Labour Economics 49 (2017) 162–178

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

Collaboration and free-riding in team contests[☆]

Mürüvvet Büyükboyacı[a,*], Andrea Robbett[b]

[a] Department of Economics, Middle East Technical University, Ankara 06800, Turkey
[b] Department of Economics, Middlebury College, United States

ARTICLE INFO

JEL classification:
C91
C92
C71
C72
D72
H41

Keywords:
Contest
Weakest-link technology
Comparative advantage
Coordination
Communication
Experiment
Team performance

ABSTRACT

The organization of team contests can enhance productivity if teammates with complementary skills are able to allocate the team's tasks efficiently, but can also suffer from free-riding incentives. We report the results of a real-effort experiment in which production requires the completion of two complementary tasks, at which workers have heterogeneous skills. We vary whether participants: compete individually; compete in teams where each member must complete each task; or compete in teams where the agents can divide tasks between them and potentially specialize in the task they do best. We report three main results. First, individuals who must work alone divide their work time in a way that is qualitatively consistent with the theoretical predictions, but allocate too little time to their weaker task. Second, there is no difference in productivity or free-riding behavior between individual contests and team contests where teammates cannot specialize. Finally, and most notably, when teammates can divide work tasks, they allocate more time to the tasks they are best at and experience a strong productivity gain. This is true even among teams that cannot communicate – despite the potential for coordination failure or coordination on Pareto dominated equilibria – but the effect is strongest when communication is available.

1. Introduction

Many firms employ team-based incentives in production and the use of such incentives can affect worker productivity in different ways (Hamilton et al., 2003). First, team incentives can affect how much effort individuals put into their work, as team compensation introduces incentive to free-ride off of co-workers' efforts (Alchian and Demsetz, 1972) or, alternatively, may encourage individuals to work harder so as not to let down their teammates (Winter, 2014). A large body of theoretical and experimental work has studied the motivations of workers in team production environments, with experimental evidence regularly finding that free-riding is lower than predicted, especially when teams are engaged in a competition with another team (as reviewed later in this section). Second, the use of teams can affect *how* workers direct their efforts. If workers have complementary skills, then organizing them into teams may enable workers to allocate more time to the tasks at which they personally excel (Lazear, 1998). While productivity gains from worker complementarities is considered a primary advantage of organizing workers into teams (Lazear, 2000), this effect is largely unstudied in the experimental literature.

We conduct a real effort experiment that allows us to separately examine each of these influences in team contests. We consider an environment in which a work unit – either an individual or a team – must complete two separate tasks in order to produce. The existence of multiple tasks, which can be performed by any team member, is both a natural feature of the workplace and allows for potential complementaries in worker skills.[1] Further, team members may not necessarily be assigned which task to do but may choose for themselves how to divide

---

[1] For instance, a team working to serve coffee to customers must both take the customers' orders and prepare their drinks; a consulting team working to produce a report must both analyze data and write up the results, etc. We note that this environment has relevance to the usual examples of group contests, such as military conflicts, research and development competitions, electoral campaigns, and lobbying.

[☆] This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska–Curie grant agreement 2015ABH67840007.

[*] Corresponding author.
E-mail address: muruvvet@metu.edu.tr (M. Büyükboyacı).

https://doi.org/10.1016/j.labeco.2017.11.001
Received 13 December 2016; Received in revised form 24 October 2017; Accepted 7 November 2017

the work.[2] We exogenously vary whether team members can collaborate by dividing tasks across the team or whether each teammate must instead complete both tasks in order to contribute to the team's output. While most studies assume either that all team members are engaged in the same, substitutable task or that their efforts are perfect complements, this environment allows us to study whether teams allocate work in order to exploit gains from complementary skills. If firms implement team contests with the goal of maximizing workers' output, then the question of whether teammates can self-organize and divide work efficiently is fundamental to understanding the productivity of teams. By varying teams' abilities to collaborate, we can isolate the productivity effects of team incentives and gains from complementarities, respectively. Finally, our design allows us to observe how much time team members spend free-riding during each type of contest, enabling us to distinguish between productivity differences caused by how much effort teammates exert vs. how they allocate this effort across tasks.

In order to carefully control team members' productivities at two separate, but otherwise comparable, tasks, we make several modifications to the "ball-catching task" developed by Gächter et al. (2016). In the original ball-catching game, participants must "catch" computerized balls, which move vertically down their screens, by aligning a tray under the ball as it reaches the bottom of the screen. We duplicate this game, in order to create two nearly identical tasks: one in which participants collect red-colored balls and one in which they collect blue. Participants must pay a time cost whenever they are working on one of the two ball-catching tasks, but they can freely and dynamically switch between tasks at any time or press a "neither" button, which blanks the screen and allows them to avoid the cost of working. This set-up allows us to directly measure the amount of time that the participant allocates to each task, as well as how much time is spent free-riding. Finally, and most importantly, by varying the timing and placement of the balls, we are able to induce different, precisely controlled, "skills" for each individual at each task. To the best of our knowledge, we are the first to use a real-effort task with induced heterogeneous productivity values.

Each subject participates in 6 rounds of a contest against another individual and 6 rounds of a team contest. In the individual contests, each participant's output is simply equal to the number of balls she caught fewer of (i.e., min{red catches, blue catches}). Participants in the team contests experience one of two treatment conditions, which differ in whether teammates must each complete each task in order to produce or whether specialization is possible. In the first team treatment (*output-substitutable team contests*), participants are matched into teams of two, with each teammate's output determined the same way as in individual contests and summed at the end of the period. This treatment thus introduces team incentives, but not the ability to collaborate in production and potentially specialize. In *input-substitutable team contests*, in contrast, teammates' input efforts (the number of red and blue catches) are combined. Thus, each teammate need not catch an equal number of each colored ball in order to maximize the team's payoff and can instead choose how they wish to allocate their time across each task. In the input-substitutable treatment, there exist a multiplicity of equilibria and how each teammate will allocate effort across tasks is thus an empirical question. This production environment allows teammates to potentially produce more, if they are able to coordinate on a division of labor in which each teammate specializes on the task at which they are stronger. However, it also introduces a variety of Pareto dominated equilibria and generates a potentially challenging coordination problem in which teammates must coordinate on, not just how much effort to exert, but how to allocate that effort across tasks. As a second treatment dimension, we vary whether teammates are able to communicate.

The experiment allows us to test the following hypotheses, which are predictions of the model we present in Section 2: (1) individuals working alone allocate their work time to maximize output, which causes them to spend the majority of their time on the task at which they are least productive; (2) the introduction of team incentives alone causes production to fall, as individuals devote less time to working; (3) teammates that can collaborate are more productive, as they allocate more time to their stronger tasks; (4) teammates who are able to communicate produce more, especially in the input-substitutable treatment, as it allows teammates to coordinate on the most efficient allocation of tasks.

Several main insights emerge from the experiment. First, individuals and output-substitutable teams are similarly productive. Despite the financial incentives to free-ride in team contests, but not in individual contests, we see no difference in the amount of time spent working or in the allocation of work time across tasks. Second, as predicted, individuals and teammates who cannot collaborate allocate most of their time to the task at which they are weakest; however, they still allocate too little time to their weaker task relative to the output-maximizing prediction, and thus under-produce relative to their maximum capacity. Third, despite the potential for coordination failure, *teams in which individuals can collaborate and potentially specialize produce significantly more*. Participants in the input-substitutable treatment spend the majority of their time on the task that they are best at and, as a result, are significantly more productive than either individuals or output-substitutable teams. Finally, while communication has little effect on team production when teammates cannot collaborate, it has a positive effect in collaborative teams on both output and the amount of time participants devote to their stronger task. A content analysis of the chat transcripts indicates that participants in the input-substitutable treatment are more likely to use the chat period to discuss specific production targets and that over three-quarters discuss a strategy in which one teammate specializes in each task.

Our paper is closely related to two strands of the experimental literature: behavior in team contests and work division in teams. First, there exist competing ideas about how individuals' behavior changes from an individual contest to a team contest. According to standard economic theory, individuals will generally exert less effort in team contests than they would if they were rewarded individually, since the marginal benefit of their effort is shared among teammates (Katz et al., 1990; Lee, 1995; Nitzan, 1991; Ryvkin, 2011). However, a variety of experimental work has found that individuals in team contests exert far higher effort than predicted by theory (see Sheremeta (2017), for a recent review or Bhattacharya (2016), for a recent example).[3] For instance, when directly comparing chosen effort in individual contests with team contests, both Abbink et al. (2010) and Ahn et al. (2011) find that the introduction of team compensation does not generate the predicted strong free-riding response. Similarly, in a real effort experiment where working is associated with opportunity costs, van Dijk et al. (2001) find that participants devote the same amount of effort to working regardless of whether they are paid individually or according to their team average.

---

[2] While workers in some teams may have distinct job descriptions or have their schedule fully assigned by a manager, in many other situations teammates fill identical roles and have freedom in allocating tasks. This is especially common when teammates have the same job title and either it is not feasible for an authority to fully delegate the tasks for which they are mutually responsible or no such authority exists. For instance, academics co-authoring a paper can choose which co-author will complete each section; co-workers in a coffee shop or deli often have leeway in determining how to allocate time to tasks, which could not be fully determined in advance (taking orders, preparing sandwiches, cleaning, restocking, etc.); co-teachers of young children must coordinate on tasks as they arise to manage their classroom; members of a consulting team with the same job title must choose how to divide work to meet deadlines without necessarily consulting with their superiors; partners in a business must determine how to accomplish their goals without input from any authority; etc. While this paper focuses on the ability of teammates to coordinate their allocation of time to different tasks, we note that it is possible that our findings regarding the strong productivity gains in team contests in which workers can specialize could be even further strengthened if a well-informed principal were to assign tasks.

[3] Participants in contests between individuals also regularly over-provide effort, particularly in contests using the Tullock lottery contest success function and linear effort costs (Dechenaux et al., 2015). An individual contest experiment using the same contest success function as reported in the current paper, but with a non-linear production function, finds that effort is close to the equilibrium prediction (Gächter et al., 2016).

Babcock et al. (2015) conduct a field experiment in which participants receive cash prizes if either they personally or both members of their team perform a desired behavior, and find that team incentives are *more effective* than the individual incentives. Several experiments have found that embedding a social dilemma into an intergroup contest can reduce free-riding, suggesting that team contests may be less likely to suffer from free-riding than other types of team-based incentives.[4] Several experiments have found that intra-group communication further reduces free-riding in group contests with perfect-substitutes or weakest-link production (Bornstein, 1992; Bornstein et al., 2002; Cason et al., 2017; 2012; Leibbrandt and Sääksvuori, 2012; Sutter, 2009), while inter-group communication between the competing teams decreases competitive expenditures (Cason et al., 2012; Leibbrandt and Sääksvuori, 2012).[5]

Sheremeta (2011) studies the performance of heterogeneous teams in a chosen effort experiment across three different contest rules: perfect-substitutes, weakest-link, and best-shot. He finds that weak players do not free-ride in perfect-substitute contests, as theory predicts, and suggests that the most likely explanation is that the participants come to identify with their social group (as mentioned above). In the weakest-link contests, in which group performance is determined by the lowest effort, teammates typically choose similar effort levels and successfully coordinate on the Pareto dominant outcome. In minimum effort games, more broadly, groups of two players (as in our experiment) are typically highly successful at coordinating on the Pareto dominant outcome (van Huyck et al., 1990). A pair of experiments by Brookins et al. (2015, 2017) investigate team contests in which teammates differ in their overall abilities, focusing on the question of how to sort heterogeneous agents into teams in order to generate the highest effort in team contests. When teammates' efforts are perfectly substitutable, they find that higher effort can be induced when aggregate abilities of the competing teams are more evenly matched, consistent with theoretical predictions (Brookins et al., 2015). When teammates' efforts are perfect complements, in contrast to the theory, they find that unevenly matched teams provide no more effort or, when communication is possible, significantly less effort than balanced teams (Brookins et al., 2017). In contrast to the clear positive effect of intragroup communication on output in perfect-substitute contests, they find that effectiveness of communication in weakest-link contests depends on the composition of the group, with weak teams providing less effort when they can communicate (Brookins et al., 2017). While those experiments study individuals engaged in an identical task with heterogeneous values for the outcome or costs of effort, we consider individuals engaged in multiple tasks, with heterogeneous skills at each, while holding constant their overall costs and benefits. However, our input-substitutable treatment shares features of the weakest link game: if teammates are always able to coordinate on specializing in separate tasks, then the question becomes how much of that specialized task each player will complete. We note that the higher productivity in our input-substitutable treatment cannot be attributed to reduced free-riding, as the amount of time spent working is identical to our other treatments, and can only be attributed to the productivity gains of specialization.

