Effective leadership across economic contexts

David J. Cooper [a], Giovanna d'Adda [b], Roberto A. Weber [c],*

[a] University of Iowa, United States, and University of East Anglia, England
[b] Università Degli Studi di Milano and RFF-CMCC European Institute on Economics and the Environment, Italy
[c] University of Zurich, Switzerland

A R T I C L E   I N F O

A B S T R A C T

We use a laboratory experiment to study how leaders affect workers' productivity across economic incentive contexts. In four-person groups, three group members work on a production task, with a fourth member potentially serving as a leader. We vary the economic context by changing how worker pay is determined as a function of worker outputs, comparing Revenue Sharing, Weak Link or Tournament incentives while holding constant the activity performed by workers and the incentives for leaders. A second treatment varies whether groups have Active Leaders who can exert influence through messages to workers or Passive Supervisors who exert no influence. The average effect of having an Active Leader on group output is large only under Weak Link incentives. Across all incentive contexts, we find a positive correlation between the productivity increase in output produced by an Active Leader and independent ratings of leader quality based on measures from leadership research. The nature of leaders' communication varies across incentive contexts, with comparisons between workers most common under Tournament incentives and messages about group earnings, which speak to social considerations, most common with Weak Link incentives.

Introduction

Economic perspectives on leadership in organizations have primarily viewed the role of leaders through a transactional lens, whereby managers, or "principals," design contracts and incentives to motivate the provision of effort from employees, or "agents" (Grossman & Hart, 1983; Prendergast, 1999; Laffont & Martimort, 2009). This approach—based primarily on formal theoretical analyses of how utility maximizing agents respond to varying incentive contexts—yields valuable insights into how contract design can be used to increase workers' effort provision and firm profitability. Empirical studies often, though not always, find support for the predictions of contract theory, both in laboratory experiments where researchers can carefully control the precise incentives and all features of the production context (Bull, Schotter, & Weigelt, 1987; Nalbantian & Schotter, 1997; Carpenter, Matthews, & Schirm, 2010) and in more complex settings outside the laboratory (Lazear, 2000; Nagin, Rebitzer, Sanders, & Taylor, 2002; Bandiera,

Barankay, & Rasul, 2013; Delfgaauw, Dur, Sol, & Verbeke, 2013). Under this perspective, good "leadership" is often evaluated as reliance on good "management practices" that emphasize the use of incentives and explicit performance targets (Bloom & Van Reenen, 2007).

Economists have traditionally paid less attention to other ways in which leaders can motivate followers to exert productive effort (Zehnder, Herz, & Bonardi, 2017), for example, through transformational or charismatic influences (Howell & Avolio, 1993; Bass, 1990; Antonakis, 2012). The idea that a leader, through words alone, can motivate and inspire followers to take costly actions is relatively new to economics, though recent research has started to investigate the conditions under which such forms of leadership are effective (Brandts & Cooper, 2007; Brandts, Cooper, & Weber, 2015; d'Adda, Darai, Pavanini, & Weber, 2017; Boulu-Reshef, Holt, Rodgers, & Thomas-Hunt, 2020; Antonakis, d'Adda, Weber, & Zehnder, 2022).[1] Importantly, such studies typically hold the incentive context confronting followers fixed—or vary it only slightly—while studying the impact of having a

---

* Corresponding author.
  E-mail address: roberto.weber@econ.uzh.ch (R.A. Weber).

[1] Earlier work in economics considered the possibility that communication or visible actions from a leader with private information about the returns to effort can encourage followers to exert higher effort by providing costly and credible signals about the returns from doing so (Hermalin, 1998; Potters, Sefton, & Vesterlund, 2007).

leader who communicates with followers. Thus, they generally leave unanswered the question of whether the kind of leadership that is effective in one particular context can also succeed when the economic environment confronting followers is different.[2] Given that organizational economics prescribes alternative incentive structures for different production environments, it is important to understand the degree to which the same form of leadership can succeed at motivating followers as the incentives they face vary.

We conduct a laboratory experiment in which we test the effectiveness of a very simple form of leadership on the effort and productivity of a group of followers working on a production task under varying performance incentives. These workers can exert effort to produce higher levels of output. A "leader" in our study is a participant randomly assigned the power to send messages to followers. Importantly, we hold the incentives for the leader constant across all contexts—a leader always wants more effort and higher output from followers, reflecting, for example, a situation in which a firm benefits from greater worker output and incentivizes a leader to pursue this objective.

Our primary treatment variation is in the economic context, or the incentive system confronting followers. We vary the economic context by considering three incentives for group production commonly studied in the organizational economics literature (Camerer & Weber, 2013). These affect how group workers' productivity translates into individual payments, varying the degree of competition and complementarity between workers' efforts. Specifically, *revenue sharing* incentives divide a bonus linked to collective output evenly among workers, thus making workers' efforts substitutes. This incentive scheme creates potential free-rider problems, whereby an individual worker benefits from others' efforts even when that worker does not contribute to higher output. *Weak link* incentives reward workers as an increasing function of the lowest output generated by any group member. Under weak link incentives, workers benefit from increasing their own output only when doing so raises the minimum output in the group, making workers' efforts strong complements and yielding environments characterized by multiple equilibria that can sustain either high or low worker effort. Unlike under revenue sharing, where a worker producing more output always increases the bonus for all workers, under weak link incentives the bonus only increases in response to more output by the lowest performer. *Tournament* incentives reward workers based on their relative output, thereby adding an element of competition between individuals that is not present under revenue sharing or weak link incentives. Unlike under the other incentive mechanisms, under tournament incentives a worker increasing her output can only negatively impact other team members. The basic foundation of our study is thus a laboratory production task under varying incentive schemes.

To study leadership, our second treatment dimension varies the presence of Active Leaders who can direct and motivate workers through free-form messages sent to different combinations of group members throughout the work period. In order to study the role of leaders' messages while holding constant the incentives for workers created by having another individual who benefits from their output, our control is a Passive Supervisor condition, in which the fourth group member benefits from the group's output in the same manner as Active Leaders but cannot send messages to the group. The distinction between Active Leaders and Passive Supervisors thus corresponds, for example, to stakeholders (e.g., an owner or manager) who either take an active role in leading and directing a group or who take a more passive role and leave the group of workers to direct themselves. We investigate the impact Active Leaders have on the effort exerted by individual followers and, by extension, on the collective output produced by the team.

Our primary research question is whether leadership—operationalized by an Active Leader who can potentially influence group behavior and outcomes by sending messages—influences output similarly when varying the incentives confronting followers. While the interaction between leadership and context is familiar to leadership research (Osborn, Hunt, & Jauch, 2002; Antonakis, Avolio, & Sivasubramaniam, 2003; Porter & McLaughlin, 2006), our novel contribution is to study contextual variation only in followers' economic incentives, while holding constant all other features of the production environment and what leaders and followers do. For example, in all cases, workers perform exactly the same task and the presence of interdependent incentives does not create any complementarity in how a group's workers perform their work.

It is also worth noting that our methods are rooted in an (experimental) economic approach to studying leadership. Many leadership scholars are likely to find limitations in the narrow way in which we operationalize leadership, with the simple and unidimensional task performed by followers, and with our reliance on a single measure, output, as the primary criterion with which we evaluate leadership. Nevertheless, as noted elsewhere, economic approaches to leadership research provide opportunities to bring novel perspectives and methods that yield complementary evidence to traditional leadership research (Zehnder et al., 2017; Garretsen, Stoker, & Weber, 2020). As an example of the value of such complementary approaches, we use the messages sent by leaders to obtain independent measures of leadership quality, similar to those often employed in leadership research, allowing us to evaluate the concordance of such evaluations of good leadership with our behavioral outcome-based concept of leader effectiveness. As we note further in our discussion, the contributions of our work connect to the three "C's"—context, causality, and concepts—highlighted by Garretsen et al. (2020), as potential opportunities for economic approaches to contribute to leadership research.

The rest of this paper is organized as follows. The next section describes existing relevant research, focusing on studies that investigate leader effectiveness across varying economic contexts, presents the theoretical framework underlying our experiments, and uses it to develop hypotheses. After that, we describe the methods in our study and then present the results, both in terms of the aggregate treatment effects of leadership and a closer investigation of heterogeneity in leader effectiveness. The final section provides a discussion of our results and contribution and concludes.

Variation in the economic context and leader effectiveness

Related literature

Large literatures in sociology, political science, and organizational behavior investigate the role of leadership in shaping collective outcomes and the mechanisms through which such influence takes place. In economics, by comparison, a much smaller set of studies examine leadership, typically by exploring the effectiveness of narrow forms of leadership and focusing on its influence on very specific follower behaviors and outcomes in highly abstract and stylized contexts (e.g., Hermalin, 1998; Kosfeld, 2020). Where economists study richer forms of leadership and broader outcomes, there is often little attention to the precise mechanisms through which leadership exerts influence (Chattopadhyay & Duflo, 2004; Jones & Olken, 2005).

Perhaps the greatest attention to leadership in economics has come from experimental economics, where a rapidly growing number of papers use simple laboratory studies to investigate the effectiveness of leadership in specific contexts.[3] For example, several laboratory experiments study the power of leadership to promote voluntary cooperation

---

[2] Among the small number of experimental studies that investigate the interaction between leadership and variation in the economic context are Sahin et al. (2015); Kvaløy et al. (2015) and Fest et al. (2021). We discuss these studies in detail the next section.

[3] See Cooper and Hamman (2021) and Eichenseer (2023) for recent surveys of economic experiments studying leadership.

in public good games. Both leading by example (e.g., Moxnes & Van der Heijden, 2003; Gächter & Renner, 2003) and communication from a leader (Serra-Garcia, van Damme, & Potters, 2011; Antonakis et al., 2022) have been shown to increase cooperation. Another commonly studied context is weak link coordination games, in which followers face high levels of complementarity and where experiments show that communication from a leader is effective in some cases (Brandts & Cooper, 2007; Brandts et al., 2015) but not all (Weber, Camerer, Rottenstreich, & Knez, 2001). Another line of research investigates whether leaders facilitate specific kinds of unethical behavior (Ellman & Pezanis-Christou, 2010; d'Adda et al., 2017). In more natural environments, Englmaier, Grimm, Grothe, Schindler, and Schudy (2021) show that encouraging groups to appoint a leader improves performance in complex, non-routine tasks. Collectively, these studies provide clean evidence that leaders can influence follower behaviors and collective outcomes, but each individual study typically only investigates leader influence in a single narrow context.

Very few studies address the effects of leadership across economic incentive contexts— varying the incentives faced by group members to work together, while also varying the presence or nature of leadership. Sahin, Eckel, and Komai (2015) compare the effects of leadership in two laboratory games in which followers choose numbers corresponding to contributions to a group. In a weak link game, group members are rewarded based on the lowest number chosen in the group, whereas in a linear public good game they are rewarded for higher average group contributions. Two forms of leadership are considered: "exemplars" who move first in the group, thereby setting an example for other group members (see Eichenseer (2023)) and "managers" who recommend a number for others to choose. Both forms of leadership are effective in the weak link game and neither is effective in the public goods game, though the latter null effect may be due to high levels of contribution in the baseline treatment of the public goods game possibly leaving little room for improvement. Our study also investigates how follower incentives interact with leadership, though we consider additional incentive contexts, richer forms of production, and leadership through communication.

Two other related studies use field experiments to investigate the interaction between motivating workers with performance pay—i.e., piece-rate payments for higher output—and transformational messages from leaders (Kvaløy, Nieken, & Schöttner, 2015; Fest, Kvaløy, Nieken, & Schöttner, 2021). In these experiments, workers are hired to perform simple individual online tasks. The results of these studies reveal surprising patterns. Kvaløy et al. (2015), find that performance pay decreases worker productivity, contrary to standard economic predictions, but that a motivational message from a leader reverses this negative impact of performance pay. Fest et al. (2021), find that performance pay increases worker output, but that motivational messages from a leader can backfire and reduce worker productivity unless they contain a broad spectrum of charismatic leadership techniques (Antonakis, Fenley, & Liechti, 2011; Antonakis et al., 2022). These studies show that the economic incentives facing followers and how followers respond to communication from leaders may interact in interesting and unexpected ways. Our study expands this question by making comparisons across contexts with differing economic incentives.

Additionally, many studies in the broad field of leadership research acknowledge the importance of context for understanding leader influence (Fiedler, 1978; Osborn et al., 2002; Antonakis et al., 2003; Porter & McLaughlin, 2006; Purvanova & Bono, 2009). These studies also document that the same type of leadership may have different effects in different contexts. However, even within the extensive literature on leadership, there remain open questions regarding the precise ways in which context interacts with leadership characteristics to determine the influence of leadership (Oc, 2018). Our study is novel in that it considers contextual variation only in the incentive mechanism used to reward followers—i.e., in the contract that determines their individual payments, given a profile of output levels generated in the group. All other

aspects of the task performed by followers, the tools available to leaders and the value generated by workers' outputs are identical. In contrast, in much leadership research, variation in the context also typically involves more meaningful variation in the tasks performed by followers. Varying only a single element in the context is a common approach for economic experiments, exchanging richness for the ability to precisely identify what element of the context causes any changes in behavior.

A model of worker effort provision under varying incentive contexts

Given that our experimental variation in the economic context changes only how workers are paid for their output, economic theory can provide some guidance on how such variation could affect how leadership influences the effort workers provide and the resulting group productivity. The purpose of the theory developed below is to shed some light on how messages from leaders might increase the output produced by workers and how the nature of effective communication might vary as a function of the incentive scheme.

Let $x_i \geq 0$ be the effort provided by Worker $i \in \{1, 2, 3\}$ toward a productive activity that benefits a firm and let π_i be the monetary compensation received by the worker from the firm. For simplicity, assume that a worker's output is equal to that worker's effort. Let $c(x)$ be a worker's cost of effort, monetary and non-monetary—i.e., including both the psychological effort costs and any opportunity costs of forgoing outside options by continuing to work.[4]

Given that the three workers are symmetric, we can describe the theory from the point of view of Worker 1 without loss of generality. The utility for Worker 1 is given by Equation (1). This is a very simple model in the spirit of other models (e.g., Chen and Li (2009)) that incorporate other regarding preferences based on the degree to which one cares about one's group members.

$$u_1(x_1, \pi_1, \pi_2, \pi_3) = \pi_1 - c(x_1) + G_1(\pi_2 + \pi_3) \tag{1}$$

We refer to the parameter G_1 as Worker 1's general attitude towards Workers 2 and 3 and assume that $G_1 \geq 0$. G_1 can include distinct social motives, such as Worker 1's unconditional concern toward the other workers (altruism) or a sense of shared identity with the group. Thus, in the above framework, Worker 1 selects effort based on the expected personal benefits, the cost of exerting effort and weighted considerations about the impact that this effort has on the other workers' payoffs.[5]

Our experiment holds constant the task that workers perform and makes workers' effort provision independent of what other workers are doing—meaning that $c(x_1)$ does not vary across incentive treatments or depend on what other workers are doing. The only variation is in how Worker 1's own effort and the effort of Workers 2 and 3 are combined to determine the individual workers' payoffs, $\pi_1, \pi_2,$ and π_3. Thus, in terms of the above utility model, variation in our incentive treatments changes the impact that Worker 1's own effort has on the three workers' payoffs.

