Team Incentives and Lower Ability Workers: A Real-Effort Experiment

March 29, 2024

Richard B. Freeman Xiaofei Pan Xiaolan Yang Maoliang Ye

Richard B. Freeman, Harvard University and NBER. <freeman@nber.org>

Xiaofei Pan, Department of Economics, Bryant University. <xpan@bryant.edu>

Xiaolan Yang, Shanghai International Studies University. <yangxiaolan@shisu.edu.cn>

Maoliang Ye, Corresponding authors, School of Business, Southern University of Science and Technology, Shenzhen, China. maoliang.ye@post.harvard.edu

Acknowledgements: All authors contribute equally. This research was supported by the National Natural Science Foundation of China (Grant No. 72073057, 72373061, and 72173088), the Guangdong Natural Science Foundation (Grant No. 2022A1515011965), faculty research grants from Bryant University, and from the NBER Sloan Foundation Grant G-2005-6-21. We thank participants of seminars at Brown University, University of Massachusetts-Amherst, University of Massachusetts-Lowell, Tsinghua University, Chinese University of Hong Kong (Shenzhen), Zhejiang University, Shanghai University of Economics and Finance, Harbin Institute of Technology Shenzhen, Xiamen University, Sun Yat-sen University, Central University of Finance and Economics, Shandong University, Shenzhen University, Beijing Normal University (Zhuhai), as well as Xiamen University Experimental Economics International Workshop, International Symposium on Contemporary Labor Economics, China Workshop on Organizational Economics, WEAI meeting, and ESA meetings for their helpful comments. Ziyuan Cao provided excellent research assistance.

Team Incentives and Lower Ability Workers: A Real-Effort Experiment

Abstract: Despite a large literature on team incentives, studies on teams in a purely financial sense are limited. In such environments, team members independently engage in tasks with identifiable individual contributions, while their compensation is partially linked to team outputs. We conducted an experiment of such scenario with three distribution schemes (equal sharing, individual piece rate, and winner-takes-all) and examined these schemes both with and without a team threshold. Our results showed the surprising power of equal sharing on improving team productivity compared to winner-takes-all and individual piece-rate, contradicting the predictions of the standard economics theory. Our findings reveal that the higher team output observed under equal sharing was driven by the increased productivity of less able workers. This could be attributed to an explanation of guilt aversion and cannot be illuminated by several alternative theories. We also found that participants preferred piece rate over the other schemes. Yet, the presence of a team threshold highlighted the importance of cooperation, leading to a greater preference for equal sharing. Our findings suggest that organizations with workers of varying abilities are likely to benefit from an appropriate equal sharing component tailored to their responsiveness to sharing in rewards.

Keywords: productivity, team incentive, equal sharing, threshold, guilt aversion

1. Introduction

An increasing proportion of firms organize production around teams (Porter and Beyerlein, 2000; Lazear and Shaw, 2007; Irlenbusch and Ruchala, 2008; Pierce et al., 2021).[1] Despite a large literature on team incentives, the exploration of teams in a purely financial sense is relatively limited. In such context, team members independently engage in tasks that yield identifiable individual contributions, though their compensation is linked to the other team members' output. Various industries have adopted team-based pay structures organized in a relatively independent production environment. Examples include salespeople at retail booth (Li et al., 2019), fruit pickers (Bandiera et al., 2013) and garment factory workers (Hamilton et al., 2003). Yet workers in these environments can often communicate and learn from each other and have repeated interactions, thus factors beyond these team-based pay can play a major role on team productivity (Bandiera et al., 2010).

We study an independent individual production environment that establishes the team incentive through two financial arrangements: Firstly, workers receive compensation based on both their individual output and that of their team members. Secondly, team bonuses in one condition are contingent upon whether the team's overall output meets or surpasses predetermined thresholds.[2] Therefore, this study investigates both within-team distribution schemes and team threshold in an independent production environment. Specifically, we investigate the dynamics of a two-person team whereas members receive real-time feedback on each other's performance. Such a scenario can potentially strengthen the sense of a team.

The first objective of our study is to examine the effect of the commonly employed within-team reward-sharing distribution on worker productivity. Conventional assumption of self-interest implies such reward sharing would incentivize workers to engage in free riding behaviors (see, e.g., Hardin, 1968). However, lower ability workers who hold guilt aversion preference may exhibit a different response (Charness and Dufwenberg, 2006; Dana et al., 2006; Battigalli and

---

[1] Lazear and Shaw (2007) report that the share of large firms with workers in self-managed teams rose from 27% in 1987 to 78% in 1996.

[2] Lazear and Shaw (2007) report an increasing percentage of firms that pay workers with gainsharing incentives (from 26% in 1987 to 53% in 1999), whereas workers receive additional pay for improved performance. Similar studies employing a team threshold can be found as follows. In the team treatment of Fryer et al. (2012), the average performance of team members was utilized as the threshold to determine whether they had to pay back the incentive payment. Babcock et al. (2015) examined a more stringent variation of team incentive, whereby the bonus depended on whether both members within a team met the threshold. Refer to Ye et al. (2020) for a recent survey on team cooperation/coordination mechanisms in minimum games.

Dufwenberg 2007). Rather than engaging in free riding behavior, these individuals may be motivated to work harder in order to catch up with their higher ability partners. Specifically, when the lower ability worker shares the team rewards with higher ability workers, the guilt from legging behind their more skilled partner may drive them to put in extra effort to avoid experiencing such feelings of guilt (Gill and Stone, 2015). Furthermore, the independent production environment along with identifiable individual contributions may reinforce the social norm that each worker should be compensated in accordance with their individual contributions. This can further intensify the feelings of guilt among lower-ability individuals if they lag behind their more skillful partners. Additionally, the provision of real-time feedback on their own and their teammate's performance can create a continuous pressure for lower abilities to exert greater effort (Chen and Lim, 2013). Consequently, contrary to the traditional wisdom, reward sharing within teams has the potential to outperform other distribution schemes.

The second aspect of our study examines how a team output threshold impacts their outputs. The concept of using a threshold, or a target to promote cooperation can be attributed to Holmstrom (1982), who introduced the idea of a forcing contract mechanism. According to this mechanism, workers would only share the revenue generated when target revenue is achieved; otherwise, the worker is paid a low penalty wage. Indeed, both empirical (Knez and Simester, 2001) and experimental studies (Nalbantian and Schotter, 1997; He and Miao, 2022) have explored the use of team output threshold to incentivize individuals in a group to perform better.

The third objective of our study is to uncover workers' preferences regarding compensation rules by examining communication and voting behaviors, an area that has not been extensively explored in the literature (Charness and Kuhn, 2011; Wood et al., 2023). We also investigate how team threshold modulates workers' preferences by comparing conditions with and without team threshold.

We conducted a real effort experiment to address the three objectives. In line with our motivation, individuals in our experiment performed *the slider task* independently. This task requires an individual to move sliders into the exact middle position on a computer screen during a specified period. Success in the task demands ability and effort but not special knowledge or prowess (Gill and Prowse, 2012).

Figure 1 shows the structure of the experiment. The experiment began with a *benchmark stage* in which individuals performed the slider task for piece rate pay. We take their performance at this stage as an indicator of individual ability in the task. We then formed teams by pairing two randomly selected participants and make individual earnings depend on team as well as individual output. We identify the person with the lower (or higher) benchmark output within each team as the less (or more) able person in that team.

The experiment then proceeds in three stages. In Stage 1 of our experiment, participants performed the slider task again and were paid based on the assigned distribution scheme and the threshold condition. Specifically, we randomly assigned one of three distribution schemes to each team: *equal sharing*, *piece-rate*, and a tournament style *winner-takes-all*. We further randomly assigned the teams to either the productivity Threshold condition or a No threshold condition. In the Threshold condition, team members would only be compensated if team output reached or exceeded a specified threshold. In the No Threshold condition, individuals were compensated based on the given system without any specified threshold. In Stage 2, team members engaged in a chat session where they discussed whether to continue with their randomly assigned distribution scheme or to switch to a different scheme. In Stage 3, the teams proceeded to work on the slider task under the distribution scheme they chose during Stage 2.

We find that:

(1) Teams randomly assigned to the *equal sharing* scheme significantly outperformed teams randomly assigned to *winner-takes-all* by a considerable margin. Moreover, teams assigned to the *equal sharing* scheme performed at least as good as teams randomly assigned to the *piece-rate* scheme, which contradicts the standard economics theory and previous results. For example, Nalbantian and Schotter (1997) found a lower team output under equal sharing than other incentives.

(2) The better performance observed in the equal sharing scheme can be primarily attributed to lower ability individuals, as they exhibited higher productivity compared to those in *piece-rate* or *winner-takes-all* schemes. Importantly, this enhanced performance under *equal sharing* cannot be explained by the drop out of the lower ability workers in the *winner-takes-all* scheme (Dechenaux et al., 2015), peer pressure (Mas and Moretti 2009), reciprocity (e.g., Charness and Rabin 2002) or inequity aversion (e.g. Fehr and Schmidt 1999). Instead, we show that the higher effort exerted by lower ability workers in the equal sharing scheme can be explained by their guilt aversion toward team-mates (Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007).

(3) We observe no significant overall effect of implementing a group threshold on group

productivity. However, our findings indicate that participants exhibited diverse responses to the group threshold. Specifically, lower ability workers with personal target greater than half of the group threshold demonstrated higher output levels.

(4) When provided with the opportunity to change the distribution scheme, almost all participants abandoned the least productive *winner-takes-all* scheme. Instead, the majority of participants selected *piece-rate* compensation, even if it was not the most productive scheme for teams.

(5) Imposing a team threshold increased discussion regarding cooperation through online chat messages, which subsequently shifted participants' preferences toward *equal sharing*.

The present study has two key contributions to literature. Our first contribution demonstrates that the guilt aversion of lower ability workers under equal sharing can potentially function as a way to enhance team output, which provides a new rationale for the literature on equal pay and wage compression (Lazear, 1989; Breza et al., 2018; Hjort et al., 2020; Hooper et al., 2021). This contrasts with conventional wisdom, which typically highlights the role of higher ability workers in enhancing team productivity (Hamilton et al., 2003; Bandiera et al., 2013; Cooper et al., 2021). Moreover, unlike studies on team incentive focusing on social ties (Bandiera et al. 2010; Lim and Chen, 2014), peer effect with mutual monitoring (Falk and Ichino, 2006; Mas and Moretti, 2009), non-independent production with assistance (Chan et al., 2014a), knowledge transfer (Chan et al., 2014b), or production complementarity (Lazear, 1999), our setting is a simple real-effort work environment with anonymity, observability, and complete independence. Under our context, standard economic theory typically regards individual piece rate as more effective and fails to predict the power of equal sharing, attributed to the phenomena of free riding. The attributes of anonymity and random assignment in our study enable us to distinguish ourselves from productivity sorting (Eriksson and Villeval, 2008; Dohmen and Falk, 2011) or social ties, thus highlighting the power of equal sharing on lower ability workers possessing guilt aversion. Meanwhile, while several studies have also pointed out the potential efficacy of equal sharing or guilt aversion in teams (Chan et al. 2014a; Chen and Lim 2013; Babcock et al. 2015), the unique setting of our study sets it apart from these studies, the details of which will be discussed in Section 3.

Second, we contribute to the literature on self-selection into team pay schemes by employing controlled and incentivized experiments, which is rarely adopted in existing studies (Kuhn and

Villeval, 2015; Wood et al., 2023).[3] Our random assignment of team members and threshold conditions differs from studies on endogenous team formation (Hamilton et al., 2003; Bandiera et al., 2013; Cooper et al., 2021) or studies where the distribution scheme is fixed but teams have the discretion to impose a group output threshold (e.g., He and Miao, 2022). The complete information environment regarding within-team relative abilities also distinguishes our study from Kuhn and Villeval (2015), where participants were unaware of their co-participant's ability until the end. In addition, our comparison between the Threshold and No-Threshold conditions enables us to investigate how the introduction of a team threshold influences workers' preferences for distribution schemes. Moreover, the examination of chat messages helps in gauging the motivations (e.g., social norm, fairness, and guilt aversion) underlying the selection of distribution schemes. Thus, our study also adds to the existing literature on social norm and fairness concerning compensation schemes (Charness and Kuhn, 2007; Bolton and Werner, 2016; Breza et al., 2018; Fehr et al., 2023).

Our paper proceeds as follows. Section 2 presents our experimental protocol. Section 3 presents a summary of the guilt aversion framework. Section 4 demonstrates the main results. Section 5 discusses alternative theoretical explanations. Section 6 concludes. Appendix A displays the experimental instructions with a screenshot. Appendix B shows additional results. Appendix C presents details of our guilt-aversion model. Appendix D provides a summary of the subjects' different motivations for distribution choices as observed from the chat messages.

2. Experimental Design

We conducted the experiment at Zhejiang University, China. We recruited 248 undergraduates from several majors and organized them into sessions that contained all three distribution schemes. We computerized the experiment using z-Tree (Fischbacher, 2007). In the experiment, participants worked on real-effort slider tasks (Gill and Prowse, 2012) in which participants faced a computer screen displaying 48 sliders, each on a scale from 0 to 100.[4] Initially, all sliders were positioned at

---

[3] Real world examples of firms that let workers decide their own salaries or payment schemes include Semco in Brazil, Skyline in the U.S., and Claravision in Spain (see Charness et al., 2016 for a summary). Pierce et al. (2021) provided firm-based evidence where teams self-manage their rewards and allocate compensations internally.

