Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Legitimacy, Communication, and Leadership in the Turnaround Game

Jordi Brandts, David J. Cooper, Roberto A. Weber

MANAGEMENT SCIENCE
Vol. 61, No. 11, November 2015, pp. 2627–2645
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)
http://dx.doi.org/10.1287/mnsc.2014.2021

Legitimacy, Communication, and Leadership in the Turnaround Game

Jordi Brandts
Institut d'Anàlisi Econòmica, 08193 Barcelona, Spain; and Barcelona Graduate School of Economics, 08005 Barcelona, Spain, jordi.brandts@uab.es

David J. Cooper
Department of Economics, Florida State University, Tallahassee, Florida 32306; and University of East Anglia, Norfolk NR4 7TJ, United Kingdom, djcooper@fsu.edu

Roberto A. Weber
Department of Economics, University of Zurich, CH-8006 Zurich, Switzerland, roberto.weber@econ.uzh.ch

We study the effectiveness of leaders for inducing coordinated organizational change to a more efficient equilibrium, i.e., a turnaround. We compare communication from leaders to incentive increases and also compare the effectiveness of randomly selected and elected leaders. Although all interventions yield shifts to more efficient equilibria, communication from leaders has a greater effect than incentives. Moreover, leaders who are elected by followers are significantly better at improving their group's outcome than randomly selected leaders. The improved effectiveness of elected leaders results from sending more performance-relevant messages. Our results are evidence that the way in which leaders are selected affects their legitimacy and the degree to which they influence followers. Finally, we observe that a combination of factors—specifically, incentive increases and communication from elected leaders—yields near-universal turnarounds to full efficiency.

Keywords: leadership; job selection; coordination failure; experiments; communication
History: Received March 9, 2014; accepted May 16, 2014, by Uri Gneezy, behavioral economics. Published online in *Articles in Advance* December 22, 2014.

1. Introduction

In organizations where individuals' inputs are strong complements, the lowest-performing individual often determines group performance. An assembly line moves no faster than the slowest line worker, collaborative reports or software are incomplete until the final contribution is finished, and perceptions of overall product quality are often sensitive to the worst-performing feature. Organizations with such "weak-link" production technologies are prone to being caught in inefficient productivity traps where all individuals exert low effort. Even if it is mutually beneficial for all individuals to simultaneously increase their effort and individually incentive compatible to do so *if others also increase their effort*, escaping a productivity trap can be extremely challenging because no individual can unilaterally change the outcome for the better.

Leadership has a natural role to play in facilitating the transition to a better outcome. Indeed, there exist several studies of turnarounds in experimental weak-link environments showing that leadership of various types can play a positive role. Leading by example (Cartwright et al. 2013), communication from a leader requesting greater effort by group members (Brandts and Cooper 2007, Chaudhuri and Paichayontvijit 2010), and help with commitment from high-ability types to low-ability types (Brandts et al. 2013) can all increase the odds of escaping a productivity trap.

However, aside from leadership, there are other ways to induce change in an organization toward more efficient equilibria. For example, increasing incentives to coordinate has been shown to be a powerful way of coordinating change in contexts where production exhibits high complementarities (Brandts and Cooper 2006a), including in field settings (Knez and Simester 2002). Even small changes in incentives have been found to be quite effective (Hamman et al. 2007). Therefore, an important open question in understanding the effectiveness of leadership for inducing organizational change is how effective it is relative to other interventions, such as modest increases in incentives.

Moreover, even in environments where leadership is successful on average, many leaders do not succeed. For example, in a study by Weber et al. (2001), a randomly selected leader gave a brief speech explaining the benefits of mutual coordination, but this largely failed to improve coordination in large groups. Brandts and Cooper (2007) found that allowing communication from a leader significantly increases coordination in a majority of groups, yet 9 of 27 groups experience complete coordination failure over the final five rounds of the experiment. Thus, not all leaders are equally effective, and it is important to understand what factors may allow for greater effectiveness.

Our study addresses these open questions. We use a laboratory experiment in which an initial phase of the experimental session induces coordination on an inefficient equilibrium. We then introduce interventions intended to produce a "turnaround" to a more efficient equilibrium. Our primary focus is on the effectiveness of leaders, with only the ability to communicate with other group members, to induce coordinated change toward efficiency.

Our design allows us to directly compare the effectiveness of leaders with the effects of increased incentives, by including treatments with and without incentives and treatments with and without leaders. We also examine whether leader effectiveness is enhanced by endowing leaders with greater "legitimacy." Legitimacy is a concept widely utilized, though often somewhat vaguely defined, by organizational and political scholars, sociologists, and social psychologists. For example, it is central to many early discussions of the sources of authority, such as those by philosopher David Hume and sociologist Max Weber (Hume [1748] 1960, Weber 1948). It generally refers to a property of an authority, institution, or leader that grants the entity credibility in the eyes of an audience or followers. Tyler (2006) defined legitimacy as "a psychological property of an authority, institution, or social arrangement that leads those connected to it to believe that it is appropriate, proper and just" (p. 375).[1]

Many organizations give employees a voice in who is chosen to be their manager. Voting on leaders is a common mechanism used by many organizations to endogenously choose leaders, with examples including the choice of directors for nongovernmental organizations, leadership positions in political parties such as the heads of U.S. National Committees, and leaders of small work groups in companies. This raises the question of whether it matters how the leader is chosen and whether leaders selected through procedures that afford their position with greater legitimacy are more effective for inducing organizational change. Our experiment thus compares two different mechanisms for appointing a leader, with the leader either determined at random or selected by a vote among those to be led.

Previewing the results, we find that both leadership and increased incentives induce change to higher, more efficient equilibria. However, the magnitude of the effect is stronger for allowing leaders to communicate with their groups than for simply providing increased incentives to coordinate efficiently. Thus, leadership has a positive effect, even relative to an alternative strong intervention such as increased incentives. Moreover, the effects of increased incentives and leadership are generally additive—having a leader induces change to more efficient equilibria even in the presence of increased incentives.

We also find that elected leaders are significantly better at improving their organization's situation than randomly selected leaders. In fact, their success is striking: elected leaders produce successful turnarounds almost universally, particularly when they are accompanied by increased incentives. Indeed, in the condition with elected leaders and increased incentives, every group enjoys an extended stretch of coordination at the highest level of efficiency. This dramatic success contrasts with the more variable effectiveness of randomly selected leaders in our experiment and in prior studies. To our knowledge, *no prior experiment* finds leadership to be as effective at inducing coordination on the high-effort equilibrium as our treatment with elected leaders and increased incentives.

In the second part of our analysis, we investigate what makes elected leaders more effective. Specifically, we study whether different factors can account for the increased efficacy of elected leaders. In particular, we investigate whether elections produce different—and abler—leaders than random assignment, whether elected leaders work harder or take different actions than randomly selected leaders, and whether followers merely respond differently to leaders whom they have elected. Although groups do use the information they have available on potential leaders to select candidates with specific characteristics—i.e., those who contributed more effort in the pre-turnaround stage and who are generally more knowledgeable—neither of these individual characteristics corresponds to leaders who are actually more effective. Moreover, we find no evidence that followers respond differentially to elected leaders once we control for what those leaders do.

Instead, our results suggest that elected leaders act differently than nonelected ones. The primary tool available to leaders in our experiment is the messages they send to their groups. All of these messages were recorded and analyzed to identify the frequency and efficacy of various types of messages. Leaders were given no guidance about what sorts of messages to send, but most sent messages with a clear goal of improving coordination. We find that elected leaders send more messages that induce efficient coordination, and we find that this difference in behavior accounts

[1] Weber (1948) noted the importance of legitimacy for power to be effective and how the source of a leader's authority can affect the extent to which it is legitimate. Simon (1953), in noting the difficulty in measuring political power, argued that power and authority are linked to the perceptions of its legitimacy. Lipset (1959) discussed the legitimacy of a political institution as the degree to which it is viewed as appropriate and correct and therefore likely to facilitate social choice, and DelliCarpini et al. (2004) noted that deliberation and participation can increase the legitimacy of political institutions.

for the increased effectiveness of elected leaders. One interpretation of our results is that a large part of the increased effectiveness of leaders selected through more legitimate procedures may be due to an effect of the procedure on leaders themselves—leaders who see themselves as legitimately selected to improve their group's outcome exert themselves more to exercise effective leadership (in the context of our experiment, working "harder" by sending more relevant messages).

Our research highlights the potential value of simple coordination games for studying something as complex as leadership. Although scholars in many fields have studied leadership and how to measure it (e.g., Simon 1953, Bass and Bass 2008), our approach is a relatively new one in which leadership is measured by a leader's ability to induce a clearly measurable behavioral change—in our case, a coordinated move away from an inefficient equilibrium toward increased efficiency—in a simple game representing firm production. We view these simple, controlled behavioral experiments as a valuable context in which to learn about what factors make leadership effective. Our work also highlights the value of the laboratory environment, with its high degrees of control and the ability to randomly assign groups to interventions, such as different combinations of leadership and incentives. Such a setting makes it possible to identify the relative effectiveness of distinct mechanisms, absent the selection and endogeneity problems present in the field. Moreover, we think that coordination games, such as the weak-link game and the turnaround game, are especially valuable for learning about leadership. To be effective in such games, a leader must be able to not only convince others of the appropriate course of action but also create the belief that others believe this to be the case.

Section 2 of this paper reviews relevant literature. Section 3 presents the experimental design and procedures. Section 4 develops our hypotheses. Sections 5 and 6 contain the results. Section 7 concludes.

2. Related Literature

Our paper contributes to an emerging stream of research in experimental economics on the effects of leadership. In addition to the previously mentioned papers that study the role of leadership in overcoming coordination failure, there also exist a number of examples of the power of leadership in the related context of public good games with voluntary contributions, where the focus is on leaders' ability to increase contributions by followers. One set of studies uses a "leading-by-example" design in which a leader makes a public contribution decision before the other group members. These studies find that leading by example can induce higher contributions than in the absence of such a possibility (Moxnes and Van der Hejden 2003, Güth et al. 2007, Levati et al. 2007, Potters et al. 2007, Rivas and Sutter 2011). Other experiments explore the individual characteristics that make some leaders more effective than others in obtaining public goods provision (Hamman et al. 2011, Gächter et al. 2012, Arbak and Villeval 2013).

Kocher et al. (2013) compared behavior with randomly selected versus elected leaders, motivated by examining how managers' other-regarding preferences influence management style. They studied an environment in which members of a group with a leader all have to make suggestions for choices between lotteries, with the leader making the final decision. Their results show that elected leaders are more likely to make conformist decisions, in the sense that they tend to accommodate the preferences of the other team members even if these are at odds with those of the leader.[2]

Perhaps the public goods experiment most closely related to our current work is by Levy et al. (2011), since it involves both communication and different leader selection procedures. In the two treatments directly related to our experiments, they compared the role of randomly selected and elected leaders in the context of a four-person public good game. After playing several rounds of a standard linear public good game with voluntary contributions, the players wrote "platforms"—brief text messages discussing how the game should be played. These were distributed among the group members, who then voted for a leader. In one treatment the election winner became the leader as opposed to a randomly selected individual in the other. Once selected, the leader sent a message with a suggested contribution to the other group members at the beginning of each round. Group contributions were higher with elected leaders and elected leaders' suggestions were (weakly) followed more closely by group members. However, suggested contributions were not different between the two treatments.