We used three different incentive schemes:

1) Revenue Sharing: Each worker's earnings are equal to the *average* output generated by the three workers, multiplied by a scaling term, γ_{RS}. The scaling term can be interpreted as a payrate (e.g. $2 per unit of average output). We use the scaling term, here and below, to adjust average payments to be comparable across incentive schemes. Mathematically, worker earnings are expressed as follows:

---

[4] To ensure that the worker's optimization problem is well behaved, we assume $c(x)$ is continuous, differentiable, and has strictly positive first and second derivatives. This means, in other words, that effort always has a positive marginal cost for a worker and that the marginal cost becomes larger at higher effort rates.

[5] Including the other workers' costs of effort complicates the model without affecting the resulting hypotheses.

$$\pi_{i,RS} = \gamma_{RS}\left(\frac{x_1 + x_2 + x_3}{3}\right). \tag{2}$$

Because workers benefit from output generated by other workers (i. e., workers' efforts are substitutes), a worker's effort with Revenue Sharing creates positive externalities for other workers. Conversely, due to the fact that workers only receive a share of their individual output, Revenue Sharing creates a potential incentive to free ride on the effort of the other workers (note that a worker receives the same earnings as other group members, even if her output is zero).

2) Weak Link: Each worker's earnings are increasing in the *lowest* output achieved by any single worker within the group. That is,

$$\pi_{i,WL} = \gamma_{WL} \; min(x_1, x_2, x_3) \tag{3}$$

Given that effort is costly, workers benefit from providing additional effort only if they believe others are exerting at least that amount of effort. This payoff function creates high levels of complementarity between workers' outputs, making the strategic problem facing them similar to a "weak link" coordination game, in which the "weakest" input determines the overall reward (see, for reviews, Devetag and Ortmann (2007) and Cooper and Weber (2020)).

3) Tournament: The three workers are ranked according to their output and receive fixed payments based on this rank. Specifically,

$$\pi_{i,T} = \begin{cases} \gamma_T & \text{if } x_i = max(x_1, x_2, x_3) \\ \frac{1}{2}\gamma_T & \text{if } x_i = med(x_1, x_2, x_3) \\ 0 & \text{if } x_t = min(x_1, x_2, x_3) \end{cases} \tag{4}$$

That is, the most productive worker earns a large prize, $\gamma_T > 0$, the middle producer receives half this much, and the least productive worker receives nothing. In case of a tie, the two or more workers with the same output split the corresponding payoffs evenly. Under Tournament incentives, workers are paid solely on relative rather than absolute output. This means that a worker's effort cannot increase the total earnings available to workers (as is the case under Revenue Sharing and Weak Link incentives), but instead creates a potential negative externality for other workers. Tournament incentives produce competition between workers, reflected in the determination of $\pi_{i,T}$, which provides incentives to generate higher relative output than other workers.

Our experiment also introduces—based on the experimental condition—a leader who can send messages to workers (an "Active Leader"), motivating them to exert effort. One can think of the leader as a firm owner who benefits from selling the firm's output or a manager incentivized to increase output. The leader's payment is determined identically across all contexts and given by

$$\pi_L = \gamma_L(x_1 + x_2 + x_3) \tag{5}$$

This means that the leader always benefits from increasing workers' effort and output, regardless of how workers are being paid.[6]

---

[6] We assume that leaders are not directly concerned about the payoffs of their workers, which is consistent with standard economic assumptions that firms seek to maximize expected profits. This assumption yields the property that a leader always prefers workers to exert more effort. Of course, leaders may exhibit concern for the welfare of those they lead. In our model this concern could be represented, for instance, by an additional component of the leader's utility $G_L(\pi_1 + \pi_2 + \pi_3)$ that reflects a positive concern for the overall welfare of workers (i.e. $G_L'(\pi_1 + \pi_2 + \pi_3) > 0$). If sufficiently strong, such concern might interact with the incentive context to produce circumstances in which a leader directs workers to stop working. For instance, whereas workers generally benefit collectively from additional effort under Revenue Sharing, they never do so under Tournament incentives and only do so when increased effort comes from the lowest-performing worker under Weak Link incentives. In our analysis, we investigate whether leaders request workers to stop working, and whether such requests vary across incentive contexts.

For simplicity, our model captures the varied potential content and nature of leaders' communication by assuming that an Active Leader selects a profile of messages to send to workers, $\boldsymbol{m} = \{m_1, m_2, m_3\}$, where the leader can send the same messages to all three workers ($m_1 = m_2 = m_3$) or differentiate the messages to different workers (for example, by sending one set of messages to two workers and different messages to the third worker, $m_1 = m_2 \neq m_3$). Messages can differ in their content, and thereby in whether they contain elements (e.g., charisma, leader prototypicality) likely to increase their persuasiveness to followers (Lord, Foti, & De Vader, 1984; Antonakis et al., 2011); we represent the degree to which a leader's messages contain such effective content as $L(\boldsymbol{m})$.

Having laid out this admittedly simple model, what can it say about the effects of communication from an Active Leader on Worker 1′s output? We consider three possible channels through which an Active Leader might use different kinds of messages to influence the workers' utility function shown in Equation (1) and thereby the effort exerted by workers. For each channel, we posit how leadership may influence the separate elements in Equation (1) and how such influence may vary based on the incentive context.

First, an Active Leader may offer *encouragement* that lowers the perceived marginal cost of effort, $c'(x_1)$, relative to the potential benefits from not providing additional effort. Encouragement may include statements that convince a worker that the costs of providing additional effort are not too high or that withholding additional effort has low benefits. Importantly, this channel is likely to be influential in all three contexts, as lowering the perceived costs of providing effort can increase a worker's effort provision in all cases, i.e., under Revenue Sharing, Weak Link, and Tournament incentives.

Second, an Active Leader can use *relative comparisons* to influence a worker's beliefs about others' productivity and what such beliefs imply for the worker's own *monetary* benefits from exerting additional effort. For example, under Weak Link incentives, telling Worker 1 that the other two workers have generated more output can increase the perceived return from completing an additional task, relative to cases in which Worker 1 is uncertain about others' output. This mechanism can also influence the perceived benefits of output under Tournament incentives. For example, a leader can inform Worker 1 that she is slightly ahead of or behind another worker, thereby highlighting the potential that increasing effort has for improving or maintaining Worker 1′s relative rank. Importantly, information about relative output does not directly influence a worker's expected monetary benefit from exerting additional effort under Revenue Sharing, where Worker 1 always receives a constant return $\left(\frac{\gamma_{RS}}{3}\right)$ from higher output. Thus, from the perspective of influencing beliefs about the *material* returns to effort, relative comparisons do not allow leaders to influence worker effort under Revenue Sharing, but can be helpful to the leader under Weak Link and Tournament incentives.

Finally, Active Leaders can also increase the concern that workers have for one another, G_1, by appealing to *social considerations* such as altruism, fairness or a shared sense of identity (Boehm, Dwertmann, Bruch, & Shamir, 2015). Social considerations are likely to have positive impacts on effort under Revenue Sharing and Weak Link incentives, where increasing output can produce positive impacts on other workers' payoffs. Indeed, Chen and Li (2009) and Chen and Chen (2011) demonstrate positive benefits of increased group identity in group production contexts with Revenue Sharing and Weak Link environments. In contrast, increasing G_1 under Tournament incentives is likely to lead to lower effort, given the negative externalities that Worker 1′s effort produces for other group members. Therefore, this potential channel of influence allows an Active Leader to motivate workers to exert effort under Revenue Sharing and Weak Link incentives, but is ineffective under Tournament incentives.

Hypotheses

We now state several formal hypotheses, based on the above considerations. It is worth noting that these are not the only mechanisms through which one can imagine leaders influencing effort and output, and that even these mechanisms may operate in subtly different ways than the relationships we discuss above. However, these hypotheses provide a useful framework to make precise some possible expectations about the role of leaders in our experiment and to guide our analysis of the results.

First, given that mechanisms exist through which supervisors' directives can positively influence workers' provision of effort (e.g., through encouragement, which is possible in all cases), we expect the introduction of Active Leadership by supervisors to have a (weakly) positive effect, relative to the comparable case in which the leader role is fulfilled passively.

Hypothesis 1. *Across all three incentive schemes, groups with Active Leaders will have higher average output than groups with Passive Supervisors.*

Our next hypothesis deals with the relative effectiveness of leadership across the three contexts we study. This hypothesis is based on the observation that channels of potential influence (encouragement, comparisons, and social comparisons) operating on all three elements in Equation (1) can have positive impacts under Weak Link incentives, whereas only two channels can have potential influence under the other incentive mechanisms. Whether leaders have greater influence in Revenue Sharing or Tournament incentives will depend on whether the mechanism that is potentially present only in that context (social considerations in Revenue Sharing, relative comparisons in Tournament) is more effectively employed by Active Leaders.

Hypothesis 2. *Across incentive schemes, Active Leaders will have the greatest positive influence on output under Weak Link incentives. There is no clear prediction for whether Active Leaders will have a larger effect under Revenue Sharing or Tournament incentives.*

Next, we consider potential heterogeneity in the messages leaders send to their followers. Specifically, we anticipate that leaders who employ more persuasive messages, corresponding to higher $L(\boldsymbol{m})$, are more effective at motivating worker effort.

Hypothesis 3. *Leaders whose messages contain more elements associated with persuasive leadership will have a greater positive influence on worker output under all incentive mechanisms.*

Finally, we return to the relative effectiveness, across production contexts, of the different mechanisms through which leaders can influence follower output and the resulting relative use of messages corresponding to such mechanisms. We predict that leaders confronted with motivating workers in a particular incentive context will be more likely to employ communication strategies that correspond to mechanisms that are expected to be effective in that context.

Hypothesis 4. *a) Messages that provide encouragement will be equally frequent in all three contexts. b) Messages that appeal to relative comparisons will be more frequent under Weak Link and Tournament incentives than under Revenue Sharing. c) Messages that emphasize social considerations will be more frequent under Revenue Sharing and Weak Link incentives than under Tournament incentives.*

Hypothesis 5. *a) Messages that provide encouragement will be equally effective in all three contexts. b) Messages that appeal to relative comparisons will be more effective under Weak Link and Tournament incentives than under Revenue Sharing. c) Messages that emphasize social considerations will be more effective under Revenue Sharing and Weak Link incentives than under Tournament incentives.*

Methods

Our study employs a 3 (incentive mechanism) by 2 (leader influence) design that allows us to study how leadership affects worker productivity across the incentive contexts. The experiment involves several periods in which group members engage in a real-effort work task. Our focus is on periods in which subjects are placed in four-person groups consisting of three "workers" who work on the task and a fourth subject who either observes the workers' activities or can direct the workers by sending messages. For consistency, our instructions always refer to the fourth group member as the "supervisor" (who could either be "active" or "passive"), but here we highlight the distinction in roles by referring to "Passive Supervisors" (who can only observe workers' output) and "Active Leaders" (who can send messages to workers).

The work task

Work in our experiment involves a task in which subjects count the number of zeros in 10x10 tables of randomly ordered zeros and ones (Abeler, Falk, Goette, & Huffman, 2011). Subjects have eight minutes per period to count as many tables as possible. We measure a subject's output, our primary measure of productivity, by the number of tables completed during this period.

This task is simple, easily explained to subjects, and does not require pre-existing knowledge. It is dull and serves no obvious purpose, making it unlikely that subjects are motivated to complete the task for its own sake. Fig. 1 shows a table as presented on subjects' computer screens. Subjects enter their answer in the space to the right of the table, then click "OK" to proceed to the next table. For each table, subjects have three opportunities to give the correct answer. After an incorrect answer, subjects must wait twenty seconds before they can submit a new answer. After three incorrect answers, a table is counted as incorrect and subjects start a new table. Given the complexity of the tables and the twenty-second delay after an incorrect answer, a strategy of rapid guessing is unlikely to be profitable. Good performance requires that subjects attend to the task.

Stages and periods

The experiment consists of seven periods split across three stages. Stage 1 has a single period (Period 1), Stage 2 has two periods (Periods 2 and 3), and Stage 3 contains the remaining four periods (Periods 4 – 7). Table 1 provides a description of the activities and measures collected across different parts of the experiment. Subjects are paid based on their earnings from a single randomly selected period.

Stage 1 provides a baseline measure of subjects' ability to perform the real-effort task. All subjects work on the task for a single eight-minute period and are paid a piece rate of $2 for each correct table. There are no alternative activities during this period; the software does not allow access to other windows and we restrict subjects from accessing their mobile phones. As a measure of subjects' beliefs about their relative ability, at the end of Stage 1 we ask them to guess the quartile into which their Stage 1 performance falls and reward them with $2 for a correct guess; however, subjects do not learn if their guess is correct until the end of the session.