[4] Some studies suggest that workers' output observed in slider task is not very elastic to monetary incentives (e.g., Araujo et al., 2016). Gill and Prowse (2019) suggest that real-effort tasks in general tend to produce small responses to between-subject variation in positive piece-rate incentives and show that within-subject designs lead to greater

zero. The participant used the mouse to move as many sliders as they could to exactly the middle point 50. The task is easy to understand and to do with no scope for guessing. We use participants' benchmark performance to indicate their ability at the task.

Our experiment has three distribution schemes: *equal, piece-rate,* and *winner-takes-all*. Each of these schemes is tested under both a Threshold and a No Threshold design. In the Threshold condition, we recruited a total of 150 participants who were randomly assigned to one of the three distribution schemes. Each distribution scheme had 50 participants. We had 98 participants in the No Threshold condition (34 in *equal*, 34 in *piece-rate*, and 30 in *winner-takes-all*). We randomly assigned participants into Threshold and No Threshold sessions. Each session had an average duration of approximately 30 minutes, starting with a five-minute instruction period, followed by a two-minute practice session.

The experiment followed the Figure 1 flow chart. During the benchmark stage, each participant received a piece rate payment of 0.30 RMB for each slider-bar moved to the middle point. Their performance in the benchmark stage allowed us to assess their relative abilities. To attenuate the potential ratchet effect identified by Charness et al. (2011), participants in benchmark stage were informed of the experiment's future stages without specific details.

[Insert Figure 1 Here]

Moving to Stage 1, two participants were randomly paired into a team and the teams were then randomly assigned to one of the three distribution schemes (*equal, piece-rate or winner-takes-all*) and to either the Threshold or No Threshold condition. To minimize strategic behaviors stemming from anticipations of repetitive interactions with fixed partners across stages, participants were kept unaware of the fixed nature of team composition in Stages 2 and 3. Team members were informed about the benchmark scores of their teammates so that they could assess the relative abilities of each member in performing the slider task. Each team had four minutes to slide as many bars to the middle point as they could. During the task, they observed in real time their own and their teammate's performances; therefore, they got relative performance feedback (Eriksson et al., 2009). After completing the task, participants in the No Threshold condition received their payment based on the randomly assigned sharing scheme. In the equal scheme, each member

---

effects. Despite this low variation, we remained to observe significant productivity differences across distribution schemes. Moreover, we have the same subject over multiple stages and the same team with potentially varying team sharing schemes, both of which echo Gill and Prowse' appeal for real-effort tasks with "repeated observations of effort provision from the same subjects in a short time frame."

received half of the total scores earned by the two members. Under the piece-rate scheme, individual earnings were determined proportionally to each participant's individual performance. In the *winner-takes-all* scheme, the member with the higher performance received all the scores, while the other member received no scores. Participants in the Threshold condition received the payment only if their team reached or exceeded the threshold target but received nothing if the team's output fell short of the threshold.[5] The possibility of receiving the payment or receiving nothing based on the team's performance made the Threshold/No Threshold condition a powerful determinant of team output. Furthermore, the presence of a threshold may also have an impact on team members' preference for a distribution scheme in Stage 2.[6] Note that it is difficult for one player to reach the group threshold on their own, whereas a team with two cooperating participants is more likely to achieve it. By keeping the participants unaware of how thresholds were determined, the experiment aimed to prevent strategic gaming in the benchmark stage.

We began the next part of the experiment by providing participants with the information of the average performance of teams in Stage 1 under each distribution scheme in their session. This information enabled them to evaluate whether their team might benefit from choosing a different distribution scheme. Participants were given a four-minute period to discuss their preferences for distribution schemes in the Ztree's chat box before they cast their vote for the preferred scheme. If team members both voted to switch to the same new distribution, the pair would be switched to the new scheme. This mutual agreement principle is similar to that in the literature (Kuhn and Villeval 2015; He and Miao 2022). If they did not reach an agreement, the distribution scheme remained unchanged. We recorded the discussion and coded their conversation based on four dimensions: whether the team discussed cooperation, fairness, the originally randomly assigned scheme, or differences in abilities. In Stage 3, the teams performed the slider task for four minutes under the distribution scheme they had chosen. Participants earned, on average, around 43.6 RMB

---

[5] They still receive their earnings from benchmark stage and their show-up fee.

[6] Among the 75 teams in the Threshold condition, 51 teams had a threshold set equal to the sum of the two participants' piece rate outputs in the benchmark stage. Additionally, 24 teams received a threshold set at 1.1 times the team's benchmark output. Such thresholds reflect many real-world team targets that are based on a team's baseline performance or a reasonable markup (in our case,10%) of it. It is also chosen to ensure that it is sufficiently challenging so that the threshold is unlikely to reached by any individual team member. However, these levels may not be optimal to induce maximum team productivities.

during the experiment (including a 10 RMB show-up fee)[7]. In a follow up survey, we also asked participants in the Threshold condition about whether they had set goals of individual output.

3. The Conceptual Framework: Guilt Aversion

We offer an explanation for the enhanced performance of less able workers in team settings through the lens of guilt aversion, referencing the conceptual framework outlined in Appendix C. This approach is inspired by Charness and Dufwenberg (2006) and Battigalli and Dufwenberg, (2007), but tailored to a context where team members independently perform tasks with clearly identifiable contributions, similar to the sliding bar task in our study. We posit that this independent task completion fosters a social norm where individual earnings are expected to correspond with individual output. Our assessment of workers' perceptions of this social norm and their revealed preferences regarding task selection in a subsequent stage confirmed this assumption. Consequently, in a team environment with equal sharing, these norms compel lower ability individuals with guilt aversion to enhance their productivity to avoid adversely affecting their teammate's expected earnings. This rationale aligns with Gill and Stone (2015), where agents exerting lower effort feel undeserving of equal shares, thus motivating them to align their effort with their higher-performing teammates. Our emphasis is on the disutility stemming from disappointing a partner rather than the discomfort of receiving more than one's fair share. Importantly, our simplified model sidesteps the complex higher-order beliefs inherent in general models of guilt aversion (Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007).

Our data supports the evidence of social norm of "piece rate". Queried after the experiment about attitudes toward compensation, 85% of our subjects believed that *piece-rate* is the most popular distribution scheme compared to only 13% who considered *equal* as the most popular. Building upon these beliefs, our theoretical framework predicts that lower ability workers with sufficient guilt aversion will exert higher effort in equal sharing than in the other distribution schemes (see Appendix Table C2). While several studies have pointed out the potential efficacy of equal sharing or guilt aversion in team incentives, the unique setting of our study distinguishes it from such research.[8] For instance, Chan et al. (2014a) showed that team-based equal sharing, along

---

[7] Their earnings were around the typical student wage rate of 50 RMB per hour, approximately 7 US$.

[8] In a principal-agent setting, as opposed to a team context, Charness et al. (2013) provided experimental evidence indicating workers' aversion to guilt associated with not fulfilling promises related to effort levels.

with heterogeneity in workers' abilities, improves firm performance compared to the individual-based piece-rate pay. However, our setup differs from their production environment in that their team members can assist each other and receive monetary reward from team-based commissions. Consequently, their team productivity increase can be attributed to self-interest rather than guilt. In contrast, our environment revolves around independent work without any assistance from members. Therefore, the effectiveness of equal sharing over individual piece rate in our setup cannot be explained by self-interest. Similarly, previous studies have demonstrated the efficacy of guilt version in improving team productivity, only when participants socialize with their teammates or when the identities of team members are known to each other (Chen and Lim 2013; Babcock et al. 2015). However, our setting preserves complete anonymity that prohibits any form of social interaction (at least during the task before the online chatting stage).[9] We expanded upon the work of Chen and Lim (2013) by examining the effect of guilt aversion on team output under both threshold and no threshold conditions, whereas their study focused solely on the contest scenario, where the rival's output served as a "threshold" to surpass.

After we present our results, we will further discuss in Section 5 why alternative theories such as peer-pressure, the drop-out of lower ability workers in winner-takes all, and other social preference theories fail to explain our data.

4. Experimental Findings

In this section, we present the experimental results by stages. Specifically, we examine the impact of randomized distribution schemes on Stage 1 productivity in Section 4.1. Following that, Section 4.2 explores the effect of threshold on Stage 1 productivity. Moving forward, Section 4.3 examines team members' Stage 2 chats and their correspondingly preferred distribution schemes. Finally, Section 4.4 presents the productivity in Stage 3 after teams chose their preferred distribution schemes.

4.1 The Power of Equal Sharing

We first look at group-level outcomes for randomly assigned distribution schemes in Stage 1. *Equal sharing* yields the highest team output in Stage 1 (mean = 85.3, n = 42), followed by the

[9] Participants are not allowed to reveal any identity information during their online chatting session.

piece-rate (mean = 83.6, n= 42), while the *winner-takes-all* comes in last (mean = 77.6, n = 40). The team output under *equal sharing* scheme is statistically significantly higher than the output under the *winner-takes-all* scheme (p = 0.02, two-sided Mann-Whitney test) and insignificantly higher than the output under the *piece-rate* scheme (p = 0.69, two-sided Mann-Whitney test). In addition, *equal sharing* proves to be more effective in assisting the group to meet the threshold target compared to the other schemes. We found 72% of the groups successfully reached the threshold in *equal sharing*. In comparison, 64% of the groups achieved the threshold in the *piece-rate* scheme and only 56% of the groups managed to reach the threshold in the *winner-takes-all* scheme.

Table 1 presents a linear regression analysis about the Stage 1 output of each participant, considering the participants' benchmark output, the randomly assigned distribution schemes, and whether the threshold condition was present. Columns (1) and (2) display results for the entire sample of participants, while Columns (3) and (4) separate the results for individuals who scored lower and higher in their team during the benchmark stage.[10]

[Insert Table 1 Here]

The estimated coefficients of the benchmark output in all four columns are statistically significant with positive values. This finding supports the notion that benchmark performance effectively captures individual ability differences in performing the task. Columns 1 and 2 show that team members in the *equal sharing* scheme were significantly more productive than those in the *winner-takes-all* scheme ($p < 0.05$). Furthermore, their productivity was slightly higher than, or at least comparable to, those in the piece-rate scheme, as indicated by a Wald test (p = 0.37). This outcome contradicts the standard economic theory's expectation of a free-riding motive in sharing schemes. The estimated coefficients of the Threshold dummy variables are positive but statistically insignificant.[11]

The most striking result in Table 1 is the difference in estimates between the regression in column (3) for lower ability participants and the regression in column (4) for higher ability participants. For lower ability individuals, the regression results indicate that *equal sharing* has a substantial and significant advantage over the default *winner-takes-all* ($p < 0.05$). This estimated effect of *equal sharing* is twice the magnitude of the *piece-rate* effect. Despite a relatively small

[10] Four participants have identical benchmark scores as their team partners. In our classification, all of them are considered higher ability individuals. As a result, we have a total of 126 higher ability participants and 122 of lower ability participants. It is worth noting that alternative classifications of these four participants do not affect the results.
[11] Regressions (available upon request) with two threshold dummies (1.0 and 1.1 times) yield similar results.

number of observations, this difference in effect is nearly marginally significant (p = 0.12). In contrast, for higher ability individuals, there is no difference in the estimated coefficients of different modes of compensation. This implies that the higher team output in *equal sharing* can be attributed to the higher performance of the lower ability workers.

In Appendix Table B1, we find similar effects of distribution schemes on "approximate output," defined as the number of slider bars moved to the position between 48 and 52 or between 49 and 51 rather than exactly at 50. Notably, the difference in approximate output between *equal sharing* and *piece-rate* becomes marginally significant (p = 0.10, see Column 3) for approximate output positioned between 48 and 52.

4.2 Heterogeneous Responses to Team Thresholds

Our experiment used the Threshold condition to reinforce the team link between members by linking their pay not only to their individual output but also to the output of their teammates. We established the thresholds by considering the sum of the benchmark scores of both team members (and in some cases, set the thresholds at 1.1 times of this sum). The rationale behind this approach was to ensure that the thresholds were challenging enough to be beyond the reach of a single individual but attainable through the combined effort of both team members. Indeed, 69% of the teams with the sum of two members' benchmark scores as their team threshold were able to achieve it, while only 54% of teams managed to reach the more challenging threshold set at 10% above the sum. Furthermore, the teams that fell short of reaching the threshold came close, with an average value of 93.6% for the ratio of team productivity over the respective threshold. Overall, these findings suggest that participants perceived the thresholds as reasonably attainable. Meanwhile, the significant number of teams that failed to reach the thresholds indicate that they still presented a substantial level of difficulty.[12]

As shown in Table 1, imposing a group threshold only has a small and statistically insignificant average effect on productivity. To gain insight into the thoughts processes of the participants in a threshold setting and to explore potential variation in their responses to the group threshold, we included a question in the post-experiment questionnaire: "Did you set yourself a specific target

[12] It is possible that an alternative slightly lower threshold level could induce a higher team output. Thus, the effect of our threshold result does not conclude the overall influence of thresholds. It is worthwhile to study the threshold elasticity of team output in future work.

based on the team threshold?" The question provided four answer-options: "I tried to achieve half of the threshold", "I tried to achieve more than half of the threshold", "I tried to achieve less than half of the threshold", and "I had no target at all." The results, as depicted in Figure 2, shows that across incentive schemes, at least two-thirds of the higher ability participants targeted more than half of the team threshold (77% in *winner-takes-all*, 82% in *equal*, and 67% in *piece-rate*). A smaller portion of them aimed for exactly half of the threshold or did not have a specific target. In contrast, a smaller proportion of lower ability participants targeted more than half of the team threshold (27% in *winner-takes-all*, 27% in *equal,* and 38% in *piece-rate*). Additionally, a substantial proportion of lower ability participants said they had not thought about a personal target at all.

