Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

職場におけるただ乗り問題、ワークプレイス・デモクラシー、そして 労働者の私的行動の抑制に関する考察（リアル・エフォート実験からの事実）

亀井 憲樹、ケイティ・タベロ

2023 年 4 月 27 日
DP2023-011
https://ies.keio.ac.jp/publications/22729/

Keio University

CALAMVS GLADIO FORTIOR
1858

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
27 April, 2023

職場におけるただ乗り問題、ワークプレイス・デモクラシー、そして労働者の私的行動の抑制に関する考察（リアル・エフォート実験からの事実）

亀井 憲樹、ケイティ・タベロ

IES Keio DP2023-011
2023 年 4 月 27 日
JEL Classification: C92, D02, D72, H41
キーワード: ワークプレイスデモクラシー、モラルハザード、実験、ただ乗り、チームワーク

【要旨】

企業組織において「チーム」は典型的な意思決定ユニットであり、また作業ユニットでもある。本論文では、斬新なリアル・エフォートタスク実験をもとに行動データを収集することで、（A）どの程度、チームがグループ最適の実現のためにメンバーの私的行動を抑制することを選び、また（B）ワークプレイス・デモクラシーが労働生産性向上にどう寄与するか考察した。経済実験に参加した被験者は３人１組のチームに割り振られ、他の２チームとグループを構成し、グループでのレベニュー・シェアのもと、リアル・エフォートタスクに取り組んだ。各メンバーは、タスク期間中に、他のメンバーに知られることなく、怠業、つまり、テトリス・ゲームをプレイすることが出来た。本研究では、各グループに怠業インセンティブを減らすポリシーがランダムに導入される『非民主的な』トリートメントと、民主的に導入の可否が決定される『民主的』トリートメントの２種類が実施された。実験結果によると、非民主的トリートメントに比べ、民主的トリートメントで統計的に有意なレベルの強い労働時間当たりの生産が実現した。この正の効果はポリシー導入の可否によらず観測された。このことは民主的カルチャーが人々の行動に直接正の作用を及ぼしたことを意味する。一方で、ワークプレイス・デモクラシーにより、（強い労働による疲労を受けた）労働者の怠業時間が増大した。しかしながら、労働時間当たり生産性の増大を受け、『民主的』トリートメントで生産が減少することはなかった。

亀井 憲樹
慶應義塾大学経済学部
〒108-8345 東京都港区三田2-15-45
kenju.kamei@keio.jp

ケイティ・タベロ
ダラム大学
英国ダラム郡ミルヒル・レーン　ダラム大学ビジネススクール（郵便番号 DH1 3LB）
katy.tabero@durham.ac.uk

謝辞：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 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.

Free Riding, Democracy and Sacrifice in the Workplace:

Evidence from a Real Effort Experiment

April 2023

Kenju Kamei[#1] and Katy Tabero[#2]

[#1] Faculty of Economics, Keio University, 2-15-45, Mita, Minato-ku, Tokyo 108-8345, Japan. Email: kenju.kamei@keio.jp.

[#2] Department of Economics and Finance, Durham University, Mill Hill Lane, Durham, DH1 3LB, United Kingdom. Email: katy.tabero@durham.ac.uk.

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.

Keywords: workplace democracy, moral hazard, experiment, free riding, teamwork

JEL classification codes: C92, D02, D72, H41

Acknowledgement: 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 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.

1. Introduction

Teams are increasing popular in firms as decision-making and work units (e.g., Kamei and 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 and Demsetz, 1972; Marschak and 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 and Dewatripont, 2004) — a typical example of this is moral hazard in groups (e.g., Alchian and 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 and Ryan, 1985, 2000), but by also providing workers with opportunities to signal their willingness to cooperate with their peers through democratic processes (e.g., Connelly *et al.*, 2011; Bergh *et al.*, 2014), 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, e.g., their cooperation behaviors are heterogeneous (e.g., Fischbacher *et al.*, 2001), they are easily discouraged by seeing their peers free ride (e.g., Fischbacher and Gächter, 2010); or many tend to cooperate but by less than others (e.g., Thöni and 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 and 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., Prendergast 1999 for personnel economics, and Chaudhuri 2011 for experimental, literature). 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" is referred 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 pro-social 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'

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

cooperativeness (e.g., Tyran and Feld, 2006; Dal Bó *et al.*, 2010; Sutter *et al.*, 2010; Kamei, 2016). Among others, Tyran and Feld (2006), Dal Bó *et al.* (2010) and Dal Bó *et al.* (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 and Markussen, forthcoming; 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 find 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 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 defence (see, e.g., Smith and Bliege, 2000; Hawkes and Bliege, 2002; Sosis and Alcorta, 2003; Sosis and Bressler, 2003; Hagen and Bryant, 2003; Lannaccone, 1992). The mechanism is described as follows: sacrificing serves as a costly signal of one's own quality (e.g., Gintis *et al.*, 2001; Bliege and Smith, 2005), 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 and 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

[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 dataset 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.

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 non-work 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, e.g., 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 five 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 4s in a matrix whose cells contain 1s, 2s, 3s, or 4s. 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, 1739; Rousseau, 1755): all the three members of a team 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

[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] Knew 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.

(thereby earning some earnings 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 (forthcoming) 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., Falk *et al.*, 2006; Abeler *et al.*, 2009; Kamei and Markussen, forthcoming), 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 and 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 dataset to study the reasoning behind sacrifice decisions. A well-established coding exercise is applied to the communication logs in order 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, i.e., 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. Appendix A summarizes the experimental procedure and includes instructions used in the experiment.

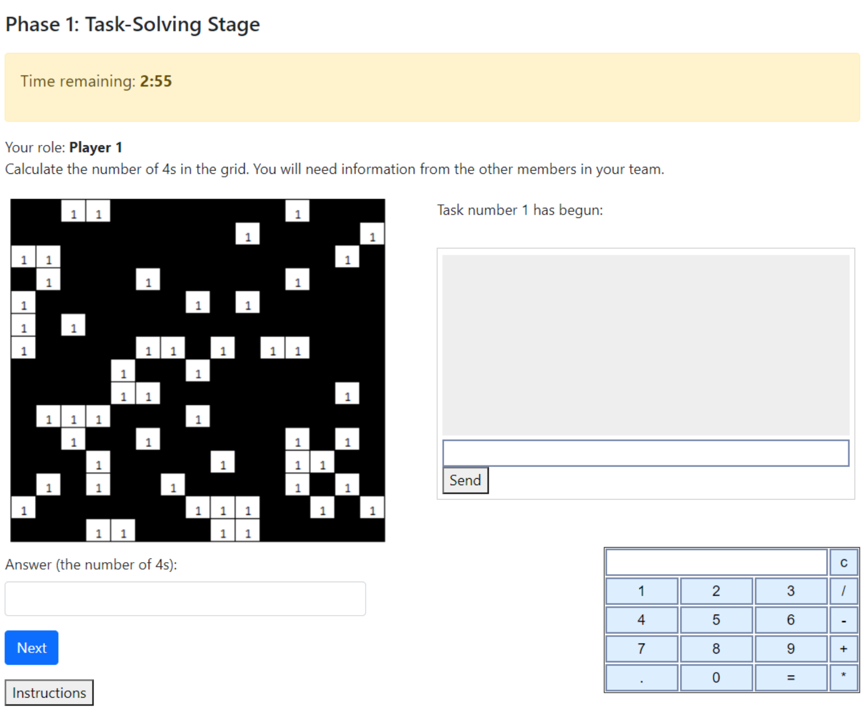
2.1. A Collaborative Real Effort Task

Three members in a respective team collaboratively solve a variant of the counting task ("collaborative counting task"). The original "counting task" (e.g., Falk *et al.*, 2006; Abeler *et al.*, 2009; Kamei and Markussen, forthcoming) is an individual real effort task in which subjects independently count the number of 0s in a matrix that contains 0s and 1s. 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 4s. Collaboration is required to find the correct answer, because only number *k*s 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 4s accurately after that. For example, if the numbers of 1s, 2s, and 3s are, respectively, 32, 14, and 43, then the correct answer (the number of 4s) is: 225 – 32 – 14 – 43 = 136. A calculator is available on each subject's computer screen. How to calculate the number of 4s, 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 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 1s, 2s, 3s, and 4s 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; 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, post-experiment. 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 2.1.

Figure 1: *A Screen for Collaborative Counting Task*



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

[8] This very rarely happened in the experiment. All three members submitted the same answers in 96.9% of teams' submissions in the experiment (3,176 out of 3,278 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.

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 three minutes.[11] While they can answer as many questions as they wish, they are not informed whether they answer each question correctly during the three-minute 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 minutes – 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, i.e., 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 2.a). The screen is then switched to the Tetris site (Figure 2.b). No one, including their teammates, are 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 per minute 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

[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 *et al.*, 2015; Kamei and Tabero, 2022).

[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 emails 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 in Kamei and Markussen (forthcoming) where an activity alternative to solving a real effort task is to watch a funny video. Subjects in Kamei and Markussen (forthcoming) received a small return per minute watching the videos.

with Kamei and Markussen (forthcoming) that adopted comedy 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 (forthcoming), 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 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 McNamara *et al.*, 2022; Doellgast and Marsden, 2018; Gibbons and Henderson, 2013, for examples). 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):

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

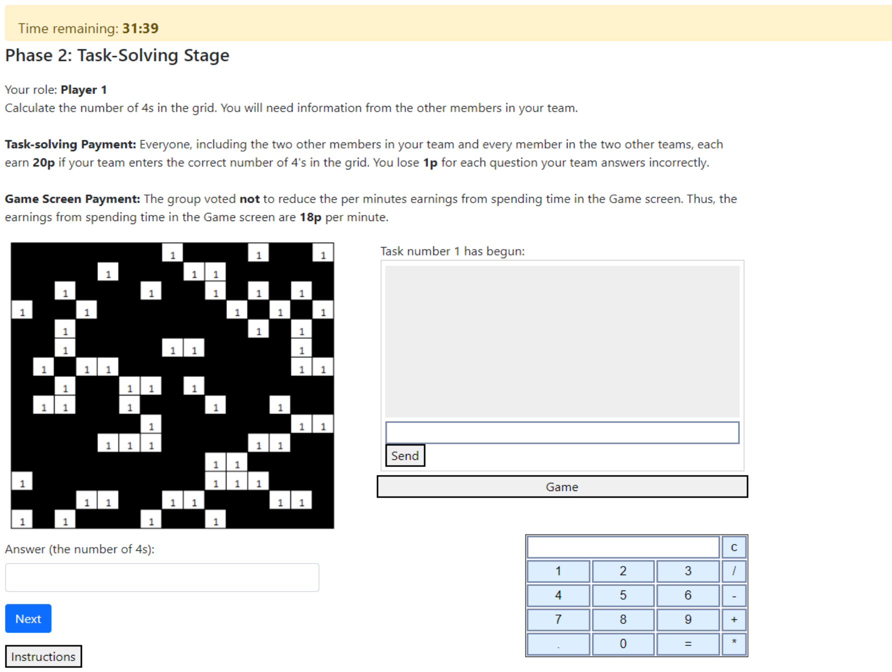
where c_k and ic_k are the numbers of, respectively, correct and incorrect answers by team *k*, g_i is the time [minutes] that member *i* spends in the Game screen, and *r* is per minute return from shirking. Notice that their work time is $35 - g_i$ as they are not allowed to work while playing Tetris. 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-minute 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.

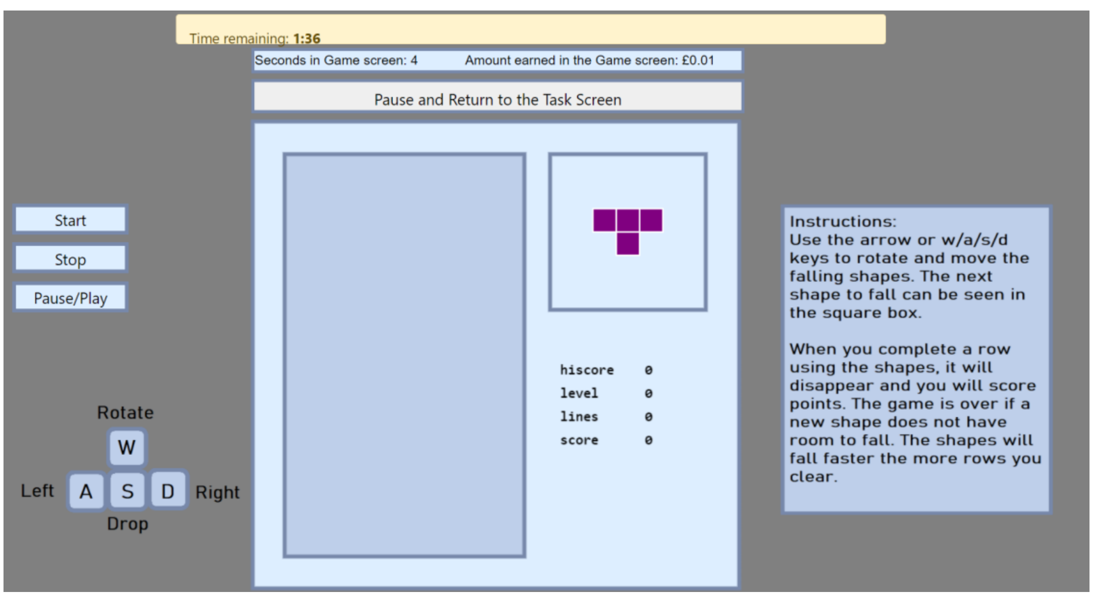
At the beginning of the main task-solving phase, the return from staying in the Game screen (*r*) could decrease from 18 to 16 pence per minute. Notice that the size of the incentive change is very small at only two pence. This means that the reduction policy can be thought of as a non-deterrent sanction policy, i.e., a policy that *does not* alter the privately optimal behaviors of workers in a group (e.g., Tyran and Feld, 2006; Kamei, 2016). 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.

[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.

Figure 2: *A Screen Image for Collaborative Counting Task in the Main Task-Solving Phase*



(a) Work site



(b) Game site

The process to implement the reduction policy differs by treatment. 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

[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).

voting procedure follows three steps:

Step 1. The three members in each team are given three minutes to discuss, using an electronic chat window (e.g., Kamei 2019b; Luhan *et al.*, 2007), whether they want to reduce the per minute earnings from staying in the Game screen. The communication contents are not revealed to any other team. See Appendix A.3 for a screen image of this step.

Step 2. After the three-minute 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.

Table 1: *Treatments, Distribution of Votes and Institutional Outcomes*

Treatment and institutional outcome	Condition in which the policy is/isn't implemented	# subjects	# of subjects in pro-reduction teams	# of subjects in anti-reduction 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 pro-reduction teams" and "# of subjects in anti-reduction teams" columns are based on the results of voting in the experiment.

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 Appendix B, a calculation 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; Dal Bó *et al.*, 2019; Sutter *et al.*, 2010; Tyran and Feld, 2006; Kamei, 2016). In a workplace setting, Kamei and Markussen (forthcoming) model this effect such that workplace democracy lowers workers' marginal effort costs. A model similar to Kamei and Markussen (forthcoming) supports the hypothesis above; a decrease in the marginal effort costs driven by democracy results in hard work among teams (see Appendix B for the detail). Part of the dividend of democracy can also be attributed to signaling effects (e.g., Tyran and Feld, 2006; Kamei, 2019a; Jensen and Markussen, 2022).

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; Dal Bó *et al.*, 2019; Tyran and Feld, 2006). By design, pro-reduction 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 pro-reduction 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 anti-policy units with this method. The detail of the re-weighting method along with the data will be provided in Section 3.