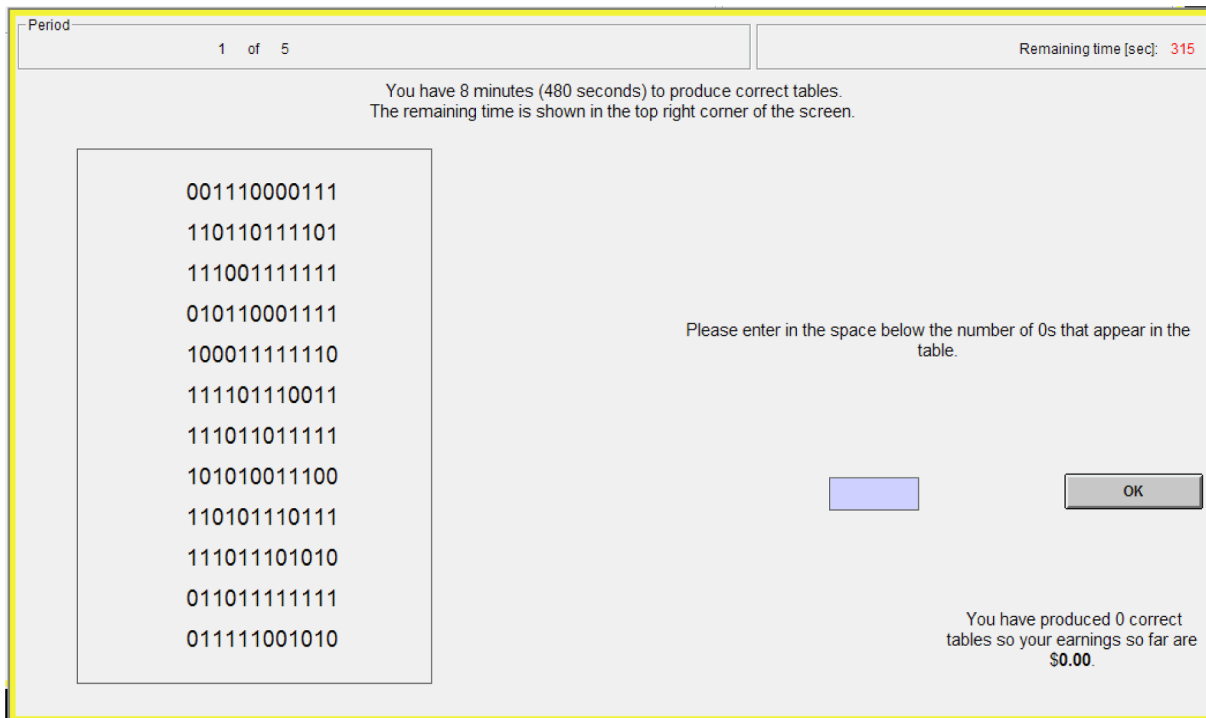For Stages 2 and 3, subjects are randomly assigned to fixed four-

Fig. 1. Interface for real-effort task.

Table 1
Description of experimental stages and activities.

Stage (Periods)	Manipulations	Activity	Measures
Stage 1 (Period 1)	None	Subjects work individually, under piece rate incentives Subjects can only work on task	Baseline ability measure Beliefs about ability

Stage 2 (Periods 2–3)	1) Incentive Scheme:
	  a) Revenue Sharing
	  b) Weak-link
	  c) Tournament
		Subjects placed in four-person groups (3 workers and 1 Passive Supervisor)
		Workers can work on task or quit (and perform outside activity)
			Baseline group performance under different incentive schemes (with a Passive Supervisor)

Stage 3 (Periods 4–7)	1) Incentive Scheme:
	  a) Revenue Sharing
	  b) Weak-link
	  c) Tournament
	2) Leader Type
	  a) Active Leader
	  b) Passive Supervisor
		Subjects remain in same four-person groups as in Stage 2
		Workers again work on task or quit
		Active leaders can send messages to workers
			Group performance under varying incentive schemes and under Active Leaders vs. Passive Supervisors
			Messages sent by Active Leaders (subsequently independently coded)

Post-experiment	None	Survey	Demographics, risk, time, and social preferences

person groups and to an incentive scheme that determines how the workers are rewarded for output in the group. Subjects remain in the same group for all remaining periods. One randomly selected group member is assigned to the leadership role, with the remaining subjects assigned the role of "workers."

In Stage 2, all groups have a Passive Supervisor who observes the workers' progress in real-time but has no mechanism for influencing their behavior. Workers perform the same task as in Stage 1, again for eight minutes. However, they are now paid based on their own and other group members' outputs—each group is assigned to one of the three incentive mechanisms: Revenue Sharing, Weak Link or Tournament, as described in Section 2.2; we provide additional details below.

Stage 3 proceeds almost identically to Stage 2. Each group contains the same three workers, who continue to work on the same task as before and are rewarded according to the same incentive mechanism as in Stage 2. The only distinction between Stages 2 and 3 is that, as a

treatment manipulation, for some groups the subject in the leader role switches from being a Passive Supervisor to being an Active Leader who can send messages to followers. We describe the implementation of this manipulation in more detail below.

Forgoing work

In Stages 2 and 3, we also provide workers the opportunity to forgo working, reflecting the possibility for workers in an organization to shirk ("quiet quitting") or pursue their personal interests. Before the beginning of each period, workers are asked if they want to begin the period by working or not working and have twenty seconds to enter and confirm a choice. If they do not make a choice during this time, the default is to start the period working. Workers can also decide to stop working at any point during the period. After entering a guess for the number of 1's in a table, either correctly or incorrectly, workers are

asked whether they want to continue working. If they choose to stop working, they must wait for twenty seconds before being able to confirm their choice. After quitting, workers have access to a limited web browser and two simple computer games.[7] If a worker quits, either at the beginning of or during a period, that worker cannot return to work in that period. To model utility from leisure, subjects receive a payment of $0.02 for each second they spend *not* working, which accrues from the moment they confirm a decision to quit. This yields a maximum payment of $9.60 for *not* working, meaning that workers face a substantial monetary opportunity cost for continuing to work.[8] This private payment from "leisure" is not used in calculations of group output or performance incentives, but is received only by the individual worker. Importantly, however, any tables that a worker completes prior to quitting are used to determine the output-based bonus payment for their group and a worker who quits also benefits from these group incentives. That is, workers are credited for any work that they do, regardless of whether or not they quit.[9]

Manipulation: Incentive context

We use three different incentive schemes in Stages 2 and 3, with a particular group experiencing the same incentive mechanism throughout these two stages. We implement the three incentive mechanisms described in Section 2.2 as follows[10]:

- **Revenue Sharing:** Each worker receives $2 times one-third of the total group output. That is, $\pi_{i,RS} = \$2\left(\frac{x_1+x_2+x_3}{3}\right)$.
- **Weak Link:** Each worker is paid $2 times the minimum of the total group output. That is, $\pi_{i,WL} = \$2\ min(x_1, x_2, x_3)$.
- **Tournament:** The worker who produces the most output in a period receives a payment of $24, the worker who produces the second most output receives $12 and the remaining worker receives $0. Specifically,

$$\pi_{i,T} = \begin{cases} \$24 & \text{if } x_i = max(x_1, x_2, x_3) \\ \$12 & \text{if } x_i = med(x_1, x_2, x_3) \\ \$0 & \text{if } x_t = min(x_1, x_2, x_3) \end{cases}$$

In case of a tie, the two (or three) workers split the corresponding payoffs equally. For example, if two workers tie for the highest output, then both those workers earn $18 for the period.

All subjects received detailed information on how to calculate both roles' payoffs prior to Stage 2. It is also important to note that our design holds the incentives for the leaders (or supervisors) constant across all incentive mechanisms and across Stages 2 and 3. Specifically, they are paid $2 times the average of total group output, or $\pi_{i,RS} = \$2\left(\frac{x_1+x_2+x_3}{3}\right)$. Note that this is the same payment as workers receive under Revenue Sharing, reflecting the property that, like under Revenue Sharing, the Active Leader or Passive Supervisor always benefits by obtaining higher output.

Manipulation: Passive Supervisors versus Active Leaders in Stage 3

In Stage 2, the fourth person in a group always serves the role of Passive Supervisor. A Passive Supervisor can observe the actions of the three workers—specifically, how many tables each worker completes, whether a worker indicates a desire to stop working, and whether a worker has confirmed a decision to quit work. Passive Supervisors cannot communicate with workers and have no other actions they can take. To prevent other subjects from discerning who is a supervisor by their idleness (and to limit boredom for subjects in this role), Passive Supervisors have the option of playing two computer games, solitaire or minesweeper, during these periods.

Our primary focus is to investigate how leadership affects group productivity as a function of the incentive context facing followers. To identify the impact of leadership, we assign a subset of supervisors to become Active Leaders in Stage 3. The remaining supervisors from Stage 2 remain as Passive Supervisors in Stage 3. The difference in output between groups with Active Leaders and Passive Supervisors in Stage 3, within an incentive mechanism, thus causally identifies the impact of leadership.

We provide Active Leaders with a single instrument, communication, through which they can exercise influence on the followers. Active Leaders can send free-form messages to workers via electronic chat throughout the period, including during a two-minute period before production begins, the eight-minute period during which workers produce output, and the time when the feedback is displayed. The lone exception occurs once a worker confirms a decision *not* to work; the leader can then no longer send messages to this worker until feedback is provided at the end of the period. Leaders can choose which workers receive any message, allowing them to send messages to the entire group, to pairs of workers, or to individual workers. Workers can only view messages sent to them and can see the list of recipients for any particular message. Leaders are provided with no directions about how to use messages, beyond being instructed not to identify themselves and to avoid offensive language. This means that the instructions do not provide any guidance on what messages to send to workers or whether any information provided must be truthful. Workers cannot send messages to either the leader or to any other workers in their group.

The Passive Supervisors continue to serve in the same role in Stage 3 as they had in Stage 2—observing workers' activities but without any ability to communicate with or influence the workers. These groups serve as a control, to identify baseline levels of group output absent any impacts of leadership. For this purpose, we can compare a group with an Active Leader in Stage 3 with other groups in Stage 3 under the same incentive mechanism.

We include Passive Supervisors—rather than, for example, having no fourth participant in a group and only three workers—for several reasons. First, because workers' efforts generate a positive externality for the earnings of Active Leaders (which is necessary to provide the leaders with incentives to motivate the group), having a control condition with no leader or supervisor would vary the potential externality from workers' efforts. When workers' effort increases payoffs for an additional individual—either a Passive Supervisor or an Active Leader—they might work harder because someone else benefits from their effort. Our

---

[7] We use a restricted web browser to limit subjects' ability to communicate with each other by accessing their email or social media accounts. We also eliminate any ability to search for information about experiments by accessing a search engine. Available websites were chosen to appeal to a variety of interests and to work within the constraints of the z-tree program. The available websites were sportsillustrated.com, gocomics.com, weather.gov, wikipedia.com, tmz. com, imdb.com, miniclip.com, washingtonpost.com, mtv.com, iheartradio. com, eonline.com, and usatoday.com. During the time period these experiments were run, none of these websites had paywalls. The two computer games were solitaire and minesweeper. Workers who quit split time roughly evenly between using the web browser and playing games.

[8] We added the leisure payment following pilot sessions in which it became clear that subjects were very reluctant to quit working solely for the browsing and game options. An alternative interpretation of the payment is that a worker can shirk at work while allocating effort to an outside job (e.g., freelancing) that provides a fixed hourly wage.

[9] Under weak link incentives, if one worker quits without working, it follows that the output-based bonus of the other workers is equal to $0.

[10] The above implementations correspond to the incentive schemes in equations (2) through (4), with $\gamma_{RS} = \gamma_{WL} = \$2$ and $\gamma_T = \$24$.

design holds such concerns constant.[11] Second, it is desirable for Active Leaders to understand the group members' tasks, the group dynamics and the information they receive as leaders at the time they adopt their leadership function. Toward this end, the role of Passive Supervisors in Stage 2 serves as training in which participants who will become Active Leaders can learn about the environment in which they are to operate. For comparability, we believe it is important to have a similar role in groups that are to have no Active Leader in Stage 3. Having such groups operate with a Passive Supervisor in Stage 2 and then removing this fourth group member in Stage 3 altogether would make the comparison between treatments more challenging.

Instructions and feedback

Subjects are provided with an initial set of instructions, focused on describing the task, prior to Stage 1 (see Appendix B for the full instructions). These instructions inform subjects that there will be two further stages, but provide no details about what will occur in Stages 2 and 3. Specifically, subjects are not informed, until after Stage 1, that they will subsequently be randomly assigned to groups or roles. Instructions for Stages 2 and 3 are provided at the same time. These include a discussion of how payoffs are determined for all roles, including a detailed description of the incentive mechanism, and how the communication software works (when relevant). All subjects receive instructions for both roles relevant for their group—i.e., for workers and either for Active Leaders or for Passive Supervisors. Workers know, in Stage 2, whether their group will have an Active Leader for Stage 3. Subjects complete a brief comprehension quiz following the instructions for Stages 2 and 3, checking their ability to calculate payoffs for all roles. Subjects receive additional instructions at the beginning of Stage 3, reminding them about any changes for Stage 3 (e.g., the introduction of an Active Leader, when applicable).

During a period, both Active Leaders and Passive Supervisors receive detailed, real-time, information on how much output each worker has produced and any declared intentions by a worker to quit working. None of this information is available to other workers during a period. At the end of each period, all group members observe a feedback screen summarizing the output (tables completed), earnings from output, time spent not working, earnings from time spent not working, and total earnings for each worker in the group. All feedback screens also show the leader or supervisor's earnings. To ensure workers are aware of the value of their outside option, they receive additional feedback telling them their earnings per second spent working. This makes it possible to compare their marginal earnings from time spent working on the task with the \$0.02 payment per second of *not* working.[12]

Primary and secondary measures

The primary outcome measure in our experiment is the amount of output generated in a group in a period, t, $\sum_{i \in \{1,2,3\}} x_{it}$, a measure of group productivity. From the perspective of experimental economics, this quantitative, objective outcome measure is a natural variable for evaluating whether a leader is effective at motivating workers.

We also obtain independent codings of the messages generated by Active Leaders, *m*. We do so in two ways. First, to obtain a measure of whether a leader's messages contain elements acknowledged in leadership research as related to effective leadership, we directly evaluated the quality of Active Leaders' messages using the kind of approach often employed in leadership research. Specifically, independent coders rated supervisors' leadership ability using the scale shown in Table 2, adapted from questions developed by Antonakis et al. (2011) to measure charismatic leadership.[13] Each leader's messages were rated independently by three coders, who read all the messages sent by a leader in Stage 3 and rated the degree to which they agreed (from 0 to 8) that the leader's statements reflected the corresponding statement in Table 2. Coders knew when each message was sent, who received it (e.g., before Period 4 to all three workers), and which incentive scheme the workers faced. They were not told anything about outcomes, although it was sometimes possible to infer this information from the content of the messages. To minimize the effects of measurement error, we use the average over the three codings.[14] Our measure of the quality of each Active Leader's messages, $L(m)$, is the average of the ratings across these seven items.

Second, we additionally obtained measures of the content of Active Leader's messages, in order to study whether leaders employed different communication strategies across incentive contexts. We describe the method for coding the content of messages in the Results section.

