Received: 20 May 2023 | Accepted: 24 November 2023

DOI: 10.1111/jems.12570

ORIGINAL ARTICLE

Journal of Economics & Management Strategy WILEY

Free riding, democracy, and sacrifice in the workplace: Evidence from a real-effort experiment

Kenju Kamei[1] | **Katy Tabero**[2]

[1]Faculty of Economics, Keio University, Tokyo, Japan

[2]Department of Economic, Social and Political Sciences, University of Southampton, Southampton, UK

Correspondence
Kenju Kamei, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan.
Email: kenju.kamei@gmail.com

Funding information
Murata Science Foundation, Grant/Award Number: Year 2021; ESRC North East Doctoral Training Partnership, Grant/Award Number: ES/P000762/1

Abstract

Teams are increasingly popular decision-making and work units in firms. This paper uses a novel real-effort experiment to show that (a) some teams in the workplace reduce their members' private benefits to achieve a group optimum in a social dilemma and (b) such endogenous choices by themselves enhance their work productivity (per-work-time production)—a phenomenon called the "dividend of democracy." In the experiment, worker subjects are randomly assigned to a team of three, and they then jointly solve a collaborative real-effort task under a revenue-sharing rule in their group with two other teams, while each individual worker can privately and independently shirk by playing a Tetris game. Strikingly, teams exhibit significantly higher productivity (per-work-time production) when they can decide whether to reduce the return from shirking by voting than when the policy implementation is randomly decided from above, irrespective of the policy implementation outcome. This means that democratic culture directly affects behavior. On the other hand, the workers under democracy also increase their shirking, presumably due to enhanced fatigue owing to the stronger productivity. Despite this, democracy does not decrease overall production thanks to the enhanced work productivity.

1 | INTRODUCTION

Teams are increasing popular in firms as decision-making and work units (e.g., Kamei & Tabero, 2022). However, team decision-making and teamwork feature a coordination problem that involves complexities relating to imperfect information, monitoring, and agency costs (e.g., Alchian & Demsetz, 1972; Marschak & Radner, 1972). Thus, maintaining motivation among workers is particularly difficult when teams are involved in the workplace, and their private interests conflict with group interests (e.g., Bolton & Dewatripont, 2004)—a typical example of this is moral hazard in groups (e.g., Alchian & Demsetz, 1972; Holmstrom, 1982). Democratic culture may help mitigate conflict within and across the teams by not only enhancing their self-determination and intrinsic motivation to cooperate (e.g., Deci & Ryan, 1985, 2000), but by also providing workers with opportunities to signal their willingness to cooperate with their peers through democratic processes (e.g., Bergh et al., 2014; Connelly et al., 2011), thereby making it easy to achieve the group optimum. In such environments, workers may decide to collectively decrease temptations by reducing their private gains for the sake of group interests. But how large could the effects of workplace democracy per se on productivity potentially be? Precisely what motivates workers' sacrificial behaviors?

How to overcome moral hazard in groups is an important, active question in economics and management. A large body of research spanning several decades has found that workers have difficulty cooperating with each other when free-riding incentives are sufficiently strong in a social dilemma (e.g., Ledyard, 1995; Zelmer, 2003). Specifically, prior experimental research suggests that while some people demonstrate conditional willingness to cooperate, groups usually cannot sustain cooperation for various reasons, for example, their cooperation behaviors are heterogeneous (e.g., Fischbacher et al., 2001), they are easily discouraged by seeing their peers free ride (e.g., Fischbacher & Gächter, 2010); or many tend to cooperate but by less than others (e.g., Thöni & Volk, 2018). This echoes theoretical research that describes why moral hazard arises among workers when their effort levels are not perfectly observable (e.g., Alchian & Demsetz, 1972; Holmstrom, 1982).[1] Both the theoretical and empirical literature therefore discuss that some institutional solutions, such as corporate culture and working environments, competition (e.g., internal job ladder, tournament), punishment and rewards, monitoring, and sorting, are required to assist collaboration and cooperation in the workplace (see, e.g., Chaudhuri, 2011 for experimental economics literature; Prendergast, 1999 for personnel economics). This study contributes to the large body of literature by investigating the impact of workplace democracy, workers' behavioral reactions to a reduction in incentives to shirk, as well as the reasoning behind their voluntary sacrificial behaviors in the workplace.

This study is the first to experimentally measure the so-called "dividend of democracy" when the decision-making and work units are teams. The "dividend of democracy" refers to an effect that democracy directly has on the behavior of those involved. The role of democratic culture on worker behavior has been actively studied in the literature in economics and management for the last two decades (see Dal Bó, 2010 for a survey). In particular, prior experiments in economics have shown that democracy in implementing a prosocial policy boosts cooperation in experimental games, such as public goods or prisoner's dilemma games, as it directly affects people's own behavior and beliefs on their peers' cooperativeness (e.g., Dal Bó et al., 2010; Kamei, 2016; Sutter et al., 2010; Tyran & Feld, 2006). Among others, Tyran and Feld (2006) and Dal Bó et al. (2010, 2019) provide methods to isolate the dividend of democracy from selection bias, showing that the dividend of democracy is large. Scholars have recently started to study the applicability of such a dividend of democracy in a workplace setting by using a design with real-effort tasks, but the results surprisingly showed that democracy per se may not have strong effects in real-effort settings (e.g., Dal Bó et al., 2019; Kamei & Markussen, 2023; Melizzo et al., 2014). While all prior experiments on democracy used individuals as the decision-making unit, the present study uses teams as the decision-making unit of policy-making and task-solving for the first time, and finds a significant dividend of democracy on work productivity (per-work-time production).

The policy available to workers in this study is one to reduce material incentives to shirk. Collectively sacrificing one's benefits through fostering customs, conventions, or rules with the aim of resolving conflicting interests has been conceptually discussed in the literature in the social sciences (such as anthropology) and biology as key features of humans. Anecdotal evidence includes costly participation in religious groups and rituals, or recreational activities in societies (e.g., dance and festivals), food sharing (e.g., turtle hunting by islanders for funerary rituals), holding redistributive feasts, and attending group raids and defense (see, e.g., Hagen & Bryant, 2003; Hawkes & Bliege Bird, 2002; Iannaccone, 1992; Sosis & Alcorta, 2003; Smith & Bliege Bird, 2000; Sosis & Bressler, 2003). The mechanism is described as follows: sacrificing serves as a costly signal of one's own quality (e.g., Smith & Bliege Bird, 2005; Gintis et al., 2001), thus helping to coordinate with others to cooperate and bolster a cooperative atmosphere in dilemma situations.[2,3] Several laboratory experiments used public goods games or prisoner's dilemma games to study costly human sacrificing tendencies with high internal validity (e.g., Aimone et al., 2013; Brekke et al., 2011; Grimm & Mengel, 2009). The findings are that some groups (individuals) do collectively (voluntarily) decide to reduce their private returns, thereby enhancing welfare. However, to the best of the authors' knowledge, sacrificing has not been studied in the workplace context using a naturally occurring, real-effort task, although recently there has been a theoretical attempt to characterize the effects of sacrificing in the workplace (Bisetti et al., 2022).[4]

While sacrifice has received less attention in the workplace so far, it is becoming more and more relevant due to a surge in remote working (potentially boosting shirking) triggered by the Covid-19 crisis and technological advances. A broad range of examples of unobserved shirking activities and countermeasure policies are readily available in the modern workplace. For example, cyberloafing is a typical and costly issue whereby employees covertly use their computer or internet access for personal use during work time. The issue is especially serious when they are not in an office. The employer may decide to introduce measures to counter employees' cyberloafing, for example, by monitoring their use of the internet, imposing internet restriction policies and penalties for breaching them, or placing technical restrictions on employees' access to certain nonwork websites.[5] While such policies can simply be imposed from above by managerial staff or teams, the policies can also be enacted through decentralized decision-making. For instance, a factory may produce mechanical parts by assigning workers to several teams to take advantage of specialization. When their environment is democratic and they recognize that cyberloafing undermines productivity, they may democratically decide to enact a restriction policy across the teams, with an aim of improving the performance in the factory if they believe that productivity impacts their material benefits, such as their wages, bonuses, or rewards.[6] Similar scenarios are common across various employment relationships, for example, a branch in a consulting firm, or a sales office for products (e.g., cars). Another related example is "moonlighting" by which employees work multiple jobs, sometimes simultaneously and/or without the permission of their main employer.[7] For example, an employee may commit to working 5 days per week while secretly working for another firm to earn more by shirking the main job. Alternatively, an employee may hold a secondary side job that takes place outside of their primary work hours, but spend time during those hours contributing to their secondary job, such as responding to e-mails, advertising, or checking their website. The increase in remote working in recent years makes monitoring more difficult. Policies to make working on the side difficult and materially unbeneficial (e.g., through using a screen-capture tool and work-time tracking) may be considered if such free riding significantly undermines production in the main workplace.

This paper conducts an experiment with a novel "collaborative" real-effort task. In the experiment, worker subjects are randomly assigned to a team of three, and three teams constitute a group. The real-effort task requires each team to jointly calculate the number of 4 s in a matrix whose cells contain 1, 2, 3, or 4 s. At the onset of the experiment, each team member is assigned a number, player 1, 2, or 3, such that they have different numbers from each other. The matrix that player k is allocated includes only number ks while the other three numbers are blacked out. Each member counts their assigned numbers, shares the counting outcome, and jointly calculates the final answer, on the condition that their remuneration is based on revenue sharing in the group. To mimic the conflict between work and shirk (or another activity) in the real workplace, each member is allowed to privately and independently play a computer game, Tetris. A shirker can privately earn some material returns from gaming on top of psychologically enjoying Tetris. The incentive structure is therefore similar to the so-called stag-hunt game (e.g., Hume, 2000 [1739]; Rousseau, 2009 [1755]): all three members must work on counting to earn a reward as a team from the collaborative task, but each member has an incentive to deviate to gaming (thereby earning some reward privately). This setup is parallel to the real-world examples of modern distractions at work, such as cyberloafing and moonlighting. Before the task-solving phase begins, a policy that reduces the incentive to play Tetris ("reduction policy," hereafter) is implemented in a group either *democratically* (by voting) or *autocratically* (randomly by the computer without voting). The two treatments (democratic, or autocratic) are designed using a between-subjects design.

In addition to the contribution to the literature on workplace democracy, this research is novel in two additional aspects. First, this study provides significant methodological contributions with the newly used "collaborative" counting task and gaming as a real activity. While much research has been conducted using real-effort tasks, a significant issue has been reported by Araujo et al. (2016) that workers' incentive elasticity of outputs may be too small with the real-effort tasks used. Recently, Corgnet et al. (2015) and Kamei and Markussen (2023) allowed subjects to use, respectively, internet browsers and comedy videos, as real leisure activities. Both of the papers showed that such activities enhance incentive elasticity in experiments. The present paper adds to the literature by using gaming as a real, but controlled, leisure activity for the first time in a computerized real-effort experiment. Further, the members of each team jointly work on a *collaborative* counting task. While an individual counting zeros task is widely used in the literature (e.g., Abeler et al., 2011; Falk et al., 2006; Kamei & Markussen, 2023), the use of a collaborative version is the first attempt in the literature, to the authors' knowledge. This design is meaningful as collaboration is a central aspect of teamwork in many firms and organizations, and the new task is designed to explicitly simulate the coordination structure. Notice the stark difference in the game structure between the standard counting task and the collaborative counting task. The new collaborative one is a coordination game: individuals earn from the team task only when all three members work by spending time counting and communicating accurately and effectively. The new task allows researchers to study coordination in the structure of a stag-hunt game in a natural way, even outside the research agenda of organizational economics and management.