3. Policy Preferences, and the Dividend of Democracy

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. 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 (= 38/93×100%) voted for the reduction policy. As a result of majority voting, the policy was enacted in 38.7% (= 36/93×100%) of groups in the ENDO treatment. Table 1 also shows a clear pattern of selection bias. In the ENDO treatment, the percentage of pro-reduction 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, pro-reduction 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) and Dal Bó *et al.* (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 three minutes 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 $p < 0.01$ ($z = 4.230$) than teams which voted against the policy (the average number of correct answers by anti-reduction teams was 0.414). This pattern holds regardless of the institutional outcome, i.e., whether the policy was enacted or not (Appendix Figure C.1). This means that pro-reduction teams may have characteristics different from anti-reduction teams. As shown in Appendix Figure C.1, the performance of teams in the EXO treatment was somewhere in the middle of the pro- and anti-reduction teams (was similar to that of anti-reduction 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. 0.594 means that if a team, i.e., all three members, worked the entire 35 minutes 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," i.e., the distribution of votes. This paper follows Dal Bó *et al.* (2019) calculating the re-weighted productivity with the 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 pro-reduction 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 re-weighted work productivity in the ENDO treatment found using these steps is still quite large – i.e., 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 pro-reduction teams are underrepresented in the ENDO treatment, i.e., 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, i.e., 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, i.e., democratic culture, affects their work motivation directly, which is consistent with the idea that democracy directly enhances intrinsic motivations to work (e.g., Deci and Ryan, 1985, 2000).

In sum, the reweighted dividend of democracy without the reduction policy (i.e., 0.539 versus 0.431) was of almost a similar magnitude to the one in groups with the reduction policy (0.529 versus 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 reports test results for the dividend of democracy on work productivity using all of the data. In order 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 to 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

[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.

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]

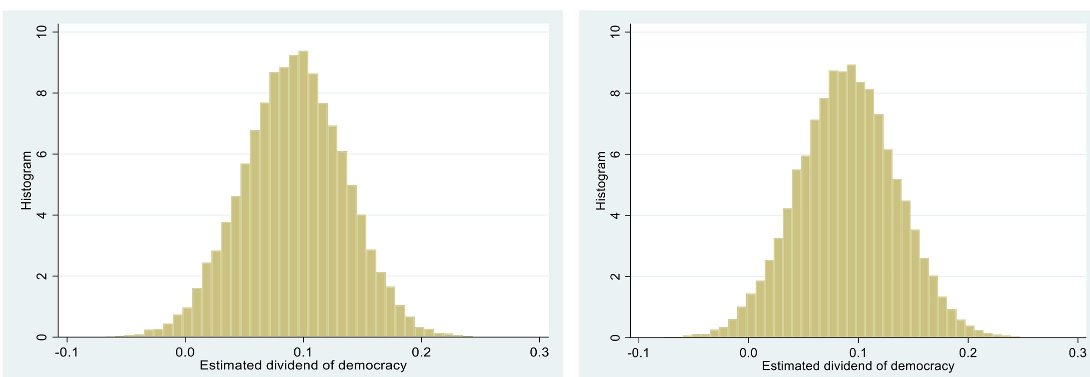
Result 2: *(a) There is strong evidence that democracy significantly boosted work productivity, defined as the production per minute spent working. (b) The positive dividend of democracy persisted throughout the task-solving phase.*

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:#1		
(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)#2	**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 anti-reduction teams) shown in rows i and ii 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). #1 The number of correct answers per minute of teamwork #2 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).

Figure 3: *Dividends of Democracy for Work Productivity*

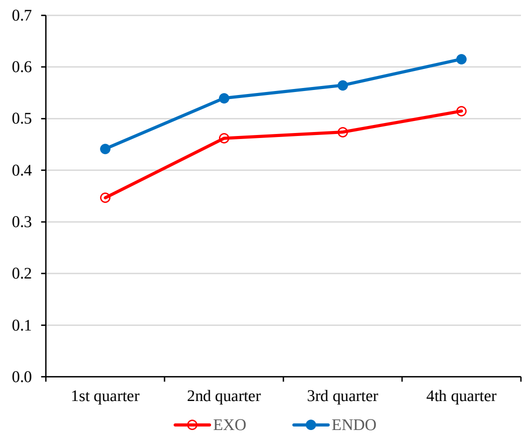


(i) When using original weights

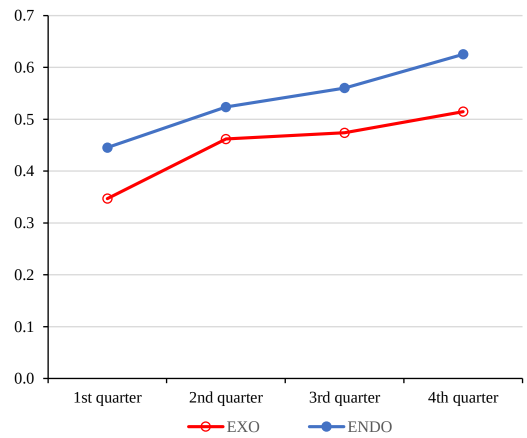
(ii) When using adjusted weights according to Dal Bó *et al.* (2019)

A. Distribution of bootstrapped dividends of democracy for productivity based on Dal Bó *et al.* (2019)

[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.



(i) When using original weights



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

B. Dividend of democracy, quarter by quarter

Notes: 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 35 task-solving phase by four (e.g., the first quarter is the first 35/4 minutes).

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., Tyran and Feld, 2006; Dal Bó *et al.*, 2010; Sutter *et al.*, 2010; Kamei, 2016), it is at odds with the finding from the 'real-effort' experiment of Kamei and Markussen (forthcoming). In Kamei and Markussen (forthcoming), subjects were assigned to a group of three and then worked individually on either the "counting task" (e.g., Falk *et al.*, 2006; Abeler *et al.*, 2009) or the "addition task" (e.g., Niederle and Vesterlund, 2007; Corgnet *et al.*, 2015) on condition that a revenue-sharing rule is in use and a funny video is available as an alternative activity. Kamei and Markussen (forthcoming) found little evidence of the effects of democratic task selection. The null result was indeed a puzzle which Kamei and Markussen (forthcoming) 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, Foster and Kamei (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, i.e., 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.

---

[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.

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) 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.

Table 3: *Work Performance and the Dividend of Democracy*

	Un-weighted			**Re-weighted**		
	All data (1)	With Policy (1a)	W/o policy (1b)	All data (2)	With Policy (2a)	W/o policy (2b)
I. Avg. 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)[#1]	0.151	**0.043****	0.949	0.331	0.747	0.285
II. Avg. 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)[#1]	0.170	**0.060***	0.983	0.330	0.671	0.336
III. Avg per member time spent in the Game screen [min.][#1]						
(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)[#1]	0.664	0.345	0.236	0.534	0.594	0.711
IV. Avg. 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)[#2]	**0.065***	**0.062***	0.555	0.138	0.498	0.150

[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.

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 anti-reduction 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).

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 in 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, i.e., 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 per minute, 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 (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 non-deterrent in the experiment. Consistent with the prior experimental evidence on exogenously introduced non-deterrent punishment (e.g., Tyran and Feld, 2006; Kamei, 2016), 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 versus 9.16), the difference was not significant at two-sided $p = 0.109$ according to the bootstrap method used in

[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 minutes, which is significantly larger at two-sided $p = 0.0015$ than the average work time by the other 87 teams where gaming was privately optimal (which was 23.12 minutes) – see Appendix Table C1.

the other tests of the paper (the difference is significant but only at the 10% level, i.e., $p = 0.0707$ 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 $p < 0.001$ (15.29 versus 28.63 minutes 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 non-deterrent 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, i.e., 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 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 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 Appendix Figure B.2. However, the analysis in the 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

[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 $p = 0.001$*** (0.0020***) 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 $p = 0.388$ 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.

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 intra-team communication logs. The detail of the coding procedure and the full lists of codes used for the present paper can be found in 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 (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/anti-policy reasoning (Code Cs) reveals clear motives behind the policy preferences. While the policy is non-deterrent, 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).

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.*

Table 4: *Significant Code Meanings and Its Impact on Voting for the Reduction Policy*

Code	Meaning	Direction
B1	Agree/Imply to count as primary behavior	(+)***
B2	Agree/Imply to game as primary behavior	(-)*
B3	Agree to hybrid behavior e.g. 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-minute phase) based on performance/needs in Phase 2	(-)**
C1	Pro-policy to deter others from switching to the game screen by reducing the return (monetary deterrence)	(+)***
C8	Anti-policy 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 e.g. attempts to encourage others or being supportive	(+)*
E5	No communication from just 1 or 2 team members	(-)***

Notes: +(-) in "Direction" means the reasoning en(dis)courages voting for the policy. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

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-minutes task-solving phases (Table 5). Three similar tendencies were observed for both the ENDO and EXO treatments. First, those who 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, e.g., Sobel (2005) and Fehr and Schmidt (2006). Second, a units' 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 cancelled out if an anti-policy team is present in a group and units have a negative view on the task-solving behavior of the anti-policy team (F13).

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, anger) about the outcome of the vote	(wt-)***
F3	Agree/Imply to count as primary behavior	(wt+)***, (p+)***
F4	Agree/Imply to game as primary behavior	(wt-)***, (p-)***

F5	Agree to hybrid behavior e.g. 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-minute 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: anti-policy 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, anger) about the policy outcome	(wt-)***, (p-)***
I4	Agree/Imply to game as primary behavior	(wt-)***, (p-)***
I5	Agree to hybrid behavior e.g. 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-minute 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 e.g. frustration, anger, disappointment	(p-***, Exo)
G5	Expression of strong positive emotion e.g. 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 e.g. expect them to take advantage	(wt-***, p-***, Endo)

Notes: wt and p in the "Direction" column indicates 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 Code Fs and Is, while only some key codes are included for the other coding categories to conserve space (Appendix D4 includes the full estimation results). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

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 non-deterrent 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 intra-team 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 in 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, i.e., 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) lead 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 and Markussen, forthcoming). 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 to 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 recent work style reform. There is a trend to transform the traditional workplace to an employee-centered workplace in many countries. For example, in the United Kingdom, some firms recently tested four-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 to 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 through 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.

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman, 2011. 2009. "Reference Points and Effort Provision." American Economic Review, 101(2), 470-492

Aimone, Jason, Laurence Iannaccone, and Michael Makowsky, 2013. "Endogenous Group Formation via Unproductive Costs." *Review of Economic Studies*, 80(4), 1215-1236.

Alchian, Armen, and Harold Demsetz, 1972 "Production, Information Costs, and Economic Organization." *American Economic Review*, 62(5), 777-795.

Araujo, Felip, Erin Carbone, Lynn Conell-Price *et al.*, 2016. "The Slider Task: an Example of Restricted Inference on Incentive Effects." *Journal of Economic Science Association* 2, 1-12.

Battaglini, Marco, Rebecca Morton, and Thomas Palfrey, 2010. "The Swing Voter's Curse in the Laboratory." *Review of Economic Studies* 77(1), 61-89.

Bednar, Jenna, Yan Chen, Tracy Liu, and Scott Page, 2012. "Behavioral Spillovers and Cognitive Load in Multiple Games: an Experimental Study." *Games and Economic Behavior*, 74, 12-31.

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

Bergh, Donald, Brian Connelly, David Ketchen, Jr and Lu M. Shannon, 2014. "Signalling Theory and Equilibrium in Strategic Management Research: An Assessment and a Research Agenda." *Journal of Management Studies*, 51(8), 1334-1360.

Bergstrom, Theodore, Lawrence Blume, and Hal Varian, 1986. "On the Private Provision of Public Goods." *Journal of Public Economics*, 29, 25-49.

Bisetti, Emilio, Benjamin Tengelsen, Ariel Zetlin-Jones, 2022. Moral Hazard in Remote Teams. *International Economic Review* 63(4), 1595-1623.

Bolton, Patrick, and Mathias Dewatripont, 2004. *Contract Theory*. MIT Press.

Brekke, Kjell, Karen Hauge, Jo Lind, and Karine Nyborg, 2011. "Playing with the Good Guys. A Public Good Game with Endogenous Group Formation." *Journal of Public Economics*, 95(9-10), 1111-1118.

Chaudhuri, Ananish, 2011. "Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature." *Experimental Economics*, 14(1), 47-83.

Cohen, Jacob, 1960 "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20, 37-46.

Connelly, Brian, Trevis Certo, Duane Ireland, and Christopher Reutzel, 2011. "Signaling Theory: A Review and Assessment." *Journal of Management*, 37(1), 39-67.

Corgnet, Brice, Roberto Hernan-Gonzalez, and Eric Schniter, 2015. "Why Real Leisure Really Matters: Incentive Effects on Real Effort in the Laboratory." *Experimental Economics*, 18, 284-301.

Dal Bó, Pedro. 2010. "Experimental Evidence on the workings of democratic institutions." In: Galiani S. and Sened I. (eds) *Institutions, property rights, and economic growth: the legacy of Douglass North*. Cambridge University Press, Cambridge

Dal Bó, Pedro, Andrew Foster, and Louis Putterman, 2010. "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review*, 100(5), 2205-2229.

Dal Bó, Pedro, Andrew Foster, and Kenju Kamei, 2019. "The Democracy Effect: a Weights-Based Identification Strategy." NEBR Working Paper 25724.Erkal, Nisvan, Lata Gangadharan, and Boon Koh, 2018. "Monetary and non-monetary incentives in real-effort tournaments." European Economic Review, 101, 528-545.

Deci, Edward, and Richard Ryan, 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.

Deci, Edward, and Richard Ryan, 2000. "The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior." *Psychological Inquiry*, 11(4), 227-268.

Elfenbein, Daniel, Ray Fisman, and Brian Mcmanus, 2012. "Charity as a Substitute for Reputation: Evidence from an Online Marketplace." *Review of Economic Studies*, 79(4), 1441-68.

Falk, Armin, David Huffman, and Konrad Mierendorff, 2006. "Incentive Properties and Political Acceptability of Workfare: Evidence from Real Effort Experiments." Working paper.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr, 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters*, 71(3), 397-404.

Fischbacher, Urs, Simon Gächter, 2010. "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments." *American Economic Review,* 100(1), 541-56.

Gintis, Herbert, Eric Smith, and Samuel Bowles, 2001. "Costly Signaling and Cooperation." *Journal of Theoretical Biology*, 213(1), 103-119.

Grimm, Veronika, and Friederike Mengel, 2009. "Cooperation in viscous populations—Experimental evidence." *Games and Economic Behavior* 66(1), 202-220.

Hagen, Edward, and Gregory Bryant, 2003. "Music and Dance as a Coalition Signaling System." *Human Nature*. 14(1), 21-51.

Hawkes, Kristen, and Rebecca Bliege, 2002. "Showing off, Handicap Signaling, and the Evolution of Men's Work." *Evolutionary Anthropology*, 11(2), 58-67.

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

Kamei, Kenju, 2016. "Democracy and Resilient Pro-social Behavioral Change: An Experimental Study." *Social Choice and Welfare*, 47(2), 359-378.

Kamei, Kenju, 2019a. "Cooperation and Endogenous Repetition in an Infinitely Repeated Social Dilemma." *International Journal of Game Theory*, 48(3), 797-834.

Kamei, Kenju, 2019b. "The Power of Joint Decision-Making in a Finitely-Repeated Dilemma." *Oxford Economic Papers*, 71(3), 600-622.

Kamei, Kenju, and Thomas Markussen, forthcoming. "Free Riding and Workplace Democracy – Heterogeneous Task Preferences and Sorting." *Management Science*.

Kamei, Kenju, Louis Putterman, and Jean-Robert Tyran, 2015. "State or Nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods." *Experimental Economics*, 18, 38-65.

Kamei, Kenju, Katy Tabero, 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.

Laurence, Iannaccone, 1992. Sacrifice and Stigma: Reducing Free-riding in Cults, Communes, and Other Collectives. *Journal of Political Economy*, 100(2), 271-291.

Ledyard, John, 1995. "Public Goods: A Survey of Experimental Research," pages 111-194 in J. Kagel and A. Roth (eds.), *Handbook of Experimental Economics*. Princeton University Press.

Luhan, Wolfgang, Martin Kocher, and Matthias Sutter, 2009. "Group polarization in the team dictator game reconsidered." *Experimental Economics* 12, 26-41.