Bracha and Fershtman (2013) study the effect of competitive incentives in an environment where workers engage in two complementary tasks: a labor-intensive task and a cognitively-demanding task. While they do not study teams, this experiment addresses how individuals allocate their time across tasks. They find that individuals engaged in a tournament tend to allocate less time to the more cognitively-demanding task. The only experiment we are aware of in which team members choose how to divide distinct tasks is reported in Cooper and Sutter (2013). In this experiment, participants are assigned to teams of two people and each member plays a takeover game (Samuelson and Bazerman, 1985) against opponents outside the team. One teammate must take the role of buyer (for whom understanding optimal bidding behavior requires demanding strategic reasoning) and the other takes the role of seller (for whom understanding whether or not it is best to accept a bid is trivial). They vary whether the roles are assigned randomly by the computer or endogenously by mutual agreement, and also vary whether the teammates are able to communicate. While they address different research questions and their environment differs (most notably, in that teammates choose how to assign two binary roles, rather than choosing how to allocate time and effort across tasks, and they do not consider contests), this experiment is similar to our design in that teammates with different strengths can choose how to most efficiently divide work. They find that allowing teams to choose their roles increases the likelihood that the more able teammate will fill the role of buyer, but this selection does not have the expected positive impact on performance: instead, teams in this treatment seem to spend too much effort (and chat time, when available) considering how to fill the roles and devote too little effort to considering the optimal bidding strategy. Since our tasks are mechanical, rather than requiring cognitive effort, we would not expect task performance in our experiment to be affected by the task allocation decision.

The paper is organized as follows. In Section 2 we sketch the theoretical framework for our experiment. In Section 3, we describe the experimental design and procedures. In Section 4, we present the results of our experiment. We conclude in Section 5. All supplementary material, including the instructions for the experiment, is presented in an Appendix.

2. A model

In this section, we provide a theoretical framework for understanding individual incentives in each of the types of contests described in the introduction: contests between individuals, contests between teams in which teammates can only combine their final output, and contests between teams in which teammates can collaborate by combining inputs. We consider the simplest environment in which collaboration is possible: we assume that there are two input "tasks," A and B, with a one-to-one perfect complements production function. Specifically, if individual i completes a_i units of Task A and b_i units of Task B, then his total output, x_i, is given by:

$$x_i = \min\{a_i, b_i\}$$

Consistent with both our real effort experiment and natural workplace limitations, we assume that agents' output is restricted by the amount of work time and their own abilities. Specifically, individual i, working only on Task A, could accomplish α_i units of Task A in the work period. If the individual instead focused only on Task B for the duration of the work period, he could accomplish β_i units of Task B. It therefore follows that an individual who wished to maximize output while working alone would devote a fraction $\beta_i/(\alpha_i + \beta_i)$ of work time to Task A and a fraction $\alpha_i/(\alpha_i + \beta_i)$ to Task B, successfully producing $x_i = a_i = b_i = \alpha_i\beta_i/(\alpha_i + \beta_i)$ units of output. We note that an individual who is stronger at Task A than at Task B would therefore need to devote a larger fraction of time to B, the task that he is weaker at, than to A – a potentially inefficient allocation of time if the individual has potential collaborators with complementary skills. In other words, the worse the individual is at Task B relative to Task A, the greater the fraction of time he must devote to Task B.

---

[4] For instance, in an early study on group contests, Schotter (1997) find that intergroup contests generate higher chosen efforts relative to other group incentives schemes. Gunnthorsdottir and Rapoport (2006) find that participants in a voluntary contribution mechanism game free-ride less when the group is simultaneously engaged in a contest with another group.

[5] One explanation for these findings comes from social identity theory (Brewer, 1979; Sherif et al., 1961; Tajfel and Turner, 1979), which suggests that the existence of a competing out-group may lead to in-group solidarity and cooperation, as well as spiteful behavior toward the out-group. A variety of recent economic experiments have addressed the question of whether social group identity influences behavior in incentivized economic games, largely finding that participants exhibit more cooperative behavior toward in-groups and more aggressive behavior toward out-groups (Charness et al., 2007; Chen and Li, 2009; Eckel and Grossman, 2005; Li et al., 2011), Goette, Huffman, and Meier, (2006).

Each agent has constant marginal opportunity costs of working on tasks A and B instead of the best alternative use of his time. Let f denote the total cost of devoting one work period to Task A (or B). In this case, each unit of Task A accomplished costs individual i f/α_i and each unit of Task B accomplished costs individual i f/β_i. In other words, the marginal cost of the task depends on the per-period opportunity cost and the fraction of that work period required for the individual to accomplish one unit of the task. Producing one unit of *output*, which requires a unit of Task A and a unit of Task B, therefore costs $f/\alpha_i + f/\beta_i$.

In the experimental implementation, we used a linear contest success function (e.g., Che and Gale, 2000; Gill and Prowse, 2012; Gächter et al., 2016). We believe this design choice is essential for the clean interpretation of our results, as it promotes subject understanding in an otherwise complex environment and focuses the attention onto *how to divide time* across tasks, rather than *how much time* to spend working overall. In this section, we describe the predictions for the linear contest success function, including for the specific parametrization of our experiment. In the Appendix, we extend our discussion to the Tullock lottery contest success function (Tullock, 1980), which is commonly used in the contest literature. While the Tullock contest generates interior predictions, rather than corner solutions, we note that the gist of the predictions regarding allocation of time across the two tasks and the incentives for free-riding and specialization in the team contests are similar in this case. Specifically, both individuals and teammates in the output-substitutable contest must devote more time to their weaker task in order to produce, the individually-optimal output level in the output-substitutable contest is less than the team-optimal output level, generating free-riding incentives, and the input-substitutable contest nests the output-substitutable contest while also generating a multiplicity of equilibria in which each teammate specializes in one task and output is greater. Given this, we believe that there is little downside to using the linear contest success function, which provides the clearest incentives for the participants.

2.1. Individual contest

We begin with the individual contest, for which the predictions are straight-forward. An individual who produces x_i units of output facing an opponent producing x_k units of output wins the contest with probability:

$$\text{Probability of Success}(x_i, x_k) = \begin{cases} \frac{50+x_i-x_k}{100} & |x_i - x_k| \leq 50 \\ 1 & x_i - x_k > 50 \\ 0 & x_k - x_i > 50 \end{cases}$$

The winner receives prize V while the loser receives nothing. Since i is working alone, the per unit costs are $f/\alpha_i + f/\beta_i$, as described above. Thus, assuming $|x_i - x_k| \leq 50$ (as must be the case in our experimental environment), i's expected payoff is given by:

$$U(x_i, x_k) = \frac{50 + x_i - x_k}{100}(V) - x_i(f/\alpha_i + f/\beta_i)$$

Given the marginal utility of output, $V/100 - f/\alpha_i - f/\beta_i$, we note that the individual works as hard as possible whenever the marginal cost, $f/\alpha_i + f/\beta_i$, is less than the marginal benefit, $V/100$, in which case individual i produces the maximum amount, $x_i = \alpha_i\beta_i/(\alpha_i + \beta_i)$. Otherwise, the individual will devote no time to either task and produce nothing. This leads to our first prediction:

Prediction (Individual contest): *As long as the marginal expected benefit, V/100, is greater than the marginal cost, agents in the individual contest will work for the full period and allocate their time to maximize total output. They will spend more time working on the task at which they are less productive and, the more polarized an individual's skills, the larger the fraction of work time that will be spent on their less productive task.*

In our experiment, $V = 200$, $f = 24$ and there exist two types of agents, Type 1 with $\alpha_i = 30$ and $\beta_i = 60$, and Type 2 with $\alpha_j = 100$ and $\beta_j = 25$. We describe in more detail in Section 3 the reasoning for the asymmetry of the two types, but for now note that both types have the same marginal cost of producing a unit of output, which is equal to 1.2 ($= 24/30 + 24/60 = 24/100 + 24/25$). The marginal benefit is equal to a 1/100 increase in the probability of winning the prize of 200, or 2. Therefore, the risk-neutral prediction for the individual contest of our experiment is that all contestants work for the full time period. In this case, Type 1 individuals will devote 2/3 of their time to Task A and 1/3 to Task B, producing 20 units of output, while Type 2 individuals will devote 1/5 of their time to Task A and 4/5 to Task B, also producing 20 units of output. Thus, consistent with prediction above, both individuals spend the majority of their time on their weaker task, and the allocation of time is more extreme for Type 2, who has more divergent skills.

2.2. Team contest with substitutable output

The predictions in the team contest with substitutable output are equally straightforward if one assumes that teammates are narrowly self-interested. In this contest, the teammates' ultimate output is combined. Specifically, team m's output X_m is given by $X_m = min\{a_i, b_i\} + min\{a_j, b_j\}$. An agent producing x_i units, paired with a teammate producing x_j, against an opposing team producing a total of X_k units wins with probability:

$$\text{Probability of Success}(x_i, x_j, X_k) = \begin{cases} \frac{50+(x_i+x_j)-X_k}{100} & |(x_i + x_j) - X_k| \leq 50 \\ 1 & (x_i + x_j) - X_k > 50 \\ 0 & X_k - (x_i + x_j) > 50 \end{cases}$$

The winning team shares the prize, V, while the losing team receives nothing. Thus, similar to the individual contest, assuming that $|(x_i + x_j) - X_k| \leq 50$ (which must be the case in our experimental implementation), i's expected payoff is given by:

$$U(x_i, x_j, X_k) = \frac{50 + (x_i + x_j) - X_k}{100}(V/2) - x_i(f/\alpha_i + f/\beta_i)$$

Note that we can think of the team's success depending on the team's total output and teammates each receiving an equal share of the prize (as denoted here), or we could think of the team's success depending on the *average* output of the team and the teammates each receiving a prize of V (e.g., Gächter et al., 2016), which generates the same marginal incentives (though different absolute payoffs). In the team contest, a narrowly self-interested individual i will devote all of his time to the tasks and produce the maximum amount whenever $V/200 > f/\alpha_i + f/\beta_i$, but otherwise produce nothing.

Prediction (Team contest with substitutable output): *As long as the marginal expected benefit, V/200, is greater than the marginal cost, agents in the team contest will work for the full period and allocate their time to maximize total output. The allocation of time will be identical to the allocation of time in the individual contest: They will spend more time working on the task at which they are less productive and, the more polarized an individual's skills, the larger the fraction of work time that will be spent on the less productive task.*

In the case that $V/100 > f/\alpha_i + f/\beta_i > V/200$, then teammates face a type of social dilemma: they maximize their own payoffs by working fully in an individual contest but fully free-ride in the team contest as long as they only consider their own private marginal benefit of working. However, if teammates were to take into account the marginal benefit to the team, they would choose to work fully in the team contest also. Thus, in this case, the optimal behavior for the individual teammate differs from the optimal behavior for the team.[6] This is the situation faced by the participants in our experiment: each faces a marginal cost of 1.2 and a marginal individual benefit of $V/200 = 200/200 = 1$, but a marginal *team* benefit of 2. Of course, individuals with other-regarding social preferences may deviate from these predications and choose to provide effort even when it is contrary to their narrow self-interest. If

---

[6] We note, however, that the total payoffs for all contest participants (i.e., the two individual contest participants or the four team contest participants) in all three types of contests are always maximized when no one exerts any costly effort.

teammates do choose to work in order to help the team, then we assume that they, like contestants in the individual contest, prefer to allocate their time across tasks in order to maximize output. Thus, Type 1 will spend 2/3 of his working time on Task A and Type 2 will spend 1/5 of her working time on Task A.

2.3. *Team contest with substitutable input*

Finally, we consider the team contest where teammates can pool their work on Task A and Task B in order to produce output. In this case, team m's output X_m is given by: $X_m = min\{(a_i + a_j), (b_i + b_j)\}$. As in the above team contest, we can think of the teammates averaging their effort on each task, or, as denoted here, totaling their effort and each receiving one half share of the ultimate prize ($V/2$) and this interpretation does not affect the predictions.

First, note that this case nests the prior team contest case. Whenever teammate j allocates her time such that $a_j = b_j$, then teammate i will also choose to allocate his time such that $a_i = b_i$. Since producing an additional unit of output has a marginal benefit of $V/200$ and a marginal cost of $f/\alpha_i + f/\beta_i$ (due to the fact that the individual must do both Task A and B to produce), he faces the same incentives as in the prior team contest. In other words, he will choose $a_i = b_i = 0$ if $V/200 < f/\alpha_i + f/\beta_i$ and he is narrowly self interested, and $a_i = b_i = \alpha_i\beta_i/(\alpha_i + \beta_i)$ if his marginal benefit of working outweighs his marginal cost.

Next, consider the best response of individual i if his teammate accomplishes unequal amounts of Task A and B, for instance, $a_j < b_j$. In this case, the marginal cost to i of producing the next $b_j - a_j$ units is relatively cheap. Since he needs only to work on Task A to complete this production, his marginal cost is simply f/α_i, which he compares to the marginal benefit, $V/200$. If he still has time left in the work period after completing $a_i = b_j - a_j$, he chooses whether to continue working, now on both tasks, by, as before, comparing $f/\alpha_i + f/\beta_i$ to $V/200$.

Putting this all together, we find that a multiplicity of outcomes can be sustained as equilibria, depending on the relative marginal costs. The team will produce nothing in equilibrium *only* if it is the case that a) $f/\alpha_i > V/200$ or $f/\beta_j > V/200$ (i.e. the teammates cannot agree on an allocation of tasks in which i works on A and j works on B) *and* b) $f/\alpha_j > V/200$ or $f/\beta_i > V/200$ (i.e. the teammates also cannot agree on an allocation of tasks in which i works on B and j works on A).