[Insert Figure 2 Here]

Finally, continuing our focus on lower ability individuals, we examined whether those who targeted over half of the threshold outperformed other lower ability team members. We regressed the Stage 1 outputs on dummy variables for whether individuals aimed to take more than half, exactly at half, or less than half of the team threshold in Table 2. The regression analysis in column (1) shows that the lower ability team members who aimed for more than half of the team threshold had significantly higher output compared to those who had no target ($p < 0.05$) and those who targeted less than half or exactly at half (Wald test: $p < 0.001$ and $p = 0.005$, respectively). Additionally, it shows that those who had no target actually outperformed those who targeted less than one half of the team threshold.[13] Since no higher ability participant aimed for less than half of the threshold, the regression analysis in column (2) focuses on comparing two groups within the higher ability category: those who targeted more than half of the threshold and those who targeted exactly half or had no specific target at all. These estimates do not reveal a distinct pattern within the higher ability group: the output of those who targeted over a half did not differ much from those who targeted exactly a half nor from those who had no specific target. To further validate these observed patterns, we expanded the sample to include individuals from the No Threshold condition and designated them as persons without an individual target. The results presented in columns (3) confirm the initial finding that individual target level has a significant

[13] When we define lower ability as those who has lower productivity in both benchmark stage and Stage one, the result becomes stronger (available upon request).

impact on individuals with lower abilities. However, the effects on individuals with high abilities are not clearly discernible in this expanded and slightly modified analysis.

[Insert Table 2 Here]

4.3 Choices and Chats for Distribution Schemes

In this subsection we first present participants' choices for distribution schemes, and then we analyze their chat messages to understand the underlying factors of their choices.

4.3.1 Choices for Distribution Schemes: Piece rate as the Most Popular Choice

Stage 2 allowed teammates to change to a distribution scheme different from the randomly assigned scheme. Table 3 shows that many teams took advantage of this opportunity.[14] The rows record their three initially assigned distribution schemes. The columns display the distribution schemes the team members ultimately end up with, both in terms of the number of teams and the proportion of the teams. These elements together illustrate the transition of team distribution schemes from the initially assigned scheme to the self-selected scheme. Panel A, which includes both the Threshold and No Threshold conditions, shows that nearly all teams initially assigned *winner-takes-all* abandoned it, with the majority opting for piece rate pay instead. Nearly 2/3rds of the teams under *equal sharing* shifted to *piece-rate*, while 1/7th shifted from *piece rate* to *equal*. Piece-rate pay was the only scheme in which the majority of members (86%) chose to stay with, making it the dominant or "attractor" distribution throughout the experiment. Panel B and Panel C show that this pattern also holds for the No Threshold and Threshold conditions respectively. Yet, compared to No threshold, more teams in Threshold chose *equal sharing*, partly because more of those originally assigned to *equal sharing* in Threshold chose to stay with it. This revealed preference for *piece-rate* scheme among most of the participants further supports the assumptions we made in building up our guilt aversion model. Specifically, the lower-ability participants may perceive that their high ability counterparts expect earnings proportional to their productivity.

[Insert Table 3 Here]

[14] Among all 124 teams, 117 of them (or 94.4%) reached unanimity regarding the distribution method for Stage 3; 71 out of the 117 teams switched to a new distribution and 46 of them sticked with the original distribution scheme.

4.3.2. Unveiling Motivations: Chat Evidence Supporting *Piece Rate* as the Most Popular Scheme

How can we best interpret the choices displayed in Table 3? It is rational economic behavior for participants to shun away from *winner-takes-all* scheme because, in terms of economic payoff, it had the lowest potential payoff. But with *equal sharing* yielding a modestly higher return than *piece-rate*, the preference for *piece-rate* must be due to something beyond choosing the most lucrative option for a team.[15]

We examine the group conversations from both the threshold and no threshold conditions. Out of the 124 groups, 46 groups provided 76 conversations revealing their motivation for their choices (see Table D1 in Appendix D). Among these 76 conversations, 57 (76%) of them indicate a motivation suggesting that higher ability individuals prefer the piece-rate scheme, or others believe they do. This finding supports the social norm perception of piece-rate pay among workers and reinforces the assumption that lower ability workers may believe their higher-ability counterparts anticipate a piece-rate pay. Specifically, 25 conversations (category A and B) out of the 57 conversations explicitly mentioned guilt, with lower ability participants proposing the piece-rate mechanism or promised to work harder if equal sharing is chosen. The remaining 32 group conversations, while not explicitly mentioning lower ability's guilt aversion, all demonstrate higher ability's preference for piece-rate. Among them, 15 conversations involve higher ability workers directly proposing piece-rate, 9 conversations show higher ability individuals sharing their knowledge with their lower ability counterpart before advocating for piece-rate, 5 conversations demonstrate both participants perceiving piece-rate as the fairest scheme,[16] and the remaining 3 conversations feature higher ability individuals persuading lower ability to choose piece-rate, portraying it as a safer choice than winner-takes-all.

4.3.3 Who Choose *Equal Sharing*

To better understand the factors influencing the choice for equal sharing, we conducted a regression analysis using a binary dependent variable, where a value of "1" represented individuals

---

[15] Babcock et al. (2015) had a similar finding: almost all individuals (97%) preferred the individual incentive than team incentive despite the higher productivity of the latter.

[16] Litereature also showed that participants perceive productivity-based pay disparity as fair when the productivity gap between team members is observable, mirroring the conditions in our environment (Bolton and Werner, 2016; Breza et al., 2018). Meanwhile, participants in Fehr et al. (2023) perceived piece rate schemes as fairer than tournaments.

choosing *equal sharing* while "0" indicated a different choice. The regression included a set of independent variables, such as personal attributes, a dummy variable indicating whether an individual was the lower ability person in a team, a dummy variable representing the Threshold condition, and the initially randomly assigned distribution scheme.

Economic logic suggests that the lower ability persons would be more inclined to favor *equal sharing* compared to higher ability persons. We test these expectations in Column 1 in Table 4. The estimated regression coefficients show that lower ability participants were indeed more likely to favor *equal sharing*. Teams that were initially compensated with a piece rate payment scheme exhibited a lower likelihood of choosing *equal sharing* compared to those initially assigned to other schemes. Additionally, participants in the threshold condition were more likely to choose *equal sharing*. Regarding gender, we find no significant difference between men and women in their preference for equal sharing. This result differs from Kuhn and Villeval (2015).[17] This disparity in results may stem from the fact that their result was influenced, at least in part, by women having more optimistic assessments of their prospective teammate's ability, whereas in our experiment, the abilities of teammates were known facts.

[Insert Table 4 here]

4.3.4 Understanding Distribution Choices through Chat Messages: Threshold versus No-Threshold

We examine the chat discussions between team members regarding the potential change in the mode of distribution. We had three research assistants (RAs) independently review the chat messages and code the messages based on whether any of the following terms/concepts appeared in the conversations: a) *Cooperation;* b) The *within-team gap in productivity*; c) O*riginal distribution* schemes in Stage 1; d) *Fairness*. While the coding is largely consistent across RAs, we apply the majority rule when they disagreed.[18]

Table 5 shows the prevalence of the four concepts in the chats based on the presence of threshold. The most discussed concept in both Threshold and No Threshold conditions was the gap

[17] Because almost no subjects choose *winner-takes-all*, we do not find a gender difference for competition preference (Niederle and Vesterlund, 2007) either.

[18] The RAs agreed unanimously from 75% to 90% of the time. The Kappa (κ) measure of an inter-rater reliability shows that the raters agreed significantly more than by chance: "Cooperation": κ= 0.46, Z=8.9, p <0.001; "Gap": κ =0.83, Z=16.1, p < 0.001; "Original assignment": κ =0.57, Z=10.9, p < 0.001; "Fairness": κ =0.65, Z=12.5, p < 0.001.

between the team members in benchmark stage. Modest proportions of conversations discussed the previously assigned distribution scheme and fairness. The most striking difference between the Threshold and No Threshold experiments was the frequency of discussions related to cooperation. Among all 75 teams in the Threshold condition, 20 (or 27%) of them discussed the importance of cooperation compared to just 2 of 49 (4%) of teams in the No Threshold condition. This discrepancy suggests that the extreme monetary incentive associated with threshold condition had a substantial impact on the participants' perception of themselves as a cohesive team.[19]

[Insert Table 5 Here]

In order to examine the potential impact of chatting about cooperation on the preference for equal sharing compensation, we reanalyze the data by including a chat-cooperation dummy variable into the independent variables of column (2) in Table 4. The estimation of this variable shows that the chat-cooperation dummy significantly explains individuals' preference for *equal sharing.* Additionally, the presence of this also reduces the coefficient of the Threshold dummy. These two results together suggest that the effect of threshold on explaining individuals' preference toward *equal sharing* occurs through the process of inducing participants to talk about the importance of cooperation in order to reach the target. One possible dynamic is that the higher ability participants need the efforts of lower ability participants to reach the target, in line with the emphasis of within-team harmony as a rationale for pay compression (Lazear, 1989). In this scenario, equal sharing can serve as a stimulus for the lower ability participant to actively contribute to the team's overall performance, per the power of equal sharing under guilt aversion.[20] In contrast, other chatting variables in Table 5 had little impact on explaining participants' preference for equal sharing.

4.3.5 Chat Evidence Revealing the Role of Threshold

In addition to the above analyses, we further leveraged detailed chat messages to assess the underlying motivations behind the choices of equal sharing scheme as we did so for the piece-rate scheme. There are 19 out of 76 conversations that supported the preference for equal sharing.

---

[19] This is reflected in their stage two choice, where 35% (Table 3, Panel C) of teams ended up choosing *equal sharing* in Threshold while only 16% (Table 3, Panel B) did so in No Threshold.

[20] In the chat messages, the higher-ability workers in 10 groups express that they would choose equal sharing due to altruism or cooperation. Among these 10 groups, the Threshold condition accounts for 90% (i.e., 9 groups), a much higher ratio than the overall ratio of groups with Threshold in our total samples. See motivation category G in Appendix Table D1.

Among them, 10 conversations showed that equal sharing was chosen out of high ability individuals' cooperation or altruism consideration, with 90% of them were in the Threshold condition. This echoes our Table 5 results that showed a significantly higher presence of cooperation consideration in the Threshold condition than in the No Threshold condition.

Similarly, teaching from the higher to the lower ability participant occurred in 13.3% of groups (or 10 out of 75 groups) in the Threshold condition but only in 2.0% of groups (or 1 out of 49 groups) in the No Threshold condition (see categories D and J in Appendix Table D1), suggesting that the need to reach the group threshold boosts the higher ability's incentive to teach. This pattern highlights the crucial role of incentive in knowledge sharing (Siemsen et al. 2007; Siemsen et al., 2008), and echoes Sandvik et al. (2020) who find that knowledge providers are more willing to share information when their own interests are linked to that of their partners, or when there is expected future monetary benefit from teaching (Cooper et al., 2021). However, among the 11 groups with teaching, the ratio that finally chose *piece rate* rather than *equal sharing* for Stage 3 is as high as 82% (i.e., 9 out of 11), even higher than the overall ratio of 71% choosing *piece rate* among all 124 groups in our sample and the ratio of 63% among all groups in the Threshold condition. This may suggest that the higher ability workers consider teaching as a favor to the lower ability and an effort to help meet the group threshold and would not further compromise. That is, they employ teaching as an alternative to choosing *equal sharing* of maintaining group cohesion and morale, without conceding their pay share.[21]

4.4. Equal Sharing in Stage 1 Has Sustained Impacts in Stage 3

Stage 3 of the experiment had participants perform the slider task with the distribution scheme they chose in Stage 2. Column (1) of Table 6 records the results of regressing the Stage 3 productivities on their performance in the benchmark, whether they were in a Threshold condition, and the initial distribution scheme. It shows that *equal sharing* in the initial random assignment was associated with higher productivity in the third stage just as in the first stage shown in Table 2. Columns (2) and (3) show further that the driving force for the higher outcome is again the better performance of the lower ability participants under *equal sharing*: Low ability participants randomly assigned to *equal sharing* in Stage 1 produce more than those randomly assigned to

[21] Cooper et al. (2021) find a similar pattern: most higher ability workers choose not to join a team if they need to share revenues with coworkers (see the treatments other than PR in their Table 2).

winner-takes-all ($p < 0.10$) and *piece rate* ($p = 0.05$). This raises the possibility that a person's early experience of a compensation system may affect their productivity in later work situations.[22] We report other results about Stage 3 in Appendix B3.