Second, the experiment is the first to investigate workers' sacrifice decisions and their reasoning in a real-effort environment. While prior research used experimental games such as public goods games to propose that some individuals will reduce their private gains in dilemma situations, showing that such decisions may lead to a Pareto improvement empirically (e.g., Aimone et al., 2013; Brekke et al., 2011; Grimm & Mengel, 2009), its validity in the workplace setting is unclear as little research used naturally occurring, real effort in their experiments. Equally important is that no research explores what may drive workers to sacrifice their private gains, because no data is available regarding their thinking. Subjects in the present experiment decide whether to reduce their private gains through communication within their team as a team decision. This design enables us to collect a unique incentive-compatible data set to study the reasoning behind sacrifice decisions. A well-established coding exercise is applied to the communication logs to uncover reasoning effectively.

The experiment results reveal some teams' preferences for sacrifice and evidence of a dividend of democracy. 40.9% of teams voted to reduce the incentive to play the game, and as a result, the reduction policy was enacted for 38.7% of groups. Teams that were involved in democratic decision-making exhibited significantly higher work productivity, that is, performance per minute of working, than those in the regime where the computer randomly decided policy implementation, whether the reduction policy was imposed or not. This means that the democratic culture per se directly affected behavior. Having said that, the workers under democracy reduced work time compared to those under autocracy, presumably due to fatigue accumulating more quickly for the former. Nevertheless, the former did not decrease team production overall thanks to the enhanced work productivity.

The present paper also provides reasoning behind workers' sacrifice decisions based on a standard coding exercise. It reveals that the units that planned to exclusively work on task-solving, believed that the reduction policy would deter others from shirking, or those that had supportive team atmospheres supported the reduction policy. It also uncovers the value of signaling through sacrificial decisions to encourage collaboration: teams who believed that other teams would complete tasks following the vote performed strongly.

The rest of the paper proceeds as follows: Section 2 summarizes the experimental design, and Section 3 reports the results. Section 4 provides insights obtained from an analysis of communication dialogues, and Section 5 concludes.

2 | EXPERIMENTAL DESIGN

The experiment is designed using a collaborative real-effort task devised for this study. At the onset of the experiment, worker subjects are randomly assigned to a team of three. The three members are then randomly assigned ID numbers, 1, 2, or 3, so that each member receives a different number from one another. Anonymity is retained such that they do not know the identity of the other members (e.g., faces, names, gender). Let us call the player who is assigned number $k \in \{1, 2, 3\}$ "player k." The team composition and the assigned ID numbers do not change for the entire experiment (partner matching). Three teams further constitute a group (each group thus has nine members). The group composition also does not change throughout. Section 2.1 explains the nature of the collaborative team real-effort task, after which Section 2.2 explains the structure of the experiment, a summary of treatments, the remuneration system, and the reduction policy that could be implemented in each group. Supporting Information Appendix A summarizes the experimental procedure and includes instructions used in the experiment.

2.1 | A collaborative real-effort task

Three members in a respective team collaboratively solve a variant of the counting task ("collaborative counting task"). The original "counting task" (e.g., Abeler et al., 2011; Falk et al., 2006; Kamei & Markussen, 2023) is an individual real-effort task in which subjects independently count the number of 0 s in a matrix that contains 0 s and 1 s. To the authors' knowledge, no collaborative version of the counting task has been devised and used in any prior experiments. In the new collaborative counting task, the three team members are provided with a 15×15 matrix, each cell of which has a randomly generated integer between 1 and 4 (each integer is independently drawn with a probability of 25%), and are then asked to submit the number of 4 s. Collaboration is required to find the correct answer, because only number ks appear on the computer screen of player k, while the other three numbers are blacked out—see Figure 1 for a screen image for player 1. Each team can find the correct answer if player k counts the number of ks correctly and shares it with their teammates, and the team calculates the number of 4 s accurately after that. For example, if the numbers 1, 2, and 3 s are, respectively, 32, 14, and 43, then the correct answer (the number of 4 s) is $225 - 32 - 14 - 43 = 136$. A calculator is available on each subject's computer screen. How to calculate the number of 4 s, and by whom, is up to each team's discretion. When the team decides on and wants to submit the answer, all three members must submit the team's joint answer on their own computer screens. Hence, in the submission stage as well they must communicate with each other about their team's decision to answer correctly. In the case of disagreement, a member can submit a

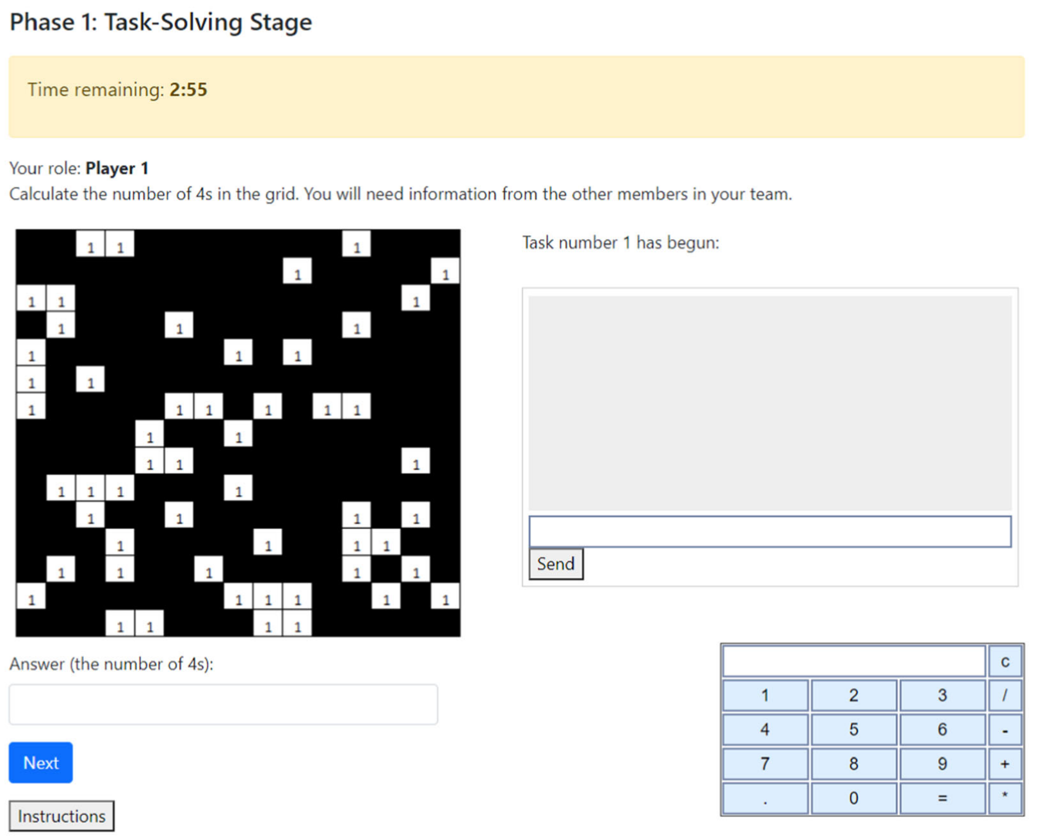


FIGURE 1 A screen for collaborative counting task. A screen image for player 1. The numbers 2, 3, and 4 s are blacked out on the screen that player 1 sees. The 15×15 matrix in this figure is for illustration only.

different answer from the others.[8] However, the answer will then be counted as incorrect. Once all three members submit an answer, a new 15×15 matrix with randomly generated 1, 2, 3, and 4 s in each cell is assigned to the team, and the process repeats.

Free-form communication is available using an electronic chat window during the entire task-solving process (Figure 1; Supporting Information Appendix A also includes the screen image of the chat window), and messages are recorded. This design piece helps the researchers study the reasoning behind members' behaviors, postexperiment. While any sort of communication, such as discussing strategy to solve the problems, sharing the number of *k*s, or chatting about unrelated matters, is allowed, subjects are prohibited from using any kind of offensive language or sharing any information that compromises anonymity.[9]

The more questions a team answers correctly, the higher the earnings they can generate in their group. Each correct answer is rewarded with 180 UK pence in the experiment. How the 180 pence are distributed within the team or the group is explained in Section 2.2.

2.2 | The experiment

There are two treatments that vary by changing the process to decide whether to enact a policy to curb members' shirking or not. A between-subjects design is used to avoid behavioral spillover (e.g., Bednar et al., 2012) or possible spill-over effects of democracy (e.g., Kamei, 2016). The experiment begins with a practice phase, which is the same for all subjects in the experiment. The main task-solving phase begins after the practice phase and differs by the treatment.[10] The practice phase plays a role in not only familiarizing subjects with the collaborative counting task, but also providing them with an opportunity to try the task and learn their ability to solve it.

In the practice phase, each team performs the collaborative counting task for 3 mins.[11] While they can answer as many questions as they wish, they are not informed whether they answer each question correctly during the 3-min period. They are instead informed of the number of correct answers at the end of the practice phase. Remuneration is based on revenue sharing in the team. This means that the money a team earns is equally divided among the three team members (each member receives 60 = 180/3 UK pence for a correct response). Each team does not interact with the other two teams in their group in this practice; nor are they informed of the performances of the other teams.

In the main task-solving phase, each team performs the collaborative counting task for a much longer duration—35 mins—with a revenue-sharing rule in their *group*. This means that the credit of each correct answer (180 UK pence) is equally shared among the three teams, that is, nine individuals as each team has three members. The marginal per-capita return is calculated as 20 (=180 × 1/9) UK pence.