If it is the case that $f/\alpha_i + f/\beta_i < V/200$ and $f/\alpha_j + f/\beta_j < V/200$, then agents will devote the entire work period to working and any outcome can be sustained where $a_i + a_j = b_i + b_j$, $a_i/\alpha_i + b_i/\beta_i = 1$, and $a_j/\alpha_j + b_j/\beta_j = 1$ (where the latter two conditions simply indicate that the fraction of time each agent devotes to Task A and to Task B sum to 1).

Finally, if it is the case that $f/\alpha_i + f/\beta_i > V/200$ and $f/\alpha_j + f/\beta_j > V/200$ but that, without loss of generality, $f/\alpha_i < V/200$ and $f/\beta_j < V/200$, [7] then the teammates find themselves in a scenario that resembles a weakest-link type game: neither agent is willing to put in the effort to produce a unit of output by herself, but i is willing to work on Task A up to $b_i \le \beta_j$ and j is willing to work on Task B up to $a_i \le \alpha_i$. In this case, any outcome such that $a_i = b_j < \min\{\alpha_i, \beta_j\}$ can be sustained as an equilibrium. In other words, there exist a multiplicity of equilibria and agents with complementary specialties can produce more by specializing in the task they are best at, provided they are able to coordinate on the Pareto dominant equilibrium.

Prediction (Team contest with substitutable input): *If the marginal expected benefit, V/200, is greater than the marginal cost of producing one unit of output, agents in the team contest will work for the full period and the allocation of time to tasks is non-unique. Otherwise, if teammate i's marginal cost of working on one task is smaller than the marginal expected benefit and teammate j's cost of working on the other task is smaller than the marginal expected benefit, then a type of weakest-link production game exists, in which any output level can be sustained as equilibrium, up to the weaker member's maximum ability on his allocated task. In either case, teammates may devote more time to their stronger task in equilibrium, and, as a result, there exist equilibria in which more total output is produced than in the individual contest or in the team contest with substitutable output.*

Consider again the specific parametrization of our experiment, where $\alpha_i = 30$, $\beta_i = 60$, $\alpha_j = 100$, and $\beta_j = 25$. Because the marginal cost of producing is greater than the marginal benefit if the teammate must produce both tasks, but the marginal cost of producing is less than the marginal benefit if each teammate only does one task, then any outcome such that $a_i = b_j < \min\{\alpha_i, \beta_j\} = 25$ and $b_i = a_j < \min\{\beta_i, \alpha_j\} = 60$ can be sustained as an equilibrium. Thus, the maximum output that could be produced in equilibrium is 60, which occurs when Type 1 (agent i) specializes in Task B, of which he can complete 60 units, and Type 2 (agent j) specializes in Task A, of which she can complete 100 units (but stops at 60). We note that there are two ways in which our output-substitutable and input-substitutable treatments differ: first, technologically, workers who devote all their time to working can produce more in the input-substitutable treatment and, second, strategically, there exist equilibria in which payoff-maximizing agents exert positive effort in the input-substitutable treatment but not in the output-substitutable treatment.[8]

Therefore, the team produces 60 units under the Pareto dominant equilibrium. If Type 2 (agent j) were motivated to continue working after completing 60 units of Task B, despite the marginal cost being greater than the marginal benefit, then she could produce an additional 8 units of output, for a total team production of 68. While the maximum joint production in the previous two types of contests (individual and team with substitutable output) was 40, it is possible to produce more in this type of environment. However, it is also possible to produce less, in equilibrium, if agents coordinate on the wrong task allocation or, out of equilibrium, if they fail to coordinate. Specifically, any output level between 0 and 60 could be sustained as an equilibrium with self-interested agents, and it is technologically feasible for the team to produce up to 68.

3. Experimental design and procedures

The experiment was designed to test the predictions of the previous section. Each experimental session had three stages.[9] In the first stage, participants completed two tasks (Task A and Task B, described in detail below), sequentially, for which they were paid a piece rate. This part allowed them to gain familiarity with the tasks and learn their skill in each. In the second and third stages of the experiment, subjects participated in a series of six individual contests against the same opponent and a series of six team contests with the same partner and opponents.[10] In half of the sessions, the individual contests came first and in half the team contests came first. The team contests varied in whether they were output-substitutable or input-substitutable and whether communication was available to the teammates. Subjects were paid their earnings from the piece rate stage and one randomly selected contest, to avoid potential wealth effects.

All parts of the experiment used the same real effort task, which is a modified version of the ball-catching game developed by

---

[7] The symmetric situation follows if $f/\alpha_j < V/200$ and $f/\beta_i < V/200$.

[8] Originally, we intended to conduct an additional control treatment in which costs were sufficiently low that all participants were incentivized to work fully across all three contest types. However, we find that there is very little free-riding in the output-substitutable treatment, even though costs are high, and that the amount of time spent free-riding is essentially identical across the team contest treatments, deeming such comparisons unnecessary in order to identify the source of the treatment differences.

[9] The instructions in a session were read in three parts: first, the first stage instructions were read, then instructions for the second (third) stage were read only after subjects completed the first (second) stage. Full instructions for the stages can be found in Appendix A.

[10] We use partner matching to allow teammates the opportunity to coordinate on specialized equilibria. The choice of keeping opponents fixed is less essential, since participants did not receive feedback on the behavior or choices of their opponents.

Gächter et al. (2016). In the original version of the game, balls fall randomly from top to bottom on the participants' screens and they can catch the balls by moving a tray underneath the ball as it reaches the bottom of the screen. Participants can control the position of the tray by clicking a "left" button or "right" button. We chose this task for two reasons. First, the authors find that participants generally have neutral opinions of the task, rather than finding it either intrinsically rewarding or costly. Since the task itself is neither particularly engaging nor tedious, it allows us to manipulate the costs and benefits associated with working on it. Second, we are able to modify the game in order to precisely control each participant's skill level and assign participants different skills in different versions of an otherwise identical task. We can thus induce different, precisely controlled, α_i and β_i for different participants.

We modify the ball-catching game in several ways. First, we create two versions of the game, one in which the balls subjects catch are red and another in which they are blue. We refer to these as the red game and blue game in the instructions and refer to them as "Task A" and "Task B" here. Second, rather than having the balls fall randomly, we hardwire precisely when and where the balls fall on the screen, in order to create four different "skill levels." In each version, a fixed number of balls always fall over the course of a 90 second period: either 100, 60, 30, or 25. The timing and placement of the balls is such that nearly any participant who is comfortable using a computer and mouse should be able to collect every single ball, with the occasional error. The balls fall in regularly spaced increments, so that someone spending x seconds on a task at which she has skill level α is expected to catch $\alpha x/90$ balls. Finally, we eliminate the cost per click that was used by Gächter et al. (2016) and instead use a cost per second, as described below, consistent with the model discussed in the previous section.

Each session consisted of 12 people and each was randomly assigned a productivity type at the start of the experiment: 6 people were "Type 1" and 6 were "Type 2." Type 1 participants had a productivity of 30 in Task A (i.e., they could catch 30 red balls in 90 s) and 60 in Task B, while Type 2 participants had a productivity of 100 in Task A and 25 in Task B. In other words, Type 1s had a comparative (and absolute) advantage on Task B while Type 2s had a comparative (and absolute) advantage on Task A. Teams were always composed of one Type 1 member and one Type 2 member, to create complementarity. Participants were told only that the balls would fall at different rates for different people, and were not told anything about the specific number of balls different people were expected to catch. Instead, we allowed them to learn about their productivity in the two tasks for themselves during the piece rate stage, which came first.

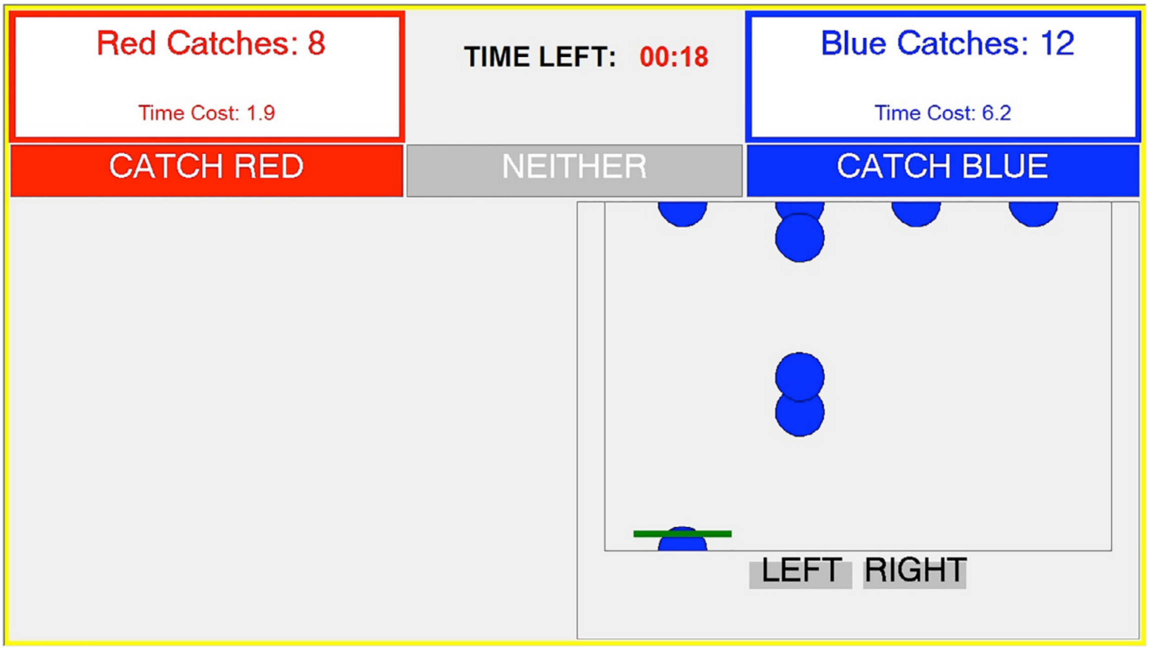
The productivity values were chosen with several considerations in mind: First, by matching a Type 1 individual with a Type 2 individual, we create a team in which collaboration and specialization could generate higher output within the given time period. Second, participants' skills are not simply mirror images of each other. This is more realistic, appears less contrived from the perspective of the subjects, and it allows for a natural feature of collaborative work environments: team members who specialize do not necessarily require the same amount of time to complete their part and, after the quicker member finishes, she can decide whether to quit working or to help the team by pitching in on both tasks. Finally, despite the seeming difference in their skills, the two types of individuals face *exactly* the same costs and potential output levels when working alone. For an individual of Type 1, the total costs are $f/30 + f/60 = f/20$ and for an individual of Type 2, the total costs are $f/100 + f/25 = f/20$. Likewise, the maximum output that each team member could produce independently is equal to 20. Therefore, the differences in task skills do not generate any *inequality* in the individual productivities of the agents.

In the contest stages, participants could choose how they wished to allocate their 90 s of time between Task A (catching red), Task B (catching blue), or neither. They made this choice dynamically, and there was no restriction on how frequently the subjects could switch between tasks. A screenshot of an English language version of the contest stage screen is shown in Fig. 1. At the beginning of the period, the screen is blank except for three buttons ("Catch Red," "Catch Blue," and "Neither"), any of which could be pressed at anytime. If the participant pressed the "Catch Red (Blue)" button, then Task A (B) appeared on the left (right) side of the screen and he or she could spend as much time as desired on the red (blue) game. When one task was open, the other disappeared from the screen. If the participant pressed the neither button, then both tasks disappeared from the screen. Whenever Task A or Task B appeared on the screen, subjects were charged a time cost of $f = 24$ tokens per 90 s period (or 0.267 tokens per second spent on a task). The consequences of these costs in each treatment were described in Section 2. They did not bear any cost when they pressed the "Neither" button and the tasks were not displayed. At the beginning of the period, participants started in the "Neither" state, so that they did not incur costs or collect balls until they actively selected one of the two tasks.

In all contests, subjects compete for a prize of 200 tokens (which is shared between the teammates in the team contests) and receive 0 tokens from losing. Regardless of whether they won the contest, the incurred time cost was subtracted from the subject's earnings. As described in the previous section, the likelihood of winning the contest depended on the number of points the individual or team and their opponents earned in the ball-catching game. Specifically, the probability that a subject or team earning x_i points wins a contest against an opponent earning x_j points is: $(x_i - x_j + 50)/100$ (as in Gill and Prowse, 2012), and Gächter et al., 2016. The computer then randomly determined the winner according to this probability. In the individual contests, the number of points the individual earns is always equal to min{red catches, blue catches}. In the output-substitutable contest, each teammate's points are determined identically to the individual contest and the teammates' points are combined at the end of the period: Team Points = min{Type 1 red catches, Type 1 blue catches} + min{Type 2 red catches, Type 2 blue catches}. In other words, teammates cannot specialize in the task they are stronger at, and must instead continue to do both tasks in order to contribute to the team's output. In the input-substitutable contest, in contrast, each teammate's catches, or inputs, are combined to determine the team's total points: Team Points = min{Type 1 red catches + Type 2 red catches, Type 1 blue catches + Type 2 blue catches}. For example, if one teammate collected 2 red and 5 blue, while the other collected 7 red and 5 blue, then the team's points in the output substitutable contest would equal the sum of the points earned individually ($2 + 5 = 7$ points), while the team's points in the input-substitutable contest would be based on the number of each color ball collected by the team (9 red and 10 blue, so 9 points).