[Inert Table 6 Here]

5. Discussions

We first discuss the more able participants in Section 5.1, then we discuss alternative explanations in Section 5.2, 5.3, 5.4 and 5.5.

5.1 Responses of the Higher Ability Participants to Equal Sharing

Since team output depends on the effort of higher ability participants as well as that of lower ability participants, our results require that higher ability participants do not strongly slack to offset the additional output of lower ability participants. If the more able considered equal sharing unfair because they earn less than they would have under *piece-rate*[23] and responded by producing less in equal sharing than in *piece-rate*, this could have readily counterbalanced the impact of guilt aversion in raising the output of the less able on team output. If, by contrast, equal sharing spurred the production of more able team members as it spurred the production of less able team members, the output of both would increase, enhancing the attraction of equal sharing.[24] Indeed, Chen and Lim (2013) report such behavior in their experiment when the team members knew each other and put out greater effort in team contests than in individual contests.[25] They attributed this to guilt aversion of both team members to letting their team down. Similarly, Babcock et al. (2015) found big productivity increases in team incentives compared to an individual incentive and they attribute this to guilt or social pressure, as individuals may be averse to being responsible for the team's failure.

---

[22] This calls for future studies with longitudinal data on workers.

[23] Gill and Stone (2015) call this *desert loss* and provide a theoretical analysis that allows *desert loss*, *desert guilt* (similar to guilt aversion in our case), and *desert elation.*

[24] In our experiment, guilt aversion under *equal sharing* can also come from higher ability participants if they feel guilty from not performing their best to benefit the lower ability partner even more. Although this aspect is beyond our formal conceptual analysis, chat messages do suggest that in three teams (all in the Threshold condition), not only the lower ability but also the higher one considers *equal sharing* a powerful team incentive to spur both of their productivity. See motivation category H in Appendix Table D1.

[25] In the context of charitable giving, Charness and Holder (2019) show that because of guilt aversion to "let down their team," participants donate more in team competition than in individual competition for matching funds even under anonymity.

5.2. Alternative Explanation 1: Do Lower Ability Workers Drop Out in Winner-takes-all?

We have demonstrated that, compared to the *winner-takes-all* scheme, *equal sharing* enhances the productivity of both the lower ability workers and the overall team. This raises the question: does the relative advantage of equal sharing stem merely from mitigating the *winner-takes-all*'s negative impact on the lower ability workers? Theoretical and experimental studies on contests have suggested a "discouragement effect": a greater degree of heterogeneity in relative ability among participants, when publicly known, results in a reduction of aggregate effort (see the literature summary by Dechenaux et al., 2015). A mechanism pertinent to our study is that a large ability gap may demotivate lower-ability participants, deemed unlikely to win, leading them to exert little effort (i.e., "drop out").

To investigate whether our result is driven by a discouragement effect, we analyze the effect of the within-team ability gap—measured as the absolute difference in output during the benchmark stage—on the productivity in Stage 1 of both the lower and higher ability workers under the *winner-takes-all* scheme, as detailed in Appendix Table B2. Column 1 reveals that a larger gap increases productivity among lower-ability workers. Column 2 further introduces a quadratic term to assess potential nonlinear effects, uncovering an inverted U-shaped relationship with a turning point at a gap of 51, exceeding the observed gap's maximum at 35. Therefore, the within-team ability gap invariably enhances the output of lower-ability workers. With the ability gap's median, standard deviation, and half of its median at 10, 8, and 5, respectively, we apply these cutoffs in Columns 3–5 by creating a dummy variable that equals to one when the ability gap has reached the cutoff point. Yet, we consistently observe the positive effect of ability gap on lower-ability workers' productivity. The same analysis in Columns 1–5 for lower ability workers is conducted for higher ability workers in Columns 6–10. The results show that the ability gap does not reduce their output.

In sum, Table C2 does not support the notion that an increasing ability gap under *winner-takes-all* schemes discourages both the lower- and higher-ability workers. This implies that the efficacy of *equal sharing* likely stems from factors beyond merely mitigating the negative aspects of *winner-takes-all*.

5.3. Alternative Explanation 2: Peer Effect?

Other forms of group interaction might also influence the behavior of more or less abled members of a team in ways that go beyond our analysis. Peer pressure (e.g., Mas and Moretti, 2009) and reference dependence (Abeler et al., 2011; Gill and Prowse, 2012) models predict that behavior depends on interactions among teammates. We expect peer influence to spur great effort in our small two-person experiment, with lower ability workers exerting more effort as they observe their better performing teammates or faced pressure from the better performers regardless of the distribution scheme. To assess whether peer pressure or reference dependence affects participants' behavior in our experiment, we added the teammate's benchmark score variable to our Table 1 regressions of Stage 1 output on independent variables. In addition, this measurement of individual ability, coupled with real time feedback of teammate's productivity, allows us to control for the impact of peer effect on productivity.

Table 7 summarizes the statistical results. Despite being imprecisely estimated, we find the ability of a randomly assigned coworker largely links to the output of lower ability workers in Stage 1 (Columns 1), but not that of the higher ability workers (Columns 2).[26] Controlling for the positive peer effect, we remain to observe a positive impact of equal sharing for lower ability workers, with the coefficient of it falls slightly from 4.367 in Table 1 to 4.063 in Table 7. Therefore, peer effects alone cannot account for the observed differences in productivity across incentive schemes.

[Insert Table 7 Here]

5.4. Alternative Explanation 3: Reciprocity?

Another explanation for lower ability workers' enhanced productivities under equal sharing is reciprocity (e.g., Charness and Rabin, 2002; Falk and Fischbacher, 2006): they reciprocate the benevolent intention of the employer or the higher ability partner for placing them into the favorable equal sharing scheme. However, in Stage 1 the distribution method is randomly assigned by the experimenter (*not* by higher ability workers) without intentions to favor the lower ability workers. Thus, reciprocity is less likely to be the explanation of the results.

5.5. Alternative Explanation 4: Inequality Aversion?

[26] Similarly, Mas and Moretti (2009) found that unlike low-ability workers who benefit from high-ability coworkers, the productivity of high-ability workers is not affected by their (low-ability) coworkers.

Inequality aversion (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) is not a good explanation, either. A well-established result in the standard neoclassical model is that the lower ability workers will show increased productivities under *piece rate* than *equal sharing* because of the free riding motive in the latter scheme. While the existence of pay inequality aversion is irrelevant for workers under *equal* sharing because of the within-team egalitarian pay, it could drive lower ability workers under *piece rate* to work even harder than the neoclassical scenario to reduce the earnings gap with their higher ability partners. Thus, inequality aversion would predict that lower ability workers would work harder and produce more under *piece rate* than under *equal* sharing, which is the opposite of our result.

In addition, tournament incentives might fail in motivating high-ability people with inequality aversion. Indeed, Bandiera et al. (2005) find that workers' productivities are lower under relative incentives when they work with their friends than under piece-rate. They also showed that this lower productivity in relative incentives only holds when workers and their friends can monitor each other. While we would not know the effectiveness of equal sharing in their environment, we show that equal sharing is more effective than winner-takes-all even if workers work with strangers in an anonymous environment.

6. Conclusions

Under our experimental protocols, we show that equal sharing raises team output by inducing greater output from lower ability workers, likely due to guilt aversion. We also find some persistent effect on improving productivity over time without adversely affecting the output of more able workers. To what extent, if at all, are our findings likely to apply to larger teams, different experimental situations, and ultimately to teams in real workplaces?

The simplicity of our experimental design has strengths in generalizing to more complicated situations. Stage 1's random assignment of compensation systems identifies the "pure effect" of those systems on behavior. The random formation of teams sidesteps the endogeneity of team formation. The ease of learning the sliding bar task guarantees that participants understand the task and can gauge the relative ability of team members, which underlies the channel of guilt aversion. The use of thresholds to create team incentives avoids tying the experiment to any specific technological or organizational mode of forming teams. Finally, allowing workers to change the mode of compensation in Stage 2 and recording their chats about mode changing gave insights into participants' thinking about cooperation, particularly under threshold conditions, that are likely to arise in any such change scenario.

Still, experiments that go beyond our design could produce results that would vary from what we obtained and offer more nuanced guidance to when equal sharing, winner-takes-all, piece rate pay, or variants thereof might incentivize lower or higher ability workers. One important factor that may affect results is the number of persons on a team: greater numbers are likely to strengthen the incentive to free ride compared to guilt aversion and thus weaken our result. Another factor that merits attention is the complexity of tasks, where the key determinant of output may be getting workers to accept tasks for which they have comparative advantage. In these situations, the experimenter will likely have acted more as a manager or team leader following more complicated instructions than in our experiment. Another factor that our experiment short-changed is the length of time the team worked together. Previous studies have involved participants who presumably knew each other prior to the experiment.

The simple individual production environment in our task makes the social norm of piece-rate pay salient. This enables us to utilize a simplified version of the guilt aversion model, employing this social norm as an approximate measurement of participants' higher-order beliefs to calculate others' expected earnings. Under more complicated production technologies, however, pay norms can become less salient or not commonly perceived, and we would need to directly measure participants' higher-order beliefs. In addition, although we have shown that the power of equal sharing can be better explained by guilt aversion rather than alternative theories, we cannot completely rule out all other potential mechanisms. We leave this to future studies.

In sum, the simplicity of our design has both positives and negatives for generalizing the results. The broadest implication for further research on teams and for managements choosing forms of compensation for teams is to be attentive to the performance of lower ability persons, whose responsiveness to sharing in rewards can be critical in team performance.

References

Abeler, Johannes, Falk Armin, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision." *American Economics Review*, 101 (2): 470–92. DOI: 10.1257/aer.101.2.470

Araujo, Felipe E., Erin Carbone, Lynn Conell-Price, Marli W. Dunietz, Ania Jaroszewicz, Rachel Landsman, Diego Lamé, Lise Vesterlund, Stephanie W. Wang, and Alistair J. Wilson. 2016. "The Slider Task: An Example of Restricted Inference on Incentive Effects." *Journal of Economic Science Association*, 2: 1–12. DOI: 10.1007/s40881-016-0025-7

Babcock, Philip, Kelly Bedard, Gary Charness, John Hartman, and Heather Royer. 2015. "Letting Down the Team? Social Effects of Team Incentives." *Journal of the European Economic Association*, 13 (5): 841–870. DOI: 10.1111/jeea.12131

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data." The Quarterly Journal of Economics, 120 (3): 917–962. DOI: 10.1093/qje/120.3.917

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2010. "Social Incentives in the Workplace." *Review of Economic Studies*, 77 (2): 417–458. DOI: 10.1111/j.1467-937X.2009.00574.x

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2013. "Team Incentives: Evidence from a Firm Level Experiment." *Journal of the European Economic Association*, 11 (5): 1079–1114. DOI: 10.1111/jeea.12028

Battigalli, Pierpaolo and Martin Dufwenberg. 2007. "Guilt in Games." American Economics Review, 97 (2): 170–176. DOI: 10.1257/aer.97.2.170

Bolton, Gary, and Peter Werner. 2016. "The Influence of Potential on Wages and Effort." *Experimental Economics*, 19: 535–561. DOI: 10.1007/s10683-015-9453-0

Bolton, Gary, and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity and Competition." *American Economic Review*, 90 (1): 166–193. DOI: 10.1257/aer.90.1.166

Breza, Emily, Supreet Kaur, and Yogita Shamdasani. 2018. "The Morale Effects of Pay Inequality." *Quarterly Journal of Economics*, 133 (2): 611–663. DOI: 10.1093/qje/qjx041

Chan, Tat Y., Jia Li, and Lamar Pierce. 2014a. "Compensation and Peer Effects in Competing Sales Teams." *Management Science*, 60 (8): 1965–84. DOI: 10.1287/mnsc.2013.1840

Chan, Tat Y., Jia Li, and Lamar Pierce. 2014b. "Learning from Peers: Knowledge Transfer and Sales Force Productivity Growth." *Marketing Science*, 33 (4): 463–84. DOI: 10.1287/mksc.2013.0831

Charness, Gary, and Martin Dufwenberg. 2006. “Promises and Partnership.” *Econometrica*, 74 (6): 1579–1601. DOI: 10.1111/j.1468-0262.2006.00719.x

Charness, Gary, Ramón Cobo-Reyes, Juan A. Lacomba, Francisco Lagos, and José María Pérez. 2016. “Social Comparisons in Wage Delegation: Experimental Evidence.” *Experimental Economics*, 19: 433–459. DOI: 10.1007/s10683-015-9448-x

Charness, Gary, Ninghua Du, Chun-Lei Yang, and Lan Yao. 2013. “Promises in Contract Design.” *European Economic Review, 64*: 194-208. DOI: 10.1016/j.euroecorev.2013.07.012

Charness, Gary, and Patrick Holder. 2019. “Charity in the Laboratory: Matching, Competition, and Group Identity.” *Management Science*, 65 (3): 1398–1407. DOI: 10.1287/mnsc.2017.2923

Charness, Gary, and Peter Kuhn. 2007. “Does Pay Inequality Affect Worker Effort? Experimental Evidence.” *Journal of Labor Economics*, 25 (4): 693–723. DOI: 10.1086/519540