There are two more distinct aspects in the main task-solving phase. First, unlike the practice phase, each member can privately shirk by playing Tetris. They can do so by simply pressing the "Game" button (Figure 2a). The screen is then switched to the Tetris site (Figure 2b). No one, including their teammates, is made aware of a member's shirking unless the member voluntarily reports their behavior using the electronic chat window. Further, the shirker earns a return by staying in the Game screen: 18 pence/min spent in the Game screen.[12] They can return to the work site from the Game site at any time. Workers are *not* allowed to work while playing Tetris, whose requirement enables the researchers to quantify shirking versus work time as their work decisions. It should be noted here that the design of gaming was carefully made to enhance external validity, as workers often have alternative activities available when shirking in the workplace rather than being inactive. An advantage of using gaming over internet browsing (Corgnet et al., 2015) as an alternative activity is the high level of control: workers may use internet browsers differently as their preferences are heterogeneous. This feature shares similarities with Kamei and Markussen (2023) that adopted comedy

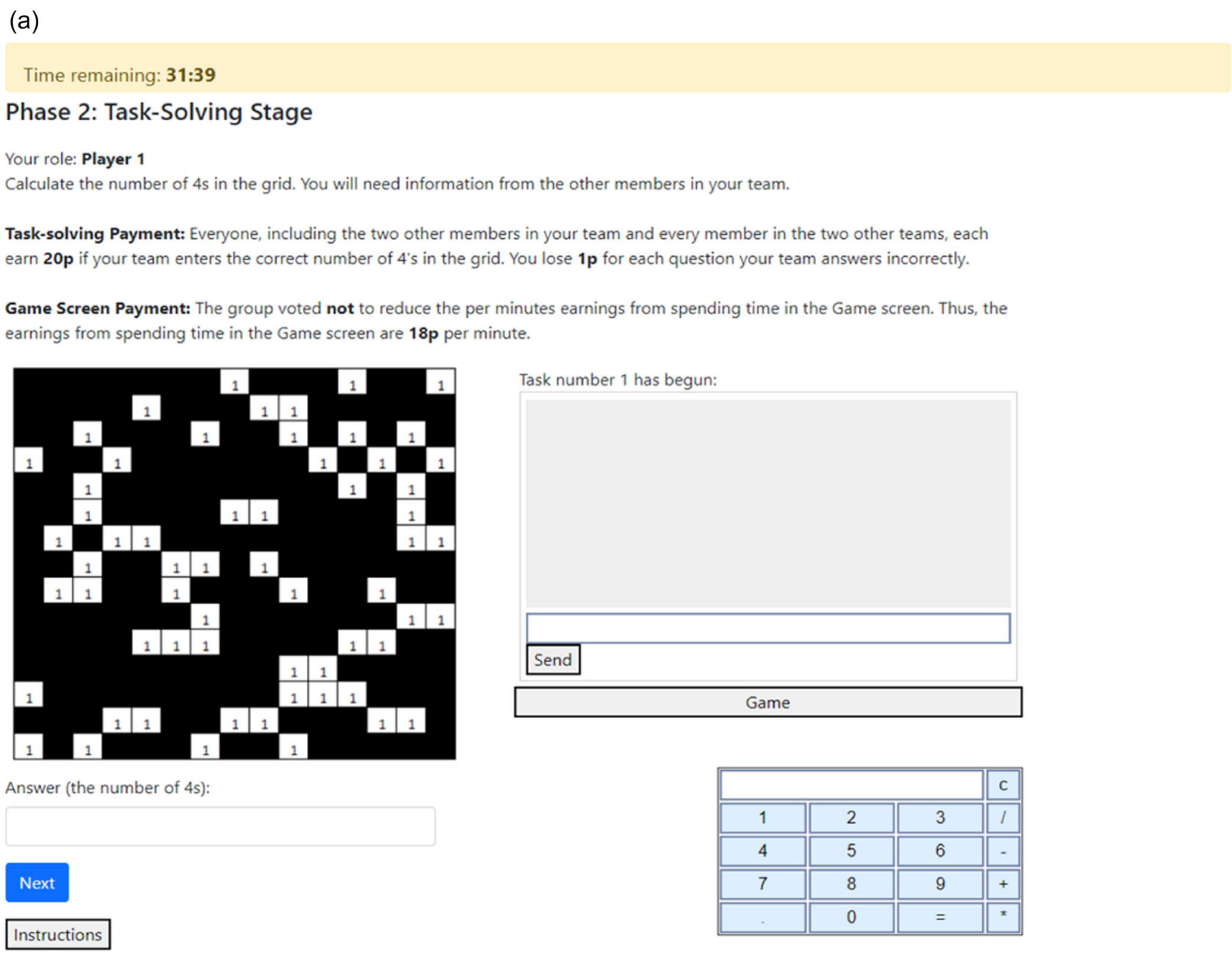


FIGURE 2 A screen image for collaborative counting task in the main task-solving phase. (a) Work site and (b) game site.

(b)

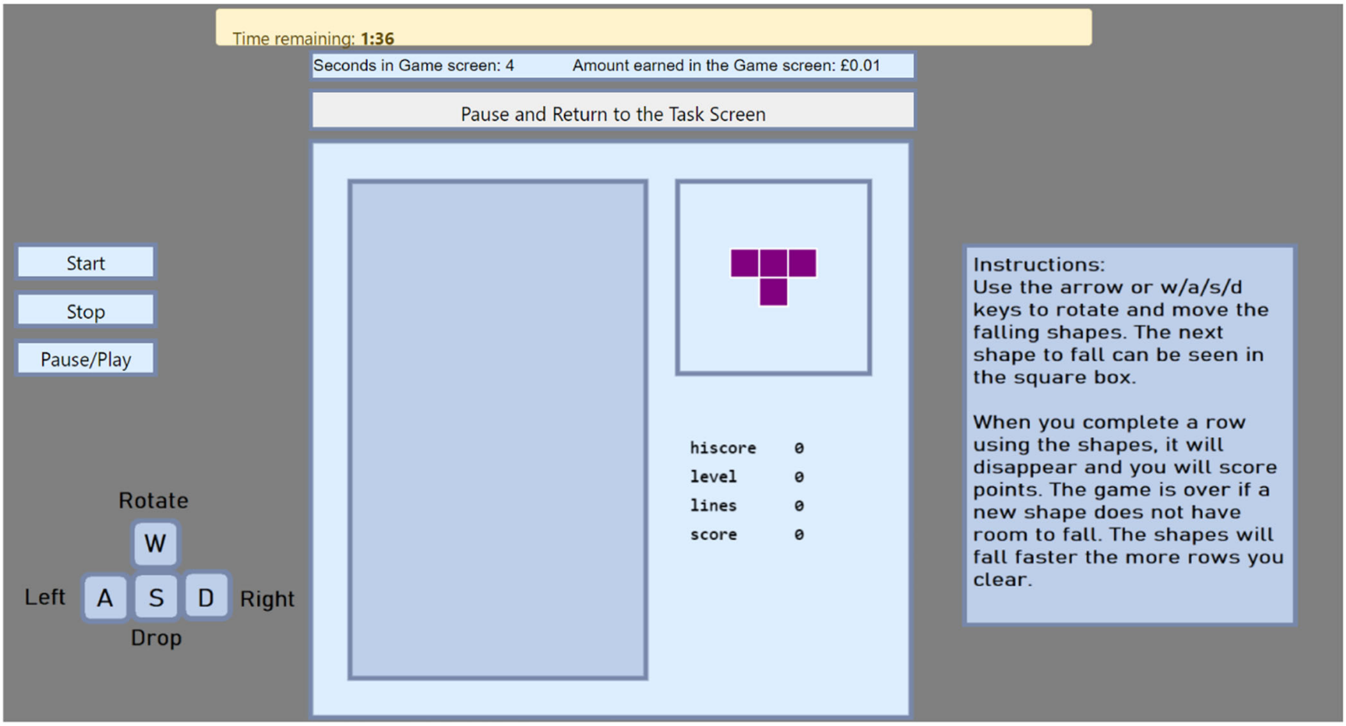


FIGURE 2 (Continued)

video clips as an alternative activity. However, using a game is better than video clips because implementation is difficult with the use of the latter. While headsets were provided to each subject in Kamei and Markussen (2023), the authors acknowledged that even a small ripple of laughter and sounds could contaminate the data. In contrast, gaming is a purely independent, quite leisure activity.

Notice that with the gaming option, the incentive structure of the team task in the main phase is one of the so-called "stag-hunt game" if they are highly skilled. Each team member can earn a small material gain with certainty by deviating from collaboration. However, they earn a large team payoff when all three team members work on the counting task, if each of them can count numbers sufficiently quickly.

Second, there is a penalty of three pence per incorrect answer in the main task-solving phase. This penalty is imposed on the team that commits the error, not the whole group. Such penalties are commonly used in the real workplace; for example, poor performance or mistakes can result in monetary or social sanctions, increased threat of dismissal (through escalation procedures or informal threats), or reduced pay where performance-related wages or team bonuses are in place (see, e.g., Doellgast & Marsden, 2019; Gibbons & Henderson, 2013; McNamara et al., 2022). The penalty is equally shared among the three members in the team (i.e., one penny is deducted from the payoff per team member). In short, the payoff of member i in team k can be expressed as Equation (1):

$$\pi_{k,i}(c_k, ic_k, g_i) = 20\left[\sum_{n=1}^{3} c_n\right] - ic_k + r \cdot g_i, \tag{1}$$

where c_k and ic_k are the numbers of, respectively, correct and incorrect answers by team k, g_i is the time (minutes) that member i spends in the Game screen, and r is per-minute return from shirking. Notice that their work time is $35 - g_i$ as they are not allowed to work while playing Tetris. Using the revenue-sharing rule per group and the alternative leisure opportunity, the aim is to model the work environment as a tension across teams between task-solving and gaming (i.e., social dilemma). As intended, gaming was a privately optimal option for almost all teams in the experiment sessions—see Section 3.3.

Worker subjects are not informed of how many questions their teams or other teams answer correctly during the 35-min task-solving phase.[13] Instead, at the end of the task-solving phase they learn (a) the total number of correct and incorrect responses of their own team and (b) the total number of correct responses in their group. This setup is realistic; for example, in manufacturing, the manager will learn how many defectives they have among mechanical parts produced in a given day, only after quality checks at specified intervals.

At the beginning of the main task-solving phase, the return from staying in the Game screen (r) could decrease from 18 to 16 pence/min. Notice that the size of the incentive change is very small at only two pence. This means that the reduction policy can be thought of as a nondeterrent sanction policy, that is, a policy that *does not* alter the privately optimal behaviors of workers in a group (e.g., Kamei, 2016; Tyran & Feld, 2006). As briefly reported in Section 3.3., this interpretation turns out to be correct in the experiment: gaming was a privately optimal choice for almost all teams, whether the reduction policy was in place or not, due to the strong incentives to free ride on other teams' work efforts.

The process to implement the reduction policy differs by treatment. There are two treatments, one with exogenous imposition of the reduction policy ("EXO" [Exogeneous], hereafter), and one where the policy is endogenously selected ("ENDO" [Endogenous], hereafter). In the EXO treatment, the policy is imposed in each group by the computer randomly (i.e., with a probability of 50%). By contrast, in the ENDO treatment, the policy is implemented based on majority voting by the three teams.[14] The voting procedure follows three steps:

Step 1. The three members in each team are given 3 mins to discuss, using an electronic chat window (e.g., Kamei, 2019b; Luhan et al., 2009), whether they want to reduce the per-minute earnings gained by staying in the Game screen. The communication contents are not revealed to any other team. See Supporting Information Appendix A.3 for a screen image of this step.

Step 2. After the 3-min discussion, the three members each submit their preferred decisions. If the three submit the same decision, it becomes their team vote. However, in the case of disagreement they can submit whatever they prefer, in which case whichever receives at least two members' support is implemented as their team vote.