At the start of each contest period, subjects were reminded as to the number of red and blue catches they made in the piece rate stage. At the beginning of each team contest period, subjects were also informed of the number of red and blue catches their partner made in the piece rate stage. After each contest, subjects saw how many of each colored ball they collected, the number of each colored ball caught by their teammate (in the team contests), their own points (output), and the realized outcome of the contest, but did not receive feedback on the points scored by their opponent(s). In half of the sessions, teams also had the opportunity to communicate with each other via chat box on their computer screens. Prior to the first contest period, the teammates had an opportunity to chat for 60 s. They had a second opportunity to chat with their teammate halfway through the series of six contests.

The experiment was conducted at the METU-FEAS Behavioral and Experimental Laboratory (BEL) at the Middle East Technical University (METU). Subjects were recruited by e-mail using the BEL database, which consists of undergraduate students at METU. All sessions were computerized using z-Tree (Fischbacher, 2007) and exactly 12 subjects were admitted to each. Overall, 192 subjects participated in the experiment: 96 Type 1 and 96 Type 2 individuals. There were sixteen sessions, with exactly four sessions (i.e., 48 subjects) conducted under each of the four team contests treatments: output-substitutable without commu-



Red Catches: 8
Time Cost: 1.9

TIME LEFT: 00:18

Blue Catches: 12
Time Cost: 6.2

Fig. 1. Contest screen. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Treatment conditions.

	Individual contest and output substitutable team contest	Individual contest and input substitutable team contest
No chat	4 sessions (48 individuals and 24 teams)	4 sessions (48 individuals and 24 teams)
Chat	4 sessions (48 individuals and 24 teams)	4 sessions (48 individuals and 24 teams)

Two of the four sessions in each cell reversed the order of individual and team contests.

nication, output-substitutable with communication, input-substitutable without communication, and input-substitutable with communication. For each team contest treatment, two sessions had the individual contests first and two had the team contests first. Each subject participated in only one session and sessions lasted approximately 60 min. The design is summarized in Table 1. Throughout the experiment, payoffs were described in terms of “tokens,” with 10 tokens corresponding to 1 Turkish Lira (TL). Subjects earned 25.05 TL on average, including a 5 TL participation fee.[11]

4. Results

In this section, we examine the overall productivity of individuals and teams in each type of contest (Section 4.2), and (then turn to the question of how participants of different skill-types produce and allocate their time to different tasks (Section 4.3), consider how they dynamically alter their behavior in response to the previous outcomes (Section 4.4), and (analyze chat content (Section 4.5).

4.1. Piece rate performance

We first provide a preliminary check of how each type's Task A productivity (i.e. the number of red catches per time period) and Task B productivity (number of blue catches per time period) compared to the prediction. Recall that the experiment was programmed so that Type 1s should be able to catch 30 (60) in Task A (B) and Type 2s should be able to catch 100 (25) in Task A (B). In Table 2, we summarize the performances in the first phase of the experiment, when subjects are paid according to a piece rate and complete only one task at a time. Ninety-six subjects of each type participated and the table reports one observation per subject per task. Both types appear to be doing their best in the piece rate payment scheme and they catch close to the expected number of balls in all four cases. Therefore, we conclude that our assigned productivities were successful.

Table 2
Productivity in the piece rate phase.

Type	Observations		Task A catches	Task B catches
Type 1	96	Observed:	29.28(2.06)	59.87(0.67)
		Predicted:	30	60
Type 2	96	Observed:	97.14(9.28)	24.82(1.54)
		Predicted:	100	25

Notes: The table reports the mean number of catches in the Task A piece rate period and Task B piece rate period for each assigned type. Standard deviations are reported in parentheses, with the predicted number of catches for each type reported in italics below.

4.2. Productivity of individuals and teams

We begin our analysis by looking at the overall outcomes by treatment, focusing on the measure most likely to be of interest to the firm: average output per person. In order to make comparisons across individual contests and the two types of team contests, we consider the average output “per teammate” by dividing the team's output by 2 in the team contests. Table 3 presents the average output in the individual contests and each of the types of team contests as well as the results of two-tailed Mann–Whitney U tests for treatment differences. The top panel of Table 3 compares the output in individual contests versus team contests, the middle panel reports the differences between output-substitutable and input-substitutable team contests, and the bottom panel presents the effects of communication in output-substitutable and input-substitutable contests. As a robustness check, these tests are reported at three different levels of observation: the team (i.e., one team is one observation), the match (i.e., the contestants and their opponents are one observation),

---

[11] At the time of the experiment, the minimum hourly wage in Turkey was 8.13 TL.

Table 3
Average output by treatment.

			Individual vs. Teamsignificance tests (z)		
		Output	Team-level	Match-level	Session-level
Individual		15.34	–	–	–
All teams		19.55	4.48***	4.18***	3.05***
Output-substitutable		16.15	0.47	1.27	0.11
Input-substitutable		22.97	5.65***	5.63***	3.36***
	Output		*Output-subs. vs. Input-subs. significance tests (z)*		
	Output-subs.	Input-subs.	Team-level	Match-level	Session-level
Overall	16.15	22.97	4.84***	4.66***	3.26***
No chat	15.98	21.16	2.99***	3.06***	2.31**
Chat	16.33	24.82	3.86***	3.46***	2.31**
	Output		*No chat vs. Chat significance tests (z)*		
	No chat	Chat	Team-level	Match-level	Session-level
Output-subs	15.98	16.33	0.35	0.09	0.29
Input-subs	21.16	24.82	2.05**	1.67*	1.44

$* p < .10$, $** p < .05$, $*** p < .01$. *Notes:* The significance tests report two-tailed Mann–Whitney U tests. The top panel compares output in the individual contests to output in all team contests, output-substitutable team contests, and input-substitutable team contests, respectively. The center panel compares output in the output-substitutable contests and input-substitutable contests, in all contests, contests without chat, and contests with chat, respectively. The bottom panel compares team output with and without chat in output-substitutable team contests and in input-substitutable contests, respectively. Significance results are reported at the team-level (each team of two participants is treated as one observation), match-level (each contestant individual or team and opposing individual or team is treated as one observation), or session-level (each session of 12 participants is treated as one observation).

and the session (i.e., all twelve individuals in the experimental session is one observation).[12]

Aggregating across all team contest treatments, we find that the output per teammate in the team contests is significantly greater than the output produced in the individual contest, as indicated by the second row in Table 3.[13] While participants in individual contests produce an average 15.34 units of output, teams produce 19.55 units per teammate, a difference that is significant at better than the $p = .001$ level at all levels of observation.

The next two rows of Table 3 separate the team contests by production technology and demonstrate that the difference between individual contests and team contests is being driven by the input-substitutable teams: while individuals and output-substitutable teams are similarly productive (15.34 units for individuals compared to 16.15 units for output-substitutable teammates), individuals in the input-substitutable teams are significantly more productive than individuals, producing 22.97 units on average (a difference that is significant at the $p = .0001$ level at all levels of observation). Recall that the equilibrium predictions are that participants in the individual contests produce the maximum feasible output (20), while teammates in the output-substitutable contest exert no costly effort and produce nothing. Therefore, we see substantial over-production in the output-substitutable contest – in fact, the production is no lower than in the individual contests, where participants have a dominant strategy to work fully. Output in the individual contests and in the output-substitutable contests with and without communication is significantly less than the maximum output of 20 ($p < .01$ in each case at either the team- or match-level). In the input-substitutable contest, there exist a multiplicity of equilibria and predicted per-teammate output ranges from 0 to 30, encompassing the observed output of 22.97.

Moving to the middle panel of Table 3, we see that input-substitutable teams are not only significantly more productive compared to individuals, but also compared to output-substitutable teams. This is the case both without communication (15.98 vs. 21.16) and, even more so, when teammates can chat (16.33 vs. 24.82). Again, this result is robust to each plausible level of observation. Finally, the bottom panel presents the effect of chat on output in the two types of team contests. While teammates in the output-substitutable teams are similarly productive with and without communication (16.33 with chat versus 15.98 without chat), the availability of chat has a significant positive effect on productivity in the teams where collaboration is possible (24.82 with chat versus 21.16 without). However, we note that this effect is significant only at the 10% level if we consider the match the unit of observation (and at the 15% level if we consider the session the unit of observation).

Fig. 2 graphs the average output per person over the course of the contest series and reinforces the results of Table 3: individuals and output-substitutable teams (with and without communication) all perform similarly, while input-substitutable teams produce more – *especially* when teammates are able to communicate. We see no evidence (in any condition) of a "decay" in teammates' efforts, as is often observed in public goods games where participants face a similar social dilemma Ledyard (1995), but note that the small group size, contest incentives, and real effort task may make free-riding less attractive than in the standard public goods game. More notably, team output is *positively* correlated with period in both input-substitutable treatments and in the output-substitutable treatment without communication.[14] In other

---

[12] We first consider the team as the level of observation, since participants receive feedback (only) regarding their teammates' productivity and behavior, which could influence behavior in subsequent contests. While participants do not learn about the output or other behavior of their opponents, they do learn whether they won or lost the contest, which is probabilistically determined by the difference in output (i.e., each additional unit produced by the opposing team decreases the likelihood of winning by 1/100). It is possible that participants could use this information to update slightly their beliefs about their opponent's behavior. Moving forward, we primarily treat the team as the unit of observation, but, to assuage any concerns that the results could be driven by interdependence of observations resulting from learning whether they won earlier contests, we also report results at the match-level (treating all participants in the series of contests as a single observation) as a robustness check and note any discrepancies.

[13] Given that the design was counterbalanced so that an equal number of teams participated in each contest order, and the fact that we find no significant differences in either individual ($Z = 1.362, p = .17$) or team contest performance ($Z = 0.861, p = .39$) on the basis of order, we aggregate the data across orderings. Likewise, regressing average output on an indicator for team contest and its interaction with an indicator for the team contest occurring first reveals that the effect of participating in a team contest is highly significant ($t = 3.61, p < .01$ clustering standard errors at the level of the match), but that there is *no* significant interaction with order ($t = 0.22, p = .825$). Similar results hold if the standard errors are clustered at the team-level.

[14] At the match-level, the spearman correlation coefficients and p-values are $\rho = 0.342$ ($p = .003$) for the input-substitutable treatment without communication, $\rho = 0.278$ ($p = .018$) for the input-substitutable treatment with communication, and $\rho = 0.405$ ($p = .0004$)

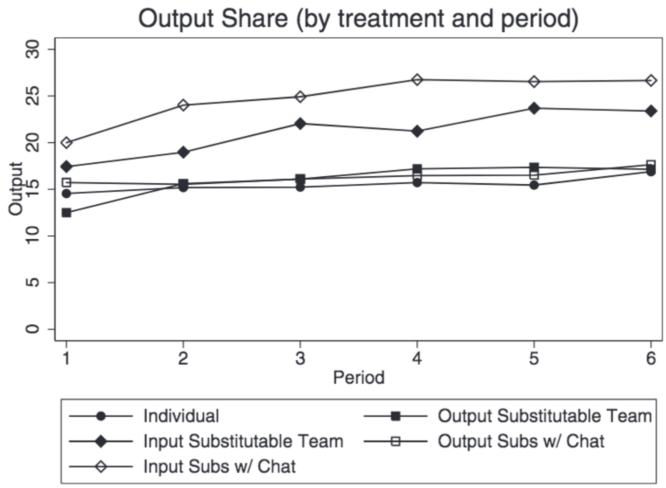


Fig. 2. Output over time by treatment.

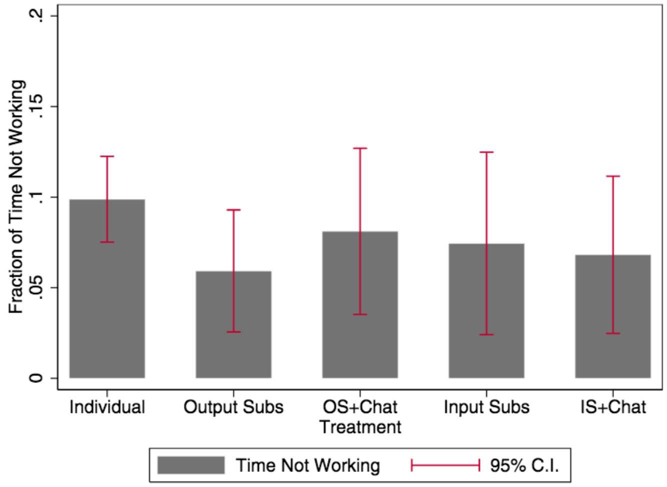


Fig. 3. Proportion of 90 s period spent on neither task. The 95% confidence intervals are calculated using the match as the unit of observation.

words, teams are moving in the direction of the Pareto optimal outcome over time. We further consider the dynamics of teammates' choices in Section 4.4.

Recall that there are two distinct reasons why teams whose members can specialize (input-substitutable teams) may achieve higher productivity in this experiment: *technologically*, teams can produce more if each individual focuses primarily on the task they are best at, and, *strategically*, teammates no longer have a dominant strategy to free-ride. However, the very similar output levels in the output-substitutable contests (where participants have a dominant strategy to spend no time working) and the individual contest (where individuals have a dominant strategy to spend all of their time working), indicates that free-riding may not be much of a factor in our team contests.

