, then i expects everyone else to participate and her participation threshold is as already derived for the analogous situation in the symmetric one-stage game: τ^C. That justifies strategy component (i). If $a_1^2 = N$, then i expects the number of participants excluding herself to be whatever number has already participated: $\#a_i^2$. In that case her expected payoff from P is $\pi(P, \frac{\#a_i^2+1}{m}; x)$ and her payoff from N is $\pi(N, \frac{\#a_i^2}{m}; x)$, and the threshold at which these are equal is $x = \chi(\#a^2)$. That justifies strategy component (h).

Consider player 1's decision in stage 3, if she is informed and $a_1^2 = N$. She expects the only other participants to be those who have already participated, so her situation is analogous to that of player $i > 1$ in case (h) and the decision rule is the same. That justifies strategy component (d). If player 1 is instead uninformed, then by μ^* she believes that x is distributed according to ϕ, regardless of others' actions, and (14) implies that she should play N. That justifies strategy component (c) and establishes the optimality of the stage 3 strategies.

In stage 2, player 1 can act only if she is informed, and in that case she believes that she is the only informed player and the situation is essentially equivalent to her initial situation in the participation equilibrium of Theorem 3. As in that game, her optimal threshold is $x = 0$. That justifies strategy component (b). A player $i > 1$ can likewise act only if she is informed. If $x \leq 0$ then N is optimal regardless of what others do. If $x > 0$ then she expects $a_1^2 = P$ and expects herself (and everyone else) to follow by participating in stage 3, regardless of her current action. Therefore she is indifferent to participation in stage 2. That justifies strategy component (f).[43]

In stage 1, player 1 expects R to earn an expected payoff equal to that in the participation equilibrium of Theorem 3, before subtracting the research cost κ, and any player $i > 1$ expects U to earn the same payoff; call this payoff π^*. Player 1's alternative is U, which pays zero because he expects no one else to do research and consequently no participation. The statement of the theorem assumes that $\kappa < \pi^*$, so R is optimal. That justifies strategy component (a). For a player $i > 1$, the alternative is R, which he expects to have no impact on anyone else's decisions but which may cause him to play differently in stage 3. If $x \leq 0$ then i does not expect to participate regardless of whether she observes x; if $x > \tau^C$ then i expects to participate regardless of whether she observes x; but if $x \in (0, \tau^C)$ then i expects everyone to participate if he plays U but expects himself to play N (i.e. free ride) while others participate if he plays R. The assumptions of the theorem require κ to exceed the expected gain from that free-riding, so U is optimal. That justifies strategy component (b). That completes the demonstration that the equilibrium strategies are optimal given equilibrium beliefs. □

Proof of Theorem 5. Assume that the prior distribution ϕ is the uniform distribution on $[-1, 1]$. By definition τ^C solves $\pi(P, 1; \tau^C) = \pi(N, \frac{m-1}{m}; \tau^C)$. Therefore, using (1):

$$\tau^C = \frac{(m-1)\alpha\beta}{m - (m-1)\alpha}$$

[43] This argument could not be used to establish that the equilibrium is trembling-hand perfect (given a similar extension of that concept to a continuous state space). In that case justifying the equilibrium would require a more complicated pattern of trembles than the present device of adding a common probability of error, $\frac{1}{k}$, at every information set.

The bounds on κ, in Theorem $5'$, reduce to:

$$\overline{\kappa} = \int_0^1 \frac{1}{2}\pi(P,1;x)dx = \int_0^1 \frac{1}{2}x dx = \frac{1}{4}$$

$$\underline{\kappa} = \int_0^{\tau^C} \frac{1}{2}\left[\pi(N,\frac{m-1}{m};x) - \pi(P,1;x)\right]dx = \int_0^{\tau^C} \frac{1}{2}\left[(x+\beta)(\frac{m-1}{m})\alpha - x\right]dx = \frac{1}{4}\frac{(\frac{m-1}{m})^2\alpha^2\beta^2}{1-(\frac{m-1}{m})\alpha}$$

matching the interval in the statement of Theorem 5. Assumption (3) implies:

$$\frac{(\frac{m-1}{m})^2\alpha^2\beta^2}{1-(\frac{m-1}{m})\alpha} < \frac{(\frac{m-1}{m})^2(\frac{m}{(m-1)(1+\beta)})^2\beta^2}{1-(\frac{m-1}{m})(\frac{m}{(m-1)(1+\beta)})} = \frac{\beta}{1+\beta} < 1$$