Muriel Niederle, and Lise Vesterlund, 2007. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics*, 122(3), 1067-1101.

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature*, 37(1), 7-63.

Samuelson, Pall, 1954. "The Pure Theory of Public Expenditure." *Review of Economics and Statistics*, 36(4), 387-389.

Smith, Eric, and Rebecca Bliege, 2000. "Turtle Hunting and Tombstone Opening: Public Generosity as Costly Signaling." *Evolution and Human Behavior*, 21(4), 245-261.

Smith, Eric, and Rebecca Bliege Bird, 2005. "Costly Signaling and Prosocial Behavior." In: *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life* (eds. Gintis, H., Bowles, S., Boyd, R., and Fehr, E.), pp. 115-148. MIT Press

Sosis, Richard, and Candace Alcorta, 2003. "Signaling, Solidarity, and the Sacred: The Evolution of Religious Behavior." *Evolutionary Anthropology*, 12(6), 264-274.

Sosis, Richard, and Eric Bressler, 2003. "Cooperation and Commune Longevity: A Test of the Costly Signaling Theory of Religion." *Cross-Cultural Research*, 37(2), 211-239.

Spence, Michael, 1973. “Job Market Signaling.” *Quarterly Journal of Economics*, 87(3), 355-374.

Sutter, Mattias, Stefan Haigner, and Martin Kocher, 2010. “Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations.” *Review of Economic Studies*, 77, 1540-1566.

Thöni, Christian, and Stefan Volk, 2018. “Conditional cooperation: Review and Refinement.” *Economics Letters*, 171, 37-40

Tyran, Jean-Robert, and Lars Feld, 2006. “Achieving Compliance when Legal Sanctions are Non-deterrent.” *Scandinavian Journal of Economics*, 108(1), 135-156.

Zelmer, Jennifer, 2003. “Linear Public Goods Experiments: A Meta-Analysis.” *Experimental Economics*, 6, 299-310.

Not for Publication

Supplementary Online Appendix to: Kamei and Tabero:

"Free Riding, Sacrifice and Democracy in the Workplace:

A Real Effort Experiment"[#]

Table of Contents

[#] Kamei: Keio University (kenju.kamei@gmail.com; kenju.kamei@keio.jp); Tabero: Durham University (katy.tabero@durham.ac.uk).

Appendix A: Experiment Procedure and Instructions Used in the Experiment

Sixty-two sessions (thirty-one sessions per treatment) were conducted online using the oTree software (Chen *et al.*, 2016) and Zoom from May 2021 through January 2022, following the same procedure as a standard laboratory experiment. All subjects' cameras were on during the session to make sure that they were alone, were paying attention to the experiment, and did not cheat when making decisions. While subjects were visible by the experimenter, they were unable to see the researcher on Zoom during the experiment. They also remained anonymous during the entire session without seeing other participants' faces, names, etc. Each session consisted of nine subjects. This means that each session consists of one group. However, full anonymity was retained in the experiment since, as already mentioned, subjects did not see the other students' names or faces, and they were recruited from a very large student population in the university. All experiment sessions were conducted using the subject pool and the experiment system in the EXEC (Centre for Experimental Economics) at the University of York. As all standard experiment protocols (such as the no deception rule) have been rigorously adopted for any experiment in this laboratory for more than 30 years, it can be assumed that subjects believed in the explanation provided in the experiment, although subjects do not see the presence of any subject physically.

All subjects were recruited using *hroot* (Bock *et al.*, 2014). Subjects did not participate in more than one session. A total of 552 students in the University of York participated in the experiment.[23] The instructions shown below were neutrally framed. Terms with positive or negative connotations, such as shirk, free ride or cooperate, were not used. Each session took between 90-120 minutes.

This part of the Appendix includes the instructions for both the ENDO and EXO treatments as follows:

A.1: Instructions for phase 1 (identical for the ENDO and EXO treatments)

A.2: Instructions for phase 2 (the EXO treatment)

A.3: Instructions for phase 2 (the ENDO treatment)

References:

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch, 2014. "hroot: Hamburg Registration and Organization Online Tool." *European Economic Review*, 71, 117-120.

Chen, Daniel, Schonger, Martin, and Wickens, Chris, 2016. "oTree - An open-source platform for laboratory, online and field experiments." *Journal of Behavioral and Experimental Finance*, 9, 88-97.

[23] Two teams' observations (six subjects) were omitted from the study. The first one was due to a subject's computer experiencing technical problems, meaning that their whole team could not participate in the task-solving phase, and the second was due to suspected cheating which was identified in the chat dialogue.

A.1: Instructions: Slides for the Practice Phase (identical for the ENDO and EXO Treatments)

At the onset of the experiment, the following instructions (PowerPoint file) were shown on Zoom and were made available on the subjects' experiment screens; they were read aloud to subjects by the researcher.

Slide 1:

Welcome

You are now taking part in a decision-making experiment. In the experiment, you will be able to earn money depending on your decisions and the decisions of other participants in some tasks. Please read the following instructions carefully.

During the experiment you are not allowed to communicate with other participants. If you have any questions, please write your questions in the Zoom chat box. We will answer your question in private. Your microphones will be automatically muted by us on Zoom. The camera of your computer should be turned off until you are put into your individual breakout rooms. In your breakout rooms, your camera should be turned on to ensure that you are still participating, but only the experimenter will be able to see you and you will not be recorded.

Slide 2:

There are **nine** participants in this experiment. You will not be made aware of any information about other participants' identity, such as their faces and names. At the onset of the experiment, each participant will be randomly assigned to **a team of three**. This means that there are three teams in the experiment. Team assignment is completely random. The three members in your team will then be randomly assigned ID numbers, 1, 2, or 3, so that each member receives a different number from the others. The player who is assigned number k in your team will become "player k."

This experiment consists of two phases. The team composition **stays the same for the two phases**. We will first explain the detail of Phase 1. We will explain Phase 2 once Phase 1 is over.

Slide 3:

Phase 1:

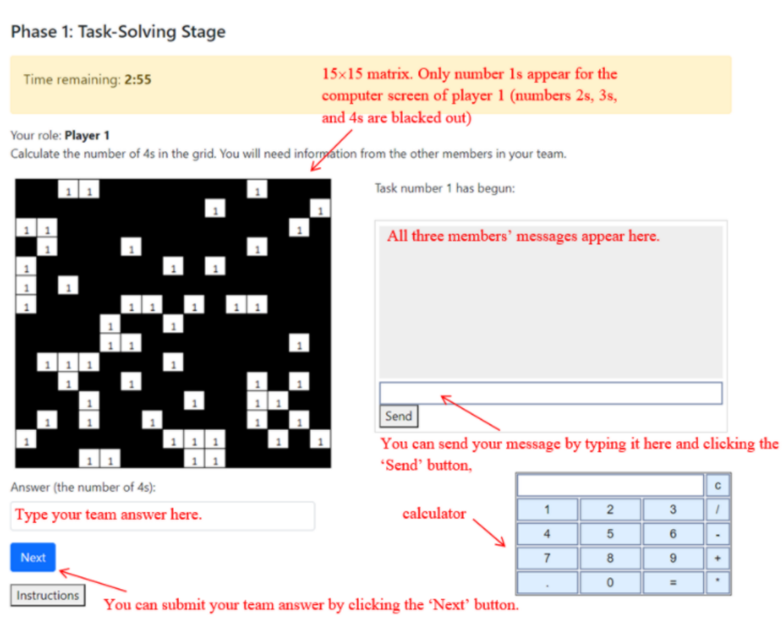
You and the other two members of your team will jointly solve **number counting tasks** for **3 minutes**. Your payment depends on the work performance of your team, not the performance of the other two teams. We will explain the nature of the number counting task, and how you are paid in this phase, in order.

1. Number Counting Task:

Your team will be asked to jointly solve as many number counting questions as possible. Each question has a table that consists of **225 randomly ordered 1s, 2s, 3s and 4s** (i.e., 15X15 matrix). Player k in your team will be shown a table on which only number ks appear while the other three numbers are blacked out, and they **will then count the number of ks**. For example, player 1 will count the number of 1s as in the following (next slide) computer screen image. You can see that on player 1's computer screen, the numbers of 2s, 3s, and 4s are blacked out.

Slide 4:

Screen Image:



Slide 5:

In order to submit a team answer, your team must jointly find **the number of 4s** included in the table. For example, if the numbers of 1s, 2s, and 3s are 43, 62, and 58, respectively, in a given table, you can find the number of 4s by the following calculation:

$$225 - 43 - 62 - 58 = 62.$$

Operationally, the three members in your team need to communicate with each other using an electronic chat window on the right side of the screen. You need to exchange information with each other regarding the numbers found on each of your own screens, although messages in the chat window are unrestricted. For example, player 1 can share the number of 1s with the two other members. The three members then jointly calculate the number of 4s through communication. A calculator is available also on each screen. In the communication stage, any kind of offensive language is prohibited and you should not share any information that compromises your anonymity.

Slide 6:

How to submit an answer: There is a blank space below the label "Answer" on each member's screen. All three members must fill the blank on their screens with the team's joint answer (the number of 4s) and click the "Submit" button. **Only when all three members submit the correct answer, will the team's answer count as correct**. The submit button **does not** need to be pressed at exactly the same time, but a new matrix will not be generated until all three members have submitted an answer for the current matrix. Hence, you need to communicate to submit the joint answer. Once an individual member has submitted an answer, they will move to a waiting room screen until both of the other two team members submit an answer. You will still be able to communicate with your team members while in the waiting room. You will not be given feedback regarding whether your team's answer is correct or not for each question. You will instead be informed of the total number of questions correctly answered at the end of the three-minute phase. Note that in the case of disagreement, each member can submit a different answer from the others. However, your team's answer is then counted as incorrect.

Once your team submits an answer, you will move on to the next number counting question and will then be shown **a new table with 1s, 2s, 3s and 4s randomly generated**. You can then solve the new question as before.

Slide 7:

2. How to earn money in Phase 1:

Payment is based on individual team performance. **Your team earns 180 pence for each correct team answer**. The payment is equally divided among the three members in your team, meaning that you and the other two members of your team each **earn 60 pence** for each question correctly answered.

Are there any questions? If you have any questions, please write your questions in the Zoom chat box now.

A.2: Instructions: Slides for the Main Task-Solving Phase in the EXO Treatment

Once Phase 1 was over, the following instructions were shown on Zoom and were made available on the subjects' experiment screens; they were read aloud to subjects by the researcher.

Slide 1:

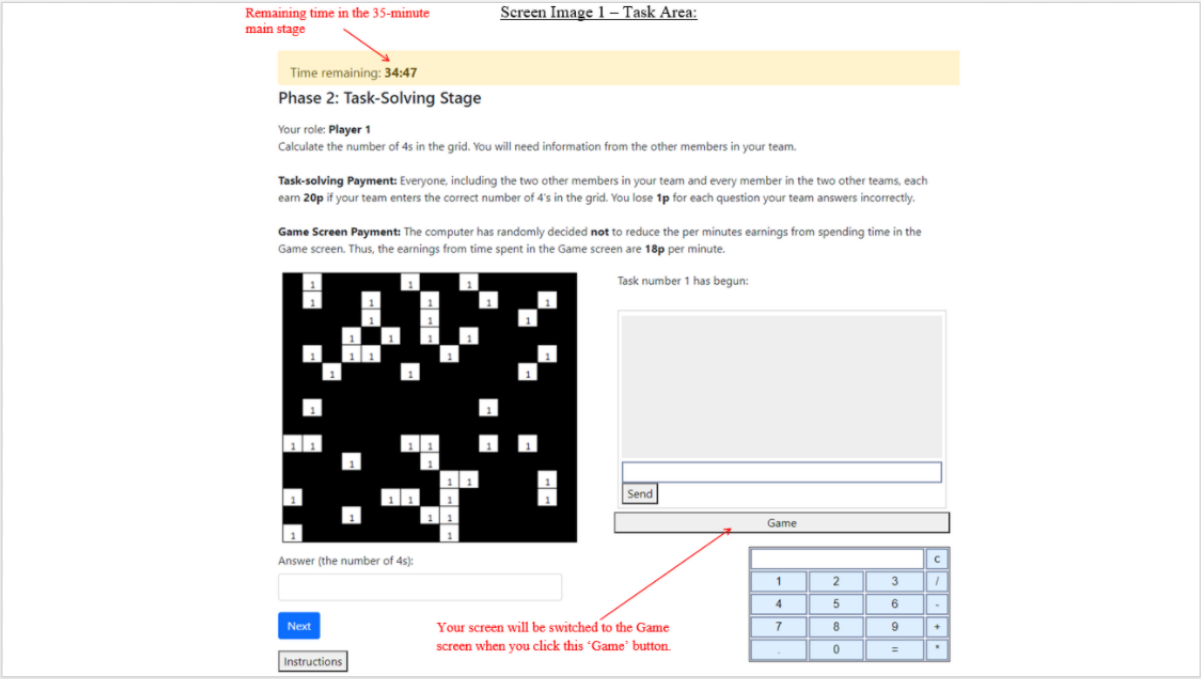
Instructions for Phase 2

As explained, you continue to be a member of the same team with two other participants in Phase 2.

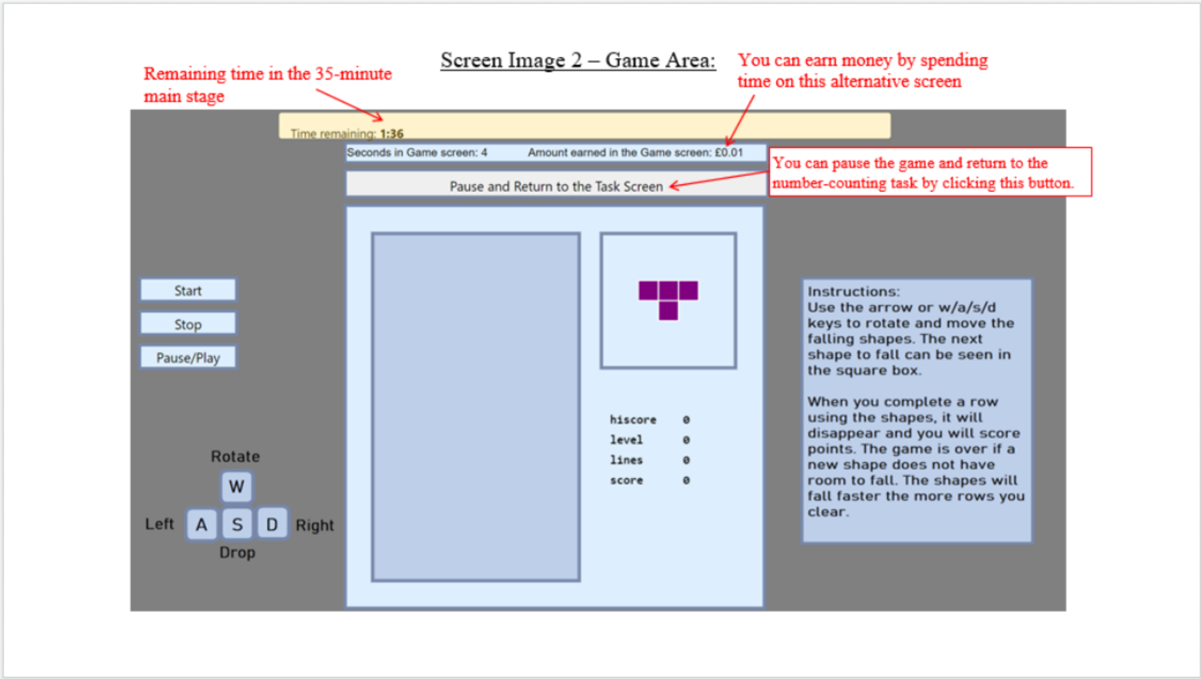
In Phase 2, you and the other two members of your team will jointly solve **number counting tasks** for **35 minutes**. Your payment depends on not only the work performance of your team **but also that of the other two teams**. The number counting task is exactly the same as the one you previously solved in Phase 1 except that an alternative activity is available.