The results of Levy et al. (2011) are consistent with the hypothesis that elected leaders are more effective than those randomly selected. However, the effect can be attributed to the content of platforms. As Levy et al. noted, "winning platforms tend to suggest and defend ... reasonable strategies for achieving efficient outcomes" (p. 43). Group members may therefore be more willing to follow elected leaders because they are associated with platforms that offer a path to

[2] Relatedly, Sutter et al. (2010) did not study leadership but focused on the impact of voting on behavior. They analyzed results from public goods games without leaders and found that the mere fact of voting on institutions makes members more cooperative compared with the same institutions being implemented exogenously. These results can be interpreted in terms of legitimate institutions triggering a cooperation premium.

cooperation, meaning that elected leaders may be more effective because they offer better platforms. In §7 we return to the relation between our results and those of these studies.

3. Experimental Design and Procedures

There were four stages for all of the experimental conditions in our design. Subjects knew the experiment had multiple stages but were not given instructions about any particular stage until reaching that stage. Stages 1 and 2 were identical in all conditions, whereas Stages 3 and 4 differed by treatment. Experimental conditions differed along two treatment dimensions in a 3 (type of leader: none, randomly selected, or elected) × 2 (incentives in the final stage: constant or increased) design.

In Stage 1 all participants answered a set of trivia questions. In Stage 2 participants played the first six-round block of the turnaround coordination game with payoffs designed to induce convergence toward the inefficient equilibrium. That is, Stage 2 is intended to reliably create a situation where firms are stuck at an inefficient equilibrium so that we can study the subsequent effectiveness of different kinds of interventions—i.e., leaders and increased incentives—for inducing a change to coordination on a more efficient equilibrium.

In Stage 3 the leader was selected for the random and elected leader conditions. Below, we describe the specific procedures employed for selecting leaders.

In Stage 4 subjects played the second and third six-round blocks of the turnaround coordination game. In the increased bonus condition, Stage 4 used payoffs more conducive to successful coordination. In the random and elected leader conditions, leaders could send all members of their groups messages in Stage 4. Stage 4 generated our primary dependent variable: the efficacy of leaders and incentives for inducing a turnaround.

Table 1 presents a timeline of the experiment. We now describe each stage in detail.

Table 1 **Timeline of Experiment**

Stage 1	Trivia quiz
Stage 2	Rounds 1–6 (block 1) of turnaround game, without leader and with payoffs designed to induce coordination failure
Stage 3	Selection of leader, either randomly or through election
Stage 4	Rounds 7–18 (blocks 2 and 3) of turnaround game, possibly with revised payoffs and communication from leader

Note. Stages 1 and 2 were identical across all treatment conditions; Stages 3 and 4 differed by condition.

3.1. Stage 1

Subjects answered general trivia questions. The purpose of this trivia quiz was to generate information that could be used for leader selection in Stage 3 of the experiment. The questions were chosen to have no relationship to the turnaround game. Our aim was to generate information that would be unrelated to aptitude for playing games, or any sort of obviously relevant skill, but that could serve as a basis for employees choosing between candidate leaders.

Subjects answered five multiple-choice general trivia questions (e.g., "What was the first U.S. state to enter the Union?"). Each question had four possible answers, and subjects averaged 1.45 correct answers. All subjects in a given session were shown the same five questions, which was common knowledge. We used two separate sets of questions to limit contamination between sessions. To give subjects an incentive to think about the questions, one subject was randomly selected to be paid \$2 at the end of the session for each correct question. Subjects did not know who the randomly selected subject was until the entire experimental session was completed, and they were not told in Stage 1 that their choices would be used for any purpose other than possibly earning a reward for correct answers. In this way, subjects' performance in the quiz could not be influenced by the desire to become an elected leader.

3.2. Stage 2

The turnaround game is a dynamic game—based on an underlying minimum-effort coordination game modeling production by workers in a firm—designed to capture important elements of the problems an organization faces when trying to escape from a performance trap (Brandts and Cooper 2006a). The game contains two distinct stages, implemented as Stages 2 and 4 of our design. Stage 2 is intended to induce coordination on an inefficient equilibrium. In Stage 4 the organization attempts a turnaround through interventions intended to produce a shift to the efficient equilibrium.

The turnaround game was played by a fixed group of five participants in the role of "employees." Stage 2 consisted of a six-round block (block 1). In each round, the five employees independently and simultaneously selected effort contributions, or the number of hours allocated to a common production task, and received a payoff based on the minimum effort expended in the group. There was no leader in Stage 2.

At the beginning of Stage 2, a common bonus rate (B) was announced, which would be valid throughout all six rounds. The bonus rate, set exogenously, determined how much additional pay each employee received for every unit increase in the group's minimum effort. In each round, all employees observed B and then—without any communication—simultaneously chose effort levels, where E_i is the effort level chosen by the ith employee. We restricted an employee's effort to be in 10-hour increments: $E_i \in \{0, 10, 20, 30, 40\}$. Intuitively, employees spend 40 hours per week on the job, and effort measures the number of these hours that they actually work hard rather than loafing.

Table 2 **Worker Payoffs for Stage 2** ($B = 6$) **and Stage 4** ($B = 10$)

Effort by worker i	Minimum effort by other workers: 0	10	20	30	40
		$B = 6$			
0	200	200	200	200	200
10	150	210	210	210	210
20	100	160	220	220	220
30	50	110	170	230	230
40	0	60	120	180	240
		$B = 10$			
0	200	200	200	200	200
10	150	250	250	250	250
20	100	200	300	300	300
30	50	150	250	350	350
40	0	100	200	300	400

Employees' payoffs for the round were determined by Equation (1), in which each additional hour of effort incurs a cost of 5. Employee i's payoff was

$$\pi_e^i = 200 - 5E_i + (B \times \min(E_1, E_2, E_3, E_4, E_5)). \qquad (1)$$

All payoffs were denominated in "experimental currency units" (ECUs). These were converted to monetary payoffs at a rate of \$1 equals 500 ECUs.

Table 2 displays the payoffs for the two bonus rates used in our experiment, $B = 6$ for Stage 2 and either $B = 6$ or $B = 10$ in Stage 4. For both values of B, the resulting game is a minimum-effort or weak-link coordination game (Van Huyck et al. 1990). All workers coordinating by selecting equal effort, for any of the five available effort levels, is a Nash equilibrium. However, coordinating on effort level 40 yields a higher payoff than coordinating, inefficiently, on effort level 0.

In Stage 2 subjects played the game induced by a bonus value of $B = 6$, shown on the upper part of Table 2. This choice of bonus was intended to generate a history of coordination failure that subjects needed to overcome in Stage 4. To see why efficient coordination is hard with $B = 6$, suppose that all five employees have previously chosen effort level 0. An employee who thinks about raising his effort from 0 to 10 faces a certain payoff reduction of 50 ECUs as a result of increased effort, whereas his maximum possible gain is only 10 ECUs beyond the 200 ECUs he gets without risk by choosing 0. For the proposed increase to have a positive expected profit, the employee must believe that the probability of the four other employees simultaneously raising their efforts from 0 to 10 equals at least 5/6. If the other four employees are assumed to choose effort levels independently, this requires a greater than 95% chance of each other employee increasing his effort to 10. The incentives to coordinate at higher effort levels are poor.

Emergence of the lowest-effort equilibrium in Stage 2 seemed likely, a priori, given these grim incentives and the results of previous research (e.g., Brandts and Cooper 2006a). Anticipating some of the results, the minimum effort was 0 in the final round of Stage 2 for 55 of the 59 groups in our experiment.

3.3. Stage 3

In Stage 3, the leader was selected for the conditions with random and elected leaders. This leader's only role was to send a message to the other subjects in the organization prior to each round of Stage 4.

3.3.1. No Leader. In conditions without a leader, there was (obviously) no need to select a leader in Stage 3. To limit any differences between conditions due to restart effects, these sessions also had a pause at the end of Stage 2. During this time, an experimenter read brief instructions that emphasized that the rules for the game would be subsequently unchanged.

3.3.2. Random Leader. This treatment follows the standard procedure in experiments of this type by selecting one participant at random to serve as the leader (e.g., Weber et al. 2001). After receiving a description of the role of the leader, subjects were told that one of the group members would be randomly selected as leader. The instructions stressed that all group members were equally likely to be chosen. Prior to resumption of the turnaround game, in Stage 4, all group members were informed of the ID number of the randomly selected leader. The leader continued to play as an employee, in addition to being able to send messages in each round, and was paid in the same way as the other employees, according to Equation (1).

3.3.3. Elected Leader. After receiving a description of the role of the leader, subjects were shown, for each member of their group, the number of correct answers from the trivia quiz in Stage 1 as well as the average effort level over the six rounds of the game in Stage 2.[3] This information was identified by randomly assigned ID numbers. Subjects then voted for an individual, using the ID numbers, to serve as the leader for the duration of Stage 4. They were free to vote for themselves. The group member receiving the most votes became the leader, with ties broken at random.

[3] The relevance of previous effort for choosing a leader is clear, but there were several good reasons for providing information about the Stage 1 quiz results prior to voting. First, quiz performance gave subjects a general measure of competence—not directly related to play of the turnaround game—that could be used in choosing a leader. As we report below, voting responded strongly to quiz performance. Second, giving subjects information only about effort levels in the first phase raises the possibility of "experimental demand" effects: when subjects only have one piece of information, this suggests that they should find some way to use this information when voting. With two pieces of information, subjects may still feel a need to use the information but have greater freedom to ignore any one piece.

Prior to resumption of the turnaround game in Stage 4, all group members were told which person (identified by ID number only) had been elected as leader. The identity of the leader was fixed for the remainder of the experiment. Group members were not given any further details about the election outcome, and information about Stage 1 and Stage 2 outcomes was not repeated. As in the random leader treatment, the elected leader continued to play as an employee, in addition to being able to send messages, and was paid the same way as the other employees.

3.4. Stage 4

In the final stage of the experiment, groups played two additional six-round blocks of the turnaround game, block 2 (rounds 7–12) and block 3 (rounds 13–18). There was a short pause between the two blocks in which the bonus rate was reiterated. In conditions with random and elected leaders, the subjects selected to be leaders for Stage 4 could send a typed message to all employees at the beginning of each round before any effort decisions were made.