Step 3. The reduction policy is implemented in the group based on majority voting. Specifically, it is implemented (not implemented) if it receives two or three supporting (opposing) team votes. All subjects in the group are informed of the vote outcome and the number of supporting votes.

Notice that as the reduction policy, despite the size of the reduction being small, may encourage teams to work harder through decreasing the material incentives to shirk, thereby leading to a higher payoff, groups may decide to decrease such private returns by voting. As summarized in Table 1, there are four possible institutional outcomes in this study.

2.3 | Theoretical predictions

Theoretical predictions on the dividend of democracy can be derived by setting a utility function for the player and then finding their utility-maximizing behavior. As shown in online Supporting Information Appendix B, a calculation

TABLE 1 Treatments, Distribution of Votes and Institutional Outcomes.

Treatment and institutional outcome	Condition in which the policy is/is not implemented	# Subjects	# Of subjects in proreduction teams	# Of subjects in antireduction teams
ENDO treatment	Voting	279	114	165
(i) Policy was implemented	At least two teams vote for the policy	108	75	33
(ii) Policy was not implemented	At least two teams vote against the policy	171	39	132
EXO treatment	By the computer	273	–	–
(i) Policy was implemented	Randomly (50% probability)	123	–	–
(ii) Policy was not implemented	Randomly (50% probability)	150	–	–
Total	–	552	114	165

Note: The numbers in the "# of subjects in proreduction teams" and "# of subjects in antireduction teams" columns are based on the results of voting in the experiment.

suggests that teams work harder with than without the reduction policy in a given institutional condition (ENDO or EXO), and that the positive effect is stronger in the ENDO than in the EXO treatment, for the following reasons. First, the positive effect of the reduction policy holds theoretically for the EXO treatment because the policy reduces the material incentives of shirking. As the reduction policy is imposed randomly in each group, in theory there are no differences in individual characteristics between the groups where the policy is imposed or not. Thus, only the material incentives matter in this treatment due to the lack of selection. Second, the positive effect is also applicable to the ENDO treatment, not only due to the beneficial effects of incentive changes, but also possible selection effects through voting. The reduction policy is enacted in the ENDO treatment only when the majority of teams support the policy. Considering that teams who are *better* at solving the collaborative counting task can be assumed to incur *smaller* effort costs for a given effort level, the beneficial effects of the policy on hard work exceed enhanced effort costs more easily for such higher-skilled teams. This means that higher-skilled teams are more likely to enact the reduction policy by voting, and to perform strongly in the ENDO treatment. In other words, the impact of the reduction policy is detected more strongly in the ENDO treatment due to selection.

It should be worth remarking here that, theoretically, the positive effect of the reduction policy does not emerge when task-solving is too costly for teams. If the return from shirking as a team is much larger than the marginal return from working, members in selfish teams will just stay in the Game screen even when the sorting effects are present in the ENDO treatment.

The main hypothesis of the paper is on the dividend of democracy summarized below:

Hypothesis. *Teams put more effort into task-solving in the ENDO than in the EXO treatment, even after controlling for possible selection effects.*

The phenomenon summarized in this hypothesis is the so-called dividend of democracy. Its mechanism lies in the democratic process that directly influences worker tendency (e.g., Dal Bó et al., 2010, 2019; Kamei, 2016; Sutter et al., 2010; Tyran & Feld, 2006). In a workplace setting, Kamei and Markussen (2023) model this effect such that workplace democracy lowers workers' marginal effort costs. A model similar to Kamei and Markussen (2023) supports the hypothesis above; a decrease in the marginal effort costs driven by democracy results in hard work among teams (see Supporting Information Appendix B for the detail). Part of the dividend of democracy can also be attributed to signaling effects (e.g., Jensen & Markussen, 2022; Kamei, 2019a; Tyran & Feld, 2006).

It should be noted that identifying the dividend of democracy requires care because of the possible selection bias already discussed (Dal Bó et al., 2010, 2019; Tyran & Feld, 2006). By design, proreduction teams are overrepresented (underrepresented) in groups where the reduction policy was (was not) endogenously enacted. As voting behavior is likely related to teams' skills and work behavior, group behaviors are not comparable between the ENDO and EXO treatments unless the distributions of votes are balanced. The present paper adopts the "weights-based identification strategy" proposed by Dal Bó et al. (2019). This estimation method uses weights under the *whole* population when calculating the average behavior in the ENDO treatment, rather than the realized vote shares in specific institutional outcomes. For instance, suppose that 50% of teams vote for the reduction policy and the policy is imposed in 50% of groups. The % of proreduction teams would be much more (less) than 50% in groups where the policy is (is not) endogenously imposed because of majority voting. Instead of the high (low) percentage in such groups, 50% is used as a weight in calculating the average behaviors of pro- and antipolicy units with this method. The detail of the reweighting method along with the data will be provided in Section 3.

3 | POLICY PREFERENCES, AND THE DIVIDEND OF DEMOCRACY

A total of 552 students (279 for the ENDO treatment and 273 for the EXO treatment) at the University of York in the United Kingdom participated in the experiment. No subjects participated in more than one session. The experiment followed standard practices in economics, such as neutral framing. Supporting Information Appendix A includes the procedure and the instructions.

Table 1 of Section 2 includes the distribution of team votes in the experiment. Consistent with the literature on voting experiments among individuals (e.g., Aimone et al., 2013; Dal Bó et al., 2010), it reveals that some teams vote to reduce their private returns from shirking. It indicates that 40.9% of teams ($=38/93 \times 100\%$) voted for the reduction policy. As a result of majority voting, the policy was enacted in 38.7% ($=36/93 \times 100\%$) of groups in the ENDO treatment. Table 1 also shows a clear pattern of selection bias. In the ENDO treatment, the percentage of proreduction teams was 69.4% (=25/36 × 100%) in groups where the policy was enacted, while the percentage was only 22.8% (=13/57 × 100%) in groups where it was not enacted. Hence, proreduction teams were overrepresented (underrepresented) in groups where the reduction policy was (was not) enacted in the ENDO treatment. This is a pattern similar to the selection bias discussed in Dal Bó et al. (2010, 2019).

In fact, teams' support for the policy was positively correlated with their performance before voting. In the practice phase, teams performed the task for only 3 mins under individual-based remuneration. The data indicate that teams which voted for the reduction policy on average answered 1.001 questions correctly in the practice phase; their performance was significantly better at two-sided $p < .01$ ($z = 4.230$) than teams which voted against the policy (the average number of correct answers by antireduction teams was 0.414). This pattern holds regardless of the institutional outcome, that is, whether the policy was enacted or not (Supporting Information Figure C.1). This means that proreduction teams may have characteristics different from antireduction teams. As shown in Supporting Information Figure C.1, the performance of teams in the EXO treatment was somewhere in the middle of the pro- and antireduction teams (was similar to that of antireduction teams) in groups where the policy was enacted (was not enacted).

In sum, selection bias must be controlled for when identifying the dividend of democracy in the data. This paper utilizes the method proposed by Dal Bó et al. (2019) to remove selection effects. Section 3 first discusses the dividend of democracy on work productivity, after which it discusses workers' effort choices in detail and their welfare consequences.

Result 1. 40.9% of teams voted for reducing returns from staying in the Game screen. As a result of majority voting, the reduction policy was enacted in 38.7% of groups in the ENDO treatment.

3.1 | Dividend of democracy on work productivity

The first key result of this study is the positive effect of democracy on work productivity. The dividend of democracy is quite strong: around 20% on average. Consider, first, groups where the reduction policy was enacted. Productivity, defined as the number of correct answers per minute of teamwork (i.e., per average time spent in the task screen by a team member), is 0.594 in the ENDO treatment. The 0.594 means that if a team, that is, all three members, worked the entire 35 mins of the task-solving phase without playing Tetris, they would be able to answer on average 20.79 (=0.594 × 35) tasks correctly. This productivity is 28.5% larger than the productivity in the EXO treatment, which is calculated as 0.462.[15] Part of the productivity increase can be attributed to selection bias as already discussed. Thus, such bias must be controlled for to isolate the dividend of democracy by adjusting the "weights," that is, the distribution of votes. This paper follows Dal Bó et al. (2019) calculating the reweighted productivity with the following two steps:

Step 1. Calculate (a) the average number of correct answers and (b) the per member average work time, using as weights the percentage of proreduction teams in the population (40.9%) rather than the percentages under the reduction regime in the ENDO treatment (69.4%).

Step 2. Calculate (a)/(b).

The reweighted work productivity in the ENDO treatment found using these steps is still quite large—that is, 0.529, 14.5% larger than that in the EXO treatment.

Consider, next, groups where the reduction policy was *not* enacted. There is also a strong effect of democracy for these groups. First, the productivity before reweighting was modestly different between the two conditions: 0.488 in the ENDO and 0.431 in the EXO treatment. However, this mild difference is due to selection, in that proreduction teams are underrepresented in the ENDO treatment, that is, these account for only 22.8% of teams (Table 1). Productivity after reweighting was large, 0.539, in the ENDO treatment. This means that the dividend of democracy is 0.108 (=0.539–0.431) correct answers per min. of teamwork, that is, a 25.1% increase in productivity. The fact that democracy strongly affects behavior irrespective of the policy implementation outcome suggests that being involved in the democratic process by itself, that is, democratic culture, affects their work motivation directly, which is consistent with the idea that democracy directly enhances intrinsic motivations to work (e.g., Deci & Ryan, 1985, 2000).

In sum, the reweighted dividend of democracy without the reduction policy (i.e., 0.539 vs. 0.431) was of almost a similar magnitude to the one in groups with the reduction policy (0.529 vs. 0.462). This underscores the strong role of democracy in improving productivity. For this reason, the two institutional outcomes (with or without the policy) are pooled to statistically test the significance of the dividend of democracy (Table 2).

TABLE 2 Dividend of democracy in work productivity.

	(A) Using original weights	(B) Using adjusted weights following Dal Bó et al.
Team production per minute of its three members' working[a]		
(a) ENDO treatment	0.536	0.535
(b) EXO treatment	0.445	0.445
(c) Dividend of democracy (=(a) − (b))	0.091	0.090
Two-sided p for H_0: (a) = (b)[b]	**0.036****	**0.046****

Notes: The overall productivity measures in rows a and b were calculated using the distribution of policy implementation in the EXO treatment (i.e., % of groups with policy: % of groups without policy = 123/273 : 150/273). The numbers in column (A) are productivity measures calculated using the original distributions of voter types under institutional outcomes (pro- or antireduction teams) shown in rows (i) and (ii) under the ENDO treatment of Table 1. The numbers in column (B) are productivity measures using the distribution of voter types in the population following the weights-based identification strategy proposed by Dal Bó et al. (2019).

[a]The number of correct answers per minute of teamwork

[b]The p values were calculated using the bootstrapping procedure described in Dal Bó et al. (2019). The number of bootstrap iterations was 20,000 (Figure 3).