An alternative activity – Tetris: In Phase 2, you can play Tetris, for example, to take a rest, whenever you like (see Screen Image 1 below). There is a **"Game"** button on the computer screen. You can choose to play Tetris at any point during the 35-minute task period by clicking the "Game" button. If clicked, your screen will change to one with a game of Tetris available (see Screen Image 2 below).

Slide 2:



Slide 3:



Slide 4:

In Tetris, you can use the arrow or w/a/s/d keys to rotate and move the falling shapes, with the aim of completing as many rows as possible. You can start or stop Tetris by clicking the 'Start' or 'Stop' button on the left side of the screen. You can also pause Tetris by clicking the 'Pause/Play' button while remaining on the Game screen (you can resume Tetris by clicking 'Pause/Play' button again).

Note that you cannot play Tetris while working on a counting task. You can return any time to the counting task screen and work on the same number counting question by clicking the "Pause and Return to the Task Screen" button on the Game screen, i.e., by pausing Tetris. You can also return to the Game screen and resume the game any time. No one, including the other two team members in your team, is informed about your decision to play Tetris unless you volunteer this information, for example in the team chat.

Slide 5:

How to earn money in Phase 2:

There are two sources for you to earn money:

(1) Earnings from the number counting task – you will be paid based on all of the participants' performances in this phase. After 35 minutes, the total number of correctly answered questions by each team are summed. Each correct answer is exchanged for **180 pence** as in Phase 1, which is equally shared among all the today's participants. There are three teams in the experiment, and each team receives **one-third of the total earnings from working on the number counting task**. This means that for each correct response, your team earns 180/3 = 60 pence, and each of the other two teams also earn 60 pence. Your team equally shares the team earnings among three members: you and the two other members in your team **will each receive 20 pence**. Note that you will also earn 20 pence for each question correctly answered by another team.

Unlike Phase 1, **there is also a penalty of 3 pence per incorrect answer** to your team; this penalty is only applied to your team's incorrect answers, and is shared equally between the three team members. This means that **you have to pay 1 penny per your team's incorrect answer**. You will only be told the total amount of correct and incorrect answers by your team, and the group, at the end of the 35-minute task-solving stage.

Slide 6:

(2) Earnings from staying in the Game screen – You will earn an amount of money also by switching to the Game screen instead of working on counting. Specifically, you will receive **18 pence** per minute spent in the Game screen. On this screen you can play Tetris. Note that the earnings from staying the Game screen will be added to your own earnings. They will not be shared with the other members of your team or with members of other teams, unlike working on the number counting task. Note that the score you get in Tetris does not affect your earnings; only the time you spend in the Game screen determines the size of payout here.

Slide 7:

Remark: Notice that if everyone stays in the Game screen for the entire 35-minute stage, each of you can earn 18 pence X 35 = 6.3 pounds in Phase 2. However, for example, if all three teams each answer two number counting questions correctly per 5 minutes, in this phase the group can correctly answer a total of 42 questions = 2 [correct answer per 5 minutes] X 7 [as the duration is 35 minutes] X 3 [teams]. This means that each participant can earn 20 pence X 42 = 8.4 pounds, which is larger than 6.3 pounds, in addition to whatever they earn from staying in the Game screen. If the three teams correctly answer more questions, it is possible for everyone to earn more than in this example. By contrast, if only your team works on the number counting task while other teams often switch to the Game screen, your team will earn less than the other teams, as revenue-sharing is used in this phase for the counting task only, not for spending time in the Game screen.

Are there any questions? If you have any questions, please write them in the Zoom chat box. Once all questions have been answered, we will explain the structure of Phase 2.

Slide 8:

Structure of Phase 2

Phase 2 proceeds with two steps:

Step 1: The computer's decision to randomly reduce the incentive to switch to the Game screen.

Before moving on to the main task-solving stage of 35 minutes, there is a possibility that the computer randomly enforces the change of payment rule from spending in the Game screen: this rule **reduces payment from 18 pence to 16 pence for each minute you spend in the Game screen**. If this policy change is made, then per minute return from being in the Game screen diminishes for every participant today.

Slide 9:

Step 2: Main Task-solving Stage

You will undertake the number counting task **for 35 minutes**. You and all other participants each get 20 pence for each question your team answers correctly, and lose 1 penny for each question your team answers incorrectly. You can also get 20 pence for each question answered correctly by another team. Instead of counting numbers, you can switch from the work screen to the Game screen, for example to play Tetris, as already discussed. You can earn points by spending time in the Game screen also – the per minute earnings differ dependent on the computer's random choice – either 16 or 18 pence per minute. The earnings from the Game screen will not be shared with anyone else in today's session.

Are there any questions? If you have any questions, please write them in the Zoom chat box

A.3: Instructions: Slides for the Main Task-Solving Phase in the ENDO Treatment

Once Phase 1 was over, the following instructions were shown on Zoom and were made available on the subjects' experiment screens; they were read aloud to subjects by the researcher.

Slide 1:

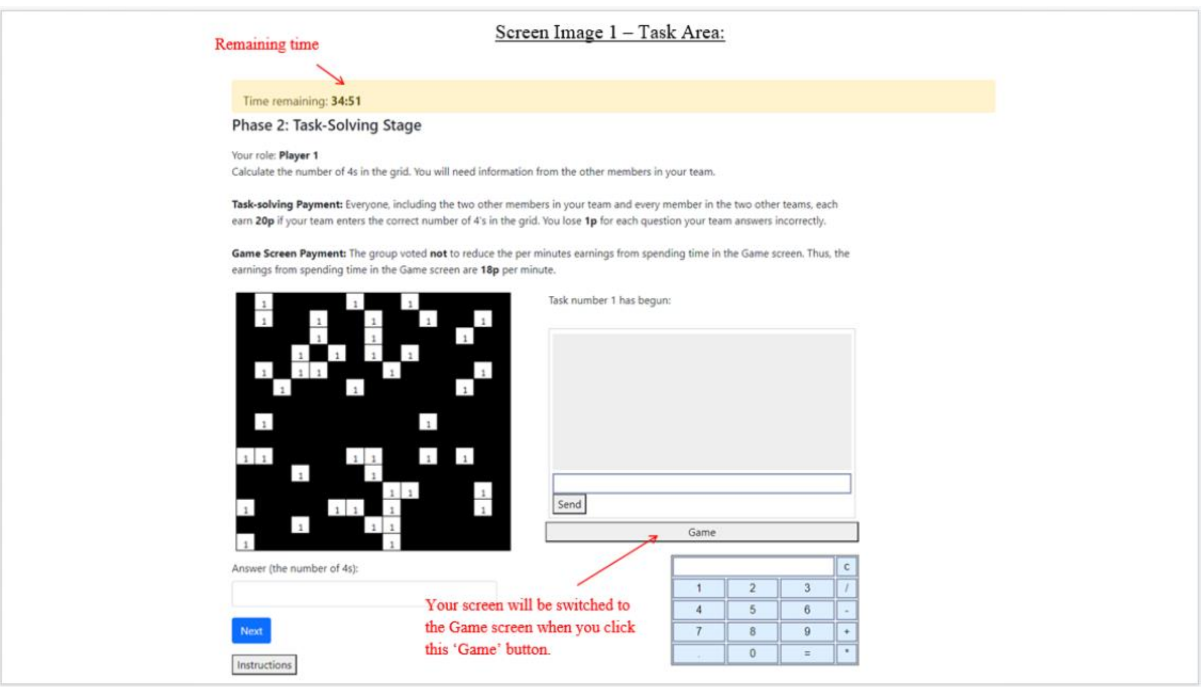
Instructions for Phase 2

As explained, you continue to be a member of the same team with two other participants in Phase 2.

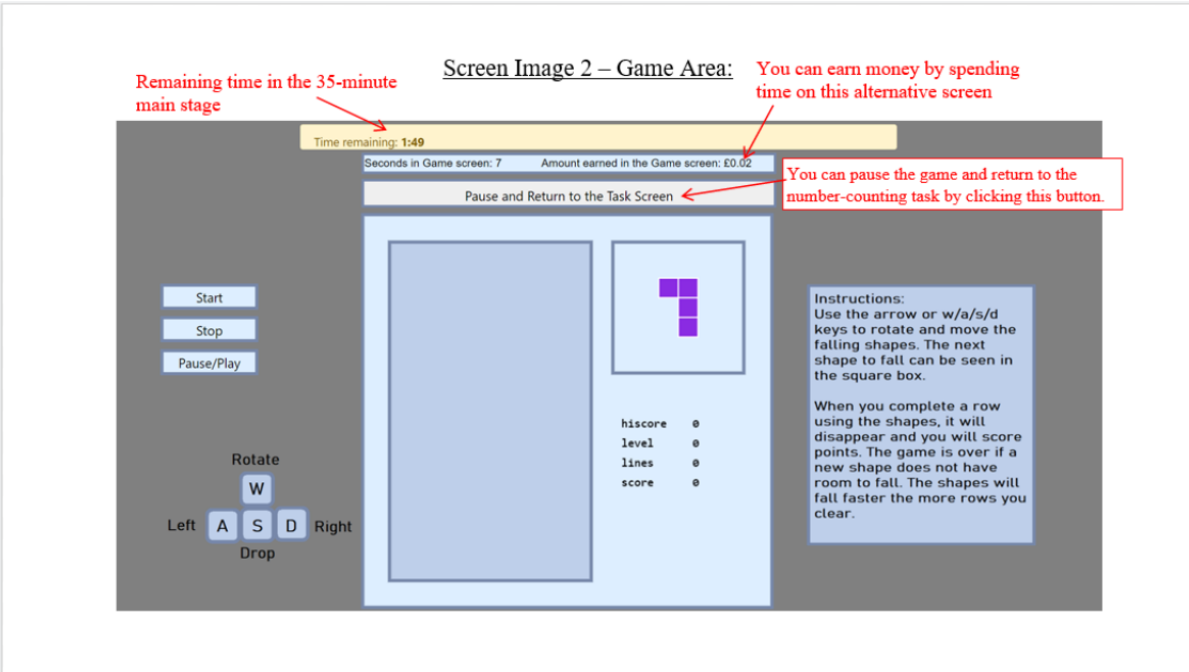
In Phase 2, you and the other two members of your team will jointly solve **number counting tasks** for **35 minutes**. Your payment depends on not only the work performance of your team **but also that of the other two teams**. The number counting task is exactly the same as the one you previously solved in Phase 1 except that an alternative activity is available.

An alternative activity – Tetris: In Phase 2, you can play Tetris, for example, to take a rest, whenever you like (see Screen Image 1 below). There is a **"Game"** button on the computer screen. You can choose to play Tetris at any point during the 35-minute task period by clicking the "Game" button. If clicked, your screen will change to one with a game of Tetris available (see Screen Image 2 below).

Slide 2:



Slide 3:



Slide 4:

In Tetris, you can use the arrow or w/a/s/d keys to rotate and move the falling shapes, with the aim of completing as many rows as possible. You can start or stop Tetris by clicking the 'Start' or 'Stop' button on the left side of the screen. You can also pause Tetris by clicking the 'Pause/Play' button while remaining on the Game screen (you can resume Tetris by clicking 'Pause/Play' button again).

Note that you <u>cannot</u> play Tetris while working on a counting task. You can return any time to the counting task screen and work on the same number counting question by clicking the "Pause and Return to the Task Screen" button on the Game screen, i.e., by pausing Tetris. You can also return to the Game screen and resume the game any time. <u>No one, including the other two team members in your team,</u> is informed about your decision to play Tetris unless you volunteer this information, for example in the team chat.

Slide 5:

How to earn money in Phase 2:

There are two sources for you to earn money:

(1) Earnings from the number counting task – you will be paid based on all of the participants' performances in this phase. After 35 minutes, the total number of correctly answered questions by each team are summed. Each correct answer is exchanged for **180 pence** as in Phase 1, which is equally shared among all the today's participants. There are three teams in the experiment, and each team receives **one-third of the total earnings from working on the number counting task**. This means that for each correct response, your team earns 180/3 = 60 pence, and each of the other two teams also earn 60 pence. Your team equally shares the team earnings among three members: you and the two other members in your team **will each receive 20 pence**. Note that you will also earn 20 pence for each question correctly answered by another team.

Unlike Phase 1, **there is also a penalty of 3 pence per incorrect answer** to your team; this penalty is only applied to your team's incorrect answers, and is shared equally between the three team members. This means that **you have to pay 1 penny per your team's incorrect answer**. You will only be told the total amount of correct and incorrect answers by your team, and the group, at the end of the 35-minute task-solving stage.

Slide 6:

(2) Earnings from staying in the Game screen – You will earn an amount of money also by switching to the Game screen instead of working on counting. Specifically, you will receive **18 pence** per minute spent in the Game screen. On this screen you can play Tetris. Note that the earnings from staying the Game screen will be added to your own earnings. They will not be shared with the other members of your team or with members of other teams, unlike working on the number counting task. Note that the score you get in Tetris does not affect your earnings; only the time you spend in the Game screen determines the size of payout here.

Slide 7:

Remark: Notice that if everyone stays in the Game screen for the entire 35-minute stage, each of you can earn 18 pence x 35 = 6.3 pounds in Phase 2. However, for example, if all three teams each answer two number counting questions correctly per 5 minutes, in this phase the group can correctly answer a total of 42 questions = 2 [correct answer per 5 minutes] x 7 [as the duration is 35 minutes] x 3 [teams]. This means that each participant can earn 20 pence x 42 = 8.4 pounds, which is larger than 6.3 pounds, in addition to whatever they earn from staying in the Game screen. If the three teams correctly answer more questions, it is possible for everyone to earn more than in this example. By contrast, if only your team works on the number counting task while other teams often switch to the Game screen, your team will earn less than the other teams, as revenue-sharing is used in this phase for the counting task only, not for spending time in the Game screen.

Are there any questions? If you have any questions, please write them in the Zoom chat box. Once all questions have been answered, we will explain the structure of Phase 2.

Slide 8:

Structure of Phase 2

Phase 2 proceeds with two steps:

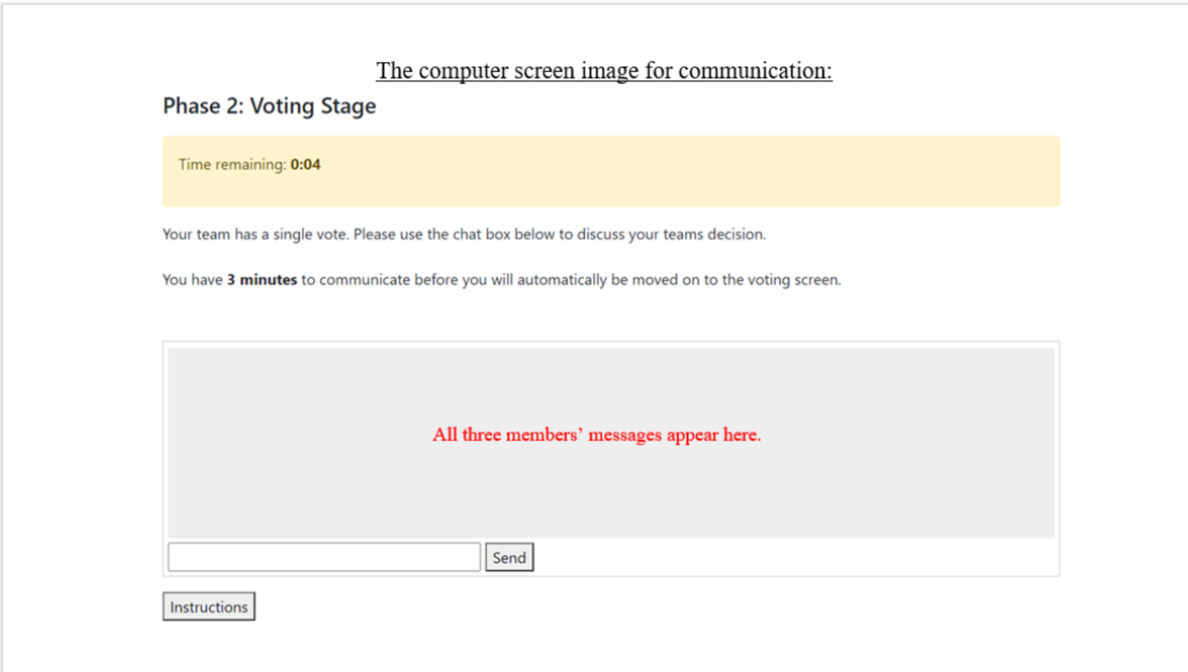
Step 1: Deciding whether to reduce the incentive to switch to the Game screen.