We also collected several secondary measures of subjects' individual characteristics. After Stage 3 concluded, subjects completed a short survey. We gathered basic demographic information and administered instruments drawn from Falk, Becker, Dohmen, Huffman, and Sunde (2023), eliciting subjects' risk preferences, time preferences, and various dimensions of their social preferences (altruism, trust, positive reciprocity, and negative reciprocity). We elicited risk preferences both via an unincentivized multiple price list and a simple survey question ("Are you generally a person who is willing to take risks or do you try to avoid them?).[15] Likewise, we administered an unincentivized multiple price list and a simple survey question ("Are you a person who is generally willing to give up something today in order to benefit from that in the future?") to measure subjects' time preferences. We elicited social preferences via a ten-item survey (e.g., "Consider the following situation: You have won \$1,000 in a lottery. Given this situation, how much would you donate to a charitable cause?"). A full copy of all survey instruments is provided in Appendix B.[16]

---

[11] It is also possible that social preferences directed toward a leader or supervisor may interact with the leader's active or passive role. For example, as noted by an anonymous reviewer, subjects may particularly resent a Passive Supervisor who benefits from the group's effort without doing anything to contribute to it, a form of negative reciprocity. We believe that such an influence, while possible, is unlikely to be strong in our experiment. The large literature on social preferences (see Cooper and Kagel (2016) for a summary) generally finds that negative reciprocity is directed towards others exhibiting *intentionally* unkind behaviors (e.g., withholding effort when they have the option to provide it). In the case of our experiment, it is unlikely that workers blame Passive Supervisors for being randomly assigned to a role over which they had no control. Instead, concerns for social welfare (maximizing the total sum of payoffs) or altruism toward a fourth group member are more likely to be germane in our study. Our design holds such concerns constant, by always having a fourth group member who benefits from workers' output in exactly the same way across conditions.

[12] Empirically, the average pay from working was \$0.034 per second with a standard deviation of \$0.028.

[13] The first three questions are adaptations of Cronshaw and Lord's (1987) General Leadership Impression questionnaire, a well-known measure of how prototypical the leader is perceived to be by the raters. The other four questions measure attributions and outcomes associated with charismatic leadership: (1) Affect for the leader; (2) Trust in the leader; (3) Leader competence; and (4) Leader influencing ability.

[14] No attempt was made to force agreement among coders, although generally their ratings were highly correlated—Cohen's quadratic weighted kappa (Cohen, 1968) equals 0.591.

[15] Due to either inconsistent choices or a minor software bug, the list-based risk measure is missing for 19 subjects. In these cases, we rely solely on the survey question. We have 18 missing observations for age. In these cases, we substituted the average value of age for the subject population.

[16] In some sessions we gathered additional information about subjects' fields of study, whether they had any sort of merit-based scholarship, their experience in leadership positions, and the Big 5 personality measure. Because we only have these measures for about half of our subjects, we do not use them in our analysis.

Table 2
Ratings of leadership quality ($L(m)$).

1. The person I am rating demonstrated *in his/her messages* leader behavior
2. The person I am rating *communicated* like a typical leader
3. The person I am rating *conveys through his/her messages an image that* fits my image of a leader
4. I like this person as a leader
5. The person I am rating appears trustworthy from his/her messages
6. The person I am rating is competent as a leader
7. The person that I am rating is able to easily influence others *through communication*

Participants and sessions

The combination of the two treatment dimensions—incentive scheme (Revenue Sharing, Weak Link, Tournament) and leadership (Passive Supervisor, Active Leaders)—results in a 3-by-2 factorial design. Assignment to one of the treatment cells occurred at the session level. For each incentive scheme, we conducted four sessions of the Passive Supervisor treatment and seven sessions of the Active Leader treatment; we oversampled the latter because we were especially interested in investigating the relationship between messages Active Leaders send and their success as leaders. There were 3 to 6 groups per session, with the differing sizes due to variation in how many subjects showed up for a session. There was no interaction between subjects in different groups, so we treat groups as independent observations. A total of 692 subjects participated across the 33 sessions. Table 3 summarizes the experimental design.

All experimental sessions took place at the xs/fs laboratory of Florida State University. Sessions were conducted from October 2012 through November 2013.[17] We used ORSEE to recruit subjects (Greiner, 2015) and zTree to implement the experiment (Fischbacher, 2007). No subject participated in more than one treatment. On average, subjects earned $24.28, which included a $10 show-up fee, earnings from one period selected at random from the three stages, and potential additional earnings from correctly guessing their quartile in Stage 1. The experiment took between 90 and 120 min to complete.

Results

The next subsection summarizes how leadership, specifically having an Active Leader, affects output across the three incentive mechanisms. This analysis provides tests of Hypotheses 1 and 2. The following subsection focuses on the relationship between the independently evaluated leadership quality of Active Leaders and the performance of their groups, testing Hypothesis 3. This section also includes a brief discussion of how group performance relates to the individual characteristics of Active Leaders. The final subsection explores the relationship between

how Active Leaders communicate, the context in which they lead and the performance of their groups, providing evidence relevant for Hypothesis 4.

Unless otherwise noted, the primary unit of analysis is a single four-person group. When referring to performance in Stage 1, we use each individual's output. Unless otherwise stated, tests of statistical significance comparing different groups at the same point in time are Mann-Whitney U tests and tests of statistical significance comparing the same group between different points in time are Wilcoxon matched-pairs signed-rank tests.

Treatment effects on output

Fig. 2 shows average group output across periods, broken down by groups assigned to have Active Leaders and Passive Supervisors in Stage 3 and separately for each incentive scheme. To allow for direct comparisons over time, Fig. 2 omits the Stage 1 (Period 1) output for individuals who would subsequently become Active Leaders or Passive Supervisors—doing so allows us to compare the performance across time of the individuals who would become workers in Stages 2 and 3.

Across all three incentive mechanisms, average individual output in Stage 1 is 7.09 tables, yielding an average three times as large when aggregating the three individuals who would subsequently be grouped together as workers. There is no significant difference between subjects in the Passive Supervisor and Active Leader treatments (7.32 vs. 6.94; $p = 0.221$), which is not surprising given that this treatment dimension was not yet revealed in Stage 1.[18]

We next consider output in Stages 2 and 3. Under Revenue Sharing incentives (left panel) average group output is very similar in Stage 2 (Periods 2 and 3) for groups subsequently assigned to have Active Leaders or Passive Supervisors. Workers' output declines slightly from Period 2 to Period 3, consistent with experimental studies of social dilemma and public goods games, where the incentive to shirk and free-ride on others' contributions similarly often lead to declining contributions over time. In Stage 3, groups with Active Leaders perform slightly better under Revenue Sharing than those with Passive Supervisors, but the difference is small and not statistically significant ($p = 0.881$).

The evolution of group output differs between Tournament incentives (right panel) and Revenue Sharing; group output remains steady over time with Tournament incentives rather than declining as occurs with Revenue Sharing. Nonetheless, the average impact of Active Leaders is quite similar in both cases. With Tournament incentives, there is little difference in Stage 2 between groups subsequently assigned Passive Supervisors versus those who will receive Active Leaders. More importantly, group output in Stage 3 is only slightly higher for groups with Active Leaders than for those with Passive Supervisors, and again the difference is not statistically significant ($p = 0.380$).

The picture is quite different under Weak Link incentives (center panel). With Passive Supervisors, we observe a strong decline in output

Table 3
Summary of observations by treatment condition.

		Leadership (Stage 3)	
		Passive Supervisor	Active Leader
Incentive Context (Stages 2 and 3)	Revenue Sharing	22 Groups 88 Subjects	35 Groups 140 Subjects
	Weak Link	23 Groups 92 Subjects	32 Groups 128 Subjects
	Tournament	22 Groups 88 Subjects	39 Groups 156 Subjects

---

[17] The study was not pre-registered. At the time these sessions were conducted, pre-registration was generally rare across the social sciences. Our primary analysis, of treatment effects on group output (reported in Section 4.1), follows directly from our design. All remaining analysis is secondary and largely exploratory.

[18] Fig. 2 omits output in Period 1 by those subjects subsequently assigned to be Passive Supervisors in Stage 2 (and possibly Active Leaders in Stage 3). However, such subjects are no more or less productive in Stage 1 than those assigned to the worker role (6.94 vs. 7.14; $p = 0.613$).
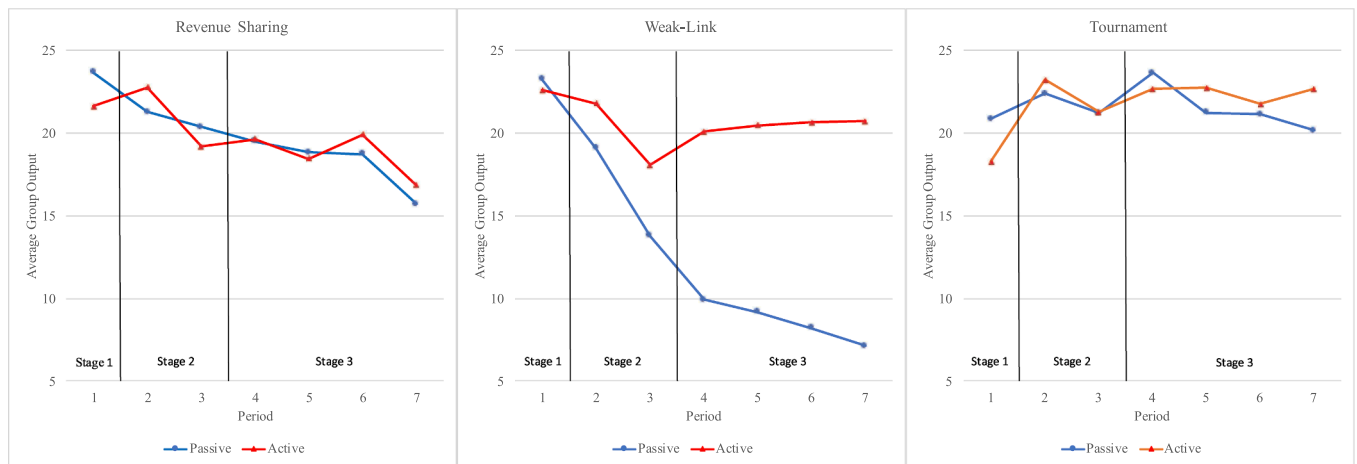
Fig. 2. Average group output.

Table 4
Panel Regressions of Group Output (Stages 2 and 3).

	(1) Revenue Sharing	(2) Weak Link	(3) Tournament
Active Leader Treatment	1.470	3.836**	0.597
	(1.523)	(1.564)	(1.324)
Stage 3	−2.636**	−7.826***	−0.239
	(1.245)	(1.028)	(0.855)
Active Leader Treatment × Stage 3	0.365	8.381***	0.450
	(1.836)	(1.574)	(1.090)
Total Worker Output in Stage 1	0.211***	0.174**	0.017
	(0.064)	(0.069)	(0.030)
Constant	5.839	4.259	20.708***
	(4.327)	(4.552)	(2.151)
Observations	342	330	366
Groups	57	55	61
R-squared	0.145	0.313	0.007

Notes: The dependent variable in all regressions is the total output in a group in a period of Stages 2 and 3 (Periods 2–7). Model 1 uses data from Revenue Sharing, model 2 uses data from Weak Link and model 3 uses data from Tournament. Active Leader (treatment) equals 1 in all periods for groups in the Active Leader treatment, including in Stage 2 when these groups have Passive Supervisors. Stage 3 equals 1 in Periods 3 through 7. Active Leader treatment × Stage 3 equals 1 for groups in the Active Leader treatment in Stage 3, when these groups have Active Leaders. Stage 1 Output is the total output produced by the three workers in a group in Stage 1 (Period 1). Standard errors (clustered at the group level) are reported in parentheses. Three (***), two (**), and one (*) stars indicate significance at the 1%, 5% and 10% levels in two-tailed tests.

over time, starting in Stage 2 and continuing through Stage 3. By the final three periods of Stage 3, average group output with Passive Supervisors is below 10. Groups that will subsequently receive Active Leaders in Stage 3 also show a decline in output in the two periods of Stage 2 (from Period 2 to Period 3), but this decline is smaller and average group output is higher than for groups that will remain with Passive Supervisors ($p = 0.015$). This difference is interesting, given that these groups' leaders cannot yet communicate with workers in Stage 2. We interpret the gap as potentially reflecting a beneficial impact of knowing there will be an Active Leader in Stage 3; anticipating that an Active Leader will be able to coordinate workers' effort to produce high output, workers may become more patient about reducing their effort level in Stage 2, thereby reducing the self-confirming downward trend in output. Indeed, under Weak Link incentives, Stage 3 output recovers and increases slightly across periods with Active Leaders who can communicate with workers. By the end of Stage 3 (Period 7), output reaches roughly its initial level with Active Leaders—output in Period 7 is not significantly different from output in Period 2 ($p = 0.681$). As a result, groups with Active Leaders have significantly greater output in Stage 3 than those with Passive Supervisors ($p < 0.001$). Stage 3 output is significantly lower under Weak Link incentives than under the other two incentive schemes (vs. RS: $p = 0.002$; vs. TN: $p < 0.001$) with Passive Supervisors, but this difference disappears with Active Leaders (vs. RS: p

$= 0.267$; vs. TN: $p = 0.728$).

Table 4 provides complementary regression analysis, using the panel structure of the data and the within-group variation in leadership functions in the Active Leader condition. The dependent variable is group output in a period. The regressions contain data from all 6 periods of Stages 2 (Periods 2–3) and 3 (Periods 4–7) for all groups and use a differences-in-differences structure to identify the impact of Active Leaders in Stage 3, while controlling for differences in output in earlier stages. As a control, we include the total output produced by the three workers in Stage 1, before the group was formed.

The first explanatory variable, Active Leader (treatment), takes on a value of 1 in all periods for groups that will receive an Active Leader in Stage 3. This measures any differences in Stage 2 between groups in this treatment condition and those assigned to have Passive Supervisors throughout the study. Consistent with our observations from Fig. 2, groups in the Weak Link condition that will go on to have Active Leaders already experience slightly higher output in Periods 2–3, reflecting the potential anticipatory effect we discuss above. In contrast, there is no difference in Stage 2 performance in Revenue Sharing and Tournament between groups that will subsequently have an Active Leader and those that will not. The second explanatory variable, Stage 3, is an indicator variable for Periods 3–7, which captures the difference between output in Stage 2 and Stage 3 for groups that retain Passive Supervisors. Again,

consistent with Fig. 2, the coefficients indicate lower average output in Stage 3 relative to Stage 2 for groups in Revenue Sharing and, more strongly, in Weak Link, with only a slight and statistically insignificant decrease in Tournament.

The key variable for our analysis is the interaction between these two variables, Active Leader treatment × Stage 3. This coefficient captures to what extent output in Stage 3 differs for groups assigned to the Active Leader condition, while controlling both for the group's history (in Stage 2) and the comparable trajectories of groups in the Passive Supervisor condition that never have Active Leaders. Consistent with our observations from Fig. 2, Active Leaders have at least small positive impacts on group output in all conditions. However, they only have a large and statistically significant positive effects under Weak Link incentives.