This is confirmed by Fig. 3, which presents the amount of time that the participants in each treatment spent *not working* on either task. In other words, we consider the proportion of the 90 s work period that neither task was displayed on the participants' screen and they did not incur any costs. (The allocation of time across Task A and Task B will be more fully explored in the following subsection.) As indicated by Fig. 3, and confirmed by pairwise Mann–Whitney tests, there are no significant differences in time spent not working in any of the contest treatments (using either the team or the match as the unit of observation). In all five treatments, participants spend, on average, 5–10% of the contest period not working on either task. Qualitatively, the percentage of time spent not working is actually highest (9.9%) in the individual contest (where no free-riding incentives exist) and lowest (5.9%) in the output-substitutable contest without chat, where we would expect the strongest free-riding incentives to exist. Thus, it appears that the differences in output levels between the output-substitutable contests and input-substitutable contests are not driven by participants in the input-substitutable contests being uniquely capable of overcoming free-riding, but instead due to teammates coordinating on a more productive division of labor, which we explore in the following section.

4.3. Production and allocation of time across tasks

Next, we turn to the behavior of the two productivity types in the individual contest and the different team contest treatments. To begin, we estimate the number of catches in Task A and in Task B in each of the types of contests, as a function of several explanatory variables.[15] The estimates are reported in Table 4 and provide several overall insights regarding each type of individual's performance in each task. Although we do not expect any effect of the subject being Type 1 or Type 2 on the number of catches in Task A and in Task B in either the individual contests or the output-substitutable contests, where individuals should always collect an equal number of each, we observe that the number of catches in Task A increases (and decreases in Task B) for Type 2 participants (i.e., the $\alpha = 100, \beta = 25$ types).[16] This result indicates that subjects perform more on their stronger task even when efficient production requires that they perform each equally. We further note, and chi-squared tests confirm, that the coefficients are similar between individual contests and output-substitutable contests ($p > .42$ for both tasks at either the team- or match-level). Moving to the estimates for the input-substitutable contests, we observe that the coefficient on the Type 2 indicator variable becomes far more extreme, indicating that subjects in input-substitutable teams move in the direction of specializing in their stronger task.[17] While we include an indicator for female as a control, we find little effect of this regressor in either Table 4 or in unreported specifications in which female is interacted with type or fully interacted with all regressors.

The number of units of each task produced by each type is presented in Fig. 4, with the vertical line at 20 showing the output-maximizing production in individual and output-substitutable contests. We begin with the individual contest, for which the amount of Task A and Task B accomplished for each type is presented in the far left of each panel of Fig. 4. In the individual contest, we expect both Type 1s and Type 2s to produce 20 units in Task A and 20 units in Task B. Instead of producing an equal number of units in the individual contest, we find that both types overproduce in their stronger task and underproduce in their weaker task. For both types, we can reject the hypothesis that the production in Task A is equal to production in Task B at all conventional

---

for the output-substitutable treatment without communication. Alternatively, if we regress output on period, clustering at the match-level, the p-values are $p = .002$ for the input-substitutable contests without communication, $p = .048$ for input-substitutable with communication, and $p = .001$ for output-substitutable without communication.

[15] We report robust standard errors clustered at the team-level. Tables are reproduced with standard errors clustered at the match-level in the Appendix and any changes in the significance of coefficients discussed are noted in the text.

[16] Clustering at the match-level, the coefficient on Type 2 remains significantly different from 0 at the 1% level in all columns except (3), where it is significant at the 5% level, as can be seen in Table 14 in the Appendix.

[17] Chi-squared tests conducted at either the team or match level confirm that the coefficient on Type 2 is significantly greater in the Input-Substitutable model of Task A production than it is in the Individual contest model ($\chi^2 \geq 40.75, p < .0001$) or in the Output-substitutable contest model ($\chi^2 \geq 21.82, p < .0001$). Likewise, the coefficient is significantly smaller in the Input-substitutable model of Task B production than in the Individual contest model ($\chi^2 \geq 71.65, p < .0001$) or in the Output-substitutable contest model ($\chi^2 \geq 54.02, p < .0001$).

Table 4
Task Performance by contest type.

	(1) Individual	(2)	(3) Output-subs	(4)	(5) Input-subs	(6)
	A	B	A	B	A	B
Type 2	8.322***	−5.518***	11.37***	−6.800***	42.60***	−34.41***
	(1.616)	(0.906)	(4.019)	(1.932)	(4.776)	(3.949)
Period	−0.463	−0.296**	−0.745**	−0.0988	−0.235	0.410
	(0.286)	(0.147)	(0.332)	(0.251)	(0.454)	(0.311)
Female	0.692	−0.241	0.869	0.341	6.314*	−3.029
	(1.466)	(0.822)	(2.015)	(1.030)	(3.546)	(2.432)
Chat			−0.0794	−0.959	−1.910	2.882
			(0.898)	(1.735)	(2.035)	(4.652)
Chat × Type 2			−7.080	2.098	15.46**	−4.995
			(4.337)	(2.262)	(7.486)	(5.561)
Constant	18.05***	22.33***	19.64***	22.75***	5.400**	43.50***
	(1.033)	(0.997)	(1.321)	(1.857)	(2.584)	(3.591)
Observations	1152	1152	576	576	576	576
Clusters	96	96	48	48	48	48
R^2	0.068	0.090	0.121	0.114	0.619	0.607

$* p < .10, ** p < .05, *** p < .01$. OLS estimates reported. Standard errors clustered by team.

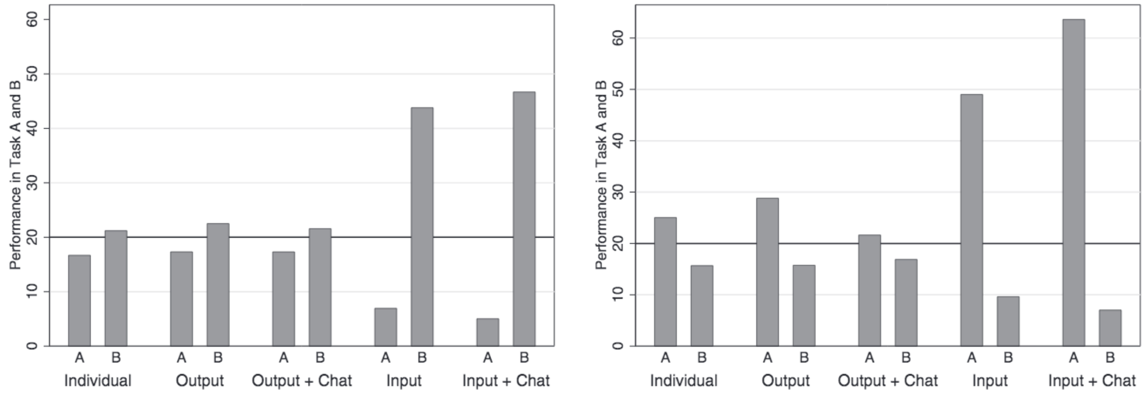


Fig. 4. Performance of Task A and Task B for Type 1 (left) and Type 2 (right) by Treatment. *Note:* The vertical line shows the output-maximizing performance in the Individual and Output-substitutable team contests.

levels.[18] Further, the gap in production across the two tasks is larger for Type 2 participants, who have a more extreme skill set, and Type 2 individuals earn significantly fewer points in the individual contests ($Z = 2.27$, $p = .023$; with the match as the level of observation, $Z = 1.73$, $p = .084$).

Next, we turn to the performance of individuals in the output-substitutable team contests, which are shown in the second and third sets of bars of each panel in Fig. 4. Contrary to the self-interested prediction, each type produces positive amounts in the output-substitutable team contest. As discussed in the Section 2 predictions, the cost of producing each unit is higher than the gain obtained and thus payoff-maximizing subjects in the output-substitutable team contest should not perform at all in either task. Nevertheless, performance in the output-substitutable team contests (with or without communication) is very similar to performance in the individual contests. As we observed in the individual contests, both types overproduce on their stronger task and underproduce on their weaker task in output-substitutable team contests. For each type, the difference in Task A and B production is significant at all reasonable levels, both with and without communication.[19]

In the output-substitutable treatments, each team can produce 40 units (20 from Type 1s and 20 from Type 2s) at maximum. We find that the mean contest points achieved in the output-substitutable team contest is 32.6 with communication and 32.0 without communication. In both cases, team output is significantly lower than the team's capacity of 40 ($Z = 6.03$, or $Z = 4.29$ with the match as the level of observation).

The final bars in Fig. 4 show the performance of individuals in the input-substitutable team contest. As expected, when collaboration is possible, each type focuses on the task that they are better at: both with and without communication, Type 1 teammates performed Task B more than Task A, while Type 2 teammates performed Task A more than Task B (in Wilcoxon signed-rank tests, $Z > 4.00$, $p < .0001$ in all four cases; with the match as the level of observation $Z > 3.05$, $p < .003$ in all four cases). In the input-substitutable treatments, we expect the team to produce 60 units if teammates specialize in the task they do best, plus an additional 8 units if Type 2 teammates subsidize the team by producing more output after reaching 60 units in Task A (as discussed by Lee, 2012). In other words, if there is a subsidy, then it should be supplied only by Type 2 teammates and Type 1 teammates should produce 0 units in Task A. However, as we can see from Fig. 4 this is not the case: with or without communication, Type 1 teammates do perform Task A instead of exclusively focusing on Task B and catch fewer than the 60 balls in Task B ($p < .0001$ in both communication conditions with one observation per team and $p < .003$ in both communication conditions with one observation per match). On the other hand, Type 2 teammates would be expected to complete 68 units of Task A and subsidize Type 1 on Task B with 8 units. As can be seen from Fig. 4, without communi-

---

[18] In Wilcoxon signed-rank tests, $Z = 6.8$ for Type 1 ($n = 96$) and $Z = 6.7$ for Type 2 ($n = 96$), with the individual as the unit of observation and $Z = 5.2$ and $Z = 5.3$, respectively with the match as the unit of observation ($n = 48$ for each type).

[19] In all four cases (i.e., Type 1 and Type 2, with and without communication), $Z > 2.57$, $p < .01$ in Wilcoxon signed-rank tests with the individual as the unit of observation (which is consistent with the team being the unit of observation, since there is exactly one member of each type in the team). If we instead consider the match as the unit of observation, $Z > 2.47$, $p < .014$ in all four cases.

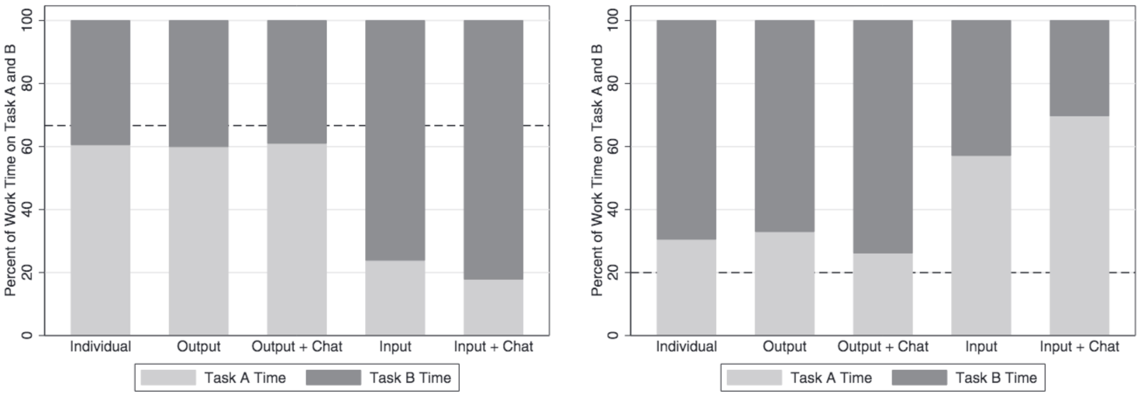


Fig. 5. Percent of work time spent on Task A and Task B for Type 1 (Left) and Type 2 (Right) by treatment. *Note:* The vertical line shows the output-maximizing allocation to Task A in the Individual and Output-substitutable team contests.

cation, Type 2 teammates also underperform on the task they are better at and we can reject the hypothesis that Type 2 teammates complete 60 units (the maximum of their Type 1 partners) of Task A at $p = .02$ (or $p = .023$ at the match-level). However, with communication, they complete 63.62 units of Task A and we cannot reject the hypotheses that their Task A productivity differs from 60 ($Z = 1.2, p = 0.22$; at match-level, $Z = 0.55, p = .58$ a) or from 68 ($Z = 0.46, p = .65$; at match-level, $Z = 0.16, p = .88$).