Aside from the leader treatments described above, Stage 4 also introduced variation in a second treatment dimension, the size of the bonus (B) for coordinating on higher minimum effort levels. In the *bonus increase treatment*, the bonus was increased from $B = 6$ to $B = 10$ for blocks 2 and 3, whereas in the *constant bonus treatment*, it remained at $B = 6$ throughout. Increasing the bonus to $B = 10$ seems a natural benchmark for the effectiveness of financial incentives; Brandts and Cooper (2006a) found $B = 10$ to have the strongest effect in a comparison of different bonus levels.[4]

Even with an increased bonus, achieving successful coordination at high effort levels remains far from trivial. The lower part of Table 2 shows the payoff table with $B = 10$. Once again, consider the payoffs for an employee who raises his effort from 0 to 10 following a history of coordination at the lowest effort level. He faces a certain payoff reduction of 50 ECUs as a result of increased effort, and his maximum possible gain is 50 ECUs beyond the 200 ECUs he gets without risk by choosing 0. To have a positive expected profit, the employee must believe that the probability of the four other employees simultaneously raising their efforts from 0 to 10 equals at least 1/2. If the other four employees are assumed to choose effort levels independently, this translates to a greater than 84% chance of *each* other employee increasing his effort. These are better incentives than exist with $B = 6$, but overcoming coordination failure remains challenging in that the group must execute a highly coordinated shift in behavior.

[4] Brandts and Cooper (2006a) increased the bonus rate from $B = 6$ to either $B = 8$, $B = 10$, or $B = 14$. The effect was largest with $B = 10$, although differences were small and not statistically significant.

3.5. Procedures

All sessions were run in the xs/fs laboratory at Florida State University (FSU); we used the software z-Tree (Fischbacher 2007). All FSU undergraduates were eligible to participate, although subjects were drawn primarily from students taking social sciences classes (economics, political science, and sociology). Subject recruitment was done using the software ORSEE (Online Recruitment System for Economic Experiments; see Greiner 2004). Subjects were guaranteed $10 for arriving on time. Average earnings across all sessions were $18.15, including the show-up fee. Average session length was about 80 minutes.

At the beginning of each session, subjects were randomly seated. Instructions were read aloud by the experimenter prior to each stage of the experiment. Before beginning play in Stage 2, the turnaround game, all subjects were asked to complete a short quiz about the payoffs and the rules of the experiment. The full text for the instructions is provided in the online appendix (available as supplemental material at http://dx.doi.org/10.1287/mnsc.2014.2021).

To facilitate comprehension of the task, the instructions used a corporate context. For example, the five players in a group were explicitly referred to as "employees" and told that they were working for a "firm." The leader was called the "firm manager." However, we avoided the use of terms with strong connotations. For example, instead of asking subjects to choose a level of "effort," they were asked to allocate time between "activity A" and "activity B," with activity A implicitly corresponding to effort.

At the beginning of each six-round block of the turnaround game, subjects were shown the bonus rate for that block. Subjects were not told what bonus rates would be in subsequent blocks.

When there was a leader (Stage 4, random or elected leader treatments), the round began with the leaders sending messages. Before any effort decisions were made, leaders saw a box in which they could type a message to all employees in their group. Leaders were given no instructions about the content of this message other than being asked to avoid messages that might identify them or with obscene or offensive content. After viewing any message from the leader, the five subjects in each firm (including the leader) simultaneously chose their effort levels for the round. While choosing, subjects saw a payoff table similar to those displayed in Table 2, showing their payoff as a function of their own effort level and the minimum effort level chosen by other firm members.

At the end of each round, subjects saw a feedback screen showing their own effort level, the minimum effort for their firm, their payoff for the round, and their running total payoff for the experiment. A separate window on the feedback screen showed subjects a

Table 3 Summary of Experimental Conditions and Data

Type of leader	Constant bonus	Bonus increase
No leader	9 groups 45 subjects	10 groups 50 subjects
Random	10 groups 50 subjects	10 groups 50 subjects
Elected	10 groups 50 subjects	10 groups 50 subjects

summary of results from earlier rounds. Subjects were *not* shown the individual effort levels selected by all five employees in their firm. The absence of feedback about individual effort levels makes it more difficult to escape coordination failure in the turnaround game (Brandts and Cooper 2006b).

At the end of the session, each subject was paid via check the earnings for all rounds played plus the $10 show-up fee. Payment was done on an individual and private basis.

Table 3 summarizes the experimental design. Within each cell we report the number of groups and subjects.[5] Across all conditions, our data includes observations from 295 subjects.

4. Hypotheses

From a purely game-theoretic point of view, adding leaders need not affect effort levels. That one of the players has been selected to be the leader does not change the game in any relevant way. The messages that leaders can send need not affect effort levels either. Since the messages are cheap talk, mutual play of any of the five effort levels remains consistent with subgame-perfect equilibrium, regardless of messages sent by the leader.[6]

There are nevertheless reasons to believe that leaders will improve coordination. Leaders, by exerting higher effort, may make it easier for other players to increase the minimum in Stage 4 of the game. In addition, leaders' messages may have a powerful effect on effort levels. Subjects presumably have little difficulty identifying the efficient equilibrium as a desirable outcome, but attempting to simultaneously coordinate a turnaround to mutual effort of level 40 is risky, especially when there is no way of knowing others' intentions and no historical precedent of successful coordination. Communication can play a natural role in making efficient coordination a focal point, by establishing an expectation that level 40 will be chosen by all employees and, perhaps even more importantly, that all employees will start choosing level 40 *in the current round*. Indeed, preceding studies have consistently found that communication has a positive effect on coordination in weak-link games and the turnaround game (e.g., Blume and Ortmann 2007, Brandts and Cooper 2007). Therefore, we anticipated that the presence of leaders would, ceteris paribus, facilitate a turnaround from the history of low effort established in Stage 2, leading to coordination at high effort levels.

Hypothesis 1. *Minimum effort in Stage* 4 *will be higher with leaders than without leaders*.

In addition to studying the effect of leadership, our design also studies how leadership effectiveness compares to the effectiveness of increased financial incentives, in the form of a higher bonus in Stage 4. Holding leadership fixed, we anticipated that the bonus rate increase would make it easier to escape the strong precedent of choosing low effort from Stage 2, consistent with the results of previous studies of the turnaround game (Brandts and Cooper 2006a, Hamman et al. 2007).

Hypothesis 2. *Minimum effort in Stage* 4 *will be higher with an increased bonus than without*.

A particularly interesting question, given our design, is how the size of the increase (if any) in minimum effort obtained by having an active leader compares with the effect of increasing the bonus rate. To our knowledge, no prior study directly compares the effectiveness of messages from a leader and increased financial incentives, where the presence of either intervention varies exogenously, as in our experiment. The closest

[5] The no leader/no bonus increase cell has one fewer group than the others. We had low attendance for the final session of this condition. Data from two sessions were dropped from the data set because, in these sessions, many of the employees did not receive the messages from their leaders. A software fix resolved this problem.

[6] Play of a babbling equilibrium may seem perverse in a weak-link game with preplay communication. If coordinating at level 40 is Pareto optimal, why would a message calling for this equilibrium not be believed? The preceding intuition can be formalized by stating that messages that are "self-signaling" and "self-committing" should be considered credible rather than treated as meaningless babble (see Farrell and Rabin 1996 and Blume and Ortmann 2007 for discussions of the relevant theory). In our game, such arguments do not resolve the problem. Consider a message calling for play of the efficient equilibrium. This is self-committing, in the sense that the leader wants to follow his own suggestion if he believes the rest of the group will do so. However, even if this makes the message credible (i.e., other group members believe the leader will choose level 40), it need not follow that others will follow the leader's suggestion. They must believe that all of the other followers will also take the leader's suggestion and choose level 40. This is a risky proposition and requires faith in mutual knowledge of the equilibrium. Making matters worse, the leader's message is not self-signaling (i.e., the leader benefits from others choosing level 40 if and only if he chooses level 40). Imagine the leader plans on choosing level 20 no matter what. If a message calling for the efficient equilibrium moved all of the followers from level 0 to level 40, the leader would benefit. The preceding example is extreme, but the general point is that the leader has an incentive to exaggerate in hopes of increasing the effort of others *even if he does not plan on following his own recommendation*.

evidence is from Brandts and Cooper (2007), who, unlike us, studied a setting in which both varying bonus rates and leader messages are controlled by a subject playing as firm manager. Their results suggest that the effect of leadership is larger than the effect of increasing the bonus rate. To provide an intuitive basis for why this might be the case, consider that the problem players face in the turnaround game is coordinating a simultaneous move to a more efficient equilibrium. Regardless of the payoff matrix (see Table 2), it is obvious that all players would benefit by coordinating on a more efficient equilibrium. As Brandts and Cooper (2006a) argued, an increase in the bonus rate is effective primarily because it serves as a coordinating device, whereby a visible incentive change leads players to mutually expect others to increase their effort at that point in time. This interpretation of the effect of a bonus increase also explains why the response to the magnitude of a bonus rate increase is minimal. However, in this regard, a leader sending a common message to the group recommending an increase in effort is a more transparent coordination device—since there is a clear message, with specific instructions, that is received by everyone—making it more likely to create the mutual expectation of increased effort. Therefore, we expected to find a larger effect from leadership than an increased bonus rate.

Hypothesis 3. *The effect of increasing the bonus rate will be smaller than the effect of having an active leader.*

Hypothesis 1 predicts a positive effect from allowing leaders to send messages to their groups. However, communication may not be a panacea for coordination problems. Even in the studies referenced above, where communication increases the likelihood of coordination at an efficient equilibrium, there are still many groups that fail to reach full efficiency (Weber et al. 2001, Brandts and Cooper 2007). This highlights the need to consider mechanisms that enhance the effectiveness of leadership. Our design addresses this by studying the possibility that changing how the leader is chosen, via election rather than random selection, may further affect the impact of leadership by yielding leaders who are more legitimate.

There are several possible mechanisms through which legitimacy may operate. First, election could select abler individuals to become leaders. In other words, allowing groups to elect leaders might create legitimacy simply through groups selecting different people, using observable characteristics, who are better at inducing turnarounds.[7] Second, elected leaders could feel greater responsibility and hence engage in more active leadership (i.e., choose higher effort levels, send more messages, or send more substantive messages). That is, a selection procedure that confers greater legitimacy to a leader may cause that leader to perceive an obligation to try harder.[8] Third, the legitimacy conferred by being elected could change how followers respond to leaders' actions, ceteris paribus.[9] That is, being elected may increase the credibility of a leader's statements, meaning that followers respond more strongly to the same statements made by an elected leader than one who is appointed at random. For all these reasons, we anticipated greater efficiency with elected leaders.

Hypothesis 4. *Minimum effort in Stage 4 will be higher with an elected leader than with a randomly selected leader.*

Conditional on finding evidence for Hypothesis 4, our analysis will also attempt to identify which of the above channels is primarily responsible for the increased effectiveness of elected leaders. To motivate our study of each channel, we present auxiliary hypotheses that further identify possible mechanisms underlying the channel through which Hypothesis 4 may operate.