Stages 2 and 3 provided workers with the ability to quit working, either at the beginning of or during a period—and thereby earn utility from "leisure." It is thus natural to consider the role of worker quitting in determining productivity, and whether such quitting interacted with the incentive treatments. Worker quitting occurred regularly. Averaging across all periods of Stages 2 and 3 for all treatments, 52 % of workers quit at some point in the period and workers averaged 5:34 min of work per eight-minute period.

Fig. 3 provides evidence on the frequency and impacts of quitting across the incentive mechanisms. The top panels show the average number of workers in a group who quit at any point in the period, either before the period starts or during the period.[19] The bottom panels show the average total minutes worked in a group in a period. Workers could not quit in Period 0, so the default number of workers quitting is 0 in Period 0 and the default number of minutes worked is 24. Both statistics are broken down by groups led in Stage 3 by Passive Supervisors and by Active Leaders.

The comparisons between Passive Supervisors and Active Leaders for either quitting or minutes worked parallel the results for output. There is little difference with Revenue Sharing and Tournament, either in Stage 2 or Stage 3. With Weak Link incentives, a large gap emerges in Stage 3, with workers quitting substantially more frequently and working fewer minutes without active leadership. As with output, the difference in minutes worked is significant in Stage 3 ($p < 0.001$).[20] In fact, the difference in output between Active Leaders and Passive Supervisors under Weak Link incentives comes entirely from quitting; if we compare productivity (output per minute worked) for groups with Passive Supervisors versus Active Leaders in Stage 3, there is no substantial difference (1.26 vs 1.22). Also as with output, we observe that differences in quitting start to emerge in Stage 2 ($p = 0.010$), before the leader can communicate with group members. However, Active Leaders in Weak Link dramatically slow the decline in time spent working on the task.

It is worth noting that the average minutes worked by a group is always far below the maximum possible worktime of 24 min in all treatments by the end of Stage 3. The highest average number of minutes worked in Period 7 is 16.7, for Tournament incentives with Passive Supervisors. Thus, the failure of Active Leaders to generate higher output with Revenue Sharing and Tournament incentives does *not* reflect a ceiling effect, as there was always plenty of room to increase output via an increase in the amount of time worked, which would have

benefitted the leaders.

Result 1. *Output is weakly higher with Active Leaders in all three contexts, but in two of the contexts (Revenue Sharing and Tournament) the positive impact of Active Leaders is negligible. The data thus only partially support* Hypothesis 1.

Result 2. *The effect of Active Leaders on output is greater for Weak Link incentives than for Revenue Sharing or Tournament incentives. The data thus supports* Hypothesis 2.

What leader characteristics are related to group performance?

Above, we show that, on average, having an Active Leader only substantively improves output with Weak Link incentives. However, even with Revenue Sharing and Tournament incentives, where having an Active Leader does not increase output on average, there is substantial variation in output between groups. The rest of our analysis explores how the characteristics of Active Leaders and the content of their messages are related to their group's performance as measured by its Stage 3 output. Therefore, in the next two sections we focus our analysis only on groups with Active Leaders.

Before proceeding, a couple of points are worth making. First, Stage 3 output is only a noisy measure of an Active Leader's effectiveness. Economists often equate good outcomes, such as high group output in Stage 3 of our experiment, with good organizational practices. However, this simplification risks making the mistake of giving leaders credit for good or bad outcomes for which they were not responsible (Weber et al., 2001). A large part of group output in Stage 3 is beyond leaders' control. The quality and intrinsic motivations of the workers may have strong effects on their group's output in Stage 3, beyond anything that Active Leaders do. Indeed, we observe a high correlation between group output in Stages 2 and 3 ($\rho = .559$) for groups with Active Leaders in Stage 3; all groups have Passive Supervisors in Stage 2, so this correlation cannot reflect any action taken by leaders in Stage 2.[21] Moreover, we have just seen that the incentive schemes, which are not controlled by the leader, influence group performance in Stages 2 and 3.

To account for variation in Stage 2 group outcomes that is beyond the control of the leader but that may affect Stage 3 output, we use a measure of leader performance that partly accounts for the luck of drawing a good group of workers or on treatment assignment.[22] For this, we use the information available when an Active Leader begins communicating with their group (i.e., at the beginning of Stage 3) to control for factors beyond the leader's control. Specifically, we regress average Stage 3 group output on the same group's average output in Stage 2 and on indicator variables for the incentive mechanism under which that group is working. Using the fitted parameters from this regression, we can predict how a group would be expected to perform in Stage 3 based solely on what occurred through the end of Stage 2. More importantly for our purposes, the residuals from these regressions (i.e. the actual average Stage 3 output for a group minus the predicted Stage 3 output) provide a measure of how a group's actual Stage 3 output departed from the output that would have been predicted for that group absent any role for the Active Leader assigned to that group. We refer to this measure as "Residual Output." In the remainder of our analysis, we use Residual Output as our primary measure of leader performance, thereby focusing on the portion of variation in group Stage 3 output that could plausibly be affected by Active Leaders. In line with this approach, Appendix Figure A2 displays the distributions of Residual Output,

---

[19] Appendix Figure A1 shows when quitting occurred across periods in Stage 3. Under Active Leaders (Figure A1a), most quitting occurred during the period, with only about 15 percent of all quitting decisions occurring before the period started. At the other end, about 45 percent of workers never quit. The distributions of quit times with Passive Supervisors (Figure A1b) are generally similar, except under Weak Link where over 40 percent of workers quit before the period starts. In Tournament, but not in the other treatments, we observe a positive and statistically significant relationship between a worker quitting and that worker's ability on the task (measured by Stage 1 output).

[20] The equivalent test statistics for quitting are $p = .058$ for Stage 2 and $p < .001$ for Stage 3.

[21] This figure, while high, is still substantially lower than the correlation between group output in Stages 2 and 3 for groups with Passive Supervisors in Stage 3 ($\rho = .805$), suggesting that Active Leaders have some influence in overcoming the history-dependence of their groups.

[22] We thank an anonymous reviewer for pointing out the need to control for the influence of Stage 2 on a group's output in Stage 3.
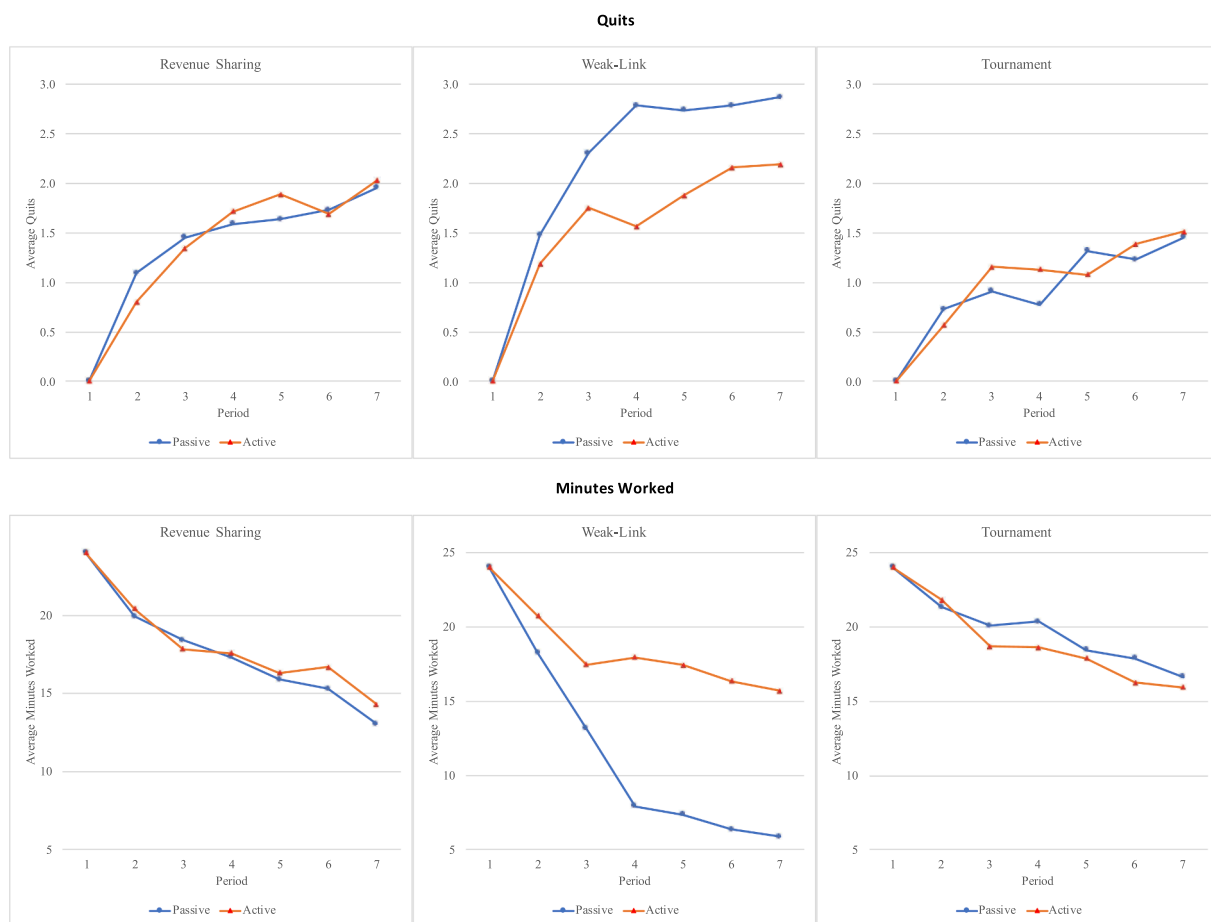
Fig. 3. Average group quits and minutes worked.

separately, for Passive Supervisors and Active Leaders. There is greater (unexplained) variation in groups led by Active Leaders than in those led by Passive Supervisors,[23] suggesting that Active Leaders may be doing something to generate this additional variation.

Following the same logic, we construct a measure of how well a leader performed at getting their group to work. We first obtain a prediction of each group's average time worked in Stage 3 by regressing average time worked by the group in Stage 3 on the group's average time worked in Stage 2 and indicator variables for the incentive mechanism confronting that group. The difference between actual and predicted time worked by the group in Stage 3 gives a measure of how a leader performed at getting their group to work that does not depend on the incentive scheme or how much the group worked in Stage 2, factors that are out of the leader's control. We refer to this measure as "Residual Time Worked."

The second important point we need to make is that leadership and leader performance are inextricably intertwined in our experiments, given the dynamic and repeated nature of the group production task. Active Leaders are constantly reacting to what their group has done and vice versa. We will therefore confine ourselves to correlational statements about the leaders' characteristics and messages and the performance of their groups, rather than claiming causality. By the same token, our survey measures of individual leaders' personal characteristics were administered immediately after the experiment concluded. This practice is common in experimental economics, reflecting an

implicit (but perhaps flawed) assumption that individual characteristics and preferences are generally stable. However, it raises the possibility that outcomes from the experiment may spill over into the survey. We therefore also confine ourselves to correlational statements relative to the survey measures of individual characteristics.

We first study the relationship between our measure of Leadership Quality (i.e., $L(\boldsymbol{m})$), obtained through independent coding of each Active Leader's messages, and Stage 3 outcomes for groups led by Active Leaders. Recall that these ratings were obtained using scales similar to those often employed in leadership research (Antonakis et al., 2011), relying on the evaluations of coders who had access to a leader's messages but not to the outcomes for a particular group.

Fig. 4 shows the relationship between Stage 3 group outcomes and the Leader Quality Rating, for every Active Leader. The blue lines are regression lines, and the text in the lower left corner of each panel shows the correlation between the outcome variable and the leadership rating, along with the (robust) p-value from a regression of the variable on the y-axis on the Leader Rating. We compare three measures of Stage 3 outcomes for a group: actual output, Residual Output, and Residual Time Worked. In all three cases, we observe positive and statistically significant relationships, indicating that raters' evaluations of the persuasiveness of a leader's messages are related to different measures of the actual effectiveness of that leader. As we expected, based on the presence of random elements beyond a leader's control in Stage 3 output, the correlation is higher for Residual Output than actual output

---

[23] The standard deviations of the residuals also differ significantly in a variance ratio test (std. dev. for Passive Supervisors = 4.94; std. dev. for Active Leaders = 6.23; $p = 0.044$).
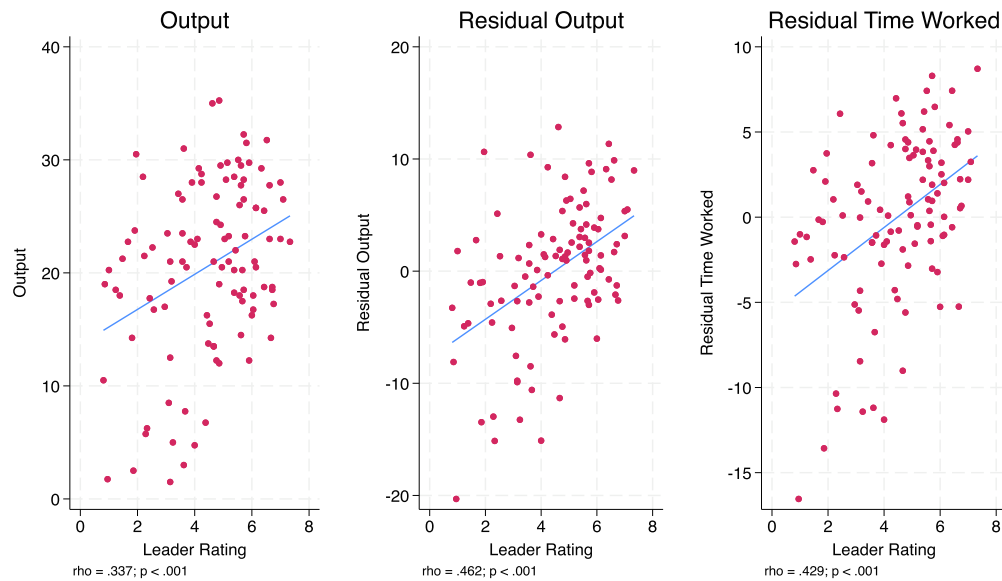
Fig. 4. Relationships between leader ratings and stage 3 outcomes.