**$p < .05$.

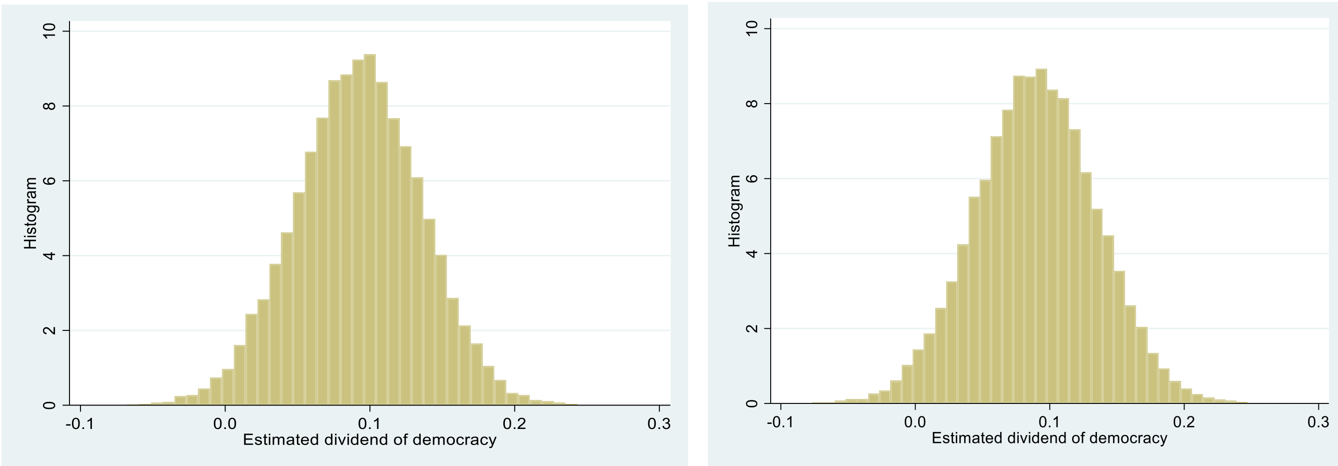
Table 2 reports test results for the dividend of democracy on work productivity using all of the data. To calculate each p value, the estimates for the dividends of democracy were calculated 20,000 times based on session-level bootstrapping.[16] Panel A of Figure 3 reports the distributions of estimated dividends of democracy. These reveal that the size and the significance of the dividend of democracy are only slightly affected by the correction of the selection bias. The overall impact is economically large: democracy boosts productivity by 20.02% (=(0.535 − 0.445)/0.445 × 100%) and it is significant at the 5% level. Hence, it can be concluded that democracy by itself strongly improves productivity.

Readers may also be interested in knowing how the dividend of democracy persists in the workplace. To answer this question, work productivity measures are calculated by splitting the data into quarters of the experiment. It first shows that experience does help improve teams' problem-solving skills, and hence their per-minute-of-teamwork performance. Panel B of Figure 3 indicates that, whether in the ENDO or EXO treatment, work productivity increased from quarter to quarter. The dynamics also reveal that higher work productivity in the ENDO treatment, relative to EXO treatment, was remarkably stable throughout the experiment. This means that fatigue (whether physical or mental) and/or monotony may not weaken the dividend of democracy in the workplace.[17]

Result 2. (a) There is strong evidence that democracy significantly boosted work productivity, defined as the production per minute spent working. (b) The positive dividend of democracy persisted throughout the task-solving phase.

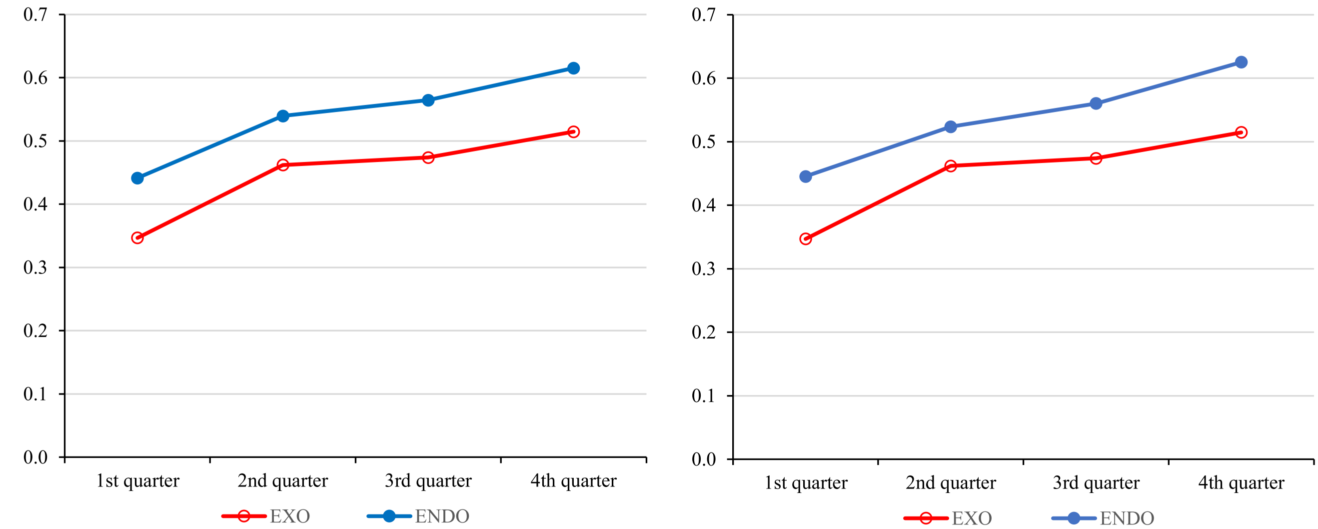
While the strong role of democracy is consistent with the findings from prior research on democracy using "experimental games," such as prisoner's dilemma and public goods games (e.g., Dal Bó et al., 2010; Kamei, 2016; Tyran & Feld, 2006; Sutter et al., 2010), it is at odds with the finding from the "real-effort" experiment of Kamei and Markussen (2023). In Kamei and Markussen (2023), subjects were assigned to a group of three and then worked individually on either the "counting task" (e.g., Abeler et al., 2011; Falk et al., 2006) or the "addition task" (e.g., Corgnet et al., 2015; Niederle & Vesterlund, 2007) on condition that a revenue-sharing rule is in use and a funny video is available as an alternative activity. Kamei and Markussen (2023) found little evidence of the effects of democratic task selection. The null result was indeed a puzzle which Kamei and Markussen (2023) were not able to explain. A similar null result for the dividend of democracy was also observed and posed as a puzzle in the real-effort experiment of Dal Bó et al. (2019) where internet surfing (e.g., Corgnet et al., 2015) was available as an alternative activity. So, why did we get a strong dividend of democracy in the present study? A likely reason is that each team member had stronger shirking opportunities in the present study. Subjects in the present experiment *jointly* solved a *collaborative* counting task as a team in a group, unlike in the prior experiments where subjects *individually* solved an *individual* real-effort task in a group. Specifically, while incentives to shirk as a decision-making unit (teams in this study or individuals in the other research) in a group are the same, each team member in the present study has additional opportunities to shirk by playing Tetris privately, that is, without notifying their other team members, whose structure features a coordination game inside the team.[18] The difference between the present and the earlier experiments suggests that the dividend of democracy may be more important in an environment where workers have stronger incentives to shirk.

(a)



(i) When using original weights (ii) When using adjusted weights according to Dal Bó *et al.* (2019)

(b)



(i) When using original weights (ii) When using adjusted weights according to Dal Bó *et al.* (2019)

FIGURE 3 Dividends of democracy for work productivity. (a) Distribution of bootstrapped dividends of democracy for productivity based on Dal Bó et al. (2019). (b) Dividend of democracy, quarter by quarter. (1) Each distribution in panel A was drawn using 20,000 estimated dividends of democracy based on bootstrap iterations. (2) The productivity measures of each quarter in panel B were calculated by splitting the duration of the 35 task-solving phase by four (e.g., the first quarter is the first 35/4 mins).

3.2 | Effort choices and welfare

The larger size of work productivity (Result 2) does not mean that democracy improves production in the workplace. Rows (I) and (II) of Table 3 report the average numbers of attempts and correct answers in the main task-solving phase. The average results are reported by the policy implementation outcome because work behaviors differed substantially by the presence of the reduction policy. It shows that teams attempted more questions and, as a result, answered more questions correctly, in the ENDO than in the EXO treatment (rows I and II). However, the positive effects of democracy are far from significant (columns 2, 2a, and 2b).

This insignificant impact, despite Result 2, was due to the workers' effort choices. As the collaborative counting task was a relatively challenging real-effort task, shirking prevailed in the experiment.[19] Workers (although insignificantly)

TABLE 3 Work performance and the dividend of democracy.

	Unweighted			**Reweighted**		
	All data (1)	**With policy (1a)**	**Without policy (1b)**	**All data (2)**	**With policy (2a)**	**Without policy (2b)**
(I) *Average number of attempts*						
(a) ENDO	19.49	25.28	14.74	18.81	20.53	17.40
(b) EXO	16.79	19.49	14.58	16.79	19.49	14.58
H_0: (a) = (b)	0.151	**0.043****	0.949	0.331	0.747	0.285
(II) *Average number of correct answers*						
(a) ENDO	12.49	16.61	9.12	11.96	13.14	11.00
(b) EXO	10.49	12.12	9.16	10.49	12.12	9.16
H_0: (a) = (b)	0.170	**0.060***	0.983	0.330	0.671	0.336
(III) *Average per member time spent in the Game screen* (minute)						
(a) ENDO	12.14	7.05	16.31	12.60	10.14	14.61
(b) EXO	11.50	8.79	13.72	11.50	8.79	13.72
H_0: (a) = (b)	0.664	0.345	0.236	0.534	0.594	0.711
(IV) *Average payoff in the main task-solving phase (pound sterling)*						
(a) ENDO	9.62	11.09	8.41	9.35	9.51	9.23
(b) EXO	8.29	8.68	7.97	8.29	8.68	7.97
H_0: (a) = (b)	**0.065***	**0.062***	0.555	0.138	0.498	0.150

Notes: The p values were calculated using the bootstrapping procedure described in Dal Bó et al. (2019). The number of bootstrap iterations was 20,000. The numbers in columns (1), (1a), and (1b) were calculated using the original distributions of voter types under institutional outcomes (pro- or antireduction teams) shown in rows (i) and (ii) of Table 1. The numbers in columns (2), (2a), and (2b) were calculated using the distribution of voter types in the population following the weights-based identification strategy developed by Dal Bó et al. (2019). The overall measures in columns (1) and (2) were calculated using the distribution of policy implementation in the EXO treatment (i.e., % of groups with policy: % of groups without policy = 123/273 : 150/273).