Finally, we look at the amount of time that each type spent on Task A or Task B in each of the treatments, shown in Fig. 5. For the individual contests, we expect Type 1 individuals to spend 67% of their time (60 s) on the task they are worse at (Task A) and 33% of their time (30 s) on the task that they are better at (Task B). We expect Type 2 individuals to spend 80% of their time (72 s) on the task they are worse at (Task B) and 20% of their time (18 s) on the task they are better at (Task A). Since participants do not always spend all of their time on the tasks, we focus on the percentage of their *working* time (i.e., time in which one task was displayed on their screen) they spend on each task. The figure shows the percentage of their working time that the participants allocate to Task A (shaded light gray) and to Task B (shaded dark gray). The dashed line depicts the percent of work time each type should spend on Task A in order to maximize output in the individual contest and the output-substitutable contest. As predicted, both types do spend significantly more time on their weaker task in individual contests (in Wilcoxon signed-rank tests, $Z = 6.11$ for Type 1 ($n = 96$) and $Z = 7.3$ for Type 2 ($n = 96$)).[20] However, they still allocate too much of their working time to their stronger task.[21]

As noted above, we see little free-riding in the output-substitutable contests and, given that participants choose to work on the tasks, we expect them to make the best use of this time and divide their work time identically to participants in the individual contests: Type 1 teammates should spend 66.7% of their time on their weaker task (Task A), while Type 2 teammates should spend 80% of their time on their weaker task (Task B). First, consider the time allocation across Task A and Task B for Type 1 team members. As can be seen from Fig. 5, both with and without communication, the time allocation is nearly identical to the individual contests: participants in the individual contests allocate 60.2% of their working time to their weaker task, compared to 60.1% in the output-substitutable contest without communication and 61.2% with communication ($p \geq .50$ in all pairwise comparisons). For Type 2 participants, there is again no difference in time allocated to their weaker task in the individual contest (69.1%) vs. the output-substitutable contest with no communication (67.5%, $Z = 0.22$ and $p = .826$). However, with communication, they succeed in allocating *more* time to their weaker task (74.1%, $Z = 2.00$ and $p = .045$; at the match-level, $Z = 1.81, p = .07$), although the fraction of time is still significantly lower than the output-maximizing level.

In the input-substitutable team contests, we predicted that team members would be able to allocate less time to their less productive tasks in favor of focusing on the task they do best. As we can see from Fig. 5, this prediction is largely confirmed. Type 1 team members spend only 24.6% (no communication) and 19.5% (with communication) of their time on their weaker task, down from over 60% in the individual and output-substitutable contests.[22] Likewise, Type 2 individuals spend only 42.3% (no communication) and 31% (with communication)[23] of their time on their weaker task, down from 67% to 74% in the individual and output-substitutable contests.[24] If the teammates had been able to perfectly coordinate, then we would expect Type 1s to spend 0% of their time on Task A and Type 2s to spend either 0% or, if they choose to subsidize the team with their remaining time, 32% on Task B.

4.4. Team and individual dynamics

In this section, we consider how individuals and teams respond dynamically to the observed outcomes of the previous period. Recall that participants in the individual contests learn only their own catches in each task, their own output, and whether they won or lost, while teammates learn their own catches and their teammate's catches in each task, their team output, and whether they won or lost.

We begin this analysis with our primary outcome variable, output per teammate. Since we can only measure this variable at the team-level in the input-substitutable contest, we have one observation per team (or individual in the individual contests) for each of the final five out of six contest periods. Table 5 presents estimates of a model in which we regress output share on the individual's or team's own output in the previous period, an indicator for whether they won in the previous period, and the period number.[25] In the individual contest and both types

---

[20] Likewise, if the match is treated as the unit of observation, $Z = 5.2$ for Type 1 ($n = 48$) and $Z = 5.94$ for Type 2 ($n = 48$).

[21] The fraction of work time allocated to the weaker task is significantly less than the output-maximizing prediction for both types at all conventional levels, taking the individual teammate of that type as the unit of observation or taking the match as the unit of observation.

[22] The difference in time spent on the weaker task is significant at the $p < .01$ level in all pairwise tests between input-substitutable and output-substitutable or individual contests, using either team or match as unit of observation.

[23] The amount of time spent on the weaker task is significantly lower with communication than without communication at $p = .03$ (or $p = .033$ at the match-level).

[24] The difference in time spent on the weaker task is significant at the $p < .01$ level in all pairwise tests between input-substitutable and output-substitutable or individual contests, using either team or match as unit of observation.

[25] In this section, we report the standard errors clustered at the team-level, but note any instance in which the significant results are altered by clustering at the match-level instead, as can be seen in Tables 11 to 13 in the Appendix. We also note that communication is available in only one of the five final periods (since participants can chat prior

Table 5
Output share dynamics.

	(1) Individual	(2) Output-subs	(3) Input-subs
Output share (t−1)	0.700***	0.581***	0.531***
	(0.0400)	(0.0683)	(0.0796)
Win (t−1) indicator	0.577**	0.886	1.510
	(0.234)	(0.544)	(0.919)
Period	−0.000111	0.0134	0.0588
	(0.0985)	(0.158)	(0.383)
Communication treatment		−0.476	1.911*
		(0.427)	(1.125)
Constant	4.544***	7.033***	9.904***
	(0.834)	(0.926)	(2.105)
Observations	960	240	240
Clusters	96	48	48
Adjusted R^2	0.518	0.411	0.354

$*\ p < .10$, $**\ p < .05$, $***\ p < .01$. OLS estimates reported. Standard errors clustered by team.

of team contests, we find that output in the current period is strongly associated with the contestant's previous period output. This persistence is consistent with other estimates of team contest dynamics, reported in Brookins et al. (2015, 2017) and Cason et al. (2017, 2012). We also observe that individual contestants respond positively to winning the previous period's contest, an event that is associated with a roughly one half unit increase in output in the current period. For teams, winning also appears positively associated with output in the subsequent period, but the effect is not statistically significant in either output-substitutable or input-substitutable contests. In the specification reported here, we do not include a lagged variable for the opponent's output, since this information was not available to subjects, but note that it is not significantly associated with output in the individual contest or either type of team contest.[26]

Since output in this experiment depends both on how participants allocate their time across tasks and how much time they spend working, we next focus only on the latter factor, specifically considering how the *amount of costly work time* an individual or team member expends is influenced by the previous period's outcomes. We note that this outcome variable is more directly comparable with other team contest experiments, in which participants only choose how much costly effort to exert. Since we can directly measure how much time each teammate spends on the two tasks, we can now analyze the individual-level dynamics of how participants respond to the overall outcomes and to the production of their teammates.

Table 6 reports regression estimates in which the amount of time the participant spent working (out of the 90-second work period) is regressed on the amount of time spent working in the previous period, an indicator for whether they won in the previous period, and controls for period, type, and female. The results are largely consistent with the estimates presented in Table 5. The amount of time a participant spends working is strongly associated with the amount of time spent working in the previous period. We again observe that participants in individual contests respond positively to winning in the previous period: in this case, a win causes them to devote approximately one and a half seconds more time to working in the next period. However, this result is only marginally significant and loses significance at the 10% level if we cluster at the match-level.[27] Consistent with the results in Table 5, winning does not significantly affect subsequent behavior in any of the team contests. In the team contests, we can also consider how participants respond to the production of their teammates in the previous period. In columns (3) and (5), we include as regressors the number of previous period catches the participant's teammate completed in the task at which the teammate is stronger (*Teammate stronger task*) and the task at which the teammate is weaker (*Teammate weaker task)*. There is indication that participants in the output-substitutable contest respond positively to the production of their teammates, especially on the teammate's weaker task. However, we caution against strong interpretations of this result, as the coefficients are only marginally significant (at the 10% level and only at the 11% level when clustering by match). While team members learn the number of catches their teammate completed, not the amount of time they worked, participants may be able to infer the teammates' work time from the information provided. If we instead consider teammate's lagged work time, instead of lagged catches, we find a similar marginally positive effect in the output-substitutable contests and again find no significant effect in the input-substitutable contests.

To more clearly assess how participants change their production in response to the previous-period production of their teammates, we next consider the number of catches the participant completes in their weaker task (i.e., Task A for Type 1 and Task B for Type 2) and in their stronger task (i.e., Task B for Type 1 and Task A for Type 2). Table 7 presents regression estimates in which catches in the weaker task (odd-numbered columns) and stronger task (even-numbered columns) are regressed on the participant's previous-period catches in each task and their teammate's previous-period catches in each task (where the teammate's "stronger" task is the same as the participant's "weaker" task). As in the previous estimates, we find that the number of catches in a given task is strongly associated with the number of catches of that task the participant completed in the previous period. Beyond this, participants clearly update their production based on their teammates' production in the previous period, and this response differs depending on whether the contest is output-substitutable or input-substitutable. In the output-substitutable contests, where the participants' performance is generally determined by their performance on the task they are worse at, we see that participants perform significantly more on their weaker task as their teammate produces more in their own weaker task. In contrast, in the input-substitutable contests, participants' responses are consistent with moving in the direction of specialization: as one's teammate performs more on their stronger task, the participant performs significantly more on their own stronger task and significantly less on their weaker task.[28]

4.5. Content analysis of communication in team contests

Finally, we turn to the content of the chats between team members. In half of the team contests, teammates had one minute to talk about the experiment at the start of the first and fourth periods. We have seen that the existence of this chat period has little impact on output-substitutable teams, which cannot coordinate their task decisions. In input-substitutable teams, the chat period has a significant positive effect on output and also shifts individuals' time allocations toward the task at which they are most productive. To better understand how participants used the communication period in each treatment, we analyze the content of the team chats.

We quantified the information transmitted in the chats by following the procedure used in Cason et al. (2017, 2012)'s analysis of communication in team contests. First, we randomly selected two sessions, one from each treatment, and developed a set of classification categories based on the messages exchanged as well as the focus of the experiment on specialization and free-riding. The coding categories are shown in Table 8. We then hired two coders who independently reviewed the messages sent within each of the teams, with the unit of observation

---

to the first and fourth periods). We therefore include an indicator for the communication condition as a control, but do not estimate the model separately for the communication condition, since communication usually not available to influence how participants respond to the previous period outcomes.

[26] All of the results discussed are unchanged clustering at the match-level.

[27] As shown in Table 12 in the Appendix.

[28] All results are significant at the same level if standard errors are clustered at the match-level.

Table 6
Time spent working dynamics.

	(1) Individual	(2) Output-subs	(3) Output-subs	(4) Input-subs	(5) Input-subs
Work time (t−1)	0.603***	0.511***	0.465***	0.601***	0.591***
	(0.0737)	(0.112)	(0.0992)	(0.124)	(0.130)
Win (t−1) indicator	1.422*	2.239	2.153	1.146	1.180
	(0.850)	(1.618)	(1.584)	(1.310)	(1.316)
Teammate weaker task (t−1)			0.407*		0.263
			(0.230)		(0.213)
Teammate stronger task (t−1)			0.146*		0.0710
			(0.0782)		(0.0789)
Communication treatment		−2.820	−2.562	0.250	0.217
		(1.732)	(1.584)	(1.542)	(1.618)
Period	0.268	−0.242	−0.215	−0.153	−0.0816
	(0.284)	(0.397)	(0.396)	(0.380)	(0.349)
Type 2	−0.700	−0.478	−0.593	−1.619	−0.338
	(1.255)	(0.876)	(1.137)	(1.132)	(2.016)
Female	−0.257	−0.0962	−0.00158	1.701*	1.767
	(1.261)	(0.971)	(1.216)	(1.005)	(1.117)
Constant	30.48***	42.87***	36.29***	32.85***	27.26***
	(6.508)	(8.915)	(8.890)	(11.36)	(9.319)
Observations	960	480	480	480	480
Clusters	96	48	48	48	48
Adjusted R^2	0.341	0.295	0.309	0.317	0.320

∗ $p < .10$, ∗∗ $p < .05$, ∗∗∗ $p < .01$. OLS estimates reported. Standard errors clustered by team.

Table 7
Catches on stronger and weaker tasks dynamics.

	(1)	(2)	(3)	(4)	(5)	(6)
	Individual		Output-subs		Input-subs	
	Weaker	Stronger	Weaker	Stronger	Weaker	Stronger
Weaker catches (t−1)	0.607***	0.0531	0.473***	−0.0246	0.564***	−0.0335
	(0.0549)	(0.116)	(0.0579)	(0.103)	(0.0776)	(0.262)
Stronger catches (t−1)	0.0254	0.468***	0.000929	0.548***	0.0150	0.564***
	(0.0205)	(0.0731)	(0.0276)	(0.0823)	(0.0217)	(0.0891)
Teammate weaker task (t−1)			0.118**	0.0394	0.0486	−0.0140
			(0.0549)	(0.114)	(0.0635)	(0.163)
Teammate stronger task (t−1)			0.0259	0.0392	−0.0561***	0.163**
			(0.0199)	(0.0355)	(0.0200)	(0.0623)
Communication treatment			−0.406	−1.254	−0.854	2.413
			(0.432)	(0.912)	(0.630)	(1.838)
Type 2	−0.481	1.762*	−0.319	0.653	0.577	6.868***
	(0.345)	(0.915)	(0.416)	(0.770)	(0.876)	(2.221)
Female	0.248	−1.087	−0.381	0.375	0.434	1.046
	(0.408)	(1.008)	(0.408)	(0.962)	(0.550)	(1.458)
Period	0.0668	0.0346	−0.00564	−0.111	−0.0362	−0.183
	(0.117)	(0.279)	(0.150)	(0.257)	(0.191)	(0.533)
Constant	5.898***	9.842***	7.221***	9.143***	4.625**	10.60
	(1.268)	(2.832)	(1.844)	(2.602)	(1.838)	(8.368)
Observations	960	960	480	480	480	480
Clusters	96	96	48	48	48	48
Adjusted R^2	0.359	0.244	0.303	0.457	0.442	0.485

∗ $p < .10$, ∗∗ $p < .05$, ∗∗∗ $p < .01$. OLS estimates reported. Standard errors clustered by team.