Before moving on to the main task-solving stage of 35 minutes, there is an option to **reduce the per minute earnings from spending time in the Game screen from 18 pence to 16 pence** in today's experiment. Each of the three teams independently vote on whether they wish to implement this reduction option. Each team is given three minutes to communicate using a chat window to decide a team vote. Each team has only one vote.

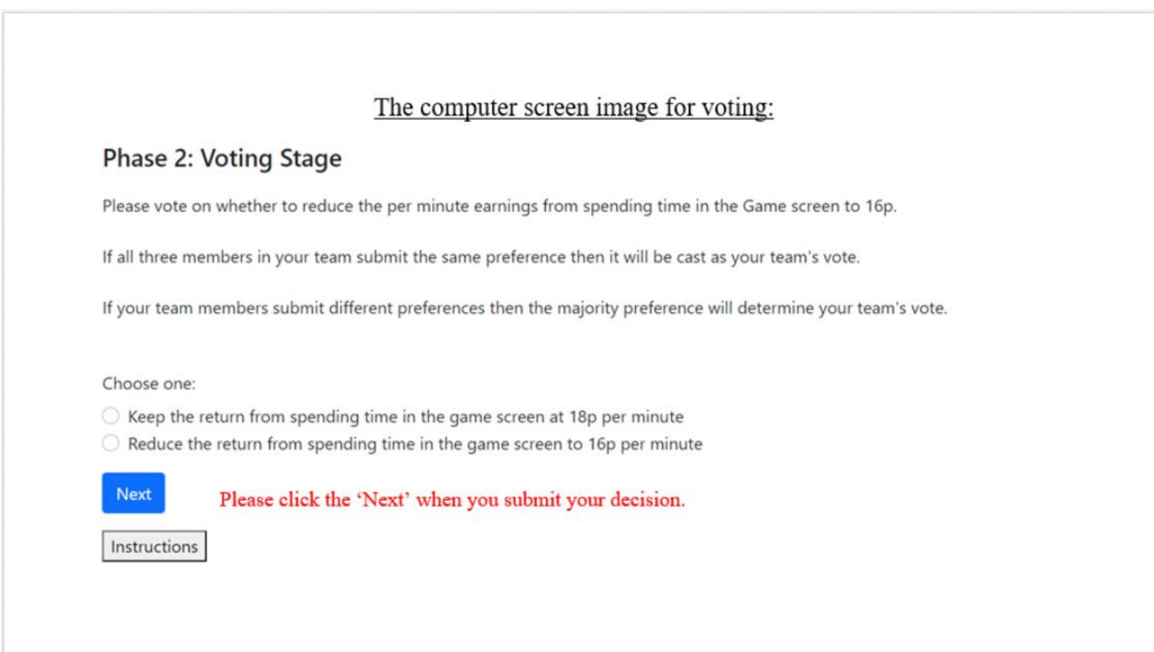
Once the 3 minutes of communication passes, the three members in your team each submit a preferred (agreed) decision. If the three members submit the same voting decision (either implement or not implement), then it becomes your team's joint voting decision. If the three members do not submit the same option, then a majority rule is applied. Whichever receives at least two members' support will be implemented as your team's vote.

After all three teams complete the voting decision, they are informed of the vote outcome. The reduction policy is implemented in today's experiment if **it receives a majority of the votes (2 or 3 votes)**. The reduction policy is not implemented if **it does not receive a majority of the votes**.

Slide 9:



Slide 10:



Slide 11:

Step 2: Main Task-solving Stage

You will undertake the number counting task **for 35 minutes**. You and all other participants each get 20 pence for each question your team answers correctly, and lose 1 penny for each question your team answers incorrectly. You can also get 20 pence for each question answered correctly by another team. Instead of counting numbers, you can switch from the work screen to the Game screen, for example to play Tetris as already discussed. You can earn points by spending time in the Game screen also – the per minute earnings differ dependent on the vote outcome – either 16 or 18 pence per minute. The earnings from the Game screen will not be shared with anyone else in today's sessions.

Are there any questions? If you have any questions, please write them in the Zoom chat box

Appendix B: The Dividend of Democracy in a Theoretical Model

This part of the appendix illustrates how sacrifice helps improve effort provision. It also studies how democracy in decision-making helps improve productivity further. The analysis can be made using a similar framework to the one used in Kamei and Markussen (forthcoming) except changing the variables. In the present paper, a group consists of three teams, and each team consists of three individual members. As the likelihood to answer a collaborative counting task depends on a team's joint effort provision, it is reasonable to assume that the payoff a team receives depends on the team's degree of effort provision $e_i \in [0,1]$. For example, if all three members put their highest effort without any shirking, $e_i = 1$. However, *e* would be considerably smaller if just one member puts in very little effort, as then the number of 4s cannot be answered accurately. On the other hand, *e* may be at an adequate level when all three execute adequate effort with some shirking. Each team's decision can be expressed as below:

$$\max_{e_i \in [0,1]} \{\pi_i(e_i|e_{-i}) = \sum_{n=1}^{3}(s_n + \mu D) \cdot e_n + r_{k\in\{\text{S,NS}\}} \cdot g_i - f(e_i)\}, \text{ where}$$

$$f(e_i) = (c_i - \delta D){e_i}^2 \text{ [cost function]; and}$$

$$e_i + g_i = 1 \text{ [allocation of effort and shirking activities].} \quad \text{(B1)}$$

Here, g_i is team *i*'s average shirking level in the phase 2 task-solving stage. Note that $e_i + g_i = 1$ because there are only two possibilities for simplicity: work or shirk. s_i is the marginal return of effort provision by team *i*, and it is assumed to be a constant (reflecting the number of tasks answered, the likelihood to answer correctly, and the earnings from correct answers and from mistakes), and r_k is marginal return of shirking which depends on their group's sacrifice decisions, i.e., *k* = S (Sacrifice) or NS (Not Sacrifice), and $r_{\text{NS}} > r_{\text{S}}$.

As in Kamei and Markussen (forthcoming), it can be assumed that being involved in democratic decision-making eases workers' effort cost ($\delta > 0$) such that $\delta < c_i$, due to either enhanced intrinsic motivation or signaling. It can also be assumed that democracy boosts team *i*'s work productivity, defined as per effort productivity, from s_i to $s_i + \mu$, where $\mu > 0$. D is an indicator variable which equals 1 in the ENDO treatment.[24]

Team *i*'s optimal effort level can be derived merely using the first-order condition for (B1) as the cost function is quadratic. The optimality condition is summarized in Equation (B2) below.

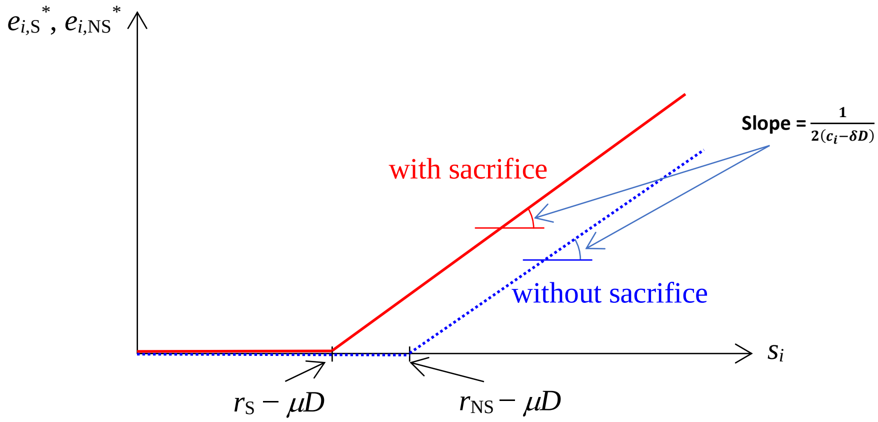
$${e_i}^* = \frac{s_i + \mu - r_k}{2(c_i - \delta D)} \text{ if } s_i > r_k - \mu\text{; and } {e_i}^* = 0 \text{ otherwise, for } k \in \{\text{S, NS}\}. \quad \text{(B2)}$$

It is clear from (B2) and the figure on the next page that sacrifice has a positive impact on teams' effort provision since $r_{\text{NS}} > r_{\text{S}}$:

[24] It can further be assumed that δ and μ depend on the policy outcome such that these parameters are larger when the policy is selected than is not selected: $\delta|_{\text{imposed}} > \delta|_{\text{not imposed}} > 0$ and $\mu|_{\text{imposed}} > \mu|_{\text{not imposed}} > 0$. This Appendix provides the theoretical result when these effects of democratic decision-making do not depend on the outcome for simplicity as the theoretical implication is similar regardless of the assumption.

$$e_{i,\mathrm{S}}{}^* = \frac{s_i+\mu-r_\mathrm{S}}{2(c_i-\delta D)} \geq e_{i,\mathrm{NS}}{}^* = \frac{s_i+\mu-r_\mathrm{NS}}{2(c_i-\delta D)}. \tag{B3}$$

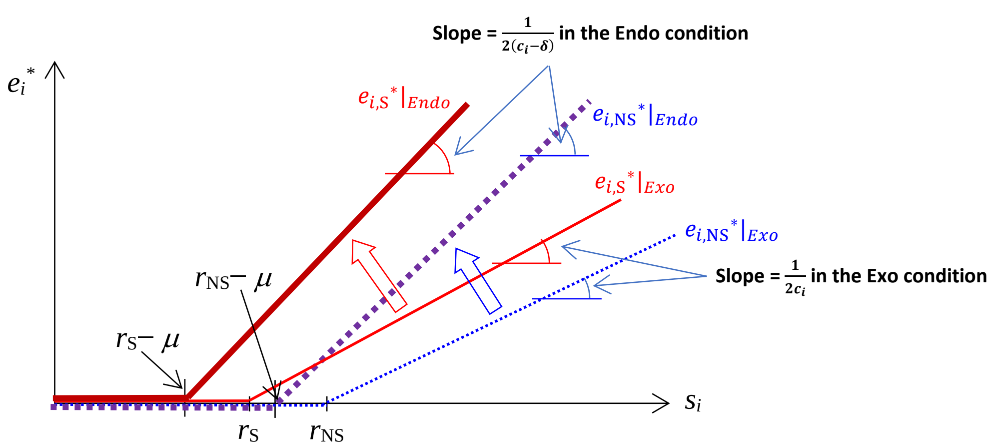
Figure B.1: *The Reduction Policy and Worker's Optimal Effort Schedule*



Further, as $\delta > 0$ and $\mu > 0$, teams in the ENDO treatment work harder than those in the EXO treatment – see the figure below:

$$e_{i,k}{}^*|_{Endo} = \frac{s_i+\mu-r_k}{2(c_i-D)} \geq e_{i,k}{}^*|_{Exo} = \frac{s_i-r_k}{2c_i} \text{ for given } k\in\{\mathrm{S,NS}\}. \tag{B4}$$

Figure B.2: *Effects of Democracy on Worker's Optimal Effort Schedule*



These analyses can be summarized as in Summary 1:

Summary 1: *(a) Teams work harder with than without the reduction policy. (b) Democracy induces the workers to work harder.*

Does the sacrifice benefit teams? This question may not be obvious because a rise in effort provision means not only an increased return from work but also a rise in the effort cost. To make the further analysis simple, assume the homogeneity in skills ($s_i = s$ for all i) and the following condition:

Assumption 1: $s > c_i$.

Assumption 1 means that the unit effort cost is not too large compared to the material return from task-solving. A calculation suggests that under Assumption 1, teams earn more when the reduction policy is imposed in their group than otherwise if $s > r_{k\in\{\mathrm{S,NS}\}}$, irrespective of whether they are in the ENDO or EXO treatment. However, if $s < r_{k\in\{\mathrm{S,NS}\}}$, sacrifice is purely harmful to welfare.

Summary 2: *Suppose Assumption 1 holds. Regardless of whether they are in the ENDO or EXO treatment, if the material benefit from working is high (low) enough that $s > r_{k\in\{\mathrm{S,NS}\}}$ ($s < r_{k\in\{\mathrm{S,NS}\}}$), teams earn more (less) when the reduction policy is imposed than is not imposed.*

Proof: Suppose first that $s > r_{k\in\{\mathrm{S,NS}\}}$, i.e., the situations are characterized by interior solutions. Then, we can show the beneficial effect of sacrifice by simply calculating the difference in the payoff between the two conditions.

$$\pi_i{}^{\mathrm{S}} = \pi_i\left(e_{i,\mathrm{S}}{}^*|e_{i,\mathrm{S}}{}^*\right) = 3s\frac{s+\mu D-r_{\mathrm{S}}}{2(c_i-\delta D)} + r_{\mathrm{S}}\cdot\left(1-\frac{s+\mu D-r_{\mathrm{S}}}{2(c_i-\delta D)}\right) - \frac{(s+\mu D-r_{\mathrm{S}})^2}{4(c_i-\delta D)}. \tag{B5}$$

$$\pi_i{}^{\mathrm{NS}} = \pi_i\left(e_{i,\mathrm{NS}}{}^*|e_{i,\mathrm{NS}}{}^*\right) = 3s\frac{s+\mu D-r_{\mathrm{NS}}}{2(c_i-\delta D)} + r_{\mathrm{NS}}\cdot\left(1-\frac{s+\mu D-r_{\mathrm{NS}}}{2(c_i-\delta D)}\right) - \frac{(s+\mu D-r_{\mathrm{NS}})^2}{4(c_i-\delta D)}. \tag{B6}$$

Then, $\pi_i{}^{\mathrm{S}} - \pi_i{}^{\mathrm{NS}} = 3s\frac{r_{\mathrm{NS}}-r_{\mathrm{S}}}{2(c_i-\delta D)} + r_{\mathrm{S}}\cdot\left(1-\frac{s+\mu D-r_{\mathrm{S}}}{2(c_i-\delta D)}\right) - r_{\mathrm{NS}}\cdot\left(1-\frac{s+\mu D-r_{\mathrm{NS}}}{2(c_i-\delta D)}\right) - \frac{(s+\mu D-r_{\mathrm{S}})^2}{4(c_i-\delta D)} + \frac{(s+\mu D-r_{\mathrm{NS}})^2}{4(c_i-\delta D)}$

$$= \frac{r_{\mathrm{NS}}-r_{\mathrm{S}}}{4(c_i-\delta D)}\{6s-4c_i-(r_{\mathrm{NS}}+r_{\mathrm{S}})+4\delta D\} > 0. \tag{B7}$$

Consider next the case of corner solutions both with and without sacrifice (i.e., $s < r_{k\in\{\mathrm{S,NS}\}}$). In this case, $e_{i,\mathrm{S}}{}^* = e_{i,NS}{}^* = 0$; written differently, $g_{i,\mathrm{S}}{}^* = g_{i,NS}{}^* = 1$. Then, from Equation (B1), $\pi_i{}^{\mathrm{S}} - \pi_i{}^{\mathrm{NS}} < 0$.