Hypothesis 5A. *Electing leaders yields a greater increase in minimum effort through the selection of abler leaders.*

Hypothesis 5B. *Electing leaders yields a greater increase in minimum effort because their election induces leaders to be more active in influencing followers (i.e., choose higher effort levels, send more messages, or send more substantive messages).*

Hypothesis 5C. *Electing leaders yields a greater increase in minimum effort because the property of being elected makes a leader's statements more likely to be followed.*

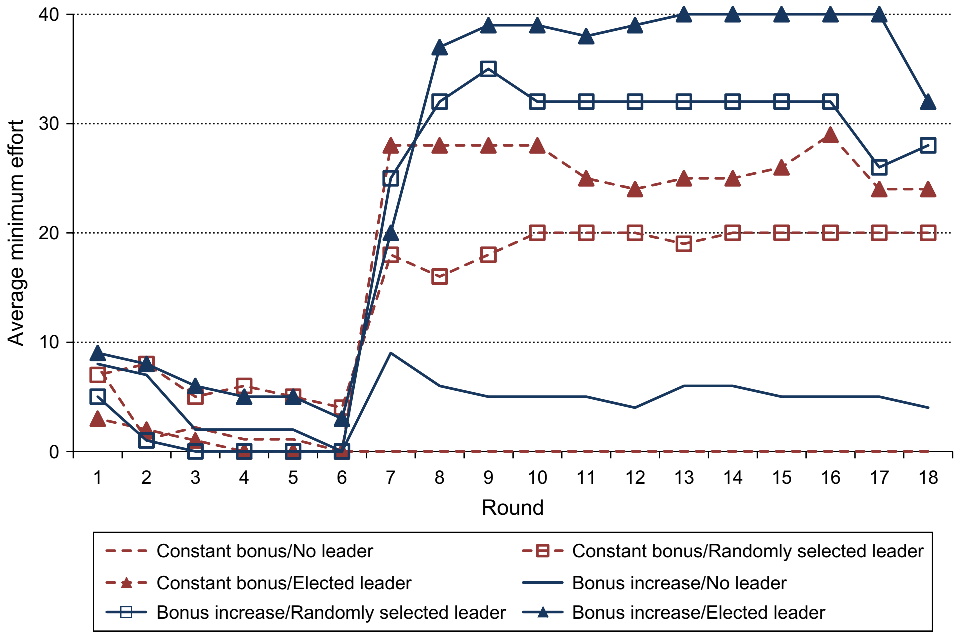
For Hypothesis 5A to hold, it must be the case that groups select particular kinds of leaders—e.g., ones with higher quiz scores or higher prior effort—and

[7] For example, in a public good environment, Hamman et al. (2011) showed that groups that can elect one of their members to determine all group members' contributions generally select those individuals who voluntarily contribute more to the public good and that such individuals make "better" leaders who mandate higher public good contributions by group members and contribute more themselves.

[8] For a related example, see Brandts et al. (2006). They studied whether individuals who are consciously selected for a task act differently than those randomly selected. Participants answered a personality questionnaire before they played a three-person game in which one of the players was selected to freely distribute a fixed pie among the three players. The comparison is between a treatment in which the person who gets to distribute the pie is randomly selected and one where one of the players elects one of the other players, after seeing the responses to the personality questionnaire. Knowingly selected distributors keep less for themselves than randomly selected ones.

[9] Weber (1948) noted the importance of legitimacy for power to be effective and how the source of a leader's authority can affect the extent to which it is legitimate. Lipset (1959) discussed the legitimacy of a political institution as the degree to which it is viewed as appropriate and correct, and therefore likely to facilitate social choice, and DelliCarpini et al. (2004) noted that deliberation and participation can increase the legitimacy of political institutions.

Figure 1 (Color online) Average Minimum Effort by Condition



that these characteristics, in turn, make leaders more effective. To test Hypothesis 5B, we need to show that there are certain actions (i.e., choosing high effort levels, sending certain types of messages) that are effective for inducing an increase in group minimum effort, and that elected leaders are more likely to take these actions. Finally, testing Hypothesis 5C requires showing that, holding constant what a leader says to the group, elected leaders are more effective than randomly appointed ones. Testing these hypotheses will require classifying message use by leaders, which we undertake in the analysis. Finally, note that Hypotheses 5A–5C need not be mutually exclusive, since all, or any combination, of the above mechanisms might make elected leaders more effective.

5. Results

Recall that "employee" refers to an individual subject in the experiment and "firm" refers to a fixed grouping of five employees. The term "effort" is used for the choice of a single employee, whereas "minimum effort" refers to the minimum of the five effort choices made by the members of the group. In each round, a group produces one observation for minimum effort and five (nonindependent) observations for effort.

In this section we investigate the effects of the various treatments on Stage 4 (rounds 7–18) individual efforts and firm minimum efforts to test Hypotheses 1–4. To better understand the impact of giving greater legitimacy to leaders via election, §6 tests Hypotheses 5A–5C. This involves comparing the behavior of elected and randomly selected leaders in Stage 2, prior to selection as a leader, and Stage 4. As an important component of this analysis, we quantify the content of leaders' messages and evaluate their frequency and effectiveness.

5.1. Treatment Effects

Figure 1 displays the average minimum effort across rounds, separately for each condition. The data are aggregated for each of the six cells in the experimental design.

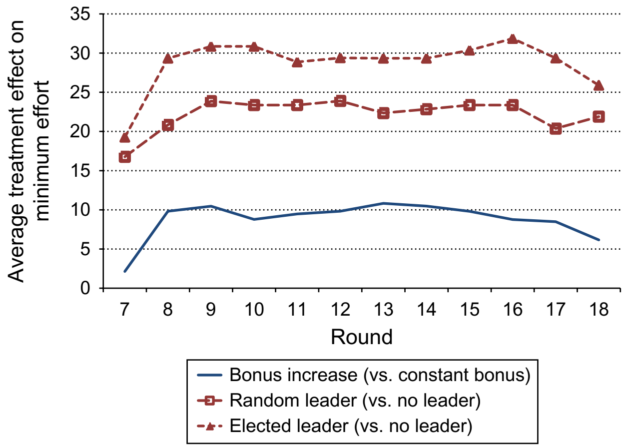
As noted previously, Stage 2 (rounds 1–6) reliably induced coordination on an inefficient outcome. The minimum effort was 0 in round 6 for 55 of 59 groups and was 0 from rounds 4 through 6 for 51 of 59 groups. Subsequent regression analysis includes controls for the variation between groups in Stage 2.

Turning to Stage 4 (rounds 7–18), the condition with no leader and a constant bonus serves as a control, showing what happens if nothing is changed from Stage 2. There is no reason to believe groups will spontaneously escape from the equilibrium productivity trap established in Stage 2 without a change in the environment. Figure 2 confirms that, in this condition, the minimum effort never rises above 0 in Stage 4.[10]

Holding the type of leader fixed (no leader, randomly selected, or elected), increasing the bonus rate for Stage 4 raises the minimum effort relative to holding the bonus rate constant. This is evident in Figure 1, when comparing each solid line to the corresponding dashed line. Increasing the bonus rate has an effect for all three types of leadership. In block 3 (the final six

[10] Individual effort in this condition shows a slight restart effect, increasing a bit between rounds 6 and 7, but this effect is small and too short-lived to affect the minimum effort for any of the groups.

Figure 2 (Color online) Average Treatment Effects on Minimum Effort



rounds), by which behavior has largely converged, the difference between increasing and not increasing the bonus rate is 5.2 versus 0 with no leader, 30.3 versus 19.8 for randomly selected leaders, and 38.7 versus 25.5 for elected leaders.

The aggregate treatment effect of increased bonuses is easily seen in Figure 2. The figure shows the *changes* in average minimum effort that correspond to the three treatment variables (bonus increase, random leader, and elected leader) with respect to the corresponding baseline conditions in which that treatment was absent. For example, the solid line labeled "Bonus increase" shows the average difference in minimum effort between groups with a bonus increase in Stage 4 ($B = 10$) and those for which the bonus remains unchanged ($B = 6$). This aggregates the effect across all three types of leadership. Figure 2 omits Stage 2 (rounds 1–6) because treatments were not introduced during this stage of the experiment. The overall effect of an increased bonus is generally about 10 units throughout most of Stage 4.

Figure 1 also shows the effects of having active leaders in each condition by comparing the unmarked lines (no leader) to the lines with squares (random leader) or triangles (elected leader) as markers. Figure 2 shows the average change in minimum effort for each type of leader, *relative to the baseline of not having a leader*, pooled across sessions with and without a change in the bonus rate for Stage 4. A few observations emerge from these figures. First, in all comparisons, adding a leader for Stage 4 increases the minimum effort relative to having no leader. This is evident from Figure 1 when comparing the lines with markers to those without. For example, in the case of constant bonuses (dashed lines), adding a randomly selected leader increases the average minimum effort by about 20 units in block 3. In all such comparisons, adding an active leader has an effect at least this large on minimum effort.

Also, the effect of having a leader is considerably larger than the effect of increasing the bonus. This can be seen most clearly in Figure 2. A bonus increase raises average minimum effort by about 10 units, but a randomly selected leader increases minimum effort by approximately 23 units, and the effect of an elected leader is close to a 30-unit increase in the average minimum effort.[11]

Another observation from Figures 1 and 2 is that elected leaders increase the average minimum effort by somewhat more than randomly selected leaders—about seven additional units of minimum effort in block 3. The magnitude of this effect is smaller than the effect of simply having a leader, but the difference is present and roughly the same in magnitude both with and without a bonus increase. Looking at Figure 2, a difference emerges by round 8 and remains fairly stable throughout Stage 4.

We have thus far based our discussion of treatment effects on a visual examination of the data. The regression reported as Model 1 in Table 4 confirms that these treatment effects are statistically significant. The data for this regression, as well as the other regressions that will be discussed later, are firm-level data taken from rounds 7–18. The dependent variable is the minimum effort from a single round for a single group—there are 12 observations per group. We use an ordered probit specification; this is appropriate given the inherently ordered and categorical nature of the dependent variable (minimum effort). Because there is correlation between observations taken from the same group, the standard errors have been corrected for clustering at the group level and are displayed in parentheses below the parameter estimates.