Charness Gary, and Peter Kuhn. 2011. “Lab Labor: What Can Labor Economists Learn from the Lab?” In *Handbook of Labor Economics*, Vol. 4A, ed. Orley Ashenfelter and David Card, 229–330. Amsterdam: Elsevier, North-Holland. DOI: 10.1016/S0169-7218(11)00409-6

Charness, Gary, Peter Kuhn, and Marie Claire Villeval. 2011. “Competition and the Ratchet Effect.” *Journal of Labor Economics*, 29 (3): 513–547. DOI: 10.1086/659347

Charness, Gary, and Matthew Rabin. 2002. “Understanding Social Preferences with Simple Tests.” *Quarterly Journal of Economics*, 117 (3): 817–869. DOI: 10.1162/003355302760193904

Chen, Hua, and Noah Lim. 2013. “Should Managers Use Team-Based Contests?” *Management Science*, 59 (12): 2823–2836. DOI: 10.1287/mnsc.2013.1743

Cooper, David J., Krista Saral, and Marie Claire Villeval. 2021. “Why Join a Team?” *Management Science*, 67 (11): 6980–6997. DOI: 10.1287/mnsc.2020.3817

Dana, Jason, Daylian M. Cain, and Robyn M. Dawes. 2006. “What You Don’t Know Won’t Hurt Me: Costly (but Quiet) Exit in Dictator Games.” *Organizational Behavior and Human Decision Processes*, 100 (2): 193–201. DOI: 10.1016/j.obhdp.2005.10.001

Dohmen, Thomas, and Armin Falk. 2011. “Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender.” *American Economic Review*, 101 (2): 556–90. DOI: 10.1257/aer.101.2.556

Dechenaux, Emmanuel, Dan Kovenock, and Roman M. Sheremeta. 2015. “A Survey of Experimental Research on Contests, All-pay Auctions and Tournaments.” *Experimental Economics,18*: 609–669. DOI: 10.1007/s10683-014-9421-0

Eriksson, Tor, Anders Poulsen, and Marie Claire Villeval. 2009. “Feedback and Incentives: Experimental Evidence.” *Labour Economics*, 16 (6): 679−688. DOI: 10.1016/j.labeco.2009.08.006

Eriksson, Tor, and Marie Claire Villeval. 2008. “Performance-Pay, Sorting and Social Motivation.” *Journal of Economic Behavior & Organization*, 68 (2): 412−421. DOI: 10.1016/j.jebo.2007.10.003

Falk, Armin, and Urs Fischbacher. 2006. “A Theory of Reciprocity.” *Games and Economic Behavior*, 54 (2): 293–315. DOI: 10.1016/j.geb.2005.03.001

Falk, Armin and Andrea Ichino. 2006. “Clean Evidence on Peer Effects.” *Journal of Labor Economics,* 24 (1): 39–57. DOI: 10.1086/497818

Fehr, Dietmar, Hannes Rau, Stefan T. Trautmann, Yilong Xu, and Christoph Becker. 2023. “Fairness Properties of Compensation Schemes.” *Working Paper*, Heidelberg University. https://drive.google.com/file/d/1ppMnEbe1zaHixhd3XEtMHvlX1cU0BfTi/view

Fehr, Ernst, and Klaus M. Schmidt. 1999. “A Theory of Fairness, Competition, and Cooperation.” *Quarterly Journal of Economics*, 114 (3): 817–868. DOI: 10.1162/003355399556151

Fischbacher, Urs. 2007. “z-Tree: Zurich Toolbox for Ready-Made Economic Experiments.” *Experimental Economics*, 10 (2): 171–178. DOI: 10.1007/s10683-006-9159-4

Fryer Jr, Roland G., Steven D. Levitt, John List, and Sally Sadoff. 2012. “Enhancing the Efficacy of Teacher Incentives Through Loss Aversion: A Field Experiment.” *NBER Working Paper*. DOI: 10.3386/w18237

Gill, David, and Victoria Prowse. 2012. “A Structural Analysis of Disappointment Aversion in a Real Effort Competition.” *American Economic Review*, 102 (1): 469–503. DOI: 10.1257/aer.102.1.469

Gill, David, and Victoria Prowse. 2019. “Measuring Costly Effort Using the Slider Task.” *Journal of Behavioral and Experimental Finance*, 21: 1–9. DOI: j.jbef.2018.11.003

Gill, David, and Rebecca Stone. 2015. “Desert and Inequity Aversion in Teams.” *Journal of Public Economics*, 123: 42–54. DOI: 10.1016/j.jpubeco.2014.12.001

Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan. 2003. “Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation.” *Journal of Political Economy*, 111 (3): 465–497. DOI: 10.1086/374182

Hardin, Garrett. 1968. “The Tragedy of the Commons.” Science 162 (3859): 1243–1248. DOI: 10.1126/science.162.3859.1243

He, Simin, and Bin Miao. 2022. "Dominated Contract in Team Production." *Working paper*, Shanghai University of Finance and Economics. DOI: 10.2139/ssrn.3804149

Hjort, Jonas, Xuan Li, and Heather Sarsons. 2020. "Across-Country Wage Compression in Multinationals." *NBER Working paper*. DOI: 10.3386/w26788

Holmstrom, Bengt. 1982. "Moral Hazard in Teams." *Bell Journal of Economics*, 13 (2): 324–40. DOI: 10.2307/3003457

Hooper, Paul L., Hillard S. Kaplan, and Adrian V. Jaeggi. 2021. "Gains to Cooperation Drive the Evolution of Egalitarianism." Nature Human Behaviour, 5 (7): 847–856. DOI: 10.1038/s41562-021-01059-y

Irlenbusch, Bernd, and Gabriele K. Ruchala. 2008. "Relative Rewards within Team-Based Compensation." *Labour Economics*, 15 (2): 141–167. DOI: 10.1016/j.labeco.2007.02.003

Knez, Marc, and Duncan Simester. 2001. "Firm-Wide Incentives and Mutual Monitoring at Continental Airlines." *Journal of Labor Economics*, 19 (4): 743–772. DOI: 10.1086/322820

Kuhn, Peter, and Marie Claire Villeval. 2015. "Are Women More Attracted to Co-operation Than Men?" *Economic Journal*, 125 (582): 115–140. DOI: 10.1111/ecoj.12122

Lazear, Edward. 1989. "Pay Equality and Industrial Politics." *Journal of Political Economy*, 97 (3): 561–80. DOI: 10.1086/261616

Lazear, Edward. 1999. "Globalisation and The Market for Team-mates." *Economic Journal, 109* (454): 15–40. DOI: 10.1111/1468-0297.00414

Lazear, Edward P., and Kathryn L. Shaw. 2007. "Personnel Economics: The Economist's View of Human Resources." *Journal of Economic Perspectives*, 21 (4): 91–114. DOI: 10.1257/jep.21.4.91

Li, Jia, Noah Lim, and Hua Chen. 2019. "Examining Salesperson Effort Allocation in Teams: A Randomized Field Experiment." *Marketing Science*, 39(6):1122–141. DOI:10.1287/mksc.2019.1163

Lim, Noah, and Hua Chen. 2014. "When Do Group Incentives for Salespeople Work?" *Journal of Marketing Research*, 51 (3): 320–334. DOI: 10.1509/jmr.13.0322

Mas, Alexandre, and Enrico Moretti. 2009. "Peers at Work." *American Economic Review*, 99 (1): 112–45. DOI: 10.1257/aer.99.1.112

Nalbantian, Haig R., and Andrew Schotter. 1997. "Productivity under Group Incentives: An Experimental Study." *American Economic Review*, 87 (3): 314–341. http://www.jstor.org/stable/2951348

Niederle, Muriel, and Lise Vesterlund. 2007. "Do Women Shy away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics*, 122 (3): 1067–1101. DOI: 10.1162/qjec.122.3.1067

Pierce, Lamar, Laura W. Wang, and Dennis J. Zhang. 2021. "Peer Bargaining and Productivity in Teams: Gender and the Inequitable Division of Pay." *Manufacturing & Service Operations Management*, 23 (4): 933–951. DOI: 10.1287/msom.2019.0821

Porter, Gayle, and Michael Beyerlein. 2000. "Historic Roots of Team Theory and Practice." In *Work Teams: Past, Present and Future*, Vol 6, Chap 1, ed. Michael M. Beyerlein, 3–24. Netherlands: Springer. DOI: 10.1007/978-94-015-9492-9_1

Sandvik, Jason J., Richard E. Saouma, Nathan T. Seegert, and Christopher T. Stanton. 2020. "Workplace Knowledge Flows." *Quarterly Journal of Economics*, 135 (3): 1635–1680. DOI: 10.1093/qje/qjaa013

Siemsen, Enno, Sridhar Balasubramanian, and Aleda V. Roth. 2007. "Incentives That Induce Task-Related Effort, Helping, and Knowledge Sharing in Workgroups." *Management Science*, 53 (10): 1533–1550. DOI: 10.1287/mnsc.1070.0714

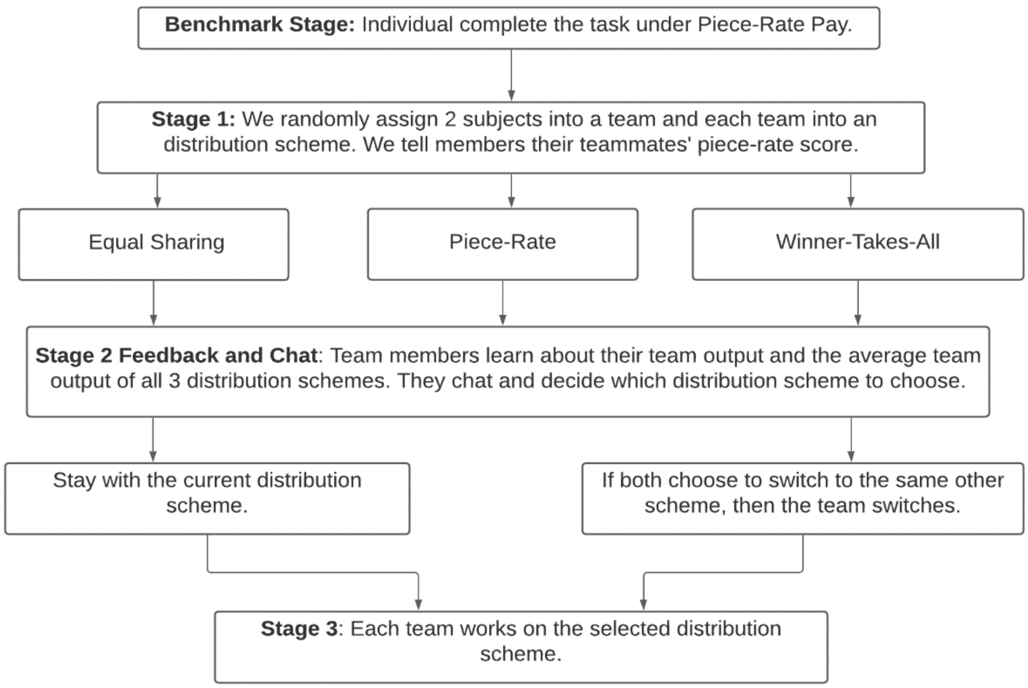
Siemsen, Enno, Aleda V. Roth, and Sridhar Balasubramanian. 2008. "How Motivation, Opportunity, and Ability Drive Knowledge Sharing: The Constraining-factor Model." *Journal of Operations Management*, 26 (3): 426–445. DOI: 10.1016/j.jom.2007.09.001

Müller, Wieland, and Andrew Schotter. 2010. "Workaholics and Dropouts in Organizations." Journal of the European Economic Association, 8 (4): 717–743. DOI: 10.1111/j.1542-4774.2010.tb00538.x

Wood, Stephen, Silvia Leoni, and Daniel Ladley. 2023. "Comparisons of the Effects of Individual and Collective Performance-related Pay on Performance: A review." *Human Resource Management Review*: 33 (4): 100982. DOI: 10.1016/j.hrmr.2023.100982