The remaining case is the situation with an interior solution under sacrifice but a corner solution without sacrifice ($r_{\mathrm{NS}} > s > r_{\mathrm{S}}$,). Then, $e_{i,\mathrm{S}}{}^* = \frac{s_i-r_{\mathrm{S}}}{2(c_i-\delta D)}$, and $e_{i,NS}{}^* = 0$), and we have the following:

$$\pi_i{}^{\mathrm{S}} = \pi_i\left(e_{i,\mathrm{S}}{}^*|e_{i,\mathrm{S}}{}^*\right) = 3s\frac{s+\mu D-r_{\mathrm{S}}}{2(c_i-\delta D)} + r_{\mathrm{S}}\cdot\left(1-\frac{s+\mu D-r_{\mathrm{S}}}{2(c_i-\delta D)}\right) - \frac{(s+\mu D-r_{\mathrm{S}})^2}{4(c_i-\delta D)}.$$

$$\pi_i{}^{\mathrm{NS}} = r_{\mathrm{NS}}.$$

In this case, it is not obvious which is larger, $\pi_i{}^{\mathrm{S}}$ or $\pi_i{}^{\mathrm{NS}}$.

□

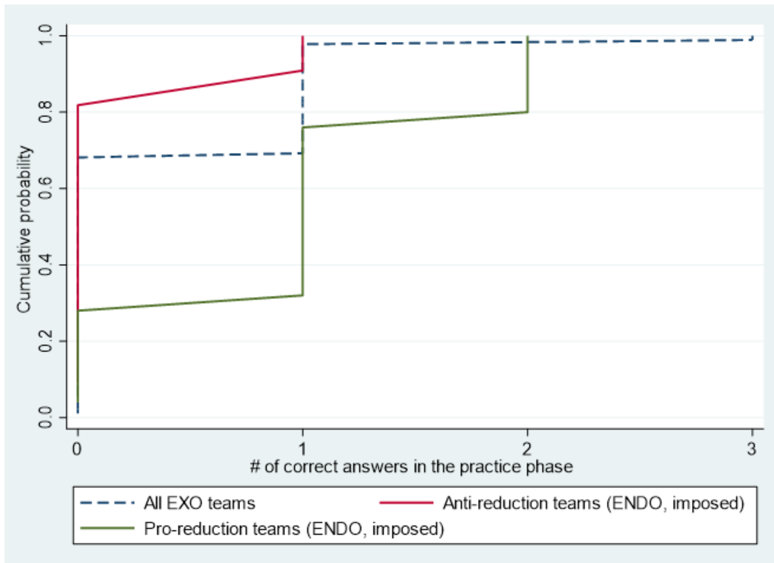
Notice that $\pi_i{}^{\mathrm{S}} - \pi_i{}^{\mathrm{NS}}$ in Equation (B7) is decreasing in c_i. This means that teams who are better at solving the collaborative counting task (i.e., teams with smaller c_i) have larger gains from the reduction policy through its strong positive impact on their effort provision.

It should be emphasized here that the beneficial effects of the reduction policy emerge when teams have sufficiently low effort costs as expressed by Assumption 1. However, introducing the policy is oppositely harmful if they are not skilled and therefore incur large costs from effort provision. This theoretical implication suggests that teams who are skilled at solving the collaborative counting task vote in favor of the reduction policy in phase 2.

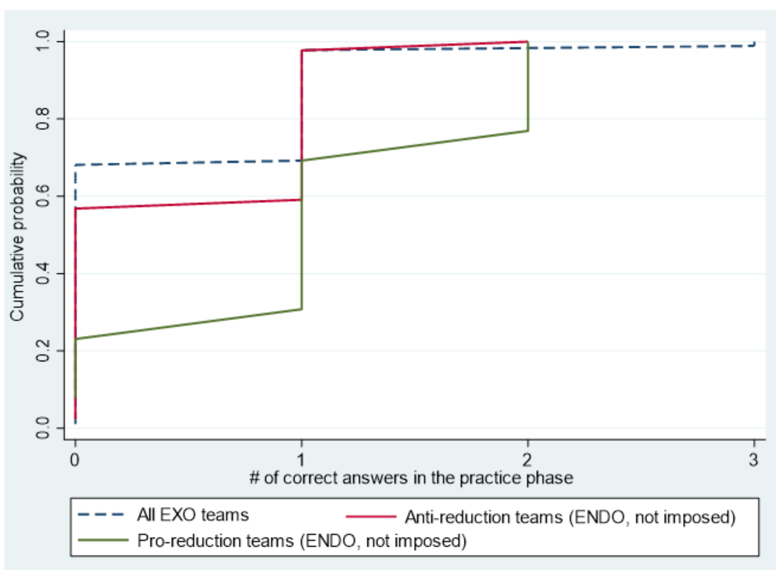
Summary 3: *Those who are better at solving the collaborative counting task in phase 1 are more likely to vote for the reduction policy in phase 2.*

Appendix C: Additional Figure and Tables

Figure C.1. *Cumulative Distribution of Performance by Voting in the Three-minutes Practice Phase*



(A) Pro- and anti-reduction teams in the groups where the policy was endogenously imposed



(B) Pro- and anti-reduction teams in the groups where the policy was not endogenously imposed

Note: The cumulative distribution of teams' performance in the EXO treatment was also drawn as a reference.

Table C.1: *Privately and Socially Optimal Choices between Task-Solving and Gaming*

[1. Which activity is privately optimal, task-solving or gaming?]

A. ENDO treatment

	A1: Under the reduction policy		A2: Without reduction policy	
	a. Task-solving is a privately optimal	b. Gaming is a privately optimal	c. Task-solving is a privately optimal	d. Gaming is a privately optimal
i. # of team	8	28	0	57
ii. Avg # of correct answer per minute	0.97	0.38	---	0.33
iii. Avg work time (min)	33.33	26.42	---	18.69

	A3: All data in the ENDO	
	e. Task-solving is a privately optimal	f. Gaming is a privately optimal
i. # of team	8	85
ii. Avg # of correct answer per minute	0.97	0.35
iii. Avg work time (min)	33.33	21.24

B. EXO treatment

	B1: With reduction policy		B2: Without reduction policy	
	a. Task-solving is a privately optimal	b. Gaming is a privately optimal	c. Task-solving is a privately optimal	d. Gaming is a privately optimal
i. # of team	3	38	1	49
ii. Avg # of correct answer per minute	0.94	0.37	1	0.30
iii. Avg work time (min)	30.74	25.86	35.00	21.00

	B3: All data in the EXO	
	e. Task-solving is a privately optimal	f. Gaming is a privately optimal
i. # of team	4	87
ii. Avg # of correct answer per minute	0.96	0.33
iii. Avg work time (min)	31.80	23.12

[2. Which activity is socially optimal, task-solving or gaming?]

A. ENDO treatment

	A1: Under the reduction policy		A2: Without reduction policy	
	a. Task-solving is a socially optimal	b. Gaming is a socially optimal	c. Task-solving is a socially optimal	d. Gaming is a socially optimal
i. # of team	28	8	29	28
ii. Avg # of correct answer per minute	0.64	0.08	0.58	0.07
iii. Avg work time (min)	32.10	13.43	26.98	10.11

	A3: All data in the ENDO

	e. Task-solving is a socially optimal	f. Gaming is a socially optimal
i. # of team	57	36
ii. Avg # of correct answer per minute	0.61	0.07
iii. Avg work time (min)	29.50	10.85

B. EXO treatment

	B1: With reduction policy		B2: Without reduction policy	
	a. Task-solving is a socially optimal	b. Gaming is a socially optimal	c. Task-solving is a socially optimal	d. Gaming is a socially optimal
i. # of team	32	9	24	26
ii. Avg # of correct answer per minute	0.51	0.07	0.58	0.07
iii. Avg work time (min)	28.50	18.09	28.81	14.33

	B3: All data in the EXO	
	e. Task-solving is a socially optimal	f. Gaming is a socially optimal
i. # of team	56	35
ii. Avg # of correct answer per minute	0.54	0.07
iii. Avg work time (min)	28.63	15.29

Appendix D: Coding Procedure and Analysis Results for the Communication Contents

D.1. Coding Procedure

Two coders were hired to judge each team's communication content for both the 3-minute communication segment prior to voting in the ENDO treatment and the 35-minute communication segments in the main task-solving phase for both the ENDO and EXO treatments, by assigning the relevant codes (summarized in section D.2). The treatments were presented as 'Treatment A' and Treatment B' to the coders, which alternated as coders completed the treatments in different orders. The coders were provided with a copy of the experiment instructions. Each coder was provided with three Excel files, termed "Coding Sheet – Treatment XY," where X indicates either "A" or "B" to designate the treatment and Y indicates the communication length, either 3 or 35 minutes (e.g., Coding Sheet – TreatmentA35). Each file had separate sections for each code type and only the relevant codes for that communication type were available. In the columns, the files contained a list of the team numbers in ascending order, which corroborated with "Segment" numbers found in the "Communication Files."

Six Communication files were also provided, comprising a sample set of ten communication segments and the remaining set of communication segments for each of the three communication combinations. Coders were instructed to first read the entire communication segment of a team and then assign as many codes as deemed appropriate in the Coding Sheet in a given teams column.

The files consisted of data from 93 teams in the ENDO treatment, each having one 3-minute and one 35-minute dialogue segment, and 91 teams from the EXO treatment, with just a 35-minute dialogue segment each, resulting in 277 dialogue segments to be coded. Coding was conducted by treatment and communication type, and further broken into four blocks as detailed below. While the coders were aware that there were two coders, they were kept anonymous from each other for the entire process and so were unable to communicate with each other.

<u>The first block (first nine days):</u>

The coding sheet, experiment instructions, and 10-segment sample communication file for the 3-minute communication in Treatment A (35-minute communication in Treatment B for the other coder) were provided on the first day. A meeting was scheduled for the same day, separately for each coder, to allow one of the researchers to explain the coding process and treatment in more detail. Coders were not made aware of the purpose of the research, subject details, or any of the analysis/results throughout the coding process.

After the sample set had been coded, a researcher met with each coder to discuss any problems or difficulties. This initial practice and feedback process took two days. After that, the researchers sent the communication file with the remaining 83 (81) dialogue segments for that block to be completed over the following seven days.

Once all 93 or 91 dependent on the coder (including the sample set) of the dialogue segments had been coded, the Coding Sheet was returned to the researchers and no further changes could be made (unless there had been some misunderstanding about the coding practice). Feedback was not given to the coders regarding their coding practice.

<u>The second block (next nine days):</u>

Once the first block was completed, the coders were given the coding sheet and 10-segment sample communication for the 35-minute communication in Treatment A (3-minute communication in Treatment A for the other coder, along with the instructions for Treatment A), and a meeting was scheduled for that day, separately for each coder, to go through the instructions and codes. The remaining procedure is the same as in the first block. Two days were given to complete the sample set, after which there was a meeting to clarify any questions. After that, seven days were given to code the remaining 83 dialogue segments. As before, no feedback was given to the coders regarding their coding practice.

The third block (next nine days):

As in the first and second block, the coders were provided with the instructions, coding sheet and 10-segment sample communication for the 35-minute communication in Treatment B (35-minute communication in Treatment A for the other coder), and a meeting was held to discuss the instructions and codes. Two days were given to complete the 10-segment sample set, before a further meeting was held to clarify any questions. The coders were allowed a further seven days to code the remaining 81 (83) dialogue segments. No feedback was given regarding the coders' coding practices.

The fourth block (final seven days):

The coding results were compared for discrepancies between the two coders' coding results. The discrepancies were then highlighted in the Excel spreadsheets and a copy was given to each coder. The coders were given a further seven days to re-evaluate these discrepancies, with the additional knowledge of one another's codes, and to either confirm or alter their initial findings. Each coder was informed that their codes would be sent to the other coder, and that they would simultaneously re-evaluate the discrepancies. Coder identity remained anonymous throughout the process (and also after the coding work) and no communication was permitted.

D.2. Full List of Codes

(a) ENDO 3-minute communication immediately before voting

Code	Description	Interpretation (not shown to coders)
Note: Coders may assign as many codes as appropriate, including assigning no codes, to a dialogue segment.		
Codes related to voting decision		
A1	Agreement among the three team members explicitly to vote for the policy	Consensus
A2	Agreement among the three team members explicitly to vote against the policy	Consensus
A3	The team's majority favored decision changes over the course of the discussion	Learning
A4	The team's majority favored decision does not change over the course of the discussion	Learning
A5	Disagreement on what to vote for at the beginning of communication which is then resolved	Disagreement
A6	One or more teammates decide to cast their own preference (leaving the majority rule to decide the team vote)	Disagreement
A7	There is an unresolved split in opinion about whether to vote for or against the policy due to strong preferences on both sides	Disagreement
A8	Teammates do not reach a consensus by the end of the 3-minute communication period for reasons other than A6 or A7	Lack of time
A9	Discuss how the other two teams may vote	Strategic
A10	Confusion about the voting rule	Confusion

Codes related to deciding what to do during the task-solving phase		
B1	Agree/Imply to count as primary behavior	Cooperative
B2	Agree/Imply to game as primary behavior	Uncooperative/free-riding
B3	Agree to hybrid behavior e.g. so many tasks/minutes before switching to the game screen	Somewhat cooperative
B4	Agree to discuss, decide and/or alter behavior during the counting task later (35-minute phase) based on performance/needs in Phase 2	Flexible strategy
B5	Suggest altering and/or discussing their behavior for the task-solving phase depending on the vote outcome	Rational
B6	Confusion about the rules in phase 2 (e.g., the revenue-sharing rule)	Confusion
Codes related to why they are pro/anti the policy		
C1	Pro-policy to deter others from switching to the game screen by reducing the return (monetary deterrence)	Monetary incentive/punishment
C2	Pro-policy to signal intention to complete the tasks to other teams	Signaling/Information Effects
C3	Pro-policy for a normative reason e.g. it is the right thing to do, it is desirable socially for their group, it is morally good	Normative reasoning
C4	Pro-policy as the policy is perceived as fair, i.e., reduces income inequality among subjects in a given group	Fairness preference
C5	Pro-policy out of spite or enjoyment of punishment	Spitefulness
C6	Pro-policy out of anticipated anger should other teams not complete tasks	Emotive reasoning/anger
C7	Pro-policy for strategic reasons (e.g., induce other teams to complete the task while they themselves do not complete the task)	Strategic
C8	Anti-policy as they intend to game for at least some of the task-solving period	Selfish
C9	Anti-policy as they like unfair distribution of income	Fairness preference (reverse)
C10	Anti-policy as they dislike punishment philosophically (e.g., dislike coercive punishment), and/or do not perceive other teams' gaming as negative	Dislike of punishment on private activity
C11	Anti-policy as they are uncertain about whether they will want to access the game screen	Uncertainty
C12	Express that the policy is not strong enough to deter others switching to the game screen (monetary)	Punishment strength
C13	Express that the policy is unlikely to affect other teams' working behavior	Policy insensitivity
Codes related to why they chose a certain behavior		
D1	Discuss ability to complete X tasks in Y minutes	Skills/Ability
D2	Believe they as a team make the most money from counting	Monetary incentive
D3	Believe they as a team make the most money from gaming	Monetary incentive
D4	Discuss their performance or comfort in Phase 1 (strong/positive)	Incentives/Rationality
D5	Discuss their performance or comfort in Phase 1 (weak/negative)	Incentives/Rationality
D6	Discuss behavior in terms of guaranteed pay (game screen) versus uncertain pay (tasks)	Uncertainty
D7	Discuss use of the game screen as a break	Team behavior: Strategizing
D8	Discuss uncertainty surrounding other teams' work choices or abilities	Uncertainty
D9	Suggest distrust of other teams e.g. expect them to take advantage	Group behavior: Distrust
D10	Discuss expected fatigue from performing the task in Phase 2 based on their experience in Phase 1	
D11	Discuss how total earnings for the whole group are maximized if all 3 teams work on the counting task	Pareto Efficiency
D12	Discuss how their vote affects the vote outcome	Pivotal voting
Codes related to team behavior		
E1	Negativity towards teammates e.g. being critical of others or discouraging	Team behavior: Negative communication