The primary independent variables in Model 1 are indicator variables identifying the treatments—e.g., whether there was a bonus increase in Stage 4, a leader (randomly selected or elected) in Stage 4, and the marginal effect of an elected leader (beyond the effect of a randomly selected leader) in Stage 4. That is, the coefficient for the indicator variable for elected leaders measures the *difference* between randomly selected and elected leaders. To capture whether the treatment effects change over time, an indicator variable for block 3 is included as an independent variable, along with interactions between indicator variables for blocks 2 and 3 and the three treatment variables. The importance of behavior in Stage 2 for outcomes in Stage 4

[11] Whereas this result corresponds to the effect of one specific bonus increase, from $B = 6$ to $B = 10$, the results of Brandts and Cooper (2006a) indicate that the effect of a bonus rate increase is generally not sensitive to the magnitude of the increase. Moreover, in their study, an increase to $B = 10$ is, if anything, more effective than larger bonus increases. Therefore, we expect that our finding about the relative importance of leadership versus bonus increases would generalize over a broad range of increases.

Table 4 Ordered Probit Analysis of Treatment Effects

Variable	Model 1	Model 2	Model 3	Model 4
Block 3 (rounds 13–18)	−0.280 (0.216)	−0.352 (0.244)	−0.291 (0.228)	−0.314 (0.312)
Block 2 × *Bonus Increase*	1.014*** (0.321)	0.903*** (0.336)	1.032*** (0.329)	1.533*** (0.335)
Block 3 × *Bonus Increase*	1.283*** (0.385)	1.281*** (0.396)	1.313*** (0.373)	1.809*** (0.434)
Block 2 × *Leader, Random or Elected*	1.683*** (0.311)	1.766*** (0.325)	1.668*** (0.303)	1.228*** (0.331)
Block 3 × *Leader, Random or Elected*	1.860*** (0.395)	1.951*** (0.409)	1.853*** (0.392)	1.431*** (0.480)
Block 2 × *Elected Leader*	0.556 (0.393)	0.796* (0.422)	0.810* (0.464)	−0.062 (0.384)
Block 3 × *Elected Leader*	0.818* (0.448)	1.183*** (0.440)	1.076** (0.479)	0.171 (0.481)
Maximum Minimum Effort (rounds 4–6)	0.041** (0.021)	0.060*** (0.021)	0.047** (0.021)	0.061*** (0.023)
Hard Quiz		−0.797** (0.350)		
Leader's Quiz Score			−0.373* (0.196)	
Leader's Average Effort Stage 2			0.004 (0.023)	
Round 7 Suggest 40 + Explanation				1.220*** (0.213)
Log likelihood	−549.43	−528.45	−533.33	−441.03

Notes. Dependent variable is group minimum effort in a round. All regressions contain 708 observations from 59 groups. Standard errors (in parentheses) are corrected for clustering at the group level.

***Statistically significant at the 1% level; **statistically significant at the 5% level; *statistically significant at the 10% level.

is self-evident, as groups who are not trapped at a minimum effort of zero in Stage 2 face an easier task entering Stage 4. To control for differences in Stage 2 outcomes, the maximum of the minimum effort from rounds 4 through 6 is included as an explanatory variable. There is a strong relationship between behavior in Stages 2 and 4 as captured by the positive and strongly significant estimate for the coefficient for this variable (*Maximum Minimum Effort* (rounds 4–6)).[12]

Looking further at the results, the effects of having a leader and increasing the bonus rate are significant at the 1% level for both blocks 2 and 3. The difference between having a randomly selected and an elected leader is significant at the 10% level for block 3, but not block 2. Although the size of the treatment effects varies slightly between blocks 2 and 3, none of these differences is statistically significant.[13] The results of Model 1 strongly support the following conclusion.[14]

Conclusion 1. *The data are consistent with Hypotheses* 1, 2, *and* 3. *Either raising the bonus or adding a leader increases minimum effort in Stage* 4. *The effect of adding a leader is greater than the effect of increasing the bonus.*

The weaker effect of having an elected versus randomly selected leader is consistent with our observation that electing leaders has a consistent effect that is smaller than the overall effect of leadership. This is summarized in the following conclusion.

Conclusion 2. *Minimum effort levels in Stage* 4 *are higher with elected leaders than with randomly selected leaders, as predicted in Hypothesis* 4, *although the marginal effect is smaller than the marginal effect of adding a randomly selected leader.*

Model 2 addresses an unexpected feature of our data. One of the two Stage 1 quizzes was significantly more difficult than the other, even though the quizzes were not designed in such a way and we expected no difference.[15] We control for whether this had an effect on Stage 4 outcomes by including an indicator variable for the more difficult quiz. Taking the harder quiz in Stage 1 has a strong negative effect on performance in Stage 4.[16] The estimates for having elected leaders and Stage 2 behavior (*Maximum Minimum Effort* (rounds 4–6)) become more significant with the inclusion of the quiz indicator, but the nature of the results is otherwise unaffected: whether a leader is elected matters, but not as much as having a leader. Since we did not anticipate an effect from quiz difficulty, we have no explanation for this unanticipated finding, but we report it and control for it (in Model 2) as an unintended treatment effect in our data. Further

[12] We experimented with various controls for Stage 2 outcomes, such as the minimum effort in round 6 or the average minimum effort across all six rounds in Stage 2. We use *Maximum Minimum Effort* in rounds 4–6 because it has the clearest relationship with Stage 4 outcomes. However, our conclusions about the treatment effects do not depend on which control for Stage 2 outcomes we use.

[13] If we do not include interactions with the identifier for block 3, the coefficients for a bonus increase and for leader are both significant at the 1% level. The difference between randomly selected and elected leaders is significant at the 10% level.

[14] We also looked at the treatment effects for waste, defined as the difference between individual effort and minimum effort for the group. Waste is a good measure of convergence, with low waste indicating strong convergence to equilibrium. Running an ordered probit regression analogous to Model 1, we find no significant treatment effects on waste.

[15] Average scores on the two quizzes were 1.99 and 1.06. This difference is significant at the 1% level ($t = 8.10$).

[16] This effect seems to reflect the difficulty of the quiz rather than performance on the quiz (i.e., the number of correct answers). These two variables (difficulty of quiz and performance on quiz) are highly correlated, but if both are included in a regression such as Model 2, the coefficient for taking the more difficult quiz remains statistically significant whereas the test score has a tiny and statistically insignificant estimate.

experiments would be needed to establish whether this represents a true effect or a statistical anomaly.[17]

6. What Makes Elected Leaders Effective?

We have found that leaders are more effective than bonus increases and that electing a leader yields leaders that are more effective. A striking aspect of our results (see Figure 1) is that elected leaders obtain nearly universal efficient coordination, particularly in the presence of increased incentives. This level of effectiveness is rare in other similar experimental studies. Thus, a more legitimate procedure for selecting a leader increases that leader's efficacy for inducing a turnaround to efficiency. An important remaining question, then, is what is the precise source through which such legitimacy operates? That is, what makes leaders—and, particularly, elected leaders—effective?

As we noted earlier, there are at least three factors that may contribute to the high degrees of effectiveness of elected leaders. First, groups may have used the information on potential leaders that was available at the time of election—accuracy of quiz responses and effort levels in Stage 2—to select abler leaders (Hypothesis 5A). In other words, elected leaders may have different observable characteristics at the time of selection, *and* these characteristics may be associated with greater efficacy. Second, the property of having been elected may lead leaders to act differently—for example, by communicating more or more persistently choosing high-effort choices (Hypothesis 5B). Finally, it is conceivable that the mere fact of being elected makes elected leaders more effective simply because followers are more likely to follow someone they elected, perhaps because they believe others are more likely to do so (Hypothesis 5C).

6.1. Who Gets Elected (and Does It Matter)?

We first consider whether groups systematically elected certain kinds of leaders and then whether these leaders tended to be more effective. The only two elements that could influence how many votes an individual received in the elected leader conditions were the two pieces of information given to all voters: how many trivia questions a subject answered correctly in Stage 1 and a subject's average effort level over block 1 (Stage 2).

The results show that, indeed, leaders in the elected leader conditions had both higher average quiz scores (2.0 versus 1.1) and Stage 2 effort (14.9 versus 10.4) than nonleaders. In the random leader conditions, we would expect leaders and nonleaders to have approximately the same characteristics, which is what we observe; leaders and nonleaders answered roughly the same number of quiz questions correctly (1.4 versus 1.5, respectively) and exerted similar effort, on average, in Stage 2 (11.5 versus 11.0). The difference between leaders and nonleaders is significant for both variables under elected leaders, but not with random leaders.[18]

However, the preceding observations do not explain the performance differences in Stage 4. Looking at the raw data, groups with leaders who answered more than the median number of questions in Stage 1 had almost the same average minimum effort in Stage 4 as those with leaders below the median (28.3 versus 28.4).[19] Things look more promising if we look at effort, as groups with leaders who had more than the median effort in Stage 2 had slightly higher average minimum effort in Stage 4 than groups with leaders below the median (29.5 versus 27.2). However, groups that achieved a minimum effort above zero at some point in rounds 4–6, the final three rounds of Stage 2, all end up having leaders with Stage 1 effort above the median. It seems likely that the positive effect of having a leader with high effort in Stage 2 is due to reverse causality, as groups that do well in Stage 2 tend to do well in Stage 4 *and* have individuals serving as leaders who have high effort in Stage 2.

Model 3 in Table 4 formally examines the impact of leader characteristics. This is a modified version of Model 1, adding controls for the leader's number of correct answers in Stage 1 and average effort across Stage 2. These variables are set equal to 0 in groups without leaders and are demeaned for groups with active leaders, with quiz scores demeaned separately for each quiz type. The parameter estimate for the leader's quiz performance is *negative* and falls just below significance at the 5% level, whereas the leader's effort in Stage 2 has no detectable effect if we control for the group's Stage 2 outcomes.[20] Controlling for the characteristics of the leaders, the effect of having an

[17] In subsequent regressions (Models 3 and 4), we did *not* include a control for quiz difficulty. Adding this variable does not affect our qualitative conclusions from these regressions.

[18] For each variable, we ran a regression with fixed effects at the group level and indicator variables for leaders in the randomly selected and elected leader treatments. The resulting parameters capture the difference between leaders and the average group member for the variable and treatment in question. See Table A.1 in the appendix for results. The correlation between quiz score and Stage 2 effort is small and not statistically significant, so the relationship between being a leader and quiz score cannot be attributed to an indirect effect of Stage 2 effort (or vice versa).

[19] The median is calculated across the 40 managers, not the 295 subjects. The median is based on the demeaned quiz score to prevent assigning almost all the managers who took the harder quiz to the lower group and the managers who took the easier quiz to the higher group.

[20] If we interact the leader characteristics with the treatment, the negative effect of the leader's quiz performance comes primarily from randomly assigned leaders. Including interactions does not change our conclusion that differing leader characteristics cannot explain the effect of having an elected leader.

elected leader becomes stronger in both blocks 2 and 3, achieving significance at the 10% level in block 2 and at the 5% level in block 3. Thus, elected leaders do well *in spite* of who is elected to be a leader.[21]

Conclusion 3. *Groups in the elected leader treatment are more likely to appoint leaders who score higher on the Stage* 1 *quiz and who expend more effort in Stage* 2, *but this does not explain the difference in minimum effort between the elected and random leader conditions. These results are inconsistent with Hypothesis* 5*A.*

6.2. Do Elected Leaders Behave Differently?

The previous subsection focused on whether electing the leader led to different types of subjects becoming leaders. This section instead asks whether being elected generates different behavior by leaders.