*$p < .1$; **$p < .05$.

shirked *more* on average in the ENDO treatment than in the EXO treatment—see columns (2), (2a), and (2b) of row (III). The higher incidence of shirking undermined the positive impact of enhanced work productivity, which resulted in the insignificant effect on the two effort output measures. Thus, this result suggests that a firm needs to have some mechanism to curb workers' effort choices beyond democracy if they want to increase production significantly, because workers' discretion to decide how much to work may partly cancel out the sustained positive dividend of democracy. However, it should be emphasized here that despite the increased shirking, team production did not decrease (instead increased although insignificantly) thanks to Result 2 under democracy. This means that the same level of production can be achieved under democracy in less work time.

One may wonder why democracy worsened shirking. One possible interpretation here is that democracy enlarged workers' motivations to earn a high payoff in the experiment. The subjects may have perceived it to be more payoff-enhancing if they worked harder for a shorter duration and then secured certain gains from staying on the Game screen once exhausted. Although it cannot be verified, this possibility may partly explain the behavior since, despite Result 2(b), subjects may quickly have feelings of fatigue if their per-minute effort levels rise. Having said that, such a reduction in work time did not work well for the workers, since, while democracy did increase the average payoff, the impact is insignificant after controlling for selection (row IV). This implies that their effort choices were not optimal. But, if this conjecture is relevant, why did perceived fatigue play a large part in the behavioral decisions of experiment subjects? A likely possibility is that Result 2 was still not enough to encourage workers to choose putting in a greater effort over shirking. This possibility is quite reasonable as discussed carefully in Section 3.3.

Result 3. Despite Result 2, democracy did not increase team production significantly, because workers under democracy decreased work time to some degree.

3.3 | Privately versus socially optimal behaviors

This experiment was designed to model a social dilemma problem, that is, conflicts among teams, in the workplace. Section 3.3 briefly checks the validity of this design setup, finding that its attempt was successful as intended. This section also tries to find an answer as to why democracy was not enough to boost team production in the experiment.

Since staying in the Game screen was remunerated with 16 or 18 pence/min, it is possible to calculate for what percentage of teams task-solving was a socially or privately optimal strategy (in the sense of material payoff maximization). In order for task-solving to be privately optimal, a team needs to be able to solve at least $0.80 = 16/20$ ($0.90 = 18/20$) tasks correctly per minute when the reduction policy is (is not) in place. A detailed look at the data (Supporting Information Appendix Table C1) indicates that gaming was a privately optimal choice for almost all teams in the EXO treatment, whether the policy was in place or not. Specifically, it is so for 95.60% of teams (87 out of 91 teams) in the EXO treatment.[20] This implies that the reduction policy was nondeterrent in the experiment. Consistent with the prior experimental evidence on exogenously introduced nondeterrent punishment (e.g., Kamei, 2016; Tyran & Feld, 2006), the effect of the reduction policy was not large in the EXO treatment. Specifically, while the average number of correct answers in the EXO treatment was larger with than without the reduction policy (12.12 vs. 9.16), the difference was not significant at two-sided $p = .109$ according to the bootstrap method used in the other tests of the paper (the difference is significant but only at the 10% level, that is, $p = .0707$ if a two-sided Mann–Whitney test is used).[21]

However, as intended, the socially optimal strategy was task-solving for many teams. In order for task-solving to be socially optimal, a team needs to be able to solve at least $0.227 \approx 16/60$ ($0.30 = 18/60$) tasks correctly per minute when the reduction policy is (is not) in place. Overall, the social optimal condition was met for 61.6% of teams (56 out of 91 teams) in the EXO treatment. Notice that task-solving is never privately optimal for teams whose task-solving is not socially optimal. Consistent with this incentive pattern, teams whose task-solving was not socially optimal spent significantly less time working on the task than the other teams at two-sided $p < .001$ (15.29 vs. 28.63 mins in the EXO treatment). The average number of correct answers per minute of working by the former was only 0.07, but that by the latter was 0.54 in the EXO treatment.

In sum, the present experiment can be thought of as exploring workers' voting and effort choice decisions under social dilemmas in the workplace when the target was a nondeterrent reduction policy.

Then, one may ask whether democracy might have altered the social dilemma situation to another one (e.g., coordination game), as arguably democracy not only enhances work productivity (Section 3.1), but also reduces effort costs in task-solving. Another look at the data, however, shows that the answer is negative. Specifically, a calculation finds that gaming was a privately optimal choice for almost all teams in the ENDO treatment, that is, 91.40% of teams (85 out of 93 teams); and task-solving was a socially optimal choice for 61.3% (57 out of 93 teams) in that treatment—see again Supporting Information Appendix Table C1. These numbers are quite similar to those in the EXO treatment already discussed.

The reason why worker behavior was characterized by Results 2 and 3 is explained by the theoretical analysis summarized in Supporting Information Appendix B. The model there assumes that, following the prior research findings, democracy eases a worker's effort cost, and it also boosts their productivity (its positive effect on work productivity is a parameter μ in that model). $\mu > 0$ was confirmed by the experiment data as summarized in Result 2. The team's optimal effort provision can then be determined by the relative strength between (a) work productivity ($s + \mu$ in the theoretical model, where s is the marginal return of effort provision by team i) and (b) the material incentives to shirk by staying in the Game screen. Theoretically, the positive value of μ (Result 2) possibly changes the materially beneficial choice from gaming to task-solving—see Supporting Information Appendix Figure B.2. However, the analysis in the Supporting Information Appendix indicates that if the impact on work productivity is not *economically* large enough, gaming is still the most materially beneficial activity even when teams have a statistically significant dividend of democracy. This is exactly what the above calculations on privately versus socially optimal choices in the experiment data demonstrate. The calculations clearly reveal that democracy did not change the underlying private incentives in the experiment. This means that additional mechanisms on top of democracy would be required to change the incentive structure so that task-solving becomes a privately optimal choice for workers, if the group wants to increase production.

4 | UNDERSTANDING SACRIFICE BEHAVIOR: COMMUNICATION CONTENTS

The present experiment provides a useful opportunity to explore workers' reasoning behind their decisions to reduce private returns from shirking as communication contents are available. While the decision data not only uncovered some subjects' preferences to reduce their private returns but also detected a significant dividend of democracy on work productivity (Section 3), it is still unclear what drove such behavioral patterns.

Two independent coders were hired to read and classify the communication contents based on their judgment of the subjects' motives. Specifically, a list of codes was designed by the authors, based on the theoretical predictions of the setup and related literature, that could potentially reflect a subject or teams reasoning and/or behavior. The list was given to the coders to assign whichever codes (including none) they deemed relevant to a given communication log. The coding procedure follows Kamei and Tabero (2022) which utilized the standard coding approach in economics to analyze teams' behavioral reasoning in the context of institutional choices based on intrateam communication logs. The detail of the coding procedure and the full lists of codes used for the present paper can be found in Supporting Information Appendix Sections D.1 and D.2.

The agreement rates and Cohen's Kappa values (Cohen, 1960) can be used to judge the consistency of the coding process between the two coders. Overall, the agreement rates (Kappa values) between the two coders were 96.9% (0.87) and 94.8% (0.78) in the ENDO and EXO treatments, respectively. The Kappa values are at least 0.4 for 92.5% and 78.0% of individual codes in the ENDO and EXO treatments, respectively (Supporting Information Appendix Section D.3). As a Kappa value of 0.4 is usually used as a threshold for a researcher to judge the reliability of coding, we use only the codes that exceed this boundary in this analysis.

Table 4 summarizes the list of codes that are found to have impacted the units' voting significantly at least at the 10% level. Their voting is clearly linked to their intention regarding what to do during the main task-solving phase (Code Bs): while units supported the reduction policy if they planned to focus on task-solving, they opposed it if they were considering using the game screen. The coding category linked to pro/antipolicy reasoning (Code Cs) reveals clear motives behind the policy preferences. While the policy is nondeterrent, those who voted in favor of it did so to deter others from shirking (Code C1). On the other hand, those who intended to game or believed that the policy was too weak to alter shirking opposed its enaction. Lastly, unsurprisingly, their views on materially beneficial behavior and team atmosphere influenced voting. Specifically, units that believed their privately optimal behavior was task-solving supported reducing the return from gaming. By contrast, units who experienced discomfort or poor performance from task-solving in the practice phase opposed such a reduction. While teams with a positive atmosphere (E2) supported the reduction policy, those with poor or lacking communication opposed it (E5).

TABLE 4 Significant code meanings and its impact on voting for the reduction policy.

Code	Meaning	Direction
B1	Agree/imply to count as the primary behavior	(+)***
B2	Agree/imply to game as the primary behavior	(−)*
B3	Agree to hybrid behavior, for example, so many tasks/minutes before switching to the game screen	(−)**
B4	Agree to discuss, decide, and/or alter behavior during the counting task later (35-min phase) based on performance/needs in Phase 2	(−)**
C1	Propolicy to deter others from switching to the game screen by reducing the return (monetary deterrence)	(+)***
C8	Antipolicy as they intend to game for at least some of the task-solving period	(−)***
C11	Express that the policy is not strong enough to deter others from switching to the game screen (monetary)	(−)**
D2	Believe they as a team make the most money from counting	(+)***
D5	Discuss their performance or comfort in Phase 1 (weak/negative)	(−)**
E2	Positivity towards teammates, for example, attempts to encourage others or being supportive	(+)*
E5	No communication from just one or two team members	(−)***

Notes: + (−) In "Direction" means the reasoning en(dis)courages voting for the policy.

*$p < .1$; **$p < .05$; ***$p < .01$.

Result 4. The units that planned to exclusively work on task-solving, believed that the reduction policy would deter others from shirking, or those that had supportive team atmospheres, voted for the reduction policy. However, those who previously experienced discomfort or poor performance from working, considered (even only potentially) using the Game screen, believed that the policy was too weak to alter peers' shirking, or had poor communication with their teammates, voted against the reduction policy.

As summarized in Result 4, units' commitment to task-solving and their intention to affect others' shirking were the drivers behind their votes in favor of the reduction policy. To explore how policy implementation outcomes affected units' behaviors, coding analyses were further performed using the communication logs of the 35-min task-solving phases (Table 5). Three similar tendencies were observed for both the ENDO and EXO treatments. First, those who

TABLE 5 Reasoning behind work choice and productivity.