($\rho = 0.462$ vs. $\rho = 0.337$; $t = 2.35$; $p = 0.010$; 103 df).[24]

The middle panel of Fig. 4 shows the relationship between Leader Ratings and Residual Output, *pooling data across all incentive conditions.* Fig. 5 explores how Leader Ratings and their relationship with Residual Output vary across incentive mechanisms. The top left panel shows the mean ratings for each incentive condition. There is some variation, with Active Leaders in Weak Link rated slightly more positively than in the other two conditions. However, we cannot reject the hypothesis that leadership ratings are identical across the three treatments, nor are any of the pairwise comparisons between incentive schemes statistically significant. The remaining panels show the relationship between Leader Ratings and Residual Output broken down by incentive mechanism. In all cases, we observe a positive and statistically significant relationship between the variation in a group's outcomes that cannot be explained by Stage 2 outcomes and ratings of leader quality based solely on the leader's messages. It is worth noting that this relationship is strongest with Weak Link incentives, the only case in which an Active Leader had a significant effect on average group output in Stage 3.

Result 3. *For Active Leaders, there is a strong positive relationship between Ratings of Leadership Quality, as identified by independent raters, and the part of group output in Stage 3 that is not explained by Stage 2 group performance. This relationship is positive in all incentive conditions. This supports* Hypothesis 3.

We also investigate whether individual characteristics of Active Leaders are correlated with their performance, measured by Residual Output in Stage 3. Appendix Figure A3 shows how a leader's gender,

their ability in performing the task in Stage 1 and several measures of their preferences collected at the end of the experiment (altruism, positive and negative reciprocity, propensity to trust, risk attitudes, and a measure of intertemporal preferences) correlate with that leader's group's outcomes (measured by Residual Output). For the two characteristics that are independent of our treatment conditions, gender and Stage 1 performance, no significant correlations are found. For example, one might expect that leaders who are themselves good at the task, as measured by Stage 1 output, might also make good leaders, but we find little evidence of a relationship between how well a leader performed the task in Stage 1 and Residual Output of the group they lead in Stage 3 ($\rho = 0.053, p = 0.590$). Turning to the preference measures, negative reciprocity ($\rho = 0.159, p = 0.086$) and trust ($\rho = 0.158, p = 0.083$) are weakly, and only marginally statistically significantly, correlated with Residual Output. Beyond the concern with spillovers mentioned above, it must be noted that these are post hoc correlational observations, which should therefore be interpreted cautiously.

The content of leaders' messages across contexts

To study how Active Leaders' communication strategies differ across incentive contexts, we quantified the content of supervisors' messages through coding. We first looked at a sample of messages and developed coding categories designed to capture common types of messages. The actual coding was done by three graduate student research assistants; these were *not* the same three RAs who provided Leader Quality Ratings. We trained the coders on how to do the coding, mainly through examples of how various messages would be coded, but did not share any hypotheses about the content of the messages with the coders. The coders were told that their job was to represent the content of the messages as accurately as possible, but not to interpret the messages. The coders were allowed to code multiple categories for a single message. Throughout our discussion of the coding exercise, we use the
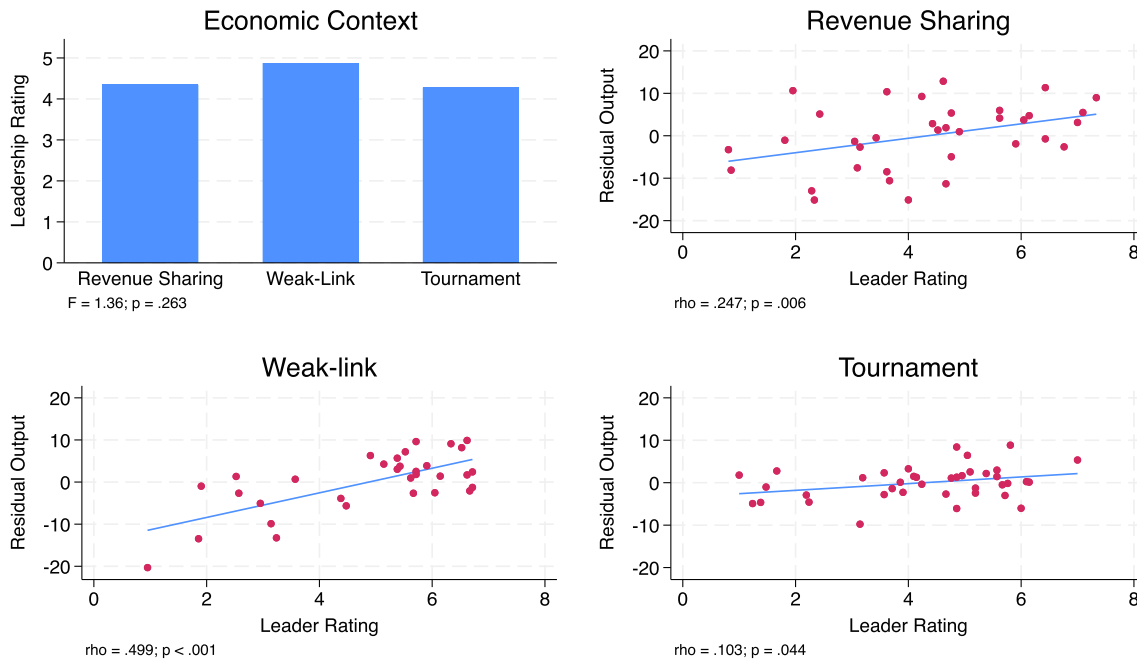
---

[24] As noted above, we cannot make any causal claims about these relationships, given that the leadership ratings are based on messages sent simultaneously with workers working on the real effort task. The coders could not observe group output, but group output could affect the content of supervisors' messages and coders might have been able to guess how the group was doing from the message content. Thus, whereas leadership ratings capture something that is closely related to Stage 3 outcomes, the direction of causality is unclear. However, one piece of evidence that Leader Ratings are not simply picking up variation in Stage 3 outcomes comes from the relationship between Stage 2 output and the Leader Ratings, which are not significantly correlated ($\rho = -0.084, p = 0.359$). Thus, even though Stage 2 output is a strong predictor of output in Stage 3, it seems largely independent of the ratings of the messages from the independent coders.

Fig. 5. Ratings of leader quality across incentive conditions.

average coding across the three RAs.[25] The coding categories were as follows:

- *Positive Feedback*: Leader praises workers for their good performance.
- *Negative Feedback*: Leader criticizes workers for their bad performance.
- *Encouragement*: Leader sends a message that includes a form of general encouragement.
- *Information About Experimental Environment*: Leader gives workers information about the experimental environment (i.e., number of periods, nature of the task, rules of the experiment, etc.).
- *Information About Behavior*: Leader gives workers information on what they or others have done in the period but does *not* explicitly compare individuals.
- *Comparison*: Leader gives information on relative performance of workers, or otherwise compares workers.
- *Advice*: Leader gives advice to an individual worker, several workers, or the entire group on how to perform the task.
- *Goals*: Leader sets a goal for an individual worker, several workers, or the entire group about how much to work.
- *Individual Earnings*: Leader emphasizes the material earnings for an individual worker.

- *Group Earnings*: Leader emphasizes the material earnings for the group or other group members.
- *Supervisor Earnings*: Leader emphasizes their own material earnings.
- *Fairness*: Leader emphasizes ethical and fairness aspects with respect to an individual worker, several workers, or the entire group's behavior.
- *Humor*: Leader employs humor in their message.
- *Nonsense*: Messages that did not have any obvious meaning to the coders.
- *Miscellaneous*: Messages that could not otherwise be categorized.

Table 5 summarizes the coding results regarding leaders' message use across contexts. For each category, the table reports the average number of messages sent per minute of time spent working by the group.[26] Dividing by the amount of time spent working is necessary to avoid confounding how likely the leader was to send various types of messages with how long the leader could communicate with active workers. We report frequencies for all three incentive schemes pooled (Column 1) and each incentive scheme separately (Columns 2–4). The fifth column reports Fleiss (1971) kappa for each category.[27] Intercoder agreement was generally high, with the exceptions coming from categories that were rarely coded (fairness and miscellaneous). The sixth column reports p-values for whether the frequency of the coding category differs significantly across treatments. This figure is based on a simple OLS regression where a group's frequency per minute worked in Stage 3 for a particular category is regressed on dummies for the incentive mechanisms (with Revenue Sharing as the omitted category) and the group's average output in Stage 2. The p-values are taken from an F-test for joint significance of the two incentive mechanism dummies. Given the large number of categories, the final column reports "sharpened" *q*-values, which correct for multiple hypothesis testing using the

---

[25] As a complementary approach, we also attempted to use content analysis based on machine learning, employing the topic modeling approach used by van der Velde and Gerpott (2023). We describe our implementation and results in Appendix C. Overall, topic modelling yields noisy results, likely due to the nature of the messages (many short, quick messages, requiring context for interpretation, rather than longer statements more amenable to extracting meaning through machine learning). However, it also confirms a few points observed from the human coding of leader messages. First, groups with high Residual Output in the Tournament treatment have leaders who communicate to workers their ranking and set goals in terms of number of tables. Under Weak Link incentives, communicating to workers their relative standing is also associated with higher Residual Output. Second, positive feedback and/or encouragement are positively associated with Residual Output across incentive schemes. Third, nonsense and humor are not conducive to higher Residual Output.

[26] This is the sum of the time spent working by the three workers, not the average.

[27] Fleiss's kappa is more appropriate than the better-known Cohen's kappa because there are three coders and a different set of coders was used for one of the sessions.

Table 5
Frequency of message categories by context.

	All	RS	WL	TN	Kappa	p-value	q-value
# Messages	0.683	0.710	0.848	0.524	n/a	0.150	0.200
Positive Feedback	0.095	0.101	0.130	0.061	0.800	0.035	0.136
Negative Feedback	0.042	0.041	0.065	0.024	0.575	0.275	0.255
Encouragement	0.108	0.156	0.126	0.051	0.579	0.064	0.146
Inf. About Exp.	0.028	0.035	0.022	0.027	0.703	0.360	0.276
Inf. About Beh.	0.066	0.049	0.119	0.036	0.407	0.226	0.255
Comparison	0.122	0.035	0.077	0.236	0.808	0.000	0.001
Advice	0.009	0.013	0.009	0.006	0.810	0.308	0.255
Goals	0.081	0.070	0.161	0.026	0.631	0.000	0.001
Individual Earnings	0.049	0.037	0.084	0.030	0.608	0.096	0.170
Group Earnings	0.028	0.025	0.061	0.003	0.525	0.000	0.001
Supervisor Earnings	0.011	0.010	0.020	0.003	0.724	0.196	0.244
Fairness	0.005	0.009	0.005	0.001	0.286	0.064	0.146
Humor	0.099	0.109	0.108	0.081	0.652	0.734	0.499
Nonsense	0.051	0.112	0.032	0.012	0.601	0.463	0.353
Miscellaneous	0.002	0.003	0.002	0.001	0.103	0.256	0.255

Note: p-values are uncorrected results of an F-test of the null hypothesis that treatment dummies are jointly equal to zero. q-values are "sharpened" q-values adjusted for multiple testing using the method of Benjamini et al. (2006) as implemented by Anderson (2008).

method of Benjamini, Krieger, and Yekutieli (2006) as implemented by Anderson (2008).

There was no shortage of messages sent by leaders, with an average of slightly more than ten messages sent per work period. These were split almost 50/50 between messages sent to the entire group and messages send to a subset of the workers. Some categories of messages, such as positive feedback, encouragement, and the use of humor were common across all incentive schemes. Positive feedback is always far more common than negative feedback. It is striking how rare it is for leaders to discuss fairness, even under revenue sharing incentives where this seems like an obvious discussion point. Most types of messages are used with roughly the same frequency across the three treatments, but there are three notable exceptions. Comparisons were far more common under Tournament incentives, whereas the use of goals and references to group earnings were more common under Weak Link incentives than under the other two incentive schemes. We thus observe some evidence that Active Leaders employ different communication strategies in different contexts.[28] Some, but not all, of these patterns provide support for Hypothesis 4. Messages of encouragement are less frequent under Tournament incentives than with either of the other two incentives; however, consistent with Hypothesis 4a these differences are not statistically significant. Consistent with Hypothesis 4c, references to total group earnings (a form of social consideration) are more frequent under Weak Link and Revenue Sharing than under Tournament incentives. Comparisons are more frequent under Tournament and Weak Link incentives than under Revenue Sharing, as predicted by Hypothesis 4b, but the latter difference is small in magnitude.[29]

Result 4. *Message usage varied by incentive context. a) Encouragement was used similarly frequently across conditions. b) Comparison was more common under Tournament incentives than under Revenue Sharing, but only slightly more common under Weak Link than under Revenue Sharing. c) Messages about Group Earnings, which speak to social considerations, were*

more frequent under Weak Link and Revenue Sharing than under Tournament incentives, but the differences for Fairness are small. These results partially support Hypothesis 4.

Fig. 6 shows, for several of the message categories, the Residual Output for individual workers that received messages in that category and those that did not.[30] Each row of the figure presents results from a different message category and each column corresponds to an incentive mechanism. The first three rows show results for the three message categories employed at different frequencies across incentive contexts: comparisons, goals and group earnings (see Table 5). The use of comparisons by leaders is correlated with higher Residual Output under both Weak Link and Tournament incentives but there is no relationship for Revenue Sharing, consistent with Hypothesis 5b. We find no statistically significant relationship with Residual Output for messages about group earnings under Revenue Sharing or Weak Link incentives, which is inconsistent with our prediction in Hypothesis 5c for messages emphasizing social considerations. Hypothesis 5a predicted that encouragement should be equally effective for all three incentive mechanisms. The fourth row of Fig. 6 shows the relationship between receiving encouragement and Residual Output under each incentive mechanism. We observe significant correlations for Weak Link and Tournament incentives, albeit weakly in the latter case, but not for Revenue Sharing.[31] Finally, we observe no significant relationships for messages about goals, though we had no a priori predictions for this category.

Result 5. *Correlations between message use and group performance varied by incentive context. a) Encouragement correlates with group performance under Weak Link and Tournament incentives, but not under Revenue Sharing. b) Comparison was correlated with group performance under Weak Link and Tournament incentives, but not under Revenue Sharing. c) There is little evidence that the use of messages about Group Earnings, which speak to social considerations, is related to performance. These results partially support* Hypothesis 5.

Viewed broadly, the results in Table 5 and Fig. 6 provide some,

---

[28] Consistent with our model, we did not expect many instances in which leaders requested that one or more followers quit working. Following an anonymous reviewer's suggestion, we had a single research assistant code the messages for suggestions that followers quit working. Such suggestions were rare (0.008). While the frequency varied across conditions, suggestions to quit were not particularly frequent under any incentive mechanism (RS: 0.007, WL: 0.014, TN: 0.002). Note that if such requests reflected leaders' concerns with workers' welfare (see footnote 6), we would expect such instances to be most frequent under TN and least frequent in RS, neither of which holds in the data.