Therefore, $\underline{\kappa} < \overline{\kappa}$. The rest of the theorem follows directly from Theorem $5'$. □

Theorem $6'$: If $\kappa < \int_{\tau^C}^{\overline{x}} \phi(x)\left[\pi(P,1;x) - \pi(N,1-\frac{1}{m};x)\right]dx$, then the following strategies s, event partition E, and assessment μ constitute a discrete sequential equilibrium.

Strategies s: $s_1^1 = R$; $t_i^2 = 1$; $t_1^3(a^2) = \tau^C$ for all a^2; $s_1^3(\cdot) = N$.

Event partition $E = \{[\underline{x}, \tau^C], (\tau^C, \overline{x}]\}$.

Assessment μ:. In stages 2 and 3, every player believes that every other player is informed. Every player who has not (yet) observed x believes that x is distributed according to its prior ϕ.

Proof: Lemma C(i) shows that $\tau^C \in (x_0, x)$, so each element of E has positive measure. Let s^* and μ^* denote the equilibrium strategy profile and assessment; s^* is E-discrete, because players take the same sequence of actions in any two states x and x' in the same element of E.

For $k = 2, 3, 4, ...$, define the sequence of E-discrete strategy profiles $s(k)$, and the implied assessments $\mu(k)$, as in the proof of Theorem $5'$. Clearly $s(k) \to s^*$ as k diverges. For each k, the construction of $s(k)$ implies that every action in stages 1 and 2 is statistically independent of x, so the posterior distribution of x under $\mu(k)$ is simply the prior ϕ. The construction of $s(k)$ implies that any player i's posterior probabiliy, under $\mu(k)$, that any given other player j is informed, is $1-k$ in stage 2, or $\frac{\Phi(\tau^C)(1-\frac{2}{k})+(1-\frac{1}{k})\frac{1}{k}}{\Phi(\tau^C)(1-\frac{2}{k})+(2-\frac{1}{k})\frac{1}{k}}$ in stage 3 (if player j chose N in stage 2), probabilities which converge to the equilibrium assessment of one as k diverges. Therefore, $\mu(k) \to \mu^*$.

It remains to show that the equilibrium strategies s^* are optimal, given the equilibrium assessment μ^*. As for Theorem $5'$, it is sufficient to establish the optimality of each player's behavior strategy at each of his information sets, assuming equilibrium play at any subsequent information sets.

We proceed by backward induction, from stage 3. A player who has not observed x believes that x is distributed according to ϕ and (14) implies that N is strictly dominant in this situation. A player who has observed $x > \tau^C$ believes from μ^* that all other players are informed and therefore will participate in stage 3 if they have not already, so Lemma C(i) implies that P is optimal. For a player who has observed $x \leq \tau^C$, (14) and (13) imply that N is optimal regardless of others' actions. That establishes the optimality of the stage 3 strategy.

In stage 2, a player can act only if he is informed. If he observed $x > \tau^C$ then he expects to play P in stage 3 and so is indifferent to her current action. If he observed $x \leq \tau^C$ then N is again optimal regardless of others' actions. That establishes the optimality of the stage 2 strategy.

In stage 1, each player expects R to earn an expected payoff of $\pi(P, 1; x) - \kappa$ if $x > \tau^C$ or $-\kappa$ if $x \leq \tau^C$. He expects U to earn an expected payoff of $\pi(N, 1 - \frac{1}{m}; x)$ if $x > \tau^C$ or zero if $x \leq \tau^C$. The initial assumption $\kappa < \int_{\tau^C}^{\overline{x}} \phi(x) \left[\pi(N, 1 - \frac{1}{m}; x) - \pi(P, 1; x)\right] dx$ thus implies that R is optimal. That establishes the optimality of s^* given μ^*. □