Code	Meaning	Direction
Codes related to reactions to vote outcome in ENDO (Code Fs) or policy outcome in EXO (Code Is)		
(ENDO treatment)		
F1	Express negative emotions (e.g., upset and anger) about the outcome of the vote	(wt−)***
F3	Agree/imply to count as the primary behavior	(wt+)***, (p+)***
F4	Agree/imply to game as the primary behavior	(wt−)***, (p−)***
F5	Agree to hybrid behavior, for example, so many tasks/minutes before switching to the game screen	(wt−)**, (p−)*
F6	Agree to discuss, decide, and/or alter behavior during the counting task later (35-min phase) based on performance/needs in Phase 2	(p−)*
F7	Express belief/hope that other teams will complete tasks following the vote	(wt+)***, (p+)**
F8	Express belief that teams will not complete tasks following the vote	(wt+)***, (p+)***
F9	Discuss the distribution of votes and predict how each team may respond to one another	(wt+)***
F13	Belief on other teams' responses: antipolicy teams will work little	(p−)***
F15	Discuss whether to change behavior based on the vote outcome	(wt+)***
(EXO treatment)		
I1	Express negative emotions (e.g., upset and anger) about the policy outcome	(wt−)***, (p−)***
I4	Agree/imply to game as the primary behavior	(wt−)***, (p−)***
I5	Agree to hybrid behavior, for example, so many tasks/minutes before switching to the game screen	(wt−)***, (p−)**
I6	Agree to discuss, decide, and/or alter behavior during the counting task later (35-min phase) based on performance/needs in Phase 2	(wt−)*
Other codes (the same codes were used for the ENDO and EXO treatments)		
G4	Expression of strong negative emotion, for example, frustration, anger, disappointment	(p−***, Exo)
G5	Expression of strong positive emotion, for example, enjoyment, things are going well	(wt+***, p+**, Exo)
D4	Discuss their performance or comfort in Phase 1 and/or so far in Phase 2 (strong/positive)	(wt+***, p+***, Endo)
D5	Discuss their performance or comfort in Phase 1 and/or so far in Phase 2 (weak/negative)	(wt−***, p−***, Endo), (wt−**, Exo)
D8	Discuss uncertainty surrounding other teams' work choices or abilities	(wt−***, p−**, Exo)
D9	Suggest distrust of other teams, for example, expect them to take advantage	(wt−***, p−***, Endo)

Notes: wt and p in the "Direction" column indicate two work performance measures: work time (minutes) and productivity defined as the number of correct answers divided by the work time. + (−) means the reasoning in(de)creases the performance measures. All significant codes are listed for Codes Fs and Is, while only some key codes are included for the other coding categories to conserve space (Supporting Information Appendix D4 includes the full estimation results).

*$p < .1$; **$p < .05$; ***$p < .01$.

reacted negatively to the implementation outcome tended to work less (F1, I1). Such negative reciprocal tendencies were unsurprising considering the large findings of other-regarding preferences—see, for example, Sobel (2005) and Fehr and Schmidt (2006). Second, a unit's plan to work on counting or engage in gaming affects performance (F3, F4, F5, F6, I4, I5, I6), similar to Result 4. Third, units' positive and negative experiences of task-solving, respectively, improve and hurt performance (G4, G5, D4, D5).

The results reveal *signaling* effects of voting on task-solving, and some nuanced evidence about the teams' dividends of democracy seen in Result 2. First, units that considered the distribution of votes to predict others' task-solving or discussed changing behavior worked longer (F9, F15). Second, units who believed that other teams would complete tasks following the vote performed strongly (F7), resonating with the idea that voting has a signaling value, thereby encouraging collaboration. Further, even units who thought that others would not respond to the reduction policy improved their performance (F8), which implies democracy directly affects behavior beyond signaling. Nevertheless, its effects are canceled out if an antipolicy team is present in a group and units have a negative view of the task-solving behavior of the antipolicy team (F13).

5 | CONCLUSION

Teams are popular decision-making and work units in organizations that feature a complex coordination problem. Overcoming moral hazard among teams in the workplace plays a crucial role in maintaining productivity in the firm, whether in the traditional work environment or in a remote working setting, such as that triggered for many by the Covid-19 crisis. The present paper investigated how frequently groups reduce the return from shirking by enacting a formal nondeterrent sanction policy, and how such endogenous choices per se improve work productivity. To achieve this, a novel real-effort experiment was designed, equipped with (a) a collaborative counting task featuring an intrateam coordination game and (b) gaming (Tetris) as a real leisure activity. The experiment results showed that around 40% of teams voted to reduce the return from staying on the Game screen. A contents analysis using teams' communication logs showed that such voting was driven by not only their commitment to work on counting but also their belief that the reduction policy would deter others from shirking.

The decision data uncovered a significant and strong dividend of democracy on work productivity. Strikingly, whether the policy was enacted or not, teams in the ENDO treatment displayed significantly higher per-work-time production than those in the EXO treatment. This means that democratic culture directly affects behavior positively. However, the workers under democracy also experienced *higher* levels of shirking, that is, the time spent on the Game screen was larger in the ENDO than in the EXO treatment, presumably driven by their enhanced fatigue due to the more intensive working in the former. This implies that while additional mechanisms that affect incentives besides democracy may be required to increase production, democracy may improve efficiency. What kinds of mechanisms would work best to instead increase production further remains for future research. Having said that, it should be emphasized here that the average production of the workers under democracy did not decrease (it increased, although insignificantly) thanks to their strong per-work-time production.

The findings on the positive dividend of democracy on work productivity have a policy implication for effective human resource and management practices. While prior research suggests that innovative human resource management involving worker participation (such as that in production sites) leads to better work performance (e.g., Ichniowski et al., 1997), it is unclear how democracy affects behavior, as earlier real-effort experiments failed to detect strong dividends of democracy (e.g., Dal Bó et al., 2019; Kamei & Markussen, 2023). Using an environment with strong shirking incentives, the present experiment suggests that organizations with a shared goal can benefit from introducing participatory decision-making with their employees or group members, by potentially improving their work productivity. Even when democracy induces the workers to work less, the improvement in productivity allows for achieving a production goal with fewer working hours.

The effect of democratic culture in achieving the same goal in less work time collaborates with recent work style reform. There is a trend to transform the traditional workplace into an employee-centered workplace in many countries. For example, in the United Kingdom, some firms recently tested 4-day work weeks to make working conditions flexible to meet the different needs of employees.[22] Having higher work productivity in a democratic environment certainly helps firms achieve the same or potentially better outcomes with fewer working hours. This boost to productivity is achieved through enhanced self-determination and signaling effects; workplace democracy provides the workers with the ability to foster trust with each other and to indicate their intentions or desire to cooperate through democratic procedures such as voting, and the recipients can then respond to these signals. Such a social exchange may be fundamental for workers to achieve collaboration by reducing uncertainty surrounding each other's behavior in a democratic workplace environment. The firm may create a positive and collaborative atmosphere and improve productivity by designing democratic systems in multiple layers and activities across the organization (Smith & Bliege Bird, 2000; Smith & Bliege Bird, 2005).

ACKNOWLEDGMENTS

This project was supported by a grant-in-aid from the Murata Science Foundation. Durham University and the Northern Ireland and North East Doctoral Training Partnership (ES/P000762/1) provided additional funding support. The authors thank John Hey and Mark Wilson (an IT manager at the University of York) for their support in recruiting subjects, and Louis Putterman and Pedro Dal Bó for helpful comments.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon request.

ORCID

Kenju Kamei http://orcid.org/0000-0002-7905-8584

ENDNOTES

[1] The difficulty in sustaining cooperation has also been widely discussed in the theoretical literature on the voluntary provision of public goods (e.g., Bergstrom et al., 1986; Samuelson, 1954).

[2] In general, actors' many decisions are characterized as costly signaling in modern societies. Examples include the job market, in which applicants invest in education or other qualifications to indicate their quality (Spence, 1973), or at the firm level by which firms indicate their quality to other firms, the market, or other stakeholders through investment in high profile board members, awards, alliances, or underpricing (see Bergh et al., 2014 for a review and examples).

[3] Empirically, people are known to choose transaction partners in dilemma situations based on factors that inform the quality of that partner. Elfenbein et al. (2012), using a novel data set composed of more than 160,000 eBay listings, successfully demonstrated that in online marketplaces, buyers tend to purchase products tied to charity, and thus sellers have incentives to use a charity program (e.g., eBay's Giving Works program) as a quality signal.

[4] Bisetti et al. (2022) propose a self-reporting mechanism in which a team's pay is based on their observed joint output and their team's self-reported performance. They prove that a team has the incentive to under-report their group's performance (sacrifice wages for all in the team) as a punishment to free-riders, thereby enabling them to achieve higher welfare.

[5] Strengthening monitoring increases the probability that cyberloafing is detected and penalties are assigned, thereby reducing workers' incentives to cyberloaf. As will be described soon, for the sake of simplicity, the present paper considers a policy to reduce material returns from shirking deterministically in the workplace in the experiment.

[6] Knez and Simester (2001) argued that work groups may voluntarily strengthen mutual monitoring within their groups to obtain a bonus through achieving a firm-level goal.

[7] Moonlighting is increasingly common in some countries because it is encouraged by the government. For example, lifetime employment was a common practice in Japan traditionally. However, the Japanese Ministry of Health, Labour and Welfare published the "Guidelines for Promotion of Side Work" and deleted the description of prohibition of subsidiary business from "The Model Rules of Employment" in 2018.

[8] This very rarely happened in the experiment. All three members submitted the same answers in 96.9% of teams' submissions in the experiment (3176 out of 3278 completed tasks in the 62 experiment sessions). The authors read through all the communication dialogues and their submitted answers, and found that almost all disagreements are errors or typos. The mean number and the mode of disagreements across all teams that disagreed were, respectively, 1.72 and 1. The size of the error rate is unsurprising because the average number of attempts for these teams was 24.14 questions, above the average of 17.81 for the experiment, which might increase potential errors in typing.

[9] The authors read through the communication dialogues and found no team to have broken the anonymity rule.

[10] The practice phase and the main task-solving phase are called "phase 1" and "phase 2" in the experiment instructions.

[11] To avoid cognitive overload, subjects are provided with instructions for the practice phase only at the beginning of the experiment. Instructions for the main task-solving phase are distributed once the practice ends. Such gradual learning approach is often taken in experiments (e.g., Ertan et al., 2009; Kamei & Tabero, 2022; Kamei et al., 2015).

[12] This return can be thought of as material returns that can be obtained from shirking in the real workplace. Shirkers may build their social network using social media or by exchanging e-mails during work time, develop skills to benefit future job prospects, complete personal tasks, or even moonlight privately as in the real-world example described in the introduction of the paper. Such activities may not only provide intrinsic satisfaction but may also provide material benefits. A similar designing approach was chosen by Kamei and Markussen (2023) where an activity alternative to solving a real-effort task is to watch a funny video. Subjects in Kamei and Markussen (2023) received a small return per minute watching the videos.

[13] As discussed later, the present paper uses the "weights-based identification strategy" proposed by Dal Bó et al. (2019) to identify the dividend of democracy. The requirement to use this method is that teams' types are independently drawn and their behavior only depends on the team's own type. It is therefore essential to avoid dynamic interactions across teams including information feedback.

[14] While another realistic voting method is a unanimity rule (consensus), this study adopted majority voting because the interpretation of data becomes complex when the unanimity rule is in use as it possibly involves strategic voting among voters (e.g., Battaglini et al., 2010; Kamei, 2019a).

[15] The average number of correct answers and average per member working/shirking time by institutional condition can be found in Table 3.

[16] Each estimate was calculated using 62 sessions randomly drawn from the set of the original 62 sessions.

[17] An analysis in Section 3.2 suggests that workers in the ENDO treatment did not accumulate fatigue with a higher work pace, as they instead increased the time spent in the Game screen.

[18] A team cannot complete a task while some member is shirking. Such shirking is also interpreted as maliciousness or lack of team spirit towards members who are motivated and are waiting for the shirker's input to find the answer.