[29] Nevertheless, even the small difference in the frequency of comparisons between Revenue Sharing and Weak-link is statistically significant (t = 2.45; p =.016).

[30] We use individual-level measures of Residual Output to reflect the fact that workers in a group may have received different messages. We construct this measure similarly to how we do so for the corresponding group-level measure of Residual Output, by regressing each individual worker's Stage 3 output on that worker's output in Stage 2 and indicators for incentive mechanism and using the difference between the actual and fitted values as the relevant measure of individual Residual Output.

[31] Not shown in the figure, because the frequency of these message categories did not differ by treatment, humor and nonsense have a consistently negative correlation with Residual Output across incentive schemes.
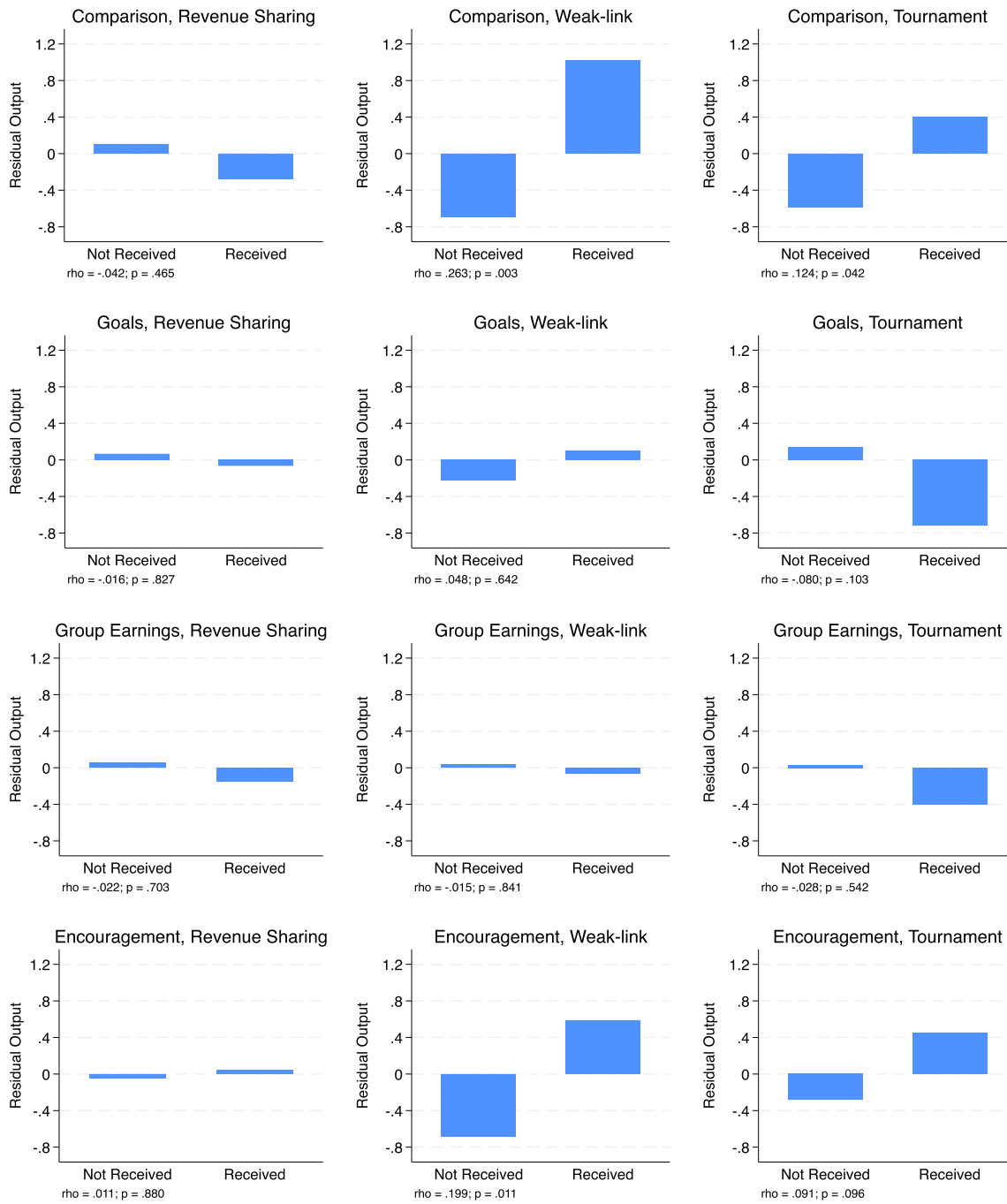
Fig. 6. Message categories used by leaders and group performance (Residual Output).

though only partial, support for the relationships in Hypotheses 4 and 5. However, it is important to again note that these findings must be interpreted cautiously, as suggestive evidence at best, given the concerns with endogeneity that we discussed earlier.

We finish our exploration of the relationships between group performance and the contents of Active Leader's messages by examining a notable feature of our experimental design, Active Leaders' ability to send messages to subsets of their workers. When we added this feature to the design, we conjectured that Active Leaders would take more advantage of it under Tournament Incentives as part of a broader strategy of making comparisons across workers about their relative performance. We find evidence supporting this conjecture. We define

"inclusiveness" as the number of workers receiving a message and study variation in inclusiveness across incentive contexts and periods of Stage 4. The left panel of Fig. 7 displays the average inclusiveness across the three incentive contexts. Inclusiveness is roughly the same under Revenue Sharing and Weak Link incentives ($p = 0.551$), but is substantially lower under Tournament incentives than under either Revenue Sharing
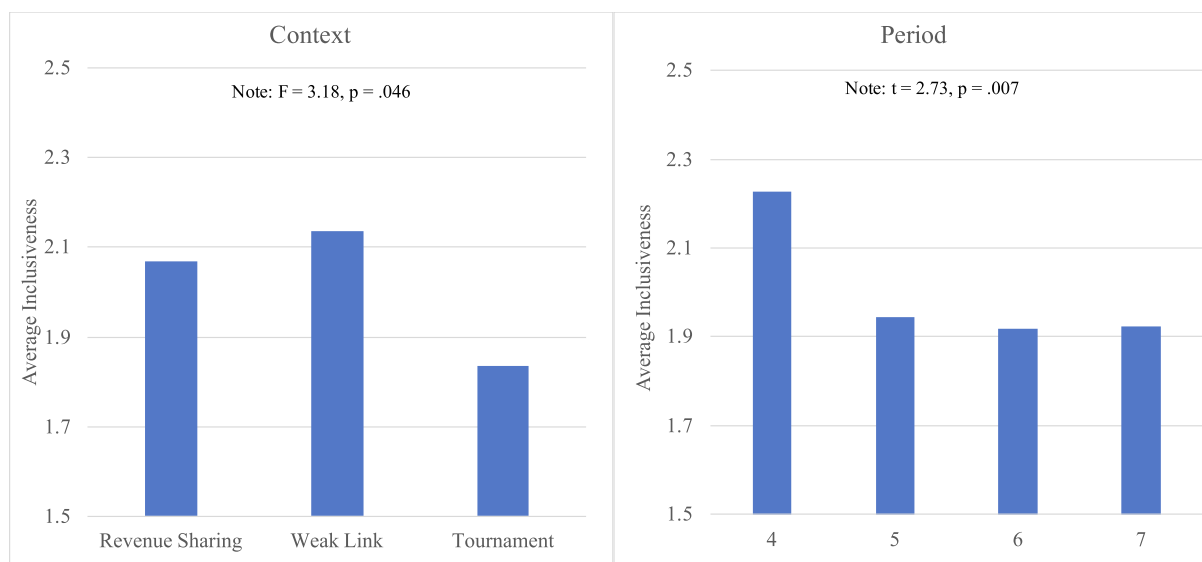
Fig. 7. Inclusiveness of messages across incentive contexts and periods.

($p = 0.061$) or Weak Link incentives ($p = 0.012$).[32]

The right panel reveals an unanticipated feature of the data: inclusiveness decreases significantly with experience. Some of this relationship is due to an overall decrease in work time (messages cannot be sent to workers who have quit), but inclusiveness decreases more rapidly than can be explained by this mechanical factor. We also note that Active Leaders send messages to the worker in their group who produced the least output more frequently (1.76 messages per minute worked) than the highest producer (1.03 messages per minute worked). That is, when non-inclusive messages are used, they tend to be targeted at low performers.

However, the data provides little evidence that non-inclusive messages increase group output as there is almost no correlation between inclusiveness and Residual Output in Stage 3 ($\rho = -0.003$).[33] This lack of a relationship need not imply that using non-inclusive messages does *not* increase group output, because the endogeneity of message inclusiveness makes it challenging to draw firm conclusions from correlational analysis. Moreover, the facts that Active Leaders use non-inclusive messages more frequently with experience and use them in a tailored fashion, targeting them differentially toward more and less productive workers, suggests that they believe such messages can be effective.

Discussion and conclusions

The question of how leaders affect the behavior of followers has been of long interest to organizational researchers, political scientists and psychologists, and is receiving growing attention among economists. However, economists' views of "leadership" have traditionally been narrow, focusing on the use of incentive mechanisms as the primary instrument through which leadership exerts effects. But this perspective

is changing, with a greater number of economists investigating alternative ways in which leaders can motivate followers. Our experiment contributes to this progression, by investigating how a richer form of leadership than that typically studied by economists interacts with the economic incentive context facing followers to influence group performance. Our work yields clear evidence that communication from leaders that is effective for increasing worker productivity in one incentive context is not effective in others. We also provide suggestive evidence that the use and effectiveness of different communication strategies varies across contexts.

A notable feature of our study is that it draws from both leadership research and economics. While starting from the perspective of experimental economics, we study a question—the role of context—rooted in organizational leadership research. We also use a design and conceptualization of leadership that is richer and more complex than that typically employed by economists. This is an important step forward because there has been minimal exchange between the two fields. We hope that our paper can provide valuable insights for each group of researchers and play a small role in facilitating further conversations between the two fields.

Lessons for economists

For economists, perhaps the most important takeaway is the simplest. Experiments in organizational economics often study the effects of varying incentive mechanisms, whereas studies of leadership in experimental economics typically focus on the impact of leadership in a single environment. Our results illustrate a basic point: variation in the outcomes produced either by an incentive mechanism or by the influence of a leader may depend on the interaction between the two factors. The relevance of this interaction is clearly illustrated in our experiment, where the average effect of leadership is large and positive only in the presence of a weak link incentive mechanism and, conversely, the detrimental impact on output of weak link incentives is entirely mitigated by leadership. Moreover, we provide evidence that the nature of leadership is flexible and adapts to the economic context, as illustrated by the differing patterns of messages sent under varying incentive mechanisms. Experimental economists tend to focus on external validity in terms of comparisons between lab and field data—for example, in evaluating the impact of incentives (Esteves-Sorenson, 2018)—but this dimension of external validity is not the only relevant one. Even in a relatively simple laboratory setting where contexts only vary along a

---

[32] One possible explanation for the difference across incentive contexts is purely mechanical. Workers cannot receive a message if they are not working. Thus, if workers spend little time working, messages are necessarily less inclusive. The data do not support this explanation. Restricting attention to Stage 3 with Active Leaders, work times are not significantly different across incentive schemes ($F = .110$; $p = .898$) and, to the extent that any difference exists, work times are highest under Tournament Incentives (see Fig. 3).

[33] Antonakis et al. (2022), provide evidence that group members receiving charismatic leader messages in the presence of one another—thereby creating common knowledge that they are all viewing the same message—can increase the effectiveness of leader communication.

single dimension, the influence of leadership differs substantially across settings. This result suggests that economists need to be cautious about drawing broad conclusions about the impacts of factors, like leadership, based on evidence from very particular settings.

We also demonstrate the potential value of measures not typically employed by economists. Many economists will be unfamiliar with the leadership ratings we employ as a measure of leadership quality, separate from group productivity, although measures of this sort are common in the leadership literature. Economists have generally been slow to embrace the value of process data in understanding why individuals and groups make choices. To the extent that economists have looked at the content of communication (Brandts, Cooper, & Rott, 2019), the focus has typically been on granular analysis of the content of single messages rather than the broad holistic approach embodied by scales like the leadership ratings we study. Yet, the leadership ratings we elicit relate to how groups perform in a way that is not tied to a specific type of content. We see substantial potential value for economists in using these sorts of leadership ratings scales to evaluate richer dimensions of leader quality and performance than those that are the typical focus in economics.

Lessons for leadership researchers

It is also worth noting how our study is distinct from typical approaches in leadership research—reflecting the origin of our study in an economics-based approach to understanding leadership (Zehnder et al., 2017)—and what leadership scholars can learn from these specific features of our study. In this regard, it is helpful to highlight how our paper contributes to the "three C's" highlighted by Garretsen et al. (2020), as potential contributions of employing economic approaches for leadership research.

First, our paper is focused on understanding the importance of a specific form of *context*—the economic incentives facing workers in a group—that is not typically studied in leadership research. However, to an organizational economist, this is often the critical factor for determining worker productivity. The incentives confronting workers also vary substantially across many real-world work settings, making an improved understanding of the interaction between this contextual feature and different forms of leadership a valuable avenue for future organizational research.

Second, we focus on very narrow and stylized *conceptualizations* of important elements of our study, including what we define as leadership, worker effort and group performance. By the standards of much leadership research in psychology and management, the operationalization of these factors in our experiment may seem highly simplified and unnatural.[34] While unnatural, the control afforded by simplicity can be a valuable tool for designing an experiment to isolate the role of some mechanisms while holding others constant. For example, by using a simple real-effort task that workers perform independently as a model for group production, we can hold constant what workers do and the direct influences they have on each other's productivity, allowing us to vary only how they are rewarded for their own and others' work. Also, because leaders can only exercise leadership through a relatively narrow form of one-way communication to followers, we can be certain that any influence leaders have on their groups operate through this channel.

Importantly, our point is not that the simplified concepts and abstractions typically employed by economists are a better way of conducting research on leadership—indeed, we fully acknowledge that economists' approaches are often too simple. Rather, our point is that there are potentially many benefits from dialogue and research that considers the tradeoffs between simple, stylized conceptualizations of

important constructs like leadership and productivity and ones that are richer and more complex. Indeed, there may be important elements of leadership and group production that our simple design ignores, and that may be critical for understanding the relationship between leadership and context. But this creates the potential for future research to borrow elements from experimental economics and more traditional forms of leadership research to address these concerns.