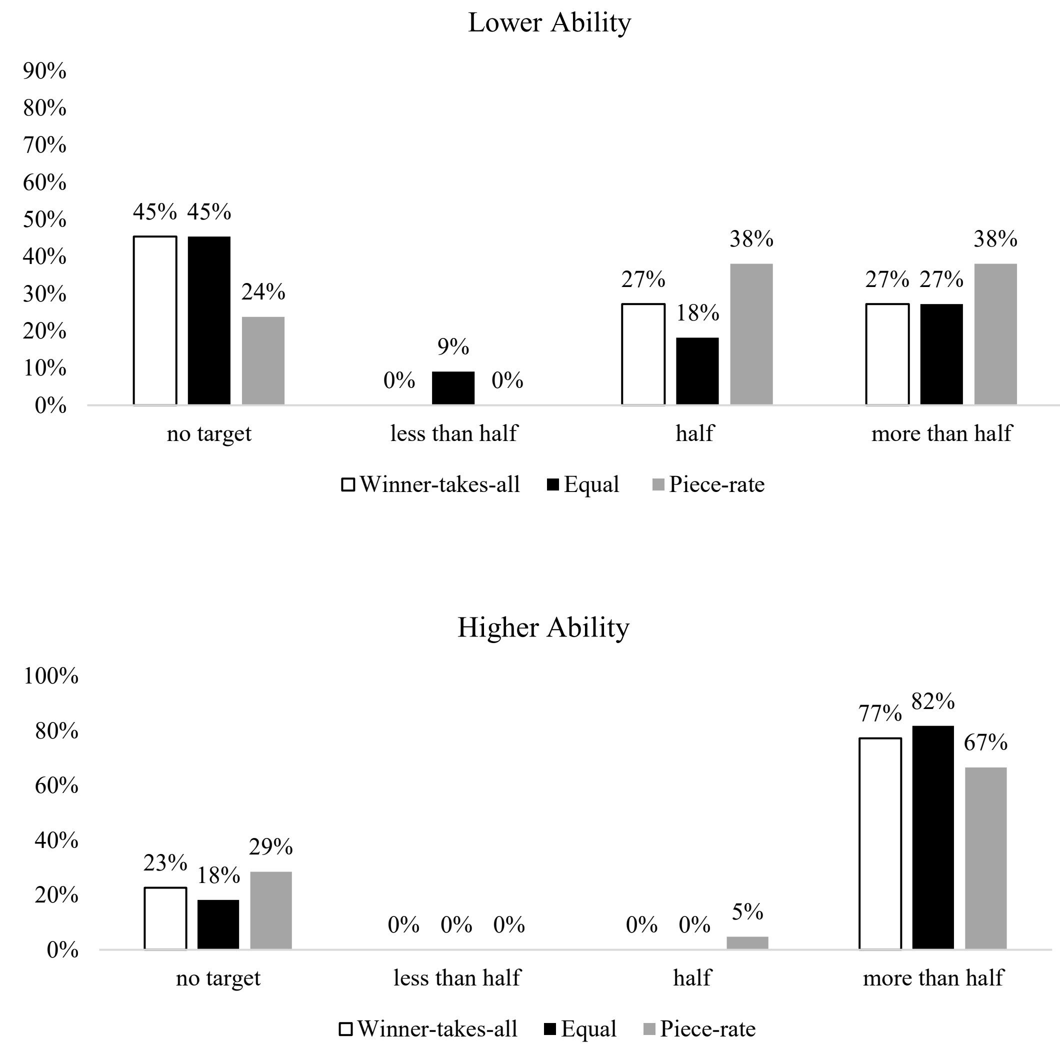
Ye, Maoliang, Jie Zheng, Plamen Nikolov, and Sam Asher. 2020. "One Step at a Time: Does Gradualism Build Coordination?" *Management Science*, 66 (1): 113–129. DOI: 10.1287/mnsc.2018.3210

Figure 1: Experiment Design Flow Chart



Note: This flow chart is the same for either the Threshold or the No Threshold condition. Participants remain under the same condition (either Threshold or No Threshold) for Stage 1 and Stage 3.

Figure 2: Individual Target Distribution in Team Production



Note: We show the individual target based on their benchmark piece-rate task performance (top panel: Lower ability; bottom panel: Higher ability). It shows the percentage of participants who either have no target, a target less than half of the team threshold, equal to half of the team threshold, or more than half of the team threshold.

Table 1: Determinants of Output in Stage 1 Task

VARIABLES	(1) All	(2) All	(3) Lower ability	(4) Higher ability
Benchmark stage output	0.726***	0.729***	0.638***	0.815***
	(0.055)	(0.056)	(0.106)	(0.065)
Equal_Stage1	2.469**	2.477**	4.367**	0.610
	(1.019)	(1.019)	(1.766)	(1.077)
Piece-rate_Stage1	1.630	1.639	2.105	1.251
	(1.044)	(1.040)	(1.802)	(0.938)
(*Winner-Take-All_stage1*)	-	-	-	-
Threshold		0.490	0.661	0.220
		(0.838)	(1.518)	(0.854)
Constant	12.070***	11.641***	13.403***	9.106***
	(2.341)	(2.476)	(4.320)	(2.893)
Equal_Stage1=Piece-rate_Stage1	p=0.37	p=0.38	p=0.12	p=0.55
Observations	248	248	122	126
R-squared	0.593	0.593	0.419	0.639

Note: The omitted category is Winner-Takes-All (in Stage 1). Robust standard errors in parentheses are clustered by group. *** p<0.01, ** p<0.05, * p<0.1. Lower ability and high ability refer to the lower ability and more able participant within the team according to their relative output in the benchmark stage. The row of "Equal_Stage1=Piece-rate_Stage1" reports the p-value of the equality test between the coefficients of Equal_stage1 and Piece-rate_stage1. In two groups where participants exhibit an equal number of outputs, all of them are categorized as higher ability. Results are highly consistent when we exclude these four participants.

Table 2: The Impact of "Individual Target" on Productivity in Stage 1

	(1)	(2)	(3)	(4)
	Threshold only		No Threshold & Threshold	
VARIABLES	Lower ability	Higher ability	Lower ability	Higher ability
Benchmark output	0.533***	0.854***	0.528***	0.697***
	(0.138)	(0.080)	(0.115)	(0.126)
Equal_Stage 1	4.169*	0.437	4.288**	0.881
	(2.130)	(1.283)	(1.860)	(1.081)
Piece-Rate_Stage 1	1.458	1.540	0.777	1.250
	(2.336)	(1.092)	(1.978)	(1.081)
Bigger-Than-Half	4.868**	-1.915	4.089**	0.373
	(2.113)	(1.480)	(1.869)	(1.054)
Half	-2.077	-2.771*	-3.103*	-3.012***
	(1.996)	(1.421)	(1.693)	(1.074)
Less-Than-Half	-5.572**		-6.291***	-
	(2.365)		(1.555)	
(*No individual target)*	-	-	-	-
Constant	17.418***	9.393**	21.707***	16.626***
	(4.375)	(4.136)	(4.328)	(5.622)
Observations	65	67	113	117
R-squared	0.465	0.720	0.369	0.481

Note: Robust standard errors clustered by team in parentheses for columns 1 and 2. Controlling for gender does not change the qualitative nature of the results. *** p<0.01, ** p<0.05, * p<0.1. Because 18 participants are missing due to software crash for the survey part of that session, regressions with survey information have smaller numbers of observations than those in Table 1.

Table 3.
Team's Stage 2 Incentive Scheme Choice Based on Stage 1 Assignment

Panel A: Choices of All Subjects

Team choice Randomly	Stage 2: Equal	Stage 2: Piece-Rate	Stage 2: Winner-Takes-All
Stage 1: Equal	15 (36%)	27 (64%)	0 (0%)
Stage 1: Piece Rate	6 (14%)	36 (86%)	0 (0%)
Stage 1: Winner-Takes-All	13 (32.5%)	25 (62.5%)	2 (2%)
Total Teams	34 (27%)	88 (71%)	2 (2%)

Panel B: Choices of Subjects in No Threshold Condition

Team choice Randomly	Stage 2: Equal	Stage 2: Piece-Rate	Stage 2: Winner-Takes-All
Stage 1: Equal	3 (18%)	14 (82%)	0
Stage 1: Piece Rate	1 (6%)	16 (94%)	0
Stage 1: Winner-Takes-All	4 (27%)	11 (73%)	0
Total Teams	8 (16%)	41 (84%)	0

Panel C: Choices of Subjects in Threshold Condition

Team choice Randomly	Stage 2: Equal	Stage 2: Piece-Rate	Stage 2: Winner-Takes-All
Stage 1: Equal	12 (48%)	13 (52%)	0
Stage 1: Piece Rate	5 (20%)	20 (80%)	0
Stage 1: Winner-Takes-All	9 (36%)	14 (56%)	2 (8%)
Total Teams	26 (35%)	47 (63%)	2 (3%)

Table 4: Determinants of Choosing Equal Sharing in Stage 2

	(1)	(2)	(3)
VARIABLES	Dependent Variable =1 if choose equal distribution		
Equal_Stage 1	-0.046	-0.046	-0.057
	(0.106)	(0.102)	(0.100)
Piece-Rate_Stage 1	-0.180*	-0.154	-0.171
	(0.098)	(0.096)	(0.106)
(Winner-takes-all_Stage1)	-		-
Lower-Ability	0.029**	0.028*	0.029*
	(0.015)	(0.015)	(0.015)
Male	0.044	0.062	0.071
	(0.053)	(0.053)	(0.052)
Threshold	0.211***	0.130	0.123
	(0.078)	(0.080)	(0.078)
Chat-cooperation		0.332***	0.386***
		(0.121)	(0.128)
Chat-gap			0.005
			(0.092)
Chat-scheme			0.045
			(0.124)
Chat-fairness			-0.189
			(0.152)
Constant	0.193**	0.163*	0.174*
	(0.092)	(0.092)	(0.091)
Observations	226	226	226
R-squared	0.088	0.162	0.177

Note: Observations are at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Summary of the Chatting Variables (mean)

Chat Concept	No Threshold (n=49 teams)	Threshold (n=75 teams)
Cooperation	4%	27%
Benchmark ability gap	31%	36%
Originally randomly assigned scheme	14%	13%
Fairness	8%	11%

Note: The table shows the proportion of the teams discussing the corresponding item.

Table 6: Determinants of Output in Self-selected Stage 3

VARIABLES	(1) All	(2) Lower ability	(3) High ability
Benchmark stage output	0.635***	0.584***	0.696***
	(0.068)	(0.111)	(0.105)
Equal_Stage 1	2.491**	3.410*	1.008
	(1.183)	(1.946)	(1.086)
Piece-rate_Stage 1	0.976	0.291	1.117
	(1.250)	(1.982)	(1.024)
(*Winner-Takes-All_Stage 1*)	-	-	-
Threshold	1.117	1.290	1.167
	(1.006)	(1.577)	(1.002)
Constant	18.039***	19.510***	15.931***
	(3.138)	(4.611)	(4.716)
Equal_Stage1=Piece-rate_Stage1	p=0.19	p=0.05	p=0.93
Observations	248	122	126
R-squared	0.474	0.323	0.500

Note: Robust standard errors in parentheses. In Column 1, the robust standard error is clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1 The omitted category is *winner-takes-all* in Stage 1 for variables "Equal_Stage1" and "Piece-rate_Stage1". The row of "Equal_Stage1=Piece-rate_Stage1" reports the p-value of the equality test between the coefficients of Equal_stage1 and Piece-rate_stage1.

Table 7: Determinants of Output in Stage 1 Controlling for Benchmark Peer Output

VARIABLES	(1) Lower ability	(2) Higher ability
Benchmark other-output	0.204	-0.013
	(0.135)	(0.057)
Benchmark self-output	0.552***	0.823***
	(0.101)	(0.073)
Equal _Stage 1	4.063**	0.607
	(1.735)	(1.080)
Piece-rate_Stage 1	1.918	1.263
	(1.762)	(0.951)
(*Winner-Takes-All_Stage 1*)	-	-
Threshold	0.798	0.191
	(1.544)	(0.855)
Constant	7.336	9.186***
	(6.479)	(2.924)
Observations	122	126
R-squared	0.437	0.640

Note: "Benchmark self-output" and "Benchmark other-output" refer to the own output and the output of the team partner in the benchmark stage, respectively. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Distributions (equal, piece-rate) are randomly assigned distribution in Stage 1.

Online Appendix (for online publications only)

Appendix A: Experimental Instructions[1]

[Note: Instructions below are based on the No Threshold condition. Additional information pertaining to the Threshold condition are *in italics*. Words in brackets [] serve as clarification for readers and are not part of the participants' instructions.]

Thank you for participating in this experiment! You have earned 10 RMB show-up fee for showing up on-time; your other earnings in today's session will be determined by your decisions (i.e., experimental earnings). Your total earnings today will be the sum of the show-up fee and your experimental earnings. Therefore, please read the instructions below carefully! Please keep silent and do not peek at others' screens. Please avoid kicking the wires below the tables! We reserve the right to ask a participant to leave without being paid for breaking these rules.

All of your decisions will be anonymous. Participants will not receive any identifying information about others either during or after this session.

You will participate in three experiments: Experiment I, II and III. These experiments are independent from each other. That is, your earnings in one experiment will not affect your earnings in the other experiments.

Your earnings from the three experiments will be added to your 10 RMB show-up fee. At the end of today's session, you will receive your total payment.

If you have questions now or at any point during the experiments, please raise your hand and an experimenter will come to answer your questions privately.

Below are the instructions for Experiment I.

Experiment I
[Benchmark Stage in main text]**:**

In this experiment, you will be asked to drag a slider bar to earn money. The slider bar starts on the left at the "0" position, and ends on the right at the "100" position. Your task is to move the slider bar to exactly the middle, to the "50" position. The number to the right side of the slider shows the exact position of the bar.

For each slider bar that moves to exactly the "50" position, you score 1 point.

In today's experiments, 1 point earns 0.3 RMB; every 10 points earns 3 RMB.

1 point =0.3RMB

[1] The original instructions are in Chinese and available upon request.

You have 4 minutes (240 seconds) to move as many slider bars as possible to position 50.

If you have any questions, please raise your hand and an experimenter will come to assist you.

Instruction for Experiment II will be distributed after Experiment I; Experiment III's instruction will be distributed after Experiment II.

If you are done with Experiment I, please lift up your head and raise your hand to let the experimenters know.

Experiment II
[Stage 1 in main text]

Your decisions in Experiment II will not influence your earnings Experiment I. Your earnings from Experiment I, II and the show-up fee will be combined to equal your total earnings.