E2	Positivity towards teammates e.g. attempts to encourage others or being supportive	Team behavior: Positive communication
E3	Discusses or suggests task-related behavior e.g. double-counting, waiting for each other, a specific method, etc.	Team behavior: Strategizing
E4	No communication of the entire team	Team behavior: No communication
E5	No communication from just 1 or 2 team members	Team behavior: Limited/poor communication
E6	Chat about topics unrelated to the experiment	Team behavior: team identity
E7	Resolve some confusion about the experiment through communication	Team behavior: Communication

(b) ENDO 35-minute communication during the main task-solving phase

Code	**Description**	**Interpretation (not shown to coders)**
Note: Coders may assign as many codes as appropriate, including assigning no codes, to a dialogue segment.		
Codes related to voting outcome		
F1	Express negative emotions (e.g., upset, anger) about the outcome of the vote	Group identity (negative)
F2	Express positive emotions (e.g., happiness) about the outcome of the vote	Group identity (positive)
F3	Agree/Imply to count as primary behavior	Cooperative
F4	Agree/Imply to game as primary behavior	Uncooperative/free-riding
F5	Agree to hybrid behavior e.g. so many tasks/minutes before switching to the game screen	Somewhat cooperative
F6	Agree to discuss, decide and/or alter behavior during the counting task later (35-minute phase) based on performance/needs in Phase 2	Flexible strategy
F7	Express belief/hope that other teams will complete tasks following the vote	Positive expectations
F8	Express belief that teams will not complete tasks following the vote	Negative expectations
F9	Discuss the distribution of votes and predict how each team may respond to one another	Signaling
F10	Belief on other teams' responses: pro-policy teams will work hard	Belief on voter type
F11	Belief on other teams' responses: anti-policy teams will work hard	Belief on voter type
F12	Belief on other teams' responses: pro-policy teams will work little	Belief on voter type
F13	Belief on other teams' responses: anti-policy teams will work little	Belief on voter type
F14	Discuss some confusion (e.g., they voted based on some misunderstanding of the experiment)	Confusion
F15	Discuss whether to change behavior based on the vote outcome	Conditional cooperation/rational
Codes related to task performance		
G1	Discuss wanting to switch to the game screen some time during the task-solving phase	Fatigue/inability
G2	Discuss difficulty/unpleasantness of task e.g. being slow, tired, bored, etc.	Fatigue/inability
G3	Enact hybrid behavior e.g. set a given number of tasks/minutes before switching to the game screen and back	
G4	Expression of strong negative emotion e.g. frustration, anger, disappointment	Emotive: Negative
G5	Expression of strong positive emotion e.g. enjoyment, things are going well	Emotive: Positive
Codes related to why they chose/are choosing a certain behavior		
D1	Discuss ability to complete X tasks in Y minutes	Rationality

D2	Believe they as a team make the most money from counting	Monetary Incentive
D3	Believe they as a team make the most money from gaming	Monetary Incentive
D4	Discuss their performance or comfort in Phase 1 and/or so far in Phase 2 (strong/positive)	Rationality
D5	Discuss their performance or comfort in Phase 1 and/or so far in Phase 2 (weak/negative)	Rationality
D6	Discuss behavior in terms of guaranteed pay (game screen) versus uncertain pay (tasks)	Uncertainty
D7	Discuss use of the game screen as a break	Team behavior: Strategizing
D8	Discuss uncertainty surrounding other teams' work choices or abilities	Uncertainty
D9	Suggest distrust of other teams e.g. expect them to take advantage	Group behavior: Distrust
D10	Compare their estimated earnings so far from task-solving and forgone earnings from not staying in the Game screen	Rationality
D11	Discuss how total earnings for the whole group are maximized if all 3 teams work on the counting task	Pareto Efficiency
Codes related to team behavior		
H1	Evident mismatch in behavior e.g. one teammate switches to the game screen against others' wishes	Team behavior: miscommunication/selfishness
H2	1 or more players discuss being trapped in the waiting screen	Team behavior: miscommunication/poor planning
H3	Positivity towards teammates e.g. attempts to encourage others or being supportive	Team behavior: Negative communication
H4	Negativity towards teammates e.g. being critical of others or discouraging	Team behavior: Positive communication
H5	Discusses or suggests task-related behavior e.g. double-counting, waiting for each other, a specific method, etc.	Team behavior: Strategizing
H6	Checking whether their teammates refrain from using the Game screen	Team behavior: Monitoring
H7	Communicate about their Tetris score or enjoying playing Tetris	Team behavior: Enjoy shirking
H8	Disagreement on what to do (count or gaming) at the beginning of communication	Team behavior: Disagreement
H9	No communication of the entire team	Team behavior: No communication
H10	No communication from just 1 or 2 team members; or ignore messages from their teammates	Team behavior: Limited/poor communication
H11	A team exclusively communicates numbers (for the counting task) throughout the 35-minute phase	Team behavior
H12	A player/s states that they have not or do not spend any time in the game screen	Strategic/Lying
H13	Chat about topics unrelated to the experiment	Team behavior: team identity
H14	Resolve some confusion about the experiment through communication	Team behavior: Communication

(c) EXO 35-minute communication during the main task-solving phase

Code	**Description**	**Interpretation (<u>not</u> shown to coders)**

Note: Coders may assign as many codes as appropriate, including assigning no codes, to a dialogue segment.		
Codes related to policy outcome		
I1	Express negative emotions (e.g., upset, anger) about the policy outcome	Group identity (negative)
I2	Express positive emotions (e.g., happiness) about the policy outcome	Group identity (positive)
I3	Agree/Imply to count as primary behavior	Cooperative
I4	Agree/Imply to game as primary behavior	Uncooperative/free-riding
I5	Agree to hybrid behavior e.g. so many tasks/minutes before switching to the game screen	Somewhat cooperative
I6	Agree to discuss, decide and/or alter behavior during the counting task later (35-minute phase) based on performance/needs in Phase 2	Flexible strategy
I7	Express belief/hope that other teams will complete tasks following the policy outcome	Positive expectations
I8	Express belief that teams will not complete tasks following the policy outcome	Negative expectations
I9	Discuss whether the method that the computer randomly decides whether the policy is implemented is fair	
I10	Discuss whether the method (computer's random choice) is accurate as described by the instructions (e.g., the computer's choice may not be random)	
I11	Discuss some confusion	
Codes related to task performance		
G1	Discuss wanting to switch to the game screen some time during the task-solving phase	Fatigue/inability
G2	Discuss difficulty/unpleasantness of task e.g. being slow, tired, bored, etc.	Fatigue/inability
G3	Enact hybrid behavior e.g. set a given number of tasks/minutes before switching to the game screen and back	Team behavior: Strategizing
G4	Expression of strong negative emotion e.g. frustration, anger, disappointment	Emotive: Negative
G5	Expression of strong positive emotion e.g. enjoyment, things are going well	Emotive: Positive
Codes related to why they chose/are choosing a certain behavior		
D1	Discuss ability to complete X tasks in Y minutes	Rationality
D2	Believe they as a team make the most money from counting	Rationality
D3	Believe they as a team make the most money from gaming	Rationality
D4	Discuss their performance or comfort in Phase 1 and/or so far in Phase 2 (strong/positive)	Rationality
D5	Discuss their performance or comfort in Phase 1 and/or so far in Phase 2 (weak/negative)	Rationality
D6	Discuss behavior in terms of guaranteed pay (game screen) versus uncertain pay (tasks)	Uncertainty
D7	Discuss use of the game screen as a break	Team behavior: Strategizing
D8	Discuss uncertainty surrounding other teams' work choices or abilities	Uncertainty
D9	Suggest distrust of other teams e.g. expect them to take advantage	Group behavior: Distrust
D10	Compare their estimated earnings so far from task-solving and forgone earnings from not staying in the Game screen	Rationality
D11	Discuss how total earnings for the whole group are maximized if all 3 teams work on the counting task	Pareto Efficiency
Codes related to team behavior		
H1	Evident Mismatch in behavior e.g. one teammate switches to the game screen against others' wishes	Team behavior: miscommunication or selfishness
H2	1 or more players discuss being trapped in the waiting screen	Team behavior: miscommunication/poor planning
H3	Positivity towards teammates e.g. attempts to encourage others or being supportive	Team behavior: Negative communication

H4	Negativity towards teammates e.g. being critical of others or discouraging	Team behavior: Positive communication
H5	Discusses or suggests task-related behavior e.g. double-counting, waiting for each other, a specific method, etc.	Team behavior: Strategizing
H6	Checking whether their teammates refrain from using the Game screen	Team behavior: Monitoring
H7	Communicate about their Tetris score or enjoying playing Tetris	Team behavior: Enjoy shirking
H8	Disagreement on what to do (count or gaming) at the beginning of communication	Team behavior: Disagreement
H9	No communication of the entire team	Team behavior: No communication
H10	No communication from just 1 or 2 team members; or ignore messages from their teammates	Team behavior: Limited/poor communication
H11	A team exclusively communicates numbers (for the counting task) throughout the 35-minute phase	Team behavior
H12	A player/s states that they have not or do not spend any time in the game screen	Strategic/Lying
H13	Chat about topics unrelated to the experiment	Team behavior: team identity
H14	Resolve some confusion about the experiment through communication	Team behavior: Communication

D.3. Agreement rates and Kappas

The average Cohen's Kappas for the initial coding were 0.67, 0.60, and 0.45 for the ENDO 3-minute dialogue segments, ENDO 35-minute dialogue segments, and EXO 35-minute dialogue segments, respectively. The reconsideration step improved the Kappas. After the independent reconsideration process, the Kappas became 0.87, 0.87, and 0.78 for the ENDO 3-minute dialogue segments, ENDO 35-minute dialogue segments, and EXO 35-minute dialogue segments, respectively.

Remark: The overall agreement rates of coding between the two coders after (before) the reconsideration process were 96.3% (90.5%), 97.5% (93.0%), and 94.9% (88.5%) for the ENDO 3-minute dialogue segments, ENDO 35-minute dialogue segments, and EXO 35-minute dialogue segments, respectively.

The following summarizes the agreement rates and the Kappas before and after the reconsideration step for each code:

(a) ENDO 3-Minute Dialogue Segments

[Agreement Rate:]

Agreement rate	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Before reconsideration	96.8%	97.8%	93.5%	80.6%	90.3%	91.4%	92.5%	96.8%	88.2%	93.5%
After reconsideration	98.9%	100.0%	100.0%	95.7%	96.8%	95.7%	95.7%	100.0%	93.5%	100.0%

Agreement rate	B1	B2	B3	B4	B5	B6
Before reconsideration	86.0%	93.5%	89.2%	88.2%	96.8%	90.3%
After reconsideration	95.7%	100.0%	95.7%	98.9%	98.9%	96.8%

Agreement rate	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
Before reconsideration	91.4%	59.1%	97.8%	96.8%	98.9%	96.8%	95.7%	92.5%	100.0%	100.0%	87.1%	94.6%	94.6%
After reconsideration	98.9%	67.7%	97.8%	100.0%	100.0%	100.0%	97.8%	97.8%	100.0%	100.0%	97.8%	97.8%	98.9%

Agreement rate	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
Before reconsideration	94.6%	77.4%	91.4%	96.8%	95.7%	87.1%	89.2%	79.6%	86.0%	91.4%	72.0%	82.8%
After reconsideration	97.8%	90.3%	95.7%	98.9%	96.8%	92.5%	97.8%	91.4%	92.5%	96.8%	81.7%	91.4%

Agreement rate	E1	E2	E3	E4	E5	E6	E7
Before reconsideration	93.5%	65.6%	81.7%	100.0%	97.8%	96.8%	92.5%
After reconsideration	97.8%	86.0%	100.0%	100.0%	98.9%	98.9%	98.9%

[Cohen's Kappa:]

Kappa	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Before reconsideration	0.93	0.96	-0.03	0.00	0.27	0.29	0.49	0.39	0.36	-0.02
After reconsideration	0.98	1.00	1.00	0.64	0.82	0.58	0.73	1.00	0.69	1.00

Kappa	B1	B2	B3	B4	B5	B6
Before reconsideration	0.69	0.80	0.75	0.61	0.65	0.00
After reconsideration	0.90	1.00	0.90	0.97	0.90	0.00

Kappa	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
Before reconsideration	0.83	0.00	0.49	-0.01	0.66	-0.01	0.58	0.85	n.a.	1.00	0.28	0.42	0.00
After reconsideration	0.98	0.22	0.49	1.00	1.00	1.00	0.82	0.96	n.a.	1.00	0.91	0.79	0.85

Kappa	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
Before reconsideration	0.85	0.55	0.65	0.82	0.86	0.63	0.58	0.53	0.31	0.51	0.31	0.04
After reconsideration	0.94	0.81	0.82	0.94	0.90	0.80	0.92	0.81	0.59	0.85	0.57	0.59

Kappa	E1	E2	E3	E4	E5	E6	E7
Before reconsideration	0.37	0.22	0.59	n.a.	0.49	0.71	0.43
After reconsideration	0.74	0.57	1.00	n.a.	0.66	0.92	0.90

(b) ENDO 35-Minute Dialogue Segments

[Agreement Rate:]

Agreement rate	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Before reconsideration	94.6%	97.8%	87.1%	80.6%	86.0%	87.1%	97.8%	97.8%	97.8%	97.8%	100.0%	97.8%	97.8%	97.8%	96.8%
After reconsideration	98.9%	100.0%	81.7%	96.8%	95.7%	93.5%	97.8%	98.9%	100.0%	98.9%	100.0%	100.0%	97.8%	98.9%	100.0%

Agreement rate	G1	G2	G3	G4	G5
Before reconsideration	88.2%	92.5%	83.9%	87.1%	88.2%
After reconsideration	100.0%	97.8%	100.0%	92.5%	96.8%

Agreement rate	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
Before reconsideration	98.9%	96.8%	96.8%	92.5%	92.5%	97.8%	86.0%	96.8%	96.8%	98.9%	94.6%
After reconsideration	98.9%	100.0%	100.0%	97.8%	97.8%	100.0%	96.8%	100.0%	100.0%	100.0%	97.8%

Agreement rate	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14
Before reconsideration	93.5%	93.5%	77.4%	83.9%	89.2%	95.7%	97.8%	89.2%	100.0%	84.9%	88.2%	97.8%	96.8%	94.6%
After reconsideration	97.8%	97.8%	91.4%	92.5%	95.7%	98.9%	98.9%	94.6%	100.0%	95.7%	93.5%	98.9%	97.8%	98.9%

[Cohen’s Kappa:]

Kappa	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Before reconsideration	0.43	0.49	0.67	0.21	0.24	-0.07	0.66	0.74	0.00	0.49	n.a.	0.00	0.49	0.00	0.39
After reconsideration	0.88	1.00	0.60	0.88	0.81	0.54	0.66	0.88	1.00	0.66	n.a.	n.a.	0.74	0.00	1.00

Kappa	G1	G2	G3	G4	G5
Before reconsideration	0.76	0.62	0.57	0.43	0.12
After reconsideration	1.00	0.90	1.00	0.71	0.75

Kappa	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
Before reconsideration	0.66	0.56	0.00	0.00	-0.02	0.66	0.40	0.65	-0.01	0.79	-0.02
After reconsideration	0.66	1.00	1.00	0.82	0.79	1.00	0.86	1.00	1.00	1.00	0.74

Kappa	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14
Before reconsideration	0.80	0.64	0.28	0.15	0.62	0.48	0.00	0.40	1.00	0.54	0.66	0.49	0.71	0.59
After reconsideration	0.93	0.88	0.75	0.59	0.86	0.88	0.66	0.75	1.00	0.88	0.81	0.79	0.82	0.92

(c) EXO 35-Minute Dialogue Segments

[Agreement Rate:]

Agreement rate	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11
Before reconsideration	100.0%	100.0%	5.4%	72.0%	87.1%	90.3%	95.7%	100.0%	100.0%	100.0%	92.5%
After reconsideration	100.0%	100.0%	41.9%	84.9%	97.8%	97.8%	96.8%	100.0%	100.0%	100.0%	98.9%

Agreement rate	G1	G2	G3	G4	G5
Before reconsideration	89.2%	92.5%	72.0%	82.8%	89.2%
After reconsideration	98.9%	97.8%	81.7%	93.5%	95.7%

Agreement rate	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
Before reconsideration	92.5%	88.2%	89.2%	93.5%	92.5%	97.8%	86.0%	98.9%	98.9%	91.4%	96.8%
After reconsideration	96.8%	95.7%	96.8%	94.6%	96.8%	98.9%	94.6%	98.9%	98.9%	97.8%	97.8%

Agreement rate	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14
Before reconsideration	92.5%	89.2%	76.3%	83.9%	68.8%	81.7%	97.8%	86.0%	100.0%	89.2%	89.2%	97.8%	95.7%	84.9%
After reconsideration	97.8%	94.6%	89.2%	92.5%	93.5%	88.2%	97.8%	95.7%	100.0%	98.9%	96.8%	98.9%	98.9%	94.6%

[Cohen’s Kappa:]

Kappa	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11
Before reconsideration	1.00	n.a.	0.00	0.05	0.13	0.00	0.00	n.a.	n.a.	n.a.	0.00
After reconsideration	1.00	n.a.	0.01	0.48	0.90	0.86	0.39	n.a.	n.a.	n.a.	0.90

Kappa	G1	G2	G3	G4	G5
Before reconsideration	0.79	0.70	0.41	0.34	0.12
After reconsideration	0.98	0.92	0.63	0.75	0.64

Kappa	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
Before reconsideration	0.50	0.59	0.44	0.22	0.49	0.88	0.54	0.88	0.00	-0.03	0.56
After reconsideration	0.75	0.83	0.81	0.27	0.78	0.94	0.78	0.88	0.00	0.66	0.74

Kappa	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14
Before reconsideration	0.82	0.39	0.41	0.10	0.28	0.07	0.49	0.47	n.a.	0.71	0.71	0.49	0.48	0.39
After reconsideration	0.95	0.68	0.73	0.55	0.87	0.51	0.49	0.87	n.a.	0.97	0.91	0.66	0.85	0.75

Note: Cohen's Kappa cannot be calculated where a code that is not used by either coder; these are marked with "n.a."