Table 8
Chat content.

	Output substitutable		Input substitutable	
Message type	Freq.	Kappa	Freq.	Kappa
State piece rate productivity	0.583**	0.50	0.833*	0.40
State costs	0.167**	0.41	0.146**	0.50
Discuss division of time	0.313***	0.80	0.229**	0.41
Set a target	0.187**	0.59	0.438**	0.58
Suggest each catch an equal number of red and blue	0.458***	0.66	0.186*	0.36
Suggest one person catch red and the other blue	0.146***	0.83	0.729***	0.90
Suggest free-riding (not working on task)	0.080***	1.00	0.146**	0.50
Number of messages	272		341	

Notes: This table reports the average frequency with which the coders classified a team's under each category. * denotes codes with Cohen's Kappa reliability between 0.3 and 0.4. ** denotes codes with Cohen's Kappa reliability between 0.4 and 0.6. *** denotes codes with Cohen's Kappa reliability above 0.6.

Table 9
Contest performance and chat content.

	(1) Output subs	(2) Input subs
Productivity	0.439	2.005
	(2.743)	(7.726)
Costs	−11.07***	−2.194
	(3.618)	(13.73)
Time division	−6.358*	7.510
	(3.063)	(6.191)
Equal	4.732**	−14.23***
	(1.902)	(4.465)
Specialize	−1.641	7.962
	(2.714)	(6.454)
Free-ride	−13.21***	−21.73
	(1.947)	(13.88)
Target	−3.629	11.39**
	(3.100)	(4.980)
Constant	34.86***	40.34***
	(1.522)	(8.871)
Observations	24	24
Adjusted R^2	0.508	0.364

* $p < .10$, ** $p < .05$, *** $p < .01$. Each observation is one team. Robust standard errors.

being both chat periods for each team (yielding 24 observations in each treatment). For each category, the team was coded as 1 if the coder deemed that the content of the chat satisfied the classification and 0 otherwise. The coder could select as many categories for each team as they believed were relevant and they were not aware of the hypotheses of the experiment.

To assess the extent of agreement between the two coders, for each category we calculate Cohen's Kappa Cohen (1960). This statistic measures the extent to which the coders' designations match, while correcting for the likelihood that the two classifications would agree by chance even if there were no relationship between them. Again following Cason et al. (2012), we consider Kappa values above 0.60 to be "substantial agreement," values between 0.41 and 0.60 to be "moderate" agreement, and denote Kappa values between 0.31 and 0.40 with a single star in Table 8 (while cautioning against strong interpretation of agreement levels below 0.40).

First, we note that teammates in both types of contests typically mention their productivities at the two tasks: Over 58% of output-substitutable teams and over 83% of input-substitutable teams discussed the number of balls they could catch. Within output-substitutable teams, the next most common topic of discussion is the output-maximizing strategy of catching an equal number of each colored ball. This strategy is rarely discussed in the input-substitutable teams, despite being an equilibrium strategy. Instead, nearly three-quarters of participants in the input-substitutable teams discuss a strategy of specialization in which one member completes each task, with many of the teams also establishing a target number of balls for the teammates to catch. While several teams mention the time costs or discuss a strategy of fully free-riding, such discussions are no more prevalent in the output-substitutable teams, in which free-riding is a dominant strategy, further underscoring our interpretation that the productivity treatment differences are not driven by differences in free-riding. Finally, there is indication that participants in the input-substitutable contests sent more messages than participants in the output-substitutable contests, although the difference is not statistically significant ($Z = 1.27$, $p = .204$).

To assess the extent to which the chat content identified in Table 8 is associated with a team's productivity, we regress the contest points earned by the team on indicators for each of the coded categories, with each team providing one observation. We caution that these regressions do not allow us to make causal inferences, since communication occurs before the first and fourth periods and behavior in the first few periods could therefore influence the later chat content. These estimates are reported in Table 9.[29] In the output-substitutable contests, we find that the team's production is significantly higher when they discuss the strategy of catching an equal number of each type of ball and is significantly lower when they discuss the time costs or mention a strategy of free-riding. Both of these associations are fairly intuitive, as teammates who focus on the time costs and identify free-riding as the dominant strategy produce far less, while those who identify and communicate the output-maximizing strategy in this environment successfully produce more. These two opposing uses of the chat stage help to explain why communication had little effect in output-substitutable contests on average. In the input-substitutable contest, in contrast, output is significantly *lower* when the teammates discuss a strategy of catching an equal number of each colored ball, while coordinating on a specific target has a strong positive effect on output. These associations are also intuitive and in line with our earlier findings. While catching an equal number of each colored ball is the output-maximizing strategy in the output-substitutable contests, teams in the input-substitutable contest who are instead able to coordinate on a different strategy can be more productive. It is therefore unsurprising that focusing on this strategy would have a positive effect in the output-substitutable contest but a negative effect in the input-substitutable contests. Furthermore, the chat reveals that teams are most successful in the input-substitutable contest when they set a specific target regarding how much each player should produce in their task of specialization.

5. Conclusion

There exist competing theories regarding how individuals' behavior will differ between individual contests and team contests. We contribute a controlled laboratory test of the productivity benefits of organizing team contests when agents have complementary skill sets. Our design allows us to isolate the effect of team incentives alone while also assessing whether teams can be more productive by allowing team members to devote their time to the tasks they do best. We set up a real effort experiment, in which workers produce by completing two complementary tasks, and vary whether participants compete as individuals, in teams where each member must independently complete each task, or in teams where either member can complete either task.

We find that the introduction of team incentives alone neither help nor hurt worker productivity. Without the possibility of collaborating in teams, output is nearly identical in the individual and team contests, despite the strong incentive to free-ride in teams: participants have a dominant strategy to work fully in the individual contest and free-ride fully in the team contest without collaboration. While we are unaware of other work demonstrating that participants in real effort experiments work equally hard in individual and team contests, we note that our results are in line with the real effort team remuneration experiment of van Dijk et al. (2001), chosen effort experiments finding below-expectation free-riding in team contests (Abbink et al., 2010; Ahn et al., 2011; Sheremeta, 2011), and work finding that the presence of an out-group leads to more cooperation with an in-group (Gunnthorsdottir and Rapoport, 2006).

In contrast, we find that teams in which workers can potentially divide the tasks between teammates experience a strong productivity gain relative to individuals or teams in which workers must complete each task independently. To the best of our knowledge, we are the first to demonstrate this benefit of team contests in a controlled laboratory experiment. This result occurs despite the potential for coordination failure in this environment, and even when teammates cannot communicate with each other. However, the productivity advantage of team contests is particularly strong when teams *can* communicate, in which case team output is 60% higher than in individual contests.

---

[29] The estimates are reported with standard errors clustered by match in Table 14 in the Appendix, which indicates that all results reported here remain significant at the $p < .05$ level.

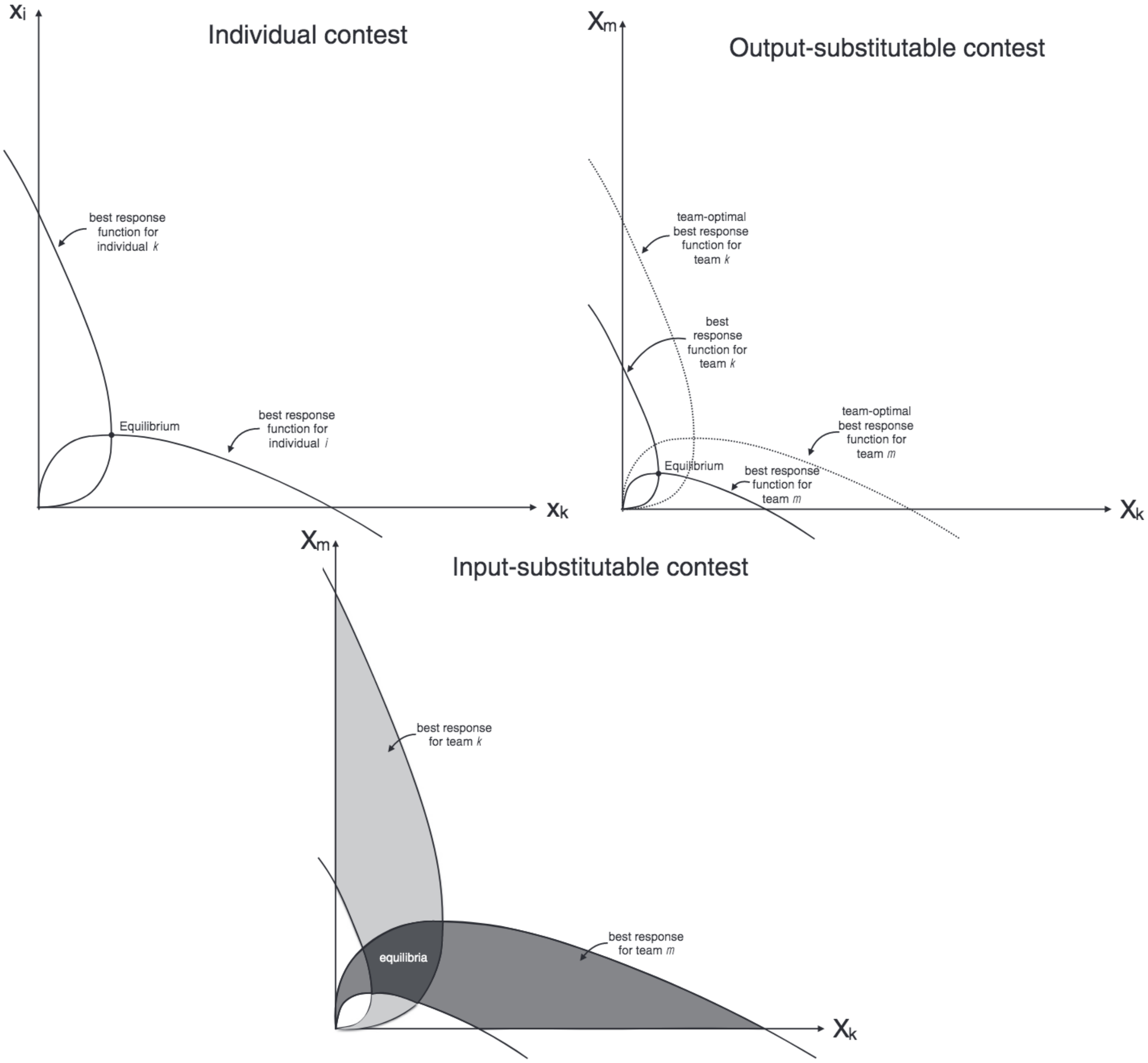


Fig. 6. Pure strategy equilibria in individual, output-substitutable, and input-substitutable contests with Tullock contest success function.

The results thus suggest that there is little downside to the organization of team contests and a potentially strong productivity advantage when workers can collaborate with co-workers who have complementary skills. In terms of the allocation of workers' time across their two tasks, we find that workers who must complete two complementary tasks choose an allocation of time that is broadly consistent with the output-maximizing predictions: participants devote significantly more time to their weaker task and the allocation of time is more extreme when workers have more disparate skills. However, they spend significantly less time on their weaker task than is required to maximize output and relatively too much time on their stronger task. The misallocation of time across tasks is aggravated when workers have more extreme skills. When teammates can collaborate, in contrast, they immediately transition to devoting more time to the task at which they are best.

This suggests that teammates are capable of organizing the division of tasks according to their capacities, and thus a principal may not be required to assign workers to tasks efficiently. However, we must also caution against the broad interpretation of these results without considering a wider variety of possible tasks. In our controlled experiment, we deliberately create two distinct, but essentially identical, production tasks. Further, we provide an objective measure of each participant's skill at each task (the result of the initial piece rate stage) and these skills are induced, rather than reflecting the participants' intrinsic abilities. In some workplace settings, however, the tasks involved may differ in degree of interest, cognitive demands, prestige, or career advancement opportunities. In such cases, ego or self-interest may prevent workers from wanting to complete the tasks most beneficial to the team and overconfidence or concerns about dignity (á la Bénabou and Tirole, 2009) may prevent workers from accurately perceiving their own relative skills, both of which may hinder a team's ability to arrive at the optimal allocation of tasks given the skills of its team members. We therefore emphasize that our findings are most directly relevant to settings in which the tasks are similar in importance or status and that the extent to which teammates coordinate on specialization in the presence of these additional considerations is an important avenue of future research.

Appendix A

A1. Extension to the Tullock contest success function

In this section, we to consider how the predictions introduced in Section 2 would vary if we considered the commonly-used Tullock contest success function instead of the linear success function. In this case, an individual or team producing x_i units of output and competing against an individual or team producing x_k units of output wins the contest with probability $x_i/(x_i+x_k)$.