In this experiment, you and one other participant will be randomly assigned into a team. Each team will then be randomly assigned into one of three distribution schemes:

Scheme 1: Equal sharing
Scheme 2: Piece-rate
Scheme 3: Winner-takes-all

That is, in this experiment, about a third of the teams will be assigned to the same distribution scheme as yours, while two thirds of the teams will be assigned to the other two schemes.

The differences between the three distribution schemes are:

Scheme 1: Equal sharing
Each of your scores will be equal to half of the total points earned by your team.

Scheme 2: Piece-rate
Your score is based on the points you earned.

Scheme 3: Winner-takes-all
The member of your team who moved the most slider bars to the "50" position earns all the points your team has collectively scored; the member who moved the fewest slider bars to the "50" position earns zero points. If you and the other member move the same number of sliders bars to position "50", the computer will randomly assign one of you to earn all the points; the other will earn zero points.

[Information in italics pertains only under Threshold condition]

Things in common across all three schemes are:

Only when your team has moved at least equal to or higher than the team productivity threshold (that will be shown on your screen) will your points be allocated according to the distribution schemes; otherwise, you will both score zero.

We use the same calculation method to calculate each group's threshold.

You have 4 minutes (240 seconds) to move as many slider bars as possible to "50".

If you are done with Experiment II, please lift your head and raise your hand to let the experimenters know.

Experiment III
[Stage 2 and Stage 3 in main text]

This is the last experiment for today's session. In this experiment, your team member is the same as the one in Experiment II, and s/he also knows that you are the same member as that in Experiment II. Your earnings in this experiment will not influence your earnings in the previous two experiments. Your earnings from this experiment will be added to your total earnings.

Similar to Experiment II, your score will be determined by the **total number of sliders bars completed.** [Pertains only under Threshold condition] *When the total number of slider bars your team has moved to "50" is equal to or bigger than the team threshold (the threshold in Experiment III is the same as the threshold determined in Experiment II), you will have a chance to earn your points; otherwise your team earns zero points.*

[Below is the chatting and choosing, the Stage 2 in Paper]

However, in this experiment, you and your team member have a chance to discuss which distribution scheme you want to use to share your earned points. You have three choices:

Scheme 1: Equal sharing
Each of your scores will be equal to half of the total points earned by your team.

Scheme 2: Piece-rate
Your score is based on the points you earned.

Scheme 3: Winner-takes-all
The member of your team who moved the most slider bars to the "50" position earns all the points your team has collectively scored; the member who moved the fewest slider bars to the "50" position earns zero points. If you and the other member move the same number of sliders bars to position "50", the computer will randomly assign one of you to earn all the points; the other will earn zero points.

However, **if and only if** both members choose the same scheme will your choice take effect; otherwise, your distribution scheme in Experiment III stays the same as that in Experiment II.

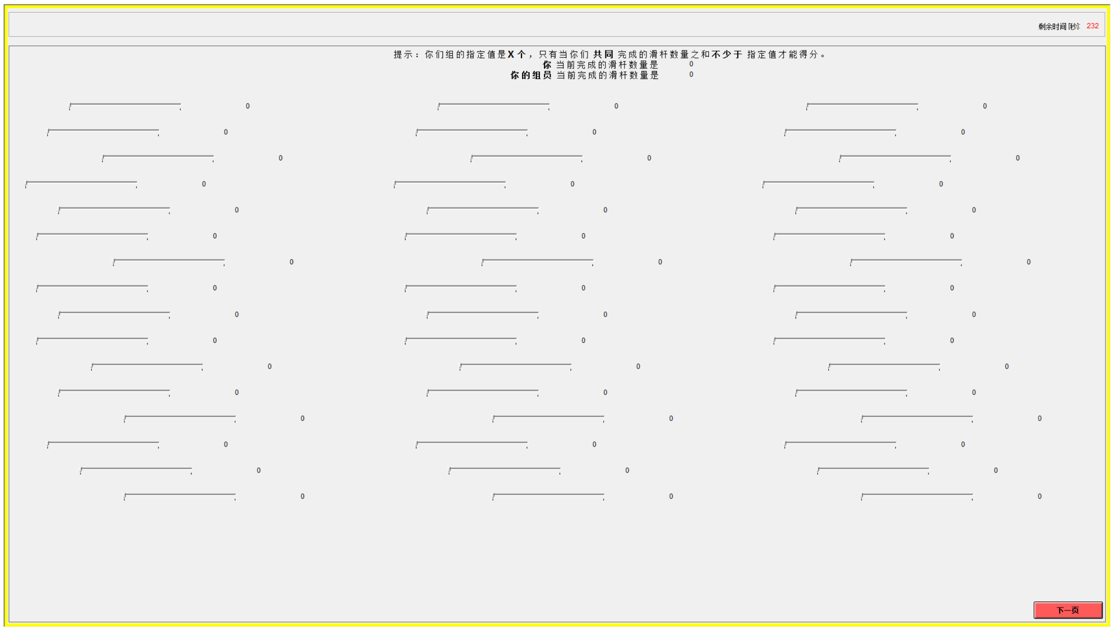
You have up to 4 minutes to discuss with your team member which distribution scheme you prefer (in the discussion window you can click Ctrl+ Space to switch to the language of Chinese simplified). During your communication with each other:

1. No revealing of your identity (e.g., age, sex, major)
2. No threatening language.

[Pertains only under Threshold condition] *To emphasize, you earn points based on the three distribution schemes only if your team's total number of slider bars moved to "50" is equal to or bigger than the team threshold (Experiment III threshold is the same as that determined in Experiment II); otherwise, both of you score zero.*

[This Task refers to Stage 3 in paper]. **You have 4 minutes (240 seconds) to move as many slider bars as possible to "50".**

Screenshot for real-effort task (per Gill and Prowse, 2012). The participants' goal is to move sliders to the middle of each bar. The first row reminds team members of the team threshold (only for the Threshold condition). The next two rows show the individual and the individual's teammate's current output level.



Appendix B: More Results

B1. Robustness Check on Approximate Output Measure

Table B1: **Determinants of Approximate Output in Stage 1 Task**

	Approximate Output (Positioned at 48–52)				Approximate Output (Positioned at 49–51)			
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	All	Lower ability	Higher ability	All	All	Lower ability	Higher ability
Benchmark stage output	0.477***	0.481***	0.471***	0.433***	0.506***	0.509***	0.514***	0.436***
	(0.050)	(0.053)	(0.092)	(0.097)	(0.054)	(0.056)	(0.090)	(0.097)
Equal_Stage1	2.061**	2.069**	3.537**	0.965	2.251**	2.258**	3.834**	1.088
	(0.838)	(0.843)	(1.560)	(0.846)	(0.871)	(0.875)	(1.547)	(0.848)
Piece-rate_Stage1	1.275	1.284	1.279	1.476*	1.524*	1.532*	1.697	1.576*
	(0.868)	(0.865)	(1.579)	(0.844)	(0.893)	(0.891)	(1.564)	(0.848)
Threshold		0.526	1.077	-0.048		0.464	0.923	-0.015
		(0.793)	(1.459)	(0.694)		(0.798)	(1.451)	(0.700)
Constant	22.127***	21.666***	20.779***	24.624***	20.673***	20.267***	18.858***	24.355***
	(2.194)	(2.547)	(4.172)	(4.569)	(2.377)	(2.693)	(4.075)	(4.570)
Equal_Stage1=Piece-rate_Stage1	P=0.35	P=0.35	P=0.10	P=0.54	P=0.39	P=0.39	P=0.12	P=0.56
Observations	248	248	122	122	248	248	122	122
R-squared	0.447	0.448	0.325	0.431	0.480	0.481	0.370	0.434

Note: We define approximate output as the number of slider bars moved to the position between 48 and 52 (Columns 1–4) or between 49 and 51 (Column 5–8). The omitted category is Winner-Takes-All (in Stage 1). Robust standard errors in parentheses are clustered by group. *** p<0.01, ** p<0.05, * p<0.1. Lower ability and high ability refer to the lower ability and more able participant within the team according to their relative output in the benchmark stage. The row of "*Equal_Stage1=Piece-rate_Stage1*" reports the p-value of the equality test between the coefficients of *Equal_Stage1* and *Piece-rate_Stage1*. In Columns 4 and 8 we drop the four participants in which both members' outputs are equal within the team.

B2. Does Within-team Ability Gap Discourage Workers in Winner-takes-all?

Table B2: Effects of Ability Gap on Lower and Higher Ability Workers' Productivity

VARIABLES	(1) Lower	(2) Lower	(3) Lower	(4) Lower	(5) Lower	(6) Higher	(7) Higher	(8) Higher	(9) Higher	(10) Higher
Benchmark stage output	1.101***	1.092***	1.104***	0.990***	0.823***	0.846***	0.843***	0.863***	0.853***	0.851***
	(0.181)	(0.188)	(0.159)	(0.175)	(0.176)	(0.082)	(0.081)	(0.073)	(0.077)	(0.092)
Threshold	2.853	3.184	2.717	2.582	2.855	0.117	0.159	0.082	0.082	0.189
	(2.805)	(2.734)	(2.633)	(2.754)	(2.946)	(1.524)	(1.563)	(1.474)	(1.547)	(1.534)
Ability gap	0.714***	1.025**				0.091	0.204			
	(0.210)	(0.447)				(0.081)	(0.232)			
Ability gap^2		-0.010					-0.004			
		(0.016)					(0.006)			
Ability gap >10			11.835***					2.988**		
			(2.465)					(1.299)		
Ability gap >8				8.886***					2.054	
				(2.788)					(1.326)	
Ability gap >5					5.700**					0.443
					(2.606)					(1.416)
Constant	-10.443	-11.902	-8.149	-3.769	2.009	6.865*	6.447*	5.764*	6.476*	7.305*
	(8.122)	(7.834)	(6.442)	(7.266)	(7.036)	(3.397)	(3.529)	(3.233)	(3.307)	(3.633)
Observations	39	39	39	39	39	41	41	41	41	41
R-squared	0.614	0.619	0.633	0.577	0.522	0.766	0.768	0.791	0.774	0.759

Note: Lower and higher refer to the lower- and higher- ability participant within the team in *winner-takes-all* scheme according to their relative output in the benchmark stage. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

B3. Comparing Stage 1 and Stage 3

In Table B2, we conduct a more detailed comparison between Stage 1, where participants are randomly assigned distribution schemes, and Stage 3, during which participants can select their preferred distribution scheme. Overall, we find participants' productivity to increase from Stage 1 to Stage 3, suggesting the premium of endogenous chosen scheme over the randomized assigned scheme, with the caveat of the potential learning effect over stages. The increment in Threshold treatment appears to be larger than that in No threshold (mean=3.45 vs 2.40, p=0.086). While this disparity between the two treatments do not differ among those who have been initially assigned to *equal* sharing mechanism (mean = 3.14 vs 3.09, p = 0.71), we discern a more substantial increase for participants randomly assigned to the other two sharing mechanisms: *piece-rate* and the *winner-takes-all*. Specifically, the average productivity increase for *piece-rate* and *winner-takes-all* combined in threshold (n= 100, mean = 3.61) exceeds the respective increment in the *No threshold* treatment (n=64, mean = 2.03, p = 0.07).

We also investigate how individual productivity changes from Stage 1 to Stage 3 when the team stays with the randomized distribution or switches to a new distribution. It is important to note that the interpretation of causality is constrained by the limited sample size and the potential endogeneity of distribution scheme changes. In the No Threshold condition, teams that chose to change their distribution scheme increased team members' output significantly, whereas those who stayed in the same distribution scheme showed much smaller increases in output. This would support a conclusion that changes raised productivity and thus allowing teams to negotiate modes of compensation among members has an economic payoff, which suggests that a market in which team members bargain over compensation schemes improves the performance of the firm.

For members in Threshold, however, the results show a different pattern. Only one group, those assigned *piece rate* in Stage 1 show a larger gain for switchers than for stayers – a result consistent with *equal pay* outscoring *piece rate*. But we see the opposite pattern for the groups assigned equal in Stage 1, with switchers having a smaller gain (3.08) than stayers (3.21), albeit a statistically insignificant difference. For teams randomly assigned to *winner-take-all*, there are big gains for those that switch (4.09) but even bigger gains for the 4 individuals that stuck with winner-take-all (5.25). What might explain the high gains to the four *winner-take-all* stay individuals is that both team members had high expectations of winning and increased their output in expectation of outdoing each other. As we did not ask those subjects why they did not switch nor why they increased their production so much, however, and do not have enough cases to explore, we cannot do more with this data but point out the problems for future research and the need for a much larger sample and probing of persons in the *winner-takes-all* stay group.