[19] The high difficulty in finding answers to the real-effort task is a crucial feature of the experiment, which was intentionally designed. Notice that if the tasks were easy, worker subjects would work hard with small output elasticity of incentive changes in this kind of real-effort experiment (Corgnet et al., 2015; Erkal et al., 2018). A challenging real-effort task and an availability of alternative activities (Tetris) were thus carefully incorporated in the design to make the output elasticity of incentives sufficiently large.

[20] Material incentives did matter for workers' effort choices. In the EXO treatment, the four teams for which task-solving was privately optimal worked on counting on average 31.80 mins, which is significantly larger at two-sided $p = .0015$ than the average work time by the other 87 teams where gaming was privately optimal (which was 23.12 mins)—see Supporting Information Table C1.

[21] The effect of the reduction policy was apparently strong in the Endo treatment (see Table 3 for the numbers). The average number of correct answers in the Endo treatment was significantly larger with than without the reduction policy at two-sided $p = .001$*** (.0020***) according to the bootstrap method (a Mann–Whitney test). However, this strong effect is just due to selection. The difference was not significant at two-sided $p = .388$ when using the bootstrap method with the distribution of votes in the population being the weights following Dal Bó et al. (2019). Recall that democracy enhanced work productivity in the experiment similarly regardless of whether the policy was imposed or not (Result 2), whose aspect makes the effect of the policy in itself small.

[22] For example, see the following BBC news: https://www.bbc.co.uk/news/business-64669987 (last accessed on April 21, 2023).

REFERENCES

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

Aimone, J. A., Iannaccone, L. R., Makowsky, M. D., & Rubin, J. (2013). Endogenous group formation via unproductive costs. *Review of Economic Studies*, *80*(4), 1215–1236.

Alchian, A., & Demsetz, H. (1972). Production, information costs, and economic organization. *American Economic Review*, *62*(5), 777–795.

Araujo, F. A., Carbone, E., Conell-Price, L., Dunietz, M. W., Jaroszewicz, A., Landsman, R., Lamé, D., Vesterlund, L., Wang, S. W., & Wilson, A. J. (2016). The slider task: An example of restricted inference on incentive effects. *Journal of the Economic Science Association*, *2*, 1–12.

Battaglini, M., Morton, R. B., & Palfrey, T. R. (2010). The swing voter's curse in the laboratory. *Review of Economic Studies*, *77*(1), 61–89.

Bednar, J., Chen, Y., Liu, T. X., & Page, S. (2012). Behavioral spillovers and cognitive load in multiple games: An experimental study. *Games and Economic Behavior*, *74*, 12–31.

Bergh, D. D., Connelly, B. L., Ketchen, Jr., D. J., & Shannon, L. M. (2014). Signalling theory and equilibrium in strategic management research: An assessment and a research agenda. *Journal of Management Studies*, *51*(8), 1334–1360.

Bergstrom, T., Blume, L., & Varian, H. (1986). On the private provision of public goods. *Journal of Public Economics*, *29*, 25–49.

Bisetti, E., Tengelsen, B., & Zetlin-Jones, A. (2022). Moral hazard in remote teams. *International Economic Review*, *63*(4), 1595–1623.

Bolton, P., & Dewatripont, M. (2004). *Contract theory*. MIT Press.

Brekke, K. A., Hauge, K. E., Lind, J. T., & Nyborg, K. (2011). Playing with the good guys. A public good game with endogenous group formation. *Journal of Public Economics*, *95*(9–10), 1111–1118.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, *14*(1), 47–83.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, *37*(1), 39–67.

Corgnet, B., Hernán-González, R., & Schniter, E. (2015). Why real leisure really matters: Incentive effects on real effort in the laboratory. *Experimental Economics*, *18*, 284–301.

Dal Bó, P. (2010). Experimental evidence on the workings of democratic institutions. In S. Galiani, & I. Sened (Eds.), *Institutions, property rights, and economic growth: The legacy of Douglass North*. Cambridge University Press.

Dal Bó, P., Foster, A., & Putterman, L. (2010). Institutions and behavior: Experimental evidence on the effects of democracy. *American Economic Review*, *100*(5), 2205–2229.

Dal Bó, P., Foster, A., & Kamei, K. (2019). *The democracy effect: A weights-based identification strategy* [NEBR Working Paper 2 5724].

Deci, E., & Ryan, R. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.

Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268.

Doellgast, V., & Marsden, D. (2019). Institutions as constraints and resources: Explaining cross-national divergence in performance management. *Human Resource Management Journal*, *29*, 199–216.

Elfenbein, D. W., Fisman, R., & Mcmanus, B. (2012). Charity as a substitute for reputation: Evidence from an online marketplace. *Review of Economic Studies*, *79*(4), 1441–1468.

Erkal, N., Gangadharan, L., & Koh, B. H. (2018). Monetary and non-monetary incentives in real-effort tournaments. *European Economic Review*, *101*, 528–545.

Ertan, A., Page, T., & Putterman, L. (2009). Who to punish? Individual decisions and majority rule in mitigating the free rider problem. *European Economic Review*, *53*, 495–511.

Falk, A., Huffman, D., & Mierendorff, K. (2006). *Incentive properties and political acceptability of workfare: Evidence from real effort experiments* [Working Paper].

Fehr, E., & Schmidt, K. (2006). The economics of fairness, reciprocity and altruism—experimental evidence and new theories. In S.-G. Kolm, & J. M. Ythier (Eds.), *Handbook of the Economics of Giving, Altruism and Reciprocity* (pp. 615–691). North Holland Publishing Co.

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, *100*(1), 541–556.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*(3), 397–404.

Gibbons, R., & Henderson, R. (2013). What Do Managers Do? (Chapter 17) Exploring Persistent Performance Differences among Seemingly Similar Enterprises. In R. Gibbons, & J. Roberts (Eds.), *The Handbook of Organizational Economics*. Princeton University Press.

Gintis, H., Smith, E. A., & Bowles, S. (2001). Costly signaling and cooperation. *Journal of Theoretical Biology*, *213*(1), 103–119.

Grimm, V., & Mengel, F. (2009). Cooperation in viscous populations—Experimental evidence. *Games and Economic Behavior*, *66*(1), 202–220.

Hagen, E. H., & Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Human Nature*, *14*(1), 21–51.

Hawkes, K., & Bliege Bird, R. (2002). Showing off, handicap signaling, and the evolution of men's work. *Evolutionary Anthropology: Issues, News, and Reviews*, *11*(2), 58–67.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, *13*(2), 324–340.

Hume, D. (2000 [1739]). A Treatise of Human Nature. Oxford University Press.

Iannaccone, L. R. (1992). Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives. *Journal of Political Economy*, *100*(2), 271–291.

Ichniowski, C., Shaw, K., & Prennushi, G. (1997). The effects of human resource management practices on productivity: A study of steel finishing lines. *American Economic Review*, *87*, 291–313.

Jensen, T., & Markussen, T. (2022). Group size, signaling and the effect of democracy. *Journal of Economic Behavior & Organization*, *187*, 258–273.

Kamei, K. (2016). Democracy and resilient pro-social behavioral change: An experimental study. *Social Choice and Welfare*, *47*(2), 359–378.

Kamei, K. (2019a). Cooperation and endogenous repetition in an infinitely repeated social dilemma. *International Journal of Game Theory*, *48*(3), 797–834.

Kamei, K. (2019b). The power of joint decision-making in a finitely-repeated dilemma. *Oxford Economic Papers*, *71*(3), 600–622.

Kamei, K., & Markussen, T. (2023). Free riding and workplace democracy—heterogeneous task preferences and sorting. *Management Science*, *69*, 3884–3904.

Kamei, K., Putterman, L., & Tyran, J. R. (2015). State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods. *Experimental Economics*, *18*, 38–65.

Kamei, K., & Tabero, K. (2022). *The individual-team discontinuity effect on institutional choices: Experimental evidence in voluntary public goods provision* [Keio-IES Discussion Paper Series 2022-015]. Institute for Economics Studies, Keio University.

Knez, M., & Simester, D. (2001). Firm-wide incentives and mutual monitoring at continental airlines. *Journal of Labor Economics*, *19*, 743–772.

Ledyard, J. (1995). Public goods: A survey of experimental research. In J. Kagel, & A. Roth (Eds.), *Handbook of experimental economics* (pp. 111–194). Princeton University Press.

Luhan, W. J., Kocher, M. G., & Sutter, M. (2009). Group polarization in the team dictator game reconsidered. *Experimental Economics*, *12*, 26–41.

Marschak, J., & Radner, R. (1972). *Economic Theory of Teams*. Yale University Press.

McNamara, T., Meloso, D., & Puncheva-Michelotti, P. (2022). 'You are free to choose … are you?' Organisational punishment as a productivity incentive in the social science literature. *Human Relations*, *75*, 322–348.

Melizzo, P., Carpenter, J., & Matthews, P. H. (2014). Workplace democracy in the lab. Industrial Relations. *Journal*, *45*(4), 313–328.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, *122*(3), 1067–1101.

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

Rousseau, J.-J. (2009 [1755]). Discourse on the Origin of Inequality. In P. Coleman (Ed.), F. Philip (Trans.). Oxford University Press.

Samuelson, P. A. (1954). The pure theory of public expenditure. *The Review of Economics and Statistics*, *36*(4), 387–389.

Smith, E. A., & Bliege Bird, R. (2000). Turtle hunting and tombstone opening. *Evolution and Human Behavior*, *21*(4), 245–261.

Smith, E., & Bliege Bird, R. (2005). Costly signaling and prosocial behavior. In H. Gintis, S. Bowles, R. Boyd, & E. Fehr (Eds.), *Moral sentiments and material interests: The foundations of cooperation in economic life* (pp. 115–148). MIT Press.

Sobel, J. (2005). Interdependent preferences and reciprocity. *Journal of Economic Literature, XLIII*, 392–436.

Sosis, R., & Alcorta, C. (2003). Signaling, solidarity, and the sacred: The evolution of religious behavior. *Evolutionary Anthropology: Issues, News, and Reviews*, *12*(6), 264–274.

Sosis, R., & Bressler, E. R. (2003). Cooperation and commune longevity: A test of the costly signaling theory of religion. *Cross-Cultural Research*, *37*(2), 211–239.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355–374.

Sutter, M., Haigner, S., & Kocher, M. G. (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies*, *77*, 1540–1566.

Thöni, C., & Volk, S. (2018). Conditional cooperation: Review and refinement. *Economics Letters*, *171*, 37–40.

Tyran, J. R., & Feld, L. P. (2006). Achieving compliance when legal sanctions are non-deterrent. *The Scandinavian Journal of Economics*, *108*(1), 135–156.

Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics*, *6*, 299–310.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kamei, K., & Tabero, K. (2023). Free riding, democracy, and sacrifice in the workplace: Evidence from a real-effort experiment. *Journal of Economics & Management Strategy*, 1–21. https://doi.org/10.1111/jems.12570