One possibility is that the observed treatment effect could be due to elected leaders choosing higher effort in Stage 4 (relative to nonleaders) than randomly appointed leaders, thus raising the minimum effort in their groups primarily through the effect of their own effort choices. Leaders in our experiment cannot lead by example (i.e., Güth et al. 2007, Gächter et al. 2012, Cartwright et al. 2013) in the standard sense, since their actions cannot be observed by others (recall that only the minimum effort for a group is reported in the feedback). However, consistent choice of a high effort level by an individual makes it easier for his or her group to coordinate at an efficient equilibrium. If elected leaders feel greater obligation to their group than randomly appointed leaders, they may be more willing to risk a low personal payoff in exchange for increasing the likelihood of successful coordination by their group.

Our results show that leaders do tend to choose higher effort levels than their followers. In round 7—the first point in the experiment at which leadership could be exercised—leaders have somewhat higher average effort than their followers (35.0 versus 30.1). This difference narrows quickly but persists throughout block 2, with an average effort of 33.2 for leaders versus 31.3 for followers. No difference is observed in block 3 (30.6 versus 30.5).[22] However, the difference between leaders and followers does not depend on whether leaders are randomly selected or elected; the difference between the average effort of leaders and followers in round 7 is 5.5 with randomly selected leaders as opposed to 4.3 with elected leaders.[23]

Conclusion 4. *Leaders tend to choose higher effort levels than their followers, but this does not differ between randomly selected and elected leaders and hence cannot explain the greater efficacy of elected leaders. This result eliminates one channel for Hypothesis* 5*B.*

Another obvious way in which leaders could differ across conditions is in the messages they send. Communication is the primary instrument leaders have available to directly influence their group members' choices. In the remainder of this section, we use content analysis to discuss whether differing message content could explain the differing efficacy of randomly selected and elected leaders.

To study leaders' communication strategies, we quantified the content of their messages by developing and implementing a coding scheme to identify properties of each message sent by the leaders. The purpose of this coding was not to capture every detail of what was said by leaders but rather to identify broad themes in the messages. After reading through the messages, the three authors agreed on the general types of messages that seemed common. The general structure of the game and the coding categories were then explained separately to two research assistants (RAs), who coded all of the messages.[24]

Coding was binary—a message was coded as a 1 if it was deemed to contain the relevant category of content and as 0 otherwise. We had no requirement on the number of codings assigned to a message; a coder could select as many or few categories as he or she deemed appropriate. The two coders coded all of the messages independently. No effort was made to force agreement among coders—the goal was to have two independent readings of each message so that any coding errors were uncorrelated. Unless otherwise noted, we use the average coding across the two coders.

Table 5 provides descriptions of all the coding categories, the percentage of messages that were coded for each category (frequency), and kappa, a common measure of intercoder agreement (Cohen 1960), for

[21] We have also looked at controlling for the number of votes received by an elected leader. The distribution of votes received by winners was the following: two votes (13 observations), three votes (4 observations), and four votes (3 observations). Adding a control for the number of votes received by an elected leader to Model 3, the parameter estimate (which equals 0.435 with a robust standard error of 0.461) is small and not close to statistical significance.

[22] Regression analysis similar to Model 1 in Table 4 confirms that leaders exert significantly more effort than followers in round 7 as well as throughout block 2 (but not block 3). This specification uses effort as the dependent variable, clusters at the group level, and adds a binary variable for leaders interacted with the variable for the block (in regressions that include data for rounds 7–18 rather than just round 7).

[23] Across block 2 these differences are 2.3 and 1.4, respectively. Regression analysis confirms the lack of significant differences.

[24] The RAs were not told any hypotheses that we had about the data, nor were they asked to make any determination about why particular messages were sent or what the effect of these messages might have been. Instead, we stressed to the RAs that their sole job was to accurately represent what the leaders had said.

Table 5 Categories of Messages for Content Analysis

Category	Description	Frequency	κ
1a	Suggested effort level 0	0.010	0.798
1b	Suggested effort level 10	0.008	1.000
1c	Suggested effort level 20	0.002	1.000
1d	Suggested effort level 30	0.005	0.799
1e	Suggested effort level 40	0.497	0.900
1f	Ambiguous suggestion, positive but not specific	0.011	0.173
2	Explanation for suggested effort	0.112	0.791
3	Appeals to mutual trust	0.014	0.776
4	Positive feedback about previous outcome	0.255	0.715
5	Negative feedback about previous outcome	0.035	0.817
6	Social banter (unrelated to game)	0.178	0.609

each category.[25] With one exception, the kappa values show substantial agreement between the coders.

Most of the categories are self-explanatory, but there are a few cases requiring additional clarification. Category 1f ("ambiguous suggestion, positive but not specific") applies to messages where leaders called for a positive effort level without specifying exactly what was to be chosen. This type of comment was rare, largely reserved for cases where firms had not coordinated on the maximum effort and the leader was simply urging the employees to choose more effort. It is the one category with a poor kappa, reflecting its vague nature.[26] Category 2 was coded for *any* explanation of why the suggested effort should be chosen, but most of these explanations consisted of appeals to the mutual benefits of coordination (e.g., "Great! Remember 40 hours = the max amount of cash in each of our pockets when we're done!"). Leaders frequently sent messages that were unrelated to the game being played, which were coded under category 6. These generally occurred in later rounds for groups that had converged to equilibrium. Common topics included jokes, discussion of how the leader planned on spending his or her earnings, and complaints that people were taking too long to make their decisions.

Most messages (88%) contained content that was relevant to play of the game. The majority of these were suggestions about what effort should be chosen. The paucity of explanations for the suggested effort level is largely driven by convergence to equilibrium. Explanations were fairly common in the early rounds (44% of messages sent in rounds 7 and 8) but rapidly died out. Employees received an explanation at least once in 73% of the groups. Positive feedback about the previous round's outcome (e.g., "You guys just did a great job! I'm so proud of you") also died out over time. Restricting attention to cases where employees had coordinated on effort level 40 in the previous round, the frequency of positive feedback fell from 56% of messages in block 2 to 34% in block 3. Once firms had converged to the efficient outcome, many leaders no longer felt the need to do much beyond telling employees to keep choosing the same effort level.

Figure 3 examines whether elected leaders communicate differently from randomly selected leaders. The upper left panel shows the frequency of relevant messages (i.e., messages that are relevant to play of the game, rather than social banter) by leader type (randomly selected versus elected). The data are grouped into three-round blocks to reduce noise in the graph. The remaining three panels graph the frequency of the three most common coding categories, other than social banter, as a function of the leader's type. These frequencies are given as percentage of total observations, including cases in which no message was sent, rather than the percentage of messages.

Looking at the upper left panel of Figure 3, elected leaders are more likely to send relevant messages than are randomly selected leaders in rounds 7–9, the critical early rounds where the largest changes in minimum efforts take place. This difference narrows somewhat in later rounds but is present throughout the experiment. Looking at specific types of comments, we see the same pattern. In rounds 7–9, elected leaders are more likely to suggest choice of effort level 40, provide an explanation for a suggested effort level, and give positive feedback. The difference between elected and randomly selected leaders consistently narrows over time but is always present.

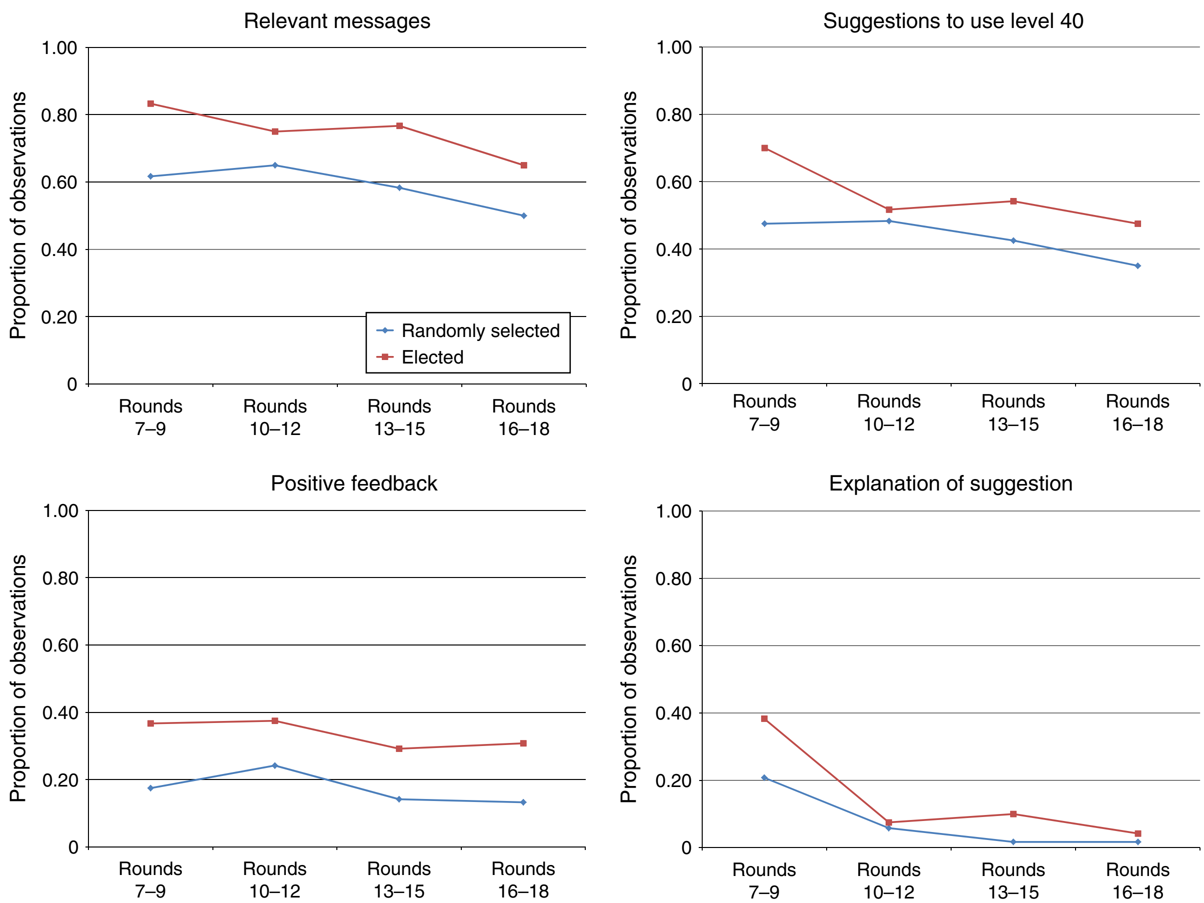
The preceding observations suggest that the superior performance of elected leaders may be due to different communication patterns. That is, in our experiment, the mechanism through which the leader selection procedure increases legitimacy and effectiveness appears to operate, at least partly, through an influence on the behavior of leaders themselves. However, before we can reach this conclusion, two things need to be established.