Table B3: Individual Productivity Changes Over Stages

Panel A: No Threshold	**Stage 3 – Stage 1**	**Switched in Stage 3?**	**Stage 3 – Stage 1**
Overall differences (n= 98)	2.40 (0.588)		
Stage 1 Equal (n=34)	3.09 (0.845)	Yes (n=28)	3.46 (0.99)
		No (n= 6)	1.33 (1.02)
Stage 1 Piece-Rate (n=34)	1.56 (0.872)	Yes (n=2)	3.5 (0.5)
		No (n=32)	1.44 (0.92)
Stage 1 Winner-takes-All (n=30)	2.47 (1.355)	Yes (n=30)	2.6 (1.36)
		No (n=0)	N/A
Panel B: Threshold	**Stage 3 – Stage 1**	**Switched in Stage 3?**	**Stage 3 – Stage1**
Overall differences (n=150)	3.45 (0.407)		
Stage 1 Equal (n=50)	3.14 (0.705)	Yes (n=26)	3.08 (1.03)
		No (n= 24)	3.21 (0.97)
Stage 1 Piece-Rate (n=50)	3.04 (0.643)	Yes (n=10)	3.6 (1.88)
		No (n=40)	2.9 (0.665)
Stage 1 Winner-takes-All (n=50)	4.18 (0.766)	Yes (n=46)	4.09 (0.82)
		No (n=4)	5.25 (1.97)

Note: n refers to the number of individuals accordingly. The number under "Stage 3 – Stage 1" column shows the average of the productivity change with the standard error in parenthesis.

Appendix C:

A Simplified Guilt-aversion Analysis of Why Less Productive Workers Raised Output in Equal Sharing

We examine the implications of guilt-aversion behavior for the *effort* of lower/higher ability participants in a two-person team. We assume that participants seek to maximize a utility function in which earnings enter positively while effort enters negatively; and where they suffer disutility from guilt aversion of letting down their team. Let e_L and e_H represent the performance/effort[2] of the lower and higher ability persons in Stage 1. Let p_L^0 and p_H^0 represent their *ex-ante beliefs* of the probability that they will outperform their partner in this stage. We assume that $p_L^0 \in [0, 0.5)$ and $p_H^0 \in (0.5,1]$ are constant and determined by their within-team relative abilities in the benchmark stage, which are common information for both members that they learn at the beginning of Stage 1.[3]

Then the lower or higher ability's *expected* earnings are: $\frac{e_L+e_H}{2}$ under equal sharing; e_L or e_H under *piece-rate,* and $p_L^0.(e_L + e_H)$ or $p_H^0.(e_L + e_H)$ under *winner-takes-all*. We represent the cost of effort with a quadratic cost function $\frac{c}{2}e^2$, where $e = e_L$ or e_H captures the effort level and $c = c_L$ or c_H captures the cost for the lower and higher abilities, respectively.

The heart of the model is the guilt aversion disutility when the team member believes that they let down the other person per the social norm argued by Charness and Dufwenberg (2006): "one central idea [in the literature on social norms] is to view a social norm as a moral expectation, which people are inclined to live up to, (for which) … guilt aversion can provide a ...kind of micro foundation." Put simply, the norm determines A's expectation, which B seeks to live up to because B would feel guilty if he did not."[4] We assume that the norm is for earnings proportionate to one's share of the team output, per *Piece-rate*. Thus, we let $\pi_H^e = e_H$ and $\pi_L^e = e_L$ represent what the

---

[2] Here we consider (expected) performance as a linear function of effort and do not distinguish between performance and effort.

[3] Although the *actual* outperforming probability may be affected by efforts, it largely centers on commonly-known relative abilities. Thus, to simplify our analysis below and to focus on discussions of guilt aversion, we assume constant *ex ante belief* of the outperforming probability. Such simplification is more likely to apply when the ability gap is large and transparent, which is the focus of our study.

[4] In our framework, there exists a nuanced divergence from the approach of Charness and Dufwenberg (2006). Specifically, player *i* perceives player *j*'s expectation to be rooted in *j*'s individually 'deserved' piece rate pay, as opposed to being based on *i*'s action.

higher and lower ability team members think they *should* deserve to earn. The earnings an individual receives in *equal* and *winner-takes-all* are based on the distribution schemes, own output, and the output of the other team member. An individual would feel guilty when the team member's actual earnings are *less* than what this member deserves but would not feel guilty when the team member receives more than or equal to this deserved amount.

Lower Ability's Guilt Aversion

Let θ (θ_L for the lower ability person and θ_H for the higher ability person) capture an individual's guilt aversion preference, with $\theta \geq 0$. Based on the distance between the high ability's earnings and the earnings they deserve, the lower ability's guilt aversion is $\theta_L * max\ (\frac{e_H - e_L}{2}, 0)$ under *equal*, 0 under *piece-rate*, and $\theta_L p_L^0 e_H$ under *winner-takes-all*. Intuitively, the lower ability believes that the higher ability expects to receive e_H. In *equal*, the higher ability receives $\frac{e_L + e_H}{2}$, thus the gap between the two is $\frac{e_H - e_L}{2}$. If the performance of the lower ability improved from the benchmark stage to exceed that of the higher ability ($e_L > e_H$ in Stage 1), guilt aversion would be 0. Taking together, we have $\theta_L * max\ (\frac{e_H - e_L}{2}, 0)$.

Since in *piece-rate*, each person earns what they contributed to the team output, the lower ability's guilt aversion is zero toward the higher ability person: $\theta_L * (e_H - e_H) = 0$. In *winner-takes-all*, the higher ability persons get $(e_H + e_L)$ if they win the tournament, which exceeds e_H, and 0 when they lose, which is lower than e_H. Thus, in *winner-takes-all* the lower ability's guilt aversion is $\theta_L p_L^0 e_H$.

High ability's Guilt Aversion

If the higher ability persons believe they will continue to perform ahead of their lower ability counterparts, they will not have any guilt under *equal.* In general, guilt aversion would be $\theta_H * max\left(\frac{e_L - e_H}{2}, 0\right)$ in *equal*. Only when e_L exceeds e_H in *equal* will they experience any guilt aversion. The higher ability will not experience any guilt aversion in *piece-rate*. In *winner-takes-all,* their guilt aversion would be the difference between what the lower ability deserves to receive (e_L) and what the lower ability actually receives; taking together, the higher ability's guilt aversion is $\theta_H p_H^0 e_L$.

The Utility Function

Taking all three parts together, the lower ability would maximize $U(e_L, e_H, c_L, \theta_L)$ and the higher ability would maximize $U(e_L, e_H, c_H, \theta_H)$ with optimal effort levels as shown in appendix Tables C1 and C2.

Table C1: Utility Function in No Threshold

	Lower Ability	Higher Ability
Equal	$\frac{e_H+e_L}{2} - \frac{c_L}{2}e_L^2 - \theta_L * \max(\frac{e_H-e_L}{2}, 0)$	$\frac{e_H+e_L}{2} - \frac{c_H}{2}e_H^2 - \theta_H * \max(\frac{e_L-e_H}{2}, 0)$
Piece-rate	$e_L - \frac{c_L}{2}e_L^2$	$e_H - \frac{c_H}{2}e_H^2$
Winner-takes-all	$p_L^0 * (e_L + e_H) - \frac{c_L}{2}e_L^2 - \theta_L p_L^0 e_H$	$p_H^0 * (e_L + e_H) - \frac{c_H}{2}e_H^2 - \theta_H p_H^0 e_L$

Table C2: Optimal Effort Level in No Threshold[5]

	Lower Ability	Higher Ability
Equal	$\frac{1+\theta_L(1-p_L^0)}{2c_L}$	$\frac{1+\theta_H(1-p_H^0)}{2c_H}$
Piece-rate	$\frac{1}{c_L}$	$\frac{1}{c_H}$
Winner-takes-all	$\frac{p_L^0}{c_L}$	$\frac{p_H^0}{c_H}$

This framework has predictions for the rank of the lower ability's effort levels in three distribution schemes:

- When $\theta_L = \frac{1}{1-p_L^0}$, *equal = piece-rate > winner-takes-all.*
- When $\theta_L > \frac{1}{1-p_L^0}$, *equal > piece-rate > winner-takes-all.*
- When $\theta_L \in [0, \frac{1}{1-p_L^0})$, *equal < piece-rate*; if $\theta_L > \frac{2p_L^0-1}{1-p_L^0}$ (it holds because $p_L^0 < 0.5$), *equal > winner-takes-all*.

[5] Under the extreme case of $p_L^0 = 0$ and $p_H^0 = 1$, i.e., the lower (or higher) ability believes that he/she has a probability of zero (or one) to outperform the teammate, the optimal effort of the lower ability under *equal* and *winner-takes-all* would be $\frac{1+\theta_L}{2c_L}$ and 0 respectively, whereas that of the higher ability would be $\frac{1}{2c_H}$ and $\frac{1}{c_H}$ respectively.

In sum, *equal* > *winner-takes-all* for all $\theta_L \geq 0$, i.e., all degrees of the lower ability's guilt aversion preference; whereas *equal* > *piece-rate iff* $\theta_L > \frac{1}{1-p_L^0}$, i.e., when lower ability's guilt aversion preference is stronger than a cutoff parameter.[6]

Threshold Condition

The Threshold condition shares the main features of the No Threshold condition with a kink at the team threshold, which produces higher or lower income depending on whether the group did or did not exceed the threshold. Let p_R represent the probability of reaching the threshold. It will be a function determined by the effort level of the two team members, as $p_R(e_L, e_H)$. The new earnings part would involve a multiplication with $p_R(e_L, e_H)$. The cost part remains the same as in No Threshold condition.

Guilt Aversion

The guilt aversion would depend on whether the team reaches the threshold. When they have reached the threshold, the part of disutility from guilt is as in No Threshold. When they fail to reach the threshold, the size of guilt aversion could depend on how much individuals believe they are responsible for this failure. We use γ to capture it, with $0 \leq \gamma \leq 1$. When $\gamma = 0$, individuals do not think they are responsible for the failure and thus do not feel guilty. Contrarily, when $\gamma = 1$, individuals believe that they are fully responsible for the failure, and they would suffer from guilt toward their partner. In this case, guilt aversion for a lower ability would be $\gamma_L \theta_L e_H$, while that for the higher ability would be $\gamma_H \theta_H e_L$. Taking together, we predict:

a. Lower and higher abilities' effort increases with γ_L or γ_H, the extent they think they are responsible for the failure to reach the threshold.
b. Lower and higher abilities' effort increases with θ_L or θ_H, the extent they feel guilty when their counterparts receive less than what they deserve.

[6] Under more general setups (e.g., without assuming *ex ante* belief of the outperforming probability), the exact condition for the effort ranks across distribution schemes can be different from what we list here. But the qualitative pattern holds that with a sufficiently strong guilt aversion preference, the low ability's productivity under *equal* sharing can exceed that under *piece-rate* or *winner-takes-all*.

Appendix D: Motivations of Distribution Choices from Chat Messages in Stage 2

Table D1: Overview of Motivations of Distribution Choices from Chat Messages

Motivation Categories	Frequency (# of teams) Total =76	Frequency (# of teams) With Threshold
Supporting the Preference of Piece Rate (PR)		
A. To avoid guilt feeling, the lower ability participant proposes *PR*.	18	13 (72%)
B. To avoid guilt feeling, the lower ability participant states he/she would work harder should *EQ* be chosen.	7	6 (86%)
C. The higher ability participant proposes *PR.* Reasons: *PR* is fair to himself (5 groups); *PR* can incentivize his partner (5 groups); unmentioned (5 groups).	15	12 (80%)
D. The higher ability teaches the task hints and then advocates *PR*.	9	8 (89%)
E. Both participants consider PR as fairest.[7]	5	2 (40%)
F. The higher ability persuades the lower ability to choose *PR* as a safer/better choice for the latter than (the default) *WTA*.	3	2 (67%)
Supporting the Preference of Equal (EQ)		
G. The higher-ability participant would choose *EQ* because of altruism or cooperation consideration (mostly with group threshold).	10	9 (90%)
H. Not only the lower ability but also the higher one considers *EQ* a powerful team incentive to spur both of their productivity.	3	3 (100%)
I. The participants who dislike competitions would choose *EQ*.[8]	2	1 (50%)
J. The higher ability teaches the task hints and then propose *EQ*.	2	2 (100%)
K. The lower ability participant proposes *EQ*. Reasons: *EQ* serves his interest (1 group); unmentioned (1 group).	2	2 (100%)

Note:
(1) *EQ*: Equal-sharing; *PR*: Piece-rate; *WTA*: Winner-takes-all.
(2) Our total sample includes 124 groups, with 75 of them (or 60.5%) under the Threshold condition and 49 (or 39.5%) under the No-threshold condition. There are 46 groups whose chats evince their choice motivations for distribution schemes, whereas other group chats do not reveal direct motive information. Some group chats may belong to multiple motivation categories.
(3) Frequency indicates the number of teams whose chats reflect the corresponding motivation categories.
(4) In two groups the lower ability participant proposes *PR* without any explanation.
(5) The original detailed chat messages (in Chinese) are available upon request.

[7] In these 5 groups, 2 low-ability participants consider *EQ* unfair for the higher-ability, while 4 participants (3 high-ability and 1 low-ability) consider *WTA* unfair.
[8] In one group, the high ability thinks so, while in another group both participants do so.

Supplementary References

Charness, Gary, and Martin Dufwenberg. 2006. "Promises and Partnership." *Econometrica*, 74 (6): 1579–1601. DOI: 10.1111/j.1468-0262.2006.00719.x

Gill, David, and Victoria Prowse. 2012. "A Structural Analysis of Disappointment Aversion in a Real Effort Competition." *American Economic Review*, 102 (1): 469–503. DOI: 10.1257/aer.102.1.469