D.4. Regression Analysis

This section reports a regression analysis to explore subjects' reasoning behind their voting decisions regarding the reduction policy and their task-solving behavior, utilizing the classified codes (see Section D.2 for the full list of codes). Following the convention in the experimental literature on team decision-making, the codes with Kappa values greater than 0.4 were used in each model.

As listed in Section D.2, five kinds of coding categories (Code As, Bs, Cs, Ds and Es) were used to classify the subjects' reasoning behind voting (the ENDO 3 Minutes Dialogue). Four kinds of coding categories (Code Fs, Gs, Ds, and Hs for the ENDO treatment; Code Is, Gs, Ds, and Hs for the EXO treatment) were used to classify their reasoning in task-solving. As each coding category classifies the same behavior just from a different angle, having all codes altogether in a regression leads to serious collinearity. The codes of one coding category are therefore included as independent variables in each model in the following analyses. Note that as shown in Section D.3, the Kappa values of almost all classified codes are more than 0.4.

(a) Voting whether to implement the reduction policy (An analysis for the ENDO 3-Minute Dialogue Segments)

The following table reports results when using Code Bs, Cs, Ds, and Es as independent variables in Models 1, 2, 3 and 4, respectively. Code As (Codes related to the voting decision), such as "A1: Agreement among the three team members explicitly to vote for the policy," have almost the same information as the teams' voting decisions. Unsurprisingly, collinearity is strong and no meaningful results are obtained when using Code As as independent variables. The results are omitted to conserve space.

Dependent variable: A dummy which equals 1 if team *i* voted for (against) the reduction policy

Model 1: Using codes related to deciding what to do during task-solving as independent variables	**Model 2**: Using codes related to why they are pro/anti the policy as independent variables	**Model 3**: Using codes related to why they chose a certain behavior as independent variables	**Model 4**: Using codes related to team behavior as independent variables

Independent variable	Coefficient estimates	Independent variable	Coefficient estimates	Independent variable	Coefficient estimates	Independent variable	Coefficient estimates
Code B1	0.31*** (0.11)	Code C1	0.57*** (0.10)	Code D1	0.07 (0.11)	Code E1	-0.21 (0.28)
Code B2	-0.26* (0.13)	Code C3	0.06 (0.05)	Code D2	0.41*** (0.11)	Code E2	0.30* (0.15)
Code B3	-0.27** (0.11)	Code C4	0.06 (0.05)	Code D3	0.04 (0.13)	Code E3	-0.11 (0.11)
Code B4	-0.27** (0.12)	Code C5	-0.21 (0.16)	Code D4	0.28 (0.17)	Code E5	-0.31*** (0.08)
Code B5	0.02 (0.11)	Code C6	0.26 (0.17)	Code D5	-0.26** (0.10)	Code E6	0.19 (0.24)
Constant	0.40*** (0.12)	Code C7	0.13 (0.11)	Code D6	-0.21 (0.13)	Code E7	-0.18 (0.21)
# of obs.	93	Code C8	-0.41*** (0.09)	Code D7	-0.16 (0.15)	Constant	0.42*** (0.08)
R-squared	0.34	Code C10	-0.10 (0.13)	Code D8	-0.03 (0.12)	# of obs.	93
		Code C11	-0.27** (0.13)	Code D9	0.19 (0.19)	R-squared	0.07
		Code C12	-0.30 (0.23)	Code D10	0.14 (0.12)		
		Code C13	0.05 (0.14)	Code D11	0.05 (0.14)		
		Constant	0.37*** (0.09)	Code D12	0.19 (0.17)		
		# of obs.	93	Constant	0.22*** (0.08)		
		R-squared	0.72	# of obs.	93		
				R-squared	0.40		

Notes: Linear regressions with robust standard errors. The numbers in parentheses are standard errors. Codes whose Kappa values are equal to or above 0.4 are used as independent variables. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

(b) Task-solving (An analysis for the ENDO/EXO 35-Minute Dialogue Segments)

Subjects' behaviors in the main task-solving phase can be characterized as (i) their work time (minutes), i.e., the duration in which they work on counting, rather than staying in the Game screen, and (ii) work productivity, i.e., the number of correct answers per minute of work time. The regression analysis below reports two versions for each coding category: one with the work time as the dependent variable, and the other with the work productivity as the dependent variable.

Subsections b1, b2, b3, and b4 below each include regression results of the ENDO treatment (Models 1 and 2) and of the EXO treatment (Models 3 and 4), side by side, to make comparison easier.

b1: Codes related to voting outcome (Code Fs) or policy outcome (Code Is))

ENDO (Code Fs)[#1]				EXO (Code Is)			
Dependent Var.: Work time (minutes)		Dependent Var.: Work productivity		Dependent Var.: Work time (minutes)		Dependent Var.: Work productivity	
Model 1		Model 2		Model 3		Model 4	
Independent variable	Coefficient estimates	Independent variable	Coefficient estimates	Independent variable	Coefficient estimates	Independent variable	Coefficient estimates
Code F1	-14.77*** (1.71)	Code F1	-0.11 (0.17)	Code I1	-8.94*** (2.98)	Code I1	-0.24*** (0.07)
Code F2	-3.01 (6.47)	Code F2	-0.03 (0.21)	Code I4	-21.71*** (2.60)	Code I4	-0.41*** (0.05)
Code F3	11.27***	Code F3	0.24***	Code I5	-9.28***	Code I5	-0.19**

Independent variable	Coefficient estimates	Independent variable	Coefficient estimates	Independent variable	Coefficient estimates	Independent variable	Coefficient estimates
	(2.70)		(0.07)		(3.09)		(0.08)
Code F4	-16.15***	Code F4	-0.25***	Code I6	-7.47*	Code I6	-0.11
	(2.90)		(0.09)		(3.92)		(0.07)
Code F5	-3.98**	Code F5	-0.09*	Code I11	2.08	Code I11	0.07
	(1.85)		(0.05)		(1.85)		(0.06)
Code F6	-3.93	Code F6	-0.28*	Cons.	27.33***	Cons.	0.43***
	(6.04)		(0.15)		(0.98)		(0.04)
Code F7	6.83***	Code F7	0.24**				
	(0.53)		(0.12)				
Code F8	5.25***	Code F8	0.28***				
	(1.03)		(0.04)				
Code F9	13.64***	Code F9	0.17				
	(0.48)		(0.10)				
Code F13	2.46	Code F13	-0.41***				
	(2.99)		(0.15)				
Code F15	4.80***	Code F15	0.07				
	(1.15)		(0.17)				
Cons.	18.47***	Cons.	0.32***				
	(2.73)		(0.06)				
# of obs.	93	# of obs.	93				
R-squared	0.46	R-squared	0.31				

b2: Codes related to task performance (Code Gs)

ENDO (Code Gs)				EXO (Code Gs)			
Dependent Var.: Work time (minutes)		Dependent Var.: Work productivity		Dependent Var.: Work time (minutes)		Dependent Var.: Work productivity	
Model 1		Model 2		Model 3		Model 4	
Independent variable	Coefficient estimates	Independent variable	Coefficient estimates	Independent variable	Coefficient estimates	Independent variable	Coefficient estimates
Code G1	-1.80	Code G1	-0.11	Code G1	-11.08***	Code G1	-0.16**
	(3.43)		(0.08)		(2.64)		(0.07)
Code G2	5.32*	Code G2	-0.01	Code G2	5.93**	Code G2	0.16***
	(3.01)		(0.07)		(2.41)		(0.06)
Code G3	2.13	Code G3	0.06	Code G3	6.66**	Code G3	0.14**
	(3.07)		(0.07)		(2.67)		(0.07)
Code G4	-1.64	Code G4	-0.16	Code G4	-3.86	Code G4	-0.22***
	(3.05)		(0.08)		(2.91)		(0.06)
Code G5	5.60	Code G5	0.03	Code G5	7.66***	Code G5	0.30**
	(5.62)		(0.11)		(2.63)		(0.13)
Cons.	21.86***	Cons.	0.46***	Cons.	26.76***	Cons.	0.39***
	(2.42)		(0.06)		(1.43)		(0.04)
# of obs.	93	# of obs.	93	# of obs.	91	# of obs.	91
R-squared	0.03	R-squared	0.06	R-squared	0.32	R-squared	0.29

b3: Codes related to why they chose/are choosing a certain behavior (Code Ds)

ENDO (Code Ds)				EXO (Code Ds)			
Dependent Var.: Work time (minutes)		Dependent Var.: Work productivity		Dependent Var.: Work time (minutes)		Dependent Var.: Work productivity	
Model 1		Model 2		Model 3		Model 4	
Independent variable	Coefficient estimates	Independent variable	Coefficient estimates	Independent variable	Coefficient estimates	Independent variable	Coefficient estimates
Code D1	19.24	Code D1	0.64**	Code D1	4.44*	Code D1	0.09
	(11.86)		(0.31)		(2.32)		(0.10)
Code D2	4.13	Code D2	-0.28***	Code D2	9.16***	Code D2	0.23***
	(2.63)		(0.06)		(2.07)		(0.08)
Code D3	-20.20***	Code D3	-0.42***	Code D3	-8.45**	Code D3	-0.11

	(1.69)		(0.04)		(3.72)		(0.09)
Code D4	7.95*** (2.42)	Code D4	0.27*** (0.06)	Code D5	-11.07** (4.71)	Code D5	-0.13 (0.10)
Code D5	-18.53*** (4.44)	Code D5	-0.43*** (0.10)	Code D6	-5.78 (4.24)	Code D6	-0.12 (-0.12)
Code D6	3.98** (1.69)	Code D6	0.07 (0.04)	Code D7	-0.97 (2.24)	Code D7	0.06 (0.07)
Code D7	3.13 (2.18)	Code D7	-0.01 (0.08)	Code D8	-14.89*** (4.36)	Code D8	-0.23** (0.09)
Code D8	0.59 (7.80)	Code D8	0.03 (0.21)	Code D10	-1.72 (2.74)	Code D10	-0.22** (0.09)
Code D9	-19.21*** (4.76)	Code D9	-0.37*** (0.12)	Code D11	9.25 (6.54)	Code D11	0.15 (0.17)
Code D10	11.06 (11.34)	Code D10	0.19 (0.29)	Cons	24.51*** (1.28)	Cons	0.36*** (0.04)
Code D11	-7.74 (5.09)	Code D11	-0.35*** (0.12)	# of obs.	91	# of obs.	91
Cons	22.42*** (1.69)	Cons	0.42*** (0.04)	R-squared	0.29	R-squared	0.19
# of obs.	93	# of obs.	93				
R-squared	0.14	R-squared	0.11				

b4: Codes related to team behavior (Code Hs)

ENDO (Code Hs)				EXO (Code Hs)			
Dependent Var.: Work time (minutes)		Dependent Var.: Work productivity		Dependent Var.: Work time (minutes)		Dependent Var.: Work productivity	
Model 1		Model 2		Model 3		Model 4	
Independent variable	Coefficient estimates	Independent variable	Coefficient estimates	Independent variable	Coefficient estimates	Independent variable	Coefficient estimates
Code H1	-5.16 (5.89)	Code H1	-0.15 (0.10)	Code H1	-1.87 (5.98)	Code H1	-0.20** (0.08)
Code H2	10.79** (4.70)	Code H2	0.17* (0.10)	Code H2	5.63 (6.58)	Code H2	0.00 (0.07)
Code H3	3.01 (3.38)	Code H3	0.02 (0.09)	Code H3	2.62 (2.09)	Code H3	0.03 (0.07)
Code H4	0.26 (6.48)	Code H4	-0.02 (0.13)	Code H4	-2.12 (4.41)	Code H4	-0.12 (0.08)
Code H5	6.70** (2.72)	Code H5	0.04 (0.09)	Code H5	3.64* (2.03)	Code H5	0.09 (0.06)
Code H6	3.90 (5.88)	Code H6	0.25** (0.12)	Code H6	2.76 (2.50)	Code H6	0.04 (0.10)
Code H7	-5.09 (7.91)	Code H7	-0.30* (0.16)	Code H7	-2.57 (1.60)	Code H7	-0.06 (0.07)
Code H8	-7.07 (4.98)	Code H8	-0.13 (0.09)	Code H8	-5.23 (3.32)	Code H8	-0.09 (0.06)
Code H9	-22.78*** (1.91)	Code H9	-0.46*** (0.05)	Code H10	-8.60* (4.80)	Code H10	-0.13* (0.08)
Code H10	-7.97 (4.91)	Code H10	-0.14 (0.09)	Code H11	8.17*** (2.62)	Code H11	0.07 (0.07)
Code H11	10.19*** (1.95)	Code H11	0.26*** (0.08)	Code H12	4.67*** (1.62)	Code H12	-0.17*** (0.06)
Code H12	10.95*** (1.99)	Code H12	0.30*** (0.05)	Code H13	-2.85 (3.44)	Code H13	-0.06 (0.08)
Code H13	2.07 (5.42)	Code H13	-0.27** (0.10)	Code H14	0.13 (2.36)	Code H14	-0.10 (0.08)
Code H14	-3.98 (3.08)	Code H14	-0.17* (0.10)	Cons.	22.97*** (2.56)	Cons.	0.43*** (0.06)
Cons.	23.27*** (1.90)	Cons.	0.46*** (0.05)	# of obs.	91	# of obs.	91

# of obs.	93	# of obs.	93	R-squared	0.57	R-squared	0.50
R-squared	0.68	R-squared	0.56				

Notes: Linear regressions with robust standard errors. The numbers in parentheses are standard errors. Work productivity is calculated as the number of correct answers divided by the duration to stay on the work site (minutes). Codes whose Kappa values are equal to or above 0.4 are used as independent variables. [#1] Code F10 was omitted despite the Kappa values being above 0.4, because only one chat dialogue was categorized as this code (making estimating its coefficient estimate impossible). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.