Proof of Theorem 6: Theorem 6 is an immediate consequence of Theorem 6′, except for the value of κ^*. It is sufficient to show that if x is uniformly distributed on $[0, 1]$, then $\int_{\tau^C}^{\overline{x}} \phi(x) \left[\pi(P, 1; x) - \pi(N, 1 - \frac{1}{m}; x)\right] dx = \left[x \frac{mx - (m-1)(x+2\beta)\alpha}{4m}\right]_{\frac{(m-1)\alpha\beta}{m-(m-1)\alpha}}^{1}$. The proof of Theorem 5 shows that $\tau^C = \frac{(m-1)\alpha\beta}{m-(m-1)\alpha}$ and the rest is a straightforward calculation. □

References

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

[2] Austen-Smith, D., "Strategic Transmission of Costly Information", *Econometrica* 62: 955-963 (1994).

[3] Blume, Andreas, Oliver Boared, and Kohei Kawamura, "Noisy Talk", manuscript, 2006.

[4] Carlsson, Hans, and Eric van Damme, "Global Games and Equilibrium Selection," *Econometrica* 61:989-1018 (1993)

[5] Case, John, *Open-Book Management*, New York: HarperCollins, 1995.

[6] Crawford, Vincent P., and Joel Sobel, "Strategic Information Transmission," *Econometrica* 50:1431-1451 (1982).

[7] Dansereau, Fred, and Francis Yammarino, *Leadership: The Multiple-level Approaches*, London: Jaipress, 1998.

[8] Daughety, Andrew F., and Jennifer Reinganum, "Hidden Talents: Complementarity and Pareto-Improving Private Information," manuscript, November 2006.

[9] Harsanyi, John., "Games with Incomplete Information Played by 'Bayesian' Players, part II: Bayesian Equilibrium Points," *Management Science* 14:320-334 (1968).

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

[11] Hermalin, Benjamin E., "Leading for the Long Term," *Journal of Economic Behavior and Organization*, 62:1-19 (2007).

[12] Hirshleifer, J., "The Private and Social Value of Information and the Reward to Inventive Activity," *American Economic Review* 61:561-74 (1971).

[13] Huck, Steffen, and Pedro Ray Biel, "Endogenous Leadership in Teams," *Journal of Institutional and Theoretical Economics*, 162:253-261 (2006).

[14] The multileader case appears in a direct antecedent of this paper, with the same title, at http://www.hss.caltech.edu/~jacksonm/dukeprogram.htm

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

[16] Komai, Mana, "Leadership in Public Projects: Vagueness or Clarity," VPI Economics WP 2002-8.

[17] Komai, Mana, and Mark Stegeman "A Theory of Leadership Based on Endogenous Assignment of Information", NBER Decentralization Conference [http://www.hss.caltech.edu/~jacksonm/dukeprogram.htm] (2004a).

[18] Komai, Mana, and Mark. Stegeman "A Theory of Leadership Based on Endogenous Assignment of Information", VPI Economics WP 2004-4 (2004b).

[19] Komai, Mana, Mark Stegeman, and Benjamin Hermalin, "Leadership and Information", *American Economic Review* 97:944-47 (2007).

[20] Kreps, David, and RobertWilson, "Sequential Equilibrium," Econometrica 50:863-94 (1982).

[21] Levy, Gilat, and Ronny Razin, "It Takes Two: An Explanation for the Democratic Peace," *Journal of the European Economic Association* 2:1-29 (2004).

[22] Milgrom, Paul., and John Roberts, *Economics, Organization, and Management*, Englewood Cliffs, New Jersey: Prentice Hall, 1992.

[23] Prendergast, Canice, "A Theory of 'Yes Men'," *American Economic Review*, 83:757-70 (1993).

[24] Rotemberg, Julio J., and Garth Saloner, "Leadership Style and Incentives," *Management Science* 39:1299-1318 (1993).

[25] Rotemberg, Julio J., and Garth Saloner, "Visionaries, Managers, and Strategic Direction," *RAND Journal of Economics*, 31:693-716 (2000).

[26] Varian, Hal, "Sequential Provision of Public Goods, *Journal of Public Economics* 53:165-86 (1992).

[27] Vesterlund, Lise, "The Informational Value of Sequential Fund-raising," *Journal of Public Economics* 87:627-657 (2003).