First, we have to provide statistical evidence that elected leaders send significantly more relevant messages than randomly selected leaders. This is complicated, because minimum effort feeds back into what messages are sent. The regressions in Table 6 address this issue. The data are drawn from rounds 7–9 of sessions with leaders. With the exception of Model 4, the dependent variable measures whether the leader suggested effort level 40 and/or provided an explanation for this suggestion. Specifically, the dependent variable

[25] Cohen's kappa equals $(p(a)-p(e))/(1-p(e))$, where $p(a)$ is the observed probability that the two coders agree and $p(e)$ is the probability of agreement by chance. Perfect agreement yields a kappa of 1, whereas kappa equals 0 if the two coders agree no more than would occur by chance.

[26] The coders had difficulty interpreting this category, with comments assigned to this category by one coder often coded under categories 4 or 6 by the other.

Figure 3 (Color online) Frequency of Relevant Messages and Categories by Leader Type



is the sum of the codings for categories 1e and 2.[27] Independent variables included in all regressions are a binary variable for whether the leader was elected, one for whether the bonus increased for Stage 4, and time-period dummy variables (not reported to save space). The regressions in Table 6 are ordinary least squares models.

The regression results strongly support our observation from Figure 3 that elected leaders send more relevant messages in rounds 7–9. Model 1 only includes the main controls. The coefficient for elected leaders is significant at the 5% level and the coefficient for a bonus increase is not statistically significant. Causality is tricky here because subjects are more likely to send messages coded for categories 1e and 2 when the minimum effort was high in the previous round. If elected leaders are unusually successful for reasons unrelated to their communication, they will tend to have higher lagged minimum efforts and hence send more relevant messages, ceteris paribus. Model 2

[27] There are three coding categories that are relevant to the games and commonly utilized by leaders (see Table 5): suggestions of effort level 40 (category 1e), explanations for this suggestion (category 2), and positive feedback (category 4). Messages coded for category 2 in rounds 7–9 are a strict subset of those coded for category 1e. Using the sum captures the idea that messages providing an explanation are a more intense version of suggesting effort level 40. (Regressions looking at the effect of messages find that both types of message lead to increased effort.) We did not include category 4 because this is almost entirely a reaction to a good outcome in the previous round, making causality murky.

Table 6 Regressions on Message Use, Coding by Leader Type

	Model 1	Model 2	Model 3	Model 4	Model 5
IV	No	No	Yes	Yes	Yes
Dependent variable	1e, 2	1e, 2	1e, 2	All relevant	1e, 2
Elected Leader	0.400**	0.361**	0.327***	0.164**	0.341***
	(0.164)	(0.140)	(0.113)	(0.076)	(0.096)
Bonus Increase	0.183	0.141	0.103	0.075	0.091
	(0.164)	(0.134)	(0.113)	(0.080)	(0.104)
Lagged Minimum Effort		0.011**	0.021***	0.017**	0.028**
		(0.005)	(0.006)	(0.008)	(0.014)
Leader's Quiz Score					−0.068
					(0.042)
Leader's Average Effort Stage 2					−0.001
					(0.007)

Notes. All regressions contain 120 observations from 40 groups and use only data from rounds 7–9. IV indicates instruments for *Lagged Minimum Effort*. All models include unreported time-period dummy variables. Standard errors (in parentheses) are corrected for clustering at the group level.

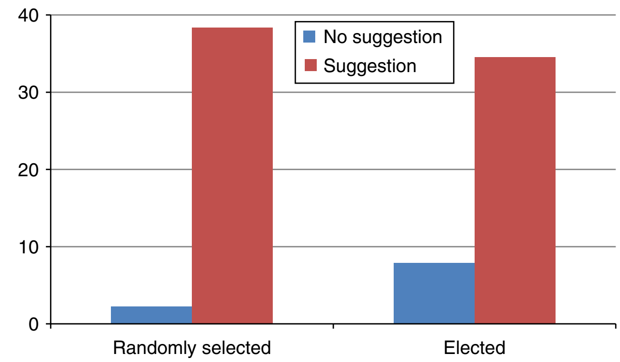
***Statistically significant at the 1% level; **statistically significant at the 5% level; *statistically significant at the 10% level.

addresses this by adding the lagged minimum effort as an independent variable. The estimate for the lagged minimum effort is positive and significant at the 5% level, but the effect on the estimates for elected leaders and bonus increases is minimal. Of course, Model 2 may still not fully address the issue of whether the effect of elected leaders is due to changes in the lagged minimum effort, because the lagged minimum effort is arguably endogenous.[28] Model 3 therefore instruments for the lagged minimum effort. The instrumental variables are the maximum of the minimum effort in rounds 4–6 for the three nonleaders in the group and the average minimum effort in rounds 1–6 for the three nonleaders. Once again, the estimates for the lagged minimum effort and elected leaders are significant, whereas the estimate for a bonus increase is not.[29]

Model 4 explores whether our conclusions are sensitive to our choice of a dependent variable. The specification is the same as in Model 3, except the dependent variable is whether any relevant category was coded (i.e., any category except for social banter). As with the narrower measure, relevant messages are more likely with higher lagged minimum efforts and for elected leaders. Model 5 explores the possibility that the greater frequency of relevant messages with elected leaders is due to the type of person who tends to be elected as a leader. Specifically, as we showed earlier, elected leaders tend to do better at the Stage 1 quiz and to have exerted more effort in Stage 2. Model 5 modifies Model 3 by controlling for these two leader characteristics. Neither of these new variables has a significant effect on the number of messages sent in categories 1e and 2, nor is the effect of having an elected leader much changed. This suggests that the fact that leaders send a greater number of messages is primarily due to their being elected, rather than the election producing leaders with different characteristics than randomly selected leaders.

Knowing that elected leaders send more relevant messages than randomly selected leaders, we next need to establish that this can explain the increase in minimum effort levels with elected leaders. Specifically, we need to show that minimum effort levels are affected by the leader's messages and that this explains the effect of having an elected leader. We therefore modified Model 1 from Table 4, our basic regression showing

Figure 4 **(Color online) Effect of Suggested Play of Level 40 by Leader Type**



Note. Observations are from rounds 7–9 with lagged minimum effort less than level 40.

the existence of a treatment effect, to account for the messages sent by leaders. As noted previously, messages are correlated with lagged outcome. To limit endogeneity concerns, the new independent variable is the sum of the codings for categories 1 and 2e *in round* 7. This is highly correlated with the use of these messages in later rounds but cannot depend on what minimum effort levels have been achieved under the leader's influence.

Even with a relatively weak measure of what messages are being sent by leaders, the result in Model 4 of Table 4 is striking. The effect of sending messages suggesting effort level 40 (and possibly explaining this suggestion) has a strong positive effect. The other treatment effects remain strong and significant, but the effect of having an elected leader disappears in terms of both statistical significance and the magnitude of the coefficient.[30] Thus, controlling for the types of messages that get sent entirely explains the difference between randomly selected and elected leaders.

Conclusion 5. *Elected leaders are more likely to send relevant messages, particularly messages suggesting and explaining use of effort level* 40. *This difference largely explains the differing performance of groups with randomly selected and elected leaders. Our findings are consistent with one of the channels described in Hypothesis* 5B.

Beyond elected leaders sending more relevant messages, it is also possible that followers might respond to these messages differently when a leader is elected, as proposed in Hypothesis 5C. Figure 4 examines this possibility. Data are taken from rounds 7–9, the first three rounds of Stage 4. This is the time frame when most changes in the effort levels are taking place and, by extension, when a leader has the greatest ability to affect outcomes. We do not include observations where

[28] Suppose that each leader has a type. If his or her type is correlated with the lagged minimum effort, the resulting estimates will be biased. Mathematically speaking, suppose the error term for individual i in round t has the form $\mu_i + \varepsilon_{it}$, where the first term is the individual-specific error term and the second term is an idiosyncratic error term. The source of concern is correlation between μ_i and the lagged minimum effort.

[29] As a different way of dealing with this problem, we also reran Models 1 and 2 using just data from round 7. The coefficient for elected leaders is positive and significant at the 10% level.

[30] This result does not depend on the specifics of how we account for relevant messages being sent. Controlling separately for categories 1e and 2 or for all relevant codings yields similar results.

the group coordinated on the efficient equilibrium in the previous round (i.e., all group members choose level 40). Groups that coordinate at level 40 virtually always stay coordinated at level 40, so there is little scope for the leader's messages to have an effect. We focus on a specific type of comment: suggestions to play level 40 (category 1e). This is the most commonly coded category and has a large impact on the group's minimum effort. Figure 4 breaks the data down by leader type (randomly selected or elected) and whether the message was coded by *either* coder for a suggestion to play level 40. Figure 4 displays the average minimum effort of the three followers for each of the resulting four cells.

Looking at Figure 4, a suggestion to play level 40 is associated with increased minimum effort for either randomly selected or elected leaders. More importantly, suggesting play of level 40 has roughly the same effect for both types of leaders. If anything, suggesting play of level 40 appears to have a slightly smaller effect with elected leaders. Thus, the positive effect of elected leaders does not seem to stem from a stronger response to the messages they send, but rather to what messages get sent.

To put the preceding analysis on firmer statistical ground, we estimated regressions of the effects of various types of messages. These are ordered probit models with the current round's minimum effort as the dependent variable and, as independent variables, binary variables for the current time period, controls for the coded content of the leader's message,[31] and the lagged minimum effort.[32] The data for this regression are firm-level data taken from rounds 7–18 with leaders (randomly selected and elected). Standard errors are corrected for clustering at the group level. The estimated effects of a suggestion to choose level 40 on minimum effort and an explanation for such a suggestion are positive and statistically significant at the 1% level. If we interact these two categories with a dummy for elected leaders, the impact of neither category differs significantly by leader type.[33]

Conclusion 6. *Relevant messages from leaders are no more effective for elected leaders than randomly selected leaders. Our findings are inconsistent with Hypothesis* 5C.

7. Final Remarks

The purpose of this paper is to study the effectiveness of leadership for inducing organizational change and compare it to the impact of changes in incentives. By varying the source of a leader's position from one that is arbitrary to one that involves the direct approval of followers, our experiment varies leaders' legitimacy and allows us to investigate how this affects leaders' effectiveness.

The presence of a leader turns out to have a stronger positive effect on performance than an increase in financial incentives. Moreover, leaders that are elected are more effective than randomly selected leaders, a condition we use as the benchmark to control for the pure effect of the presence of a leader.