Finally, our use of a controlled laboratory experiment with random assignment allows us to make clear, *causal* inferences about the effects of incentive mechanisms and leadership on group productivity. Aside from the general value of the experimental approach for establishing causality, it is also worth nothing features of our design that may seem unnatural to non-economists but that facilitate causal inference. For example, our experiment varies the presence or absence of leadership by varying only whether the fourth individual in a group can send messages to followers, but holding constant other factors like the presence of a fourth individual and the impacts that workers' productivity have on this group member. This design choice may seem like an unnatural way to add or remove the presence of "leadership"—in contrast, for example, with simply removing the fourth group member—but it allows us to avoid variation in other motivations that may be enhanced or removed when varying the external impacts of workers' output.

Importantly, we also pay special attention to not making causal claims based solely on correlational relationships that may suffer from endogeneity or reverse causality (Antonakis, Bendahan, Jacquart, & Lalive, 2010). While we have noted where strong correlations exist within the data, we fully acknowledge that untangling whether these correlations represent causal relationships is beyond the scope of the current study.

Lessons for the practice of leadership

While some level of caution is always necessary when generalizing from the laboratory, we believe that our results are relevant for leadership practice. First, we document that the effectiveness of a simple form of leadership varies based on the incentive context confronting followers. Outside the laboratory, such variation may occur due to the production technology employed by a firm or the contractual incentives employed by management. Our findings indicate that simple leadership through communication may be most effective in settings characterized by high degree of complementarities between follower actions, like under our Weak Link incentive structure (Camerer & Knez, 1997).

Our analysis also demonstrates that what leaders say to followers can be differentially effective depending on the incentive context and we provide suggestive evidence on the kinds of messages likely to be effective as incentives vary. For example, the use of relative comparisons between workers proved to be effective under Weak Link and Tournament incentives, suggesting a role in such contexts for leaders to provide information to followers about their relative productivity.

Limitations and directions for future research

We view our research as a starting point. Rather than providing definitive evidence on how leadership effectiveness varies across all relevant forms of economic contexts, we demonstrate only that such variation exists and needs to be better understood. As we note above, there may be other forms of leadership relevant for motivating worker effort in the contexts we study and other ways of operationalizing worker productivity in settings involving team production; these may even interact in different ways than what we observe in the same incentive contexts that we study. For example, whereas in our experiment workers perform their tasks independently, much team production involves interdependent activities, and it is unclear whether leadership may have alternative effects as such complementarity varies. The only way to answer such questions is with more data, and we encourage future research to help identify the precise conditions under which our

---

[34] It is worth noting, however, that they already reflect far richer versions than those employed in many economic experiments where, for example, "worker effort" might be a numerical choice, "leadership" a recommended number and "group performance" the sum of numbers chosen.

results hold and where different patterns may obtain.

There are also, of course, limitations created by the tradeoffs involved in choosing particular design features. For example, our use of Passive Supervisors held constant the impacts of workers' output on a fourth individual between conditions with and without active leadership. We view this design feature as desirable from the perspective of holding constant a potentially important social motive that may affect workers' efforts. However, it may also introduce other concerns when interpreting our results, including the possibility of an asymmetric demand effect produced when Passive Supervisors become Active Leaders in one experimental treatment (Lonati, Quiroga, Zehnder, & Antonakis, 2018; Wulff et al., 2023).

Moreover, we consider only three incentive mechanisms. But there are others, both in the organizational economics literature and in practice, that are worth studying. For example, because our interest is primarily in group incentives—where a worker's effort has potential impacts on other workers' payments—we do not investigate piece-rate incentives under which a worker is rewarded only for his or her own productivity. However, such individual incentives are widely employed in organizations and likely interact with leadership in interesting ways (Kvaløy et al., 2015; Fest et al., 2021). Moreover, we do not vary incentives for leaders, which are another potentially important consideration in understanding the interaction between economic incentives and the effects of leadership. We believe it important that future work explores the robustness of our findings to alternative designs.

It is also worth restating that some of the relationships we document are correlational, and that future work is necessary to establish whether there is a causal underlying basis. For example, the relationships we observe between the content of leaders' messages and worker productivity are based on correlations between ratings of everything a leader said over the course of the experiment and outcomes that arose synchronously in that leader's group. It is clear that, at best, the evidence we provide should be interpreted as suggestive of underlying relationships. However, given the importance of better understanding the relationship between what leaders say and the impact it has on workers operating under different economic incentives, we believe that our suggestive correlational findings represent a useful starting point from which additional research can use alternative designs to establish causality.

Finally, our objective in this research has been to contribute to an interdisciplinary approach to the study of leadership, highlighting the complementary value of economics and more traditional forms of leadership research. We believe that our study and findings achieve this objective and break new ground in understanding the interaction between variation in important economic contextual factors and the influence of leaders. However, attempting to conduct research at the intersection of two largely independent fields is challenging and we suspect that many researchers may view unfamiliar aspects of our study skeptically and question whether such features limit the validity of our results for understanding the questions that interest them. Rather than retreating to more traditional ways of studying leadership that are more closely connected to established approaches in their fields, we encourage such researchers to challenge themselves to think critically about how approaches to understanding leadership outside their own field can help them create richer research designs that are relevant for broader audiences and for obtaining more informative answers.

CRediT authorship contribution statement

David J. Cooper: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing. **Giovanna d'Adda:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing. **Roberto A. Weber:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.leaqua.2024.101788.

References

Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *American Economic Review, 101*(2), 470–492.

Anderson, M. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association, 103*(484), 1481–1495.

Antonakis, J. (2012). Transformational and charismatic leadership. *The Nature of Leadership*, 2nd ed., 256-288. Los Angeles: Sage.

Antonakis, J., Avolio, B. J., & Sivasubramaniam, N. (2003). Context and leadership: An examination of the nine-factor full-range leadership theory using the Multifactor Leadership Questionnaire. *The Leadership Quarterly, 14*(3), 261–295.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly, 21*(6), 1086–1120.

Antonakis, J., d'Adda, G., Weber, R. A., & Zehnder, C. (2022). Just words? Just speeches? On the economic value of charismatic leadership. *Management Science, 68*(9), 6355–6381.

Antonakis, J., Fenley, M., & Liechti, S. (2011). Can charisma can be taught? Tests of two interventions. *Academy of Management Learning & Education, 10*(3), 374–396.

Bass, B. M. (1990). From transactional to transformational leadership: Learning to share the vision. *Organizational Dynamics, 18*(3), 19–31.

Bandiera, O., Barankay, I., & Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association, 11*(5), 1079–1114.

Benjamini, Y., Krieger, A., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika, 93*, 491–507.

Bloom, N., & Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics, 122*(4), 1351–1408.

Boehm, S. A., Dwertmann, D. J., Bruch, H., & Shamir, B. (2015). The missing link? Investigating organizational identity strength and transformational leadership climate as mechanisms that connect CEO charisma with firm performance. *The Leadership Quarterly, 26*(2), 156–171.

Boulu-Reshef, B., Holt, C. A., Rodgers, M., & Thomas-Hunt, M. (2020). The impact of leader communication on free-riding: An incentivized experiment with empowering and directive styles. *Leadership Quarterly, 31*(3), Article 101351.

Brandts, J., & Cooper, D. J. (2007). It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure. *Journal of the European Economic Association, 5*, 1223–1268.

Brandts, J., Cooper, D. J., & Rott, C. (2019). 21. Communication in laboratory experiments. In *Handbook of Research Methods and Applications in Experimental Economics* (p. 401).

Brandts, J., Cooper, D. J., & Weber, R. A. (2015). Legitimacy, communication, and leadership in the turnaround game. *Management Science, 61*(11), 2627–2645.

Bull, C., Schotter, A., & Weigelt, K. (1987). Tournaments and piece rates: An experimental study. *Journal of political Economy, 95*(1), 1–33.

Camerer, C., & Knez, M. (1997). Coordination in organizations: A game-theoretic perspective. *Organizational Decision Making*, 158–188.

Camerer, C., & Weber, R. A. (2013). Experimental organizational economics. *The Handbook of Organizational Economics*, 213–262.

Carpenter, J., Matthews, P. H., & Schirm, J. (2010). Tournaments and office politics: Evidence from a real effort experiment. *American Economic Review, 100*(1), 504–517.

Chattopadhyay, R., & Duflo, E. (2004). Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica, 72*(5), 1409–1443.

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *American Economic Review, 99*(1), 431–457.

Chen, R., & Chen, Y. (2011). The potential of social identity for equilibrium selection. *American Economic Review, 101*(6), 2562–2589.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*, 213–220.

Cooper, D. J., and Hamman, J. (2021). Leadership and Delegation of Authority. *The Handbook of Labor, Human Resources and Population Economics*, Berlin: Springer-Verlag.

Cooper, D. J., & Kagel, J. H. (2016). Other-regarding preferences: A selective survey of experimental results. In John Kagel, & Alvin Roth (Eds.), *The Handbook of Experimental Economics* (Vol. 2). Princeton University Press.

Cooper, D. J., & Weber, R. A. (2020). Recent advances in experimental coordination games. *Handbook of Experimental Game Theory*, 149–183.

Cronshaw, S. F., & Lord, R. G. (1987). Effects of categorization, attribution, and encoding processes on leadership perceptions. *Journal of Applied Psychology, 72*(1), 97.

D'Adda, G., Darai, D., Pavanini, N., & Weber, R. A. (2017). Do leaders affect ethical conduct? *Journal of the European Economic Association, 15*, 1177–1213.

Delfgaauw, J., Dur, R., Sol, J., & Verbeke, W. (2013). Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics, 31*(2), 305–326.

Devetag, G., & Ortmann, A. (2007). When and why? A critical survey on coordination failure in the laboratory. *Experimental Economics, 10*(3), 331–344.

Englmaier, F., Grimm, S., Grothe, D., Schindler, D., and Schudy, S. (2021). The value of leadership: Evidence from a large-scale field experiment. *CESifo Working Paper No. 9273.*

Eichenseer, M. (2023). Leading by example in public goods experiments: What do we know? *The Leadership Quarterly, 34*(5), Article 101695.

Ellman, M., & Pezanis-Christou, P. (2010). Organizational structure, communication, and group ethics. *American Economic Review, 100*(5), 2478–2491.

Esteves-Sorenson, C. (2018). Gift exchange in the workplace: Addressing the conflicting evidence with a careful test. *Management Science, 64*(9), 4365–4388.

Falk, A., Becker, A., Dohmen, T., Huffman, D. B., & Sunde, U. (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science, 69*(4), 1935–2545.

Fest, S., Kvaløy, O., Nieken, P., & Schöttner, A. (2021). How (not) to motivate online workers: Two controlled field experiments on leadership in the gig economy. *The Leadership Quarterly, 32*(6), Article 101514.

Fiedler, F. E. (1978). The contingency model and the dynamics of the leadership process. *Advances in experimental social psychology, 11*, 59-112. Academic Press.

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

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378.

Gächter, S., and E. Renner (2003). Leading by Example in the Presence of Free Rider Incentives. Paper presented at a Conference on Leadership, Lyon.

Garretsen, H., Stoker, J. I., & Weber, R. A. (2020). Economic perspectives on leadership: Concepts, causality, and context in leadership research. *The Leadership Quarterly, 31*(3), Article 101410.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association, 1*(1), 114–125.

Grossman, S. J., & Hart, O. D. (1983). An analysis of the principal-agent problem. *Econometrica, 51*(1), 7–45.

Hermalin, B. (1998). Toward an economic theory of leadership: Leading by example. *American Economic Review, 88*(5), 1188–1206.

Howell, J. M., & Avolio, B. J. (1993). Transformational leadership, transactional leadership, locus of control, and support for innovation: Key predictors of consolidated-business-unit performance. *Journal of Applied Psychology, 78*(6), 891.

Jones, B., & Olken, B. (2005). Do leaders matter? National leadership and growth since world war II. *Quarterly Journal of Economics, 120*(3), 835–864.

Kosfeld, M. (2020). The role of leaders in inducing and maintaining cooperation: The CC strategy. *Leadership Quarterly, 31*(3), Article 101292.

Kvaløy, O., Nieken, P., & Schöttner, A. (2015). Hidden benefits of reward: A field experiment on motivation and monetary incentives. *European Economic Review, 76*, 188–199.

Laffont, J. J., and Martimort, D. (2009). The theory of incentives: the principal-agent model. *The Theory of Incentives*. Princeton university press.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review, 90*(5), 1346–1361.

Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management, 64*, 19–40.

Lord, R. G., Foti, R. J., & De Vader, C. L. (1984). A test of leadership categorization theory: Internal structure, information processing, and leadership perceptions. *Organizational Behavior and Human Performance, 34*(3), 343–378.

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

Nagin, D. S., Rebitzer, J. B., Sanders, S., & Taylor, L. J. (2002). Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *American Economic Review, 92*(4), 850–873.

Nalbantian, H., & Schotter, A. (1997). Productivity under group incentives: An experimental study. *American Economic Review, 87*(3), 314–341.

Oc, B. (2018). Contextual leadership: A systematic review of how contextual factors shape leadership and its outcomes. *The Leadership Quarterly, 29*(1), 218–235.

Osborn, R. N., Hunt, J. G., & Jauch, L. R. (2002). Toward a contextual theory of leadership. *The Leadership Quarterly, 13*(6), 797–837.

Porter, L. W., & McLaughlin, G. B. (2006). Leadership and the organizational context: Like the weather? *The Leadership Quarterly, 17*(6), 559–576.

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

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature, 37*(1), 7–63.

Purvanova, R. K., & Bono, J. E. (2009). Transformational leadership in context: Face-to-face and virtual teams. *The Leadership Quarterly, 20*(3), 343–357.

Sahin, S. G., Eckel, C., & Komai, M. (2015). An experimental study of leadership institutions in collective action games. *Journal of the Economic Science Association, 1*(1), 100–113.

Serra-Garcia, M., van Damme, E., & Potters, J. (2011). Hiding an inconvenient truth: Lies and vagueness. *Games and Economic Behavior, 73*(1), 244–261.

van der Velde, A., & Gerpott, F. H. (2023). When subordinates do not follow: A typology of subordinate resistance as perceived by leaders. *The Leadership Quarterly, 34*(5), Article 101687.

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

Wulff, J. N., Sajons, G. B., Pogrebna, G., Lonati, S., Bastardoz, N., Banks, G. C., & Antonakis, J. (2023). Common methodological mistakes. *The Leadership Quarterly, 34*(1), Article 101677.

Zehnder, C., Herz, H., & Bonardi, J. P. (2017). A productive clash of cultures: Injecting economics into leadership research. *The Leadership Quarterly, 28*(1), 65–85.