We also study why elected leaders are more effective. Elected leaders score better than randomly selected leaders on a task on which the election is based and also exert more effort after becoming leaders, but this does not account for their effectiveness. The mechanism through which elected leaders motivate their partners in the group to reach higher performance levels is through sending messages relevant to the play of the game. Importantly, the effectiveness of relevant messages does *not* differ between elected and randomly selected leaders. Elected leaders send relevant messages more frequently and that is why they are more effective.

Economists are trained to focus on the importance of incentives and the deep structure of games being played within organizations. However, our work stresses the "soft" parts of an organization. Having good incentives to coordinate yields efficiency improvements, but not as much as having a leader. What the leader says and how the leader is chosen are critical elements for how an organization will perform. As such, our work adds to the growing experimental literature stressing the importance of communication.[34] We also hope our work illustrates the value of marrying the analytical tools and formalism of economics with topics that have been considered more the province of management departments.

[31] These controls are limited to message categories coded in at least 10% of the observations (categories 1e, 2, 4, and 6; see Table 6). None of the omitted categories comes close to the 10% threshold, and all are sufficiently rare such that we cannot measure what, if any, effect they might have.

[32] Lagged minimum effort is included to control for omitted variable bias, as the content of messages sent in round t depends on the minimum effort in round $t-1$. Because the lagged dependent variable is included as an independent variable, parameter estimates are interpreted as measuring the effect on *changes* in the minimum effort rather than the effect on levels. Inclusion of the lagged dependent variable dictates that treatment dummies should *not* be included since these are highly correlated with the lagged dependent variable. Their inclusion does not substantively change the results.

[33] The parameter estimates for categories 1e and 2 are 0.827 and 1.167 with standard errors of 0.326 and 0.533, respectively. Both estimates are significant at the 5% level. The parameter estimates for the two interaction terms are -0.084 and 0.098 with respective standard errors of 0.384 and 0.575. Neither interaction term is statistically significant. The effects of categories 4 and 6 remain insignificant.

[34] More recent papers that emphasize the interaction between communication and economic outcomes include Charness and Dufwenberg (2006) and Charness et al. (2013) on contracting, Cooper and Kühn (2014) on firm collusion, and Brandts et al. (2013) on leadership and cooperation in teams.

The result that elected leaders are more effective than those appointed at random has important implications for the selection of leaders in firms. Indeed, whereas considerable attention is devoted to figuring out how to select the "right" people as leaders (e.g., Stoddard and Wyckoff 2009)—something that is relatively unimportant in our experiment—we highlight the importance of the *process* of selecting leaders. Follower participation is sometimes credited as a means for obtaining "buy in" by those who are to be led—but, in our case, the benefits of such follower involvement end up being at least partly the result of the effect of the selection process on leaders themselves. That is, returning to the claim from Tyler (2006) that legitimacy is "a psychological property of an authority, institution, or social arrangement that leads those connected to it to believe that it is appropriate, proper and just" (p. 375), our results suggest that legitimacy may affect the perceptions and actions of leaders themselves. This creates a new channel for enhancing the effectiveness of leaders.

More broadly, we interpret the results of our experiment as evidence that the method of leader selection affects leaders' legitimacy and hence the degree to which they consider themselves responsible to lead their groups through a turnaround. Our findings complement previous work illustrating the effects of legitimacy on decision making in groups. In both Sutter et al. (2010), where voting on institutions leads to more cooperation, and Levy et al. (2011), who found that leader platforms that have been selected by voting have a positive impact on cooperation, voting has a direct effect on the actions of those voting. In our work, voting does not affect the performance of group members directly but instead operates indirectly by making those selected through a voting procedure, i.e., group leaders, change their behavior and be more proactive in influencing followers. Thus, we provide a new account, supported by empirical evidence, for how legitimacy can affect outcomes.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mnsc.2014.2021.

Acknowledgments

Sean Collins, Glenn Dutcher, Cortney Rodet, and Krista Jabs Saral provided excellent research assistance. Jordi Brandts acknowledges financial support from the Spanish Ministerio de Ciencia e Innovación [Grant ECO2011-29847-C02-01], the Generalitat de Catalunya [Grant 2009 SGR 820], the Barcelona GSE Research A39 Recognition Program, and Consolider-Ingenio [Grant CSD2006-00016]. David Cooper acknowledges support from the National Science Foundation [Grant SES-1127704]. Roberto Weber acknowledges support from the Swiss National Science Foundation [Project 100018_140571].

Appendix

Table A.1 Characteristics of Leaders vs. Followers

Characteristic	Treatment	Leader dummy	Standard error
Quiz score	Elected	0.813***	0.233
	Random	−0.113	0.295
Stage 2 effort	Elected	4.50***	1.33
	Random	−0.46	1.51

Notes. All regressions contain 100 observations from 20 groups. All regressions include fixed effects at the group level and report robust standard errors.

***Statistically significant at the 1% level; **statistically significant at the 5% level; *statistically significant at the 10% level.

References

Arbak E, Villeval M-C (2013) Endogenous leadership: Selection and influence. *Soc. Choice Welfare* 40(3):635–662.

Bass BM, Bass R (2008) *The Bass Handbook of Leadership: Theory, Research and Managerial Applications* (Free Press, New York).

Blume A, Ortmann A (2007) The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria. *J. Econom. Theory* 132(1):274–290.

Brandts J, Cooper DJ (2006a) A change would do you good: An experimental study of how to overcome coordination failure in organizations. *Amer. Econom. Rev.* 96(3):669–693.

Brandts J, Cooper DJ (2006b) Observability and overcoming coordination failure in organizations: An experimental study. *Experiment. Econom.* 9(4):407–423.

Brandts J, Cooper DJ (2007) It's what you say not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure. *J. Eur. Econom. Assoc.* 5(6):1223–1268.

Brandts J, Güth W, Stiehler A (2006) I want YOU! An experiment studying the selection effect when assigning distributive power. *Labour Econom.* 13(1):1–17.

Brandts J, Rott C, Solà C (2014) Not just like starting over—Leadership and revivification of cooperation in groups. Barcelona GSE Working Paper 775, University of Barcelona, Barcelona, Spain.

Brandts J, Cooper D, Fatas E, Qi S (2013) Stand by me: Help and commitment in experimental coordination games. Working paper, Florida State University, Tallahassee.

Cartwright E, Gillet J, Van Vugt M (2013) Leadership by example in the weak-link game. *Econom. Inquiry* 51(4):2028–2043.

Charness G, Dufwenberg M (2006) Promises and partnerships. *Econometrica* 74(6):1579–1601.

Charness G, Du N, Yang C-L, Yao L (2013) Promises in contract design. *Eur. Econom. Rev.* 64(November):194–208.

Chaudhuri A, Paichayontvijit M (2010) Recommended play and performance bonuses in the minimum effort coordination game. *Experiment. Econom.* 13(3):346–363.

Cohen J (1960) A coefficient of agreement for nominal scales. *Educational Psych. Measurement* 20(1):37–46.

Cooper DJ, Kühn K-U (2014) Communication, renegotiation and the scope for collusion. *Amer. Econom. J.: Microeconom.* 6(2):247–278.

DelliCarpini MX, Cook FL, Jacobs LR (2004) Public deliberation, discursive participation, and citizen engagement. *Annual Rev. Political Sci.* 7:315–344.

Farrell J, Rabin M (1996) Cheap talk. *J. Econom. Perspect.* 10(3):103–118.

Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Experiment. Econom.* 10(2):171–178.

Gächter S, Nosenzo D, Renner E, Sefton M (2012) Who makes a good leader? Cooperativeness, optimism and leading-by-example. *Econom. Inquiry* 50(4):953–967.

Greiner B (2004) An online recruitment system for economic experiments. Kremer K, Macho V, eds. *Forschung und wissenschaftliches Rechnen: Beiträge zum Heinz-Billing-Preis 2003*, GWDG Bericht, Vol. 63 (Gesellschaft für wissenschaftliche Datenverarbeitung, Göttingen, Germany), 79–93.

Güth W, Vittoria Levati M, Sutter M, van der Heijden E (2007) Leading by example with and without exclusion power in voluntary contribution experiments. *J. Public Econom.* 91(5–6):1023–1042.

Hamman J, Rick S, Weber RA (2007) Solving coordination failure with "all-or-none" group-level incentives. *Experiment. Econom.* 10(3):285–303.

Hamman J, Weber R, Woon J (2011) An experimental investigation of electoral delegation and the provision of public goods. *Amer. J. Political Sci.* 55(4):738–752.

Hume D ((1748) 1960) Of the original contract. *Social Contract: Essays by Locke, Hume, and Rousseau*, with an introduction by Baker E (Oxford University Press, London), 145–166.

Knez M, Simester D (2002) Form-wide incentives and mutual monitoring at Continental Airlines. *J. Labor Econom.* 19(4):743–772.

Kocher MG, Pogrebna G, Sutter M (2013) Other-regarding preferences and management style. *J. Econom. Behav. Organ.* 88(April):109–132.

Levati MV, Sutter M, van der Heijden E (2007) Leading by example in a public goods experiment with heterogeneity and incomplete information. *J. Conflict Resolution* 51(5):793–818.

Levy DM, Padgitt K, Peart SJ, Houser D, Xiao E (2011) Leadership, cheap talk and *really* cheap talk. *J. Econom. Behav. Organ.* 77(1):40–52.

Lipset SM (1959) Some social requisites of democracy: Economic development and political legitimacy. *Amer. Political Sci. Rev.* 53(1):69–105.

Moxnes E, Van der Heijden E (2003) The effect of leadership in a public good experiment. *J. Conflict Resolution* 47(6):773–795.

Potters J, Sefton M, Vesterlund L (2007) Leading-by-example and signaling in voluntary contribution games: An experimental study. *Econom. Theory* 33(1):169–182.

Rivas MF, Sutter M (2011) The benefits of voluntary leadership in experimental public goods games. *Econom. Lett.* 112(2):176–178.

Simon HA (1953) Notes on the observation and measurement of political power. *J. Politics* 15(4):500–516.

Stoddard N, Wyckoff C (2009) *The Right Leader: Selecting Executives Who Fit* (John Wiley & Sons, Hoboken, NJ).

Sutter M, Haigner S, Kocher M (2010) Choosing the stick or the carrot? Endogenous institutional choice in social dilemma situations. *Rev. Econom. Stud.* 77(4):1540–1566.

Tyler TR (2006). Psychological perspectives on legitimacy and legitimation. *Annual Rev. Psych.* 57:375–400.

Van Huyck JB, Battalio RC, Beil RO (1990) Tacit coordination games, strategic uncertainty, and coordination failure. *Amer. Econom. Rev.* 80(1):234–248.

Weber M (1948) The meaning of discipline. Gerth HH, Mills CW, eds. and trans. *From Max Weber: Essays in Sociology* (Routledge, New York), 253–266.

Weber R, Camerer C, Rottenstreich Y, Knez M (2001) The illusion of leadership: Misattribution of cause in coordination games. *Organ. Sci.* 12(5):582–598.