In the individual contest, the contestant's expected payoff is thus given by: $U(x_i,x_k) = \frac{x_i}{x_i+x_k}(V) - x_i(f/\alpha_i + f/\beta_i)$ and the best response is: $x_i = \sqrt{\frac{Vx_k}{f/\alpha_i+f/\beta_i}} - x_k$. As is the case in the experiment, suppose that the two contestants have the same overall cost, c, of producing one unit of output, i.e., $f/\alpha_i + f/\beta_i = f/\alpha_k + f/\beta_k = c$. Then the unique Nash equilibrium is the standard Tullock prediction $x_i = x_k = \frac{V}{4c}$. The best response functions and equilibrium prediction in the individual contest are shown in the left panel of Fig. 6. Of course, as in Section 2.1, the production of the contestants is bounded above by the maximum they can produce within the allotted work time, $\alpha_i\beta_i/(\alpha_i+\beta_i)$, which may constrain output. Importantly, however, the optimal production in the Tullock contest depends on the individual contestant's belief about their opponent's ability to optimally allocate his time across tasks, whereas individual contestants in our experiment have a dominant strategy to fully produce.

The output-substitutable contest also resembles the standard Tullock contest with substitutable efforts. Specifically, an agent producing x_i units, paired with a teammate producing x_j, against an opposing team producing X_k receives the expected payoff: $U(x_i,x_k) = \frac{x_i+x_j}{x_i+x_j+X_k}(V/2) - x_i(f/\alpha_i+f/\beta_i)$, which is maximized by producing $x_i = \sqrt{\frac{VX_k}{2(f/\alpha_i+f/\beta_i)}} - x_j - X_k$. Thus, a team that is best responding to the opponent producing X_k will collectively produce $x_i + x_j = \sqrt{\frac{VX_k}{2c}} - X_k$, although the equilibrium x_i and x_j are non-unique. The team best response functions and equilibrium output, $V/8c$, are displayed in the middle panel of Fig. 6. In contrast, the "team optimal" output, which maximizes the team's joint payoffs, is twice as high. If the teammates take into account the share of the prize awarded to their teammate, as well as their own share, the team's best response function shifts out to the dashed curves in the middle panel of Fig. 6. In both cases, the optimal production for an individual teammate depends both on his beliefs about his teammate's production (given that only total team output, and not how each individual contributes to it, is unique in equilibrium) and his beliefs about the opposing team's ability to optimally allocate their time across tasks and coordinate their efforts. In contrast, teammates in our output-substitutable treatment have either a dominant strategy to fully shirk or fully produce (if their goal is to maximize team payoffs).

Finally, we consider the input-substitutable contest, which, as we observed in Section 2.3, nests the output-substitutable contest while also generating a multiplicity of equilibria in which the teammates specialize. First, we note that the minimum best response for the team is $x_i + x_j = \sqrt{\frac{VX_k}{2c}} - X_k$, i.e., the best response function in the output-substitutable contest. If the team output were lower than this, either player would wish to unilaterally produce more by working on both tasks. If the teammates specialize, the marginal cost of production is lower than if they were to complete both tasks individually and they are thus willing to produce more – conditional on their teammate matching their work. The maximum that the team would produce in equilibrium is determined by the teammate whose cost of effort at the task he is specializing in is greater. Specifically, let $c_s = f/(\max\{\min\{\alpha_i, \beta_j\}, \min\{\alpha_j, \beta_i\}\})$, i.e., the highest marginal cost of production between the two teammates if they divide the tasks most efficiently. Then the maximum the team would be willing to produce in response to X_k is $\sqrt{\frac{VX_k}{2c_s}} - X_k$. However, there also exist many other pure strategy Nash equilibria in which the amount of Task A produced by one player equals the amount of Task B produced by the other player. In other words, if player i specializes in Task A and player j specializes in Task B, any a_i, b_j such that $a_i = b_j \in [\sqrt{\frac{VX_k}{2c}} - X_k, \sqrt{\frac{VX_k}{2c_s}} - X_k]$ is a best response to X_k. The difference between this contest and the weakest-link Tullock contests studied by Sheremeta (2011), Lee (2012), and Cason et al. (2012) is that best response correspondences for the teams are bounded below (by the output-substitutable best response function) as well as above. The best response correspondences for both teams, as well as the set of Nash equilibria, are depicted in the right panel of Fig. 6.

To summarize, the primary difference between the predictions in the linear contest success function and the Tullock lottery contest success function is that the latter generates interior equilibria. Beyond this, the Tullock contest success function produces predictions that largely mirror those presented in Sections 2.1–2.3. Specifically, both individuals and teammates in the output-substitutable contest must devote more time to their weaker task in order to produce, the individually-optimal output level in the output-substitutable contest is less than the team-optimal output level, generating free-riding incentives, and the input-substitutable contest nests the output-substitutable contest while also generating a multiplicity of equilibria in which each teammate specializes in one task and output is greater.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.labeco.2017.11.001.

References

Abbink, K., Brandts, J., Herrmann, B., Orzen, H., 2010. Intergroup conflict and intra-group punishment in an experimental contest game. Am. Econ. Rev., 100 (1), 420–447.

Ahn, T.K., Isaac, R.M., Salmon, T., 2011. Rent seeking in groups. Int. J. Ind. Org. 29, 116–125.

Alchian, A.A., Demsetz, H., 1972. Production, information costs, and economic organization. Am. Econ. Rev. 62, 777–795.

Babcock, P., Bedard, K., Charness, G., Hartman, J., Royer, H., 2015. Letting down the team? Social effects of team incentives. J. Eur. Econ. Assoc. 13 (5), 841–870.

Bénabou, R., Tirole, J., 2009. Over my dead body: bargaining and the price of dignity. Am. Econ. Rev. 99 (2), 459–465.

Bhattacharya, P., 2016. Inter-team contests with power differential. J. Econ. Behav. Org. 132, 157–175.

Bornstein, G., 1992. The free-rider problem in intergroup conflicts over step-level and continuous public goods. J. Person. Soc. Psychol. 62, 597–606.

Bornstein, G., Gneezy, U., Nagel, R., 2002. The effect of intergroup competition on intra-group coordination: an experimental study. Games Econ. Behav. 41, 1–25.

Bracha, A., Fershtman, C., 2013. Competitive incentives: working harder or working smarter? Manage. Sci. 59 (4), 771–781.

Brewer, M.B., 1979. In-group bias in the minimal intergroup situation:a cognitive motivational analysis. Psychol. Bull. 86 (2), 307–324.

Brookins, P., Lightle, J. P., Ryvkin, D., An experimental study of sorting in group contests. Lab. Econ.2015. 35, 16–25.

Brookins, P., Lightle, J. P., Ryvkin, D., 2017. Sorting and communication in weak-link group contests.

Cason, T.N., Sheremeta, R.M., Zhang, J., 2017. Asymmetric and endogenous communication in competition between groups. Exp. Econ 20 (4), 946–972.

Cason, T.N., Sheremeta, R.M., Zhang, J., 2012. Communication and efficiency in competitive coordination games. Games Econ. Behav. 76, 26–43.

Charness, G., Rigotti, L., Rustichini, A., 2007. Individual behavior and group membership. Am. Econ. Rev. 97 (4), 1340–1352.

Che, Y., Gale, I., 2000. Difference-form contests and robustness of all-pay auctions. Games Econ. Behav. 30, 22–43.

Chen, Y., Li, S.X., 2009. Group identity and social preferences. Am. Econ. Rev. 99 (1), 431–457.

Cooper, D.J., Sutter, M., 2013. Role selection and team performance. IZA Discussion Paper No. 5892.

Cohen, J., 1960. A coefficient of agreement for nominal Scales. Educ. Psychol. Meas. 20, 37–46.

Dechenaux, E., Kovenock, E., Sheremeta, R., 2015. A survey of experimental research on contests, all-pay auctions and tournaments. Exp. Econ. 18 (4), 609–669.

van Dijk, F., Sonnemans, J., van Winden, F., 2001. Incentive systems in a real effort experiment. Eur. Econ. Rev. 45, 187–214.

Eckel, C.C., Grossman, P.J., 2005. Managing diversity by creating team identity. J. Econ. Behav. Org. 58 (3), 371–392.

Fischbacher, U., 2007. z-tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10, 171–178.

Gächter, S., Huang, L., Sefton, M., 2016. Combining 'real effort' with induced effort costs: the ball catching task. Exp. Econ. 19 (4), 687–712.

Gill, D., Prowse, V., 2012. A structural analysis of disappointment aversion in a real effort competition. Am. Econ. Rev. 102 (1), 469–503.

Goette, L., Huffman, D., Meier, S., 2006. The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. Am. Econ. Rev. 96 (2), 212–216.

Gunnthorsdottir, A., Rapoport, A., 2006. Embedding social dilemmas in intergroup competition reduces free-riding. Org. Behav. Hum. Decis. Process. 101, 184–199.

Hamilton, B.H., Nickerson, J.A., Owan, H., 2003. Team incentives and worker heterogeneity: an empirical analysis of the impact of teams on productivity and participation. J. Polit. Econ. 111, 465–497.

Katz, E., Nitzan, S., Rosenberg, J., 1990. Rent-seeking for pure public goods. Public Choice 65 (1), 49–60.

Lazear, E., 1998. Personnel Economics for Managers. New York: John Wiley & Sons.

Lazear, E., 2000. Personnel economics and economic approaches to incentives. HKCER Lett. 61, 1–8.

Ledyard, J.O., 1995. Public Goods: A Survey of Experimental Research. In: Kagel, J., Roth, A. (Eds.). The Handbook of Experimental Economics. Princeton University Press, Princeton.

Lee, D., 2012. Weakest-link contests with group-specific public good prizes. Eur. J. Polit. Econ. 28, 238–248.

Lee, S., 1995. Endogenous sharing rules in collective-group rent-seeking. Public Choice 85 (1-2), 31–44.

Leibbrandt, A., Sääksvuori, L., 2012. Communication in intergroup conflicts. Eur. Econ. Rev. 56, 1136–1147.

Li, S.X., Dogan, K., Haruvy, E., 2011. Group identity in markets. Int. J. Ind. Org. 29, 104–115.

Nitzan, S., 1991. Collective rent dissipation. Econ. J. 101 (409), 1522–1534.

Ryvkin, D., 2011. The optimal sorting of players in contests between groups. Games Econ. Behav. 73, 564–572.

Samuelson, W.F., Bazerman, M.H., 1985. The winners curse in bilateral negotiations. In: Smith, V. (Ed.), Research in Experimental Economics, vol. 3. Greenwich, CT: JAI Press, pp. 105–137.

Nalbantian, H., Schotter, A., 1997. Productivity under group incentives: an experimental study. Am. Econ. Rev. 87 (3), 314–341.

Sheremeta, R.M., 2017. Behavior in group contests: a review of experimental research. J. Econ. Surv. doi:10.1111/joes.12208.

Sheremeta, R.M., 2011. Perfect substitutes, best shots and weakest link contests between groups. Kor. Econ. Rev. 27 (1), 5–32.

Sherif, M., Harvey, O.J., White, B.J., Hood, W.R., Sherif, C.W., 1961. Intergroup Conflict and Cooperation: The Robber Cave Experiment. Norman, Oklahoma: University Books.

Sutter, M., 2009. Individual behavior and group membership: comment. Am. Econ. Rev. 99 (5), 2247–2257.

Tajfel, H., Turner, J., 1979. An integrative theory of intergroup conflict. In: Austin, W., Worchel, S. (Eds.), The Social Psychology of Intergroup Relations. Brooks/Cole Publ., Monterey, CA, pp. 33–48.

Tullock, G., 1980. Efficient rent seeking. In: Buchanan, J., Tollison, R., Tullock, G. (Eds.), Toward a Theory of Rent Seeking Society. A&M University Press, College Station, TX, pp. 97–112.

van Huyck, J.B., Battalio, R.C., Beil, R.O., 1990. Tacit coordination games, strategic uncertainty, and coordination failure. Am. Econ. Rev. 80 (1), 234–248.

Winter, E., 2014. Feeling Smart: Why Our Emotions are More Rational than We Think. New York: Public Affairs.

Further Reading

Chen, R., Chen, Y., 2011. The potential of social identity for equilibrium selection. Am. Econ. Rev. 101 (6), 2562–2589.

Epstein, G.S., Mealem, Y., 2009. Group specific public goods, orchestration of interest groups with free riding. Public Choice 139 (3), 357–369.

Esteban, J., Ray, D., 2001. Collective action and the group paradox. Am. Polit. Sci. Rev. 95 (3), 663–672.

Gill, D., Prowse, V., 2007. A structural analysis of disappointment aversion in a real effort competition. Am. Econ. Rev. 102 (1), 469–503.

Goette, L., Huffman, D., Meier, S., 2012a. The impact of social ties on group interactions : evidence from minimal groups and randomly assigned real groups. Am. Econ. J. 4 (1), 101–115.

Goette, L., Huffman, D., Meier, S., Sutter, M., 2012b. Competition between organizational groups:its impact on altruistic and antisocial motivations. Manage. Sci. 58 (5), 948–960.

Holmstrom, B., 1982. Moral hazard in teams. Bell J. Econ. 13 (2), 324–340.

Nitzan, S., Ueda, K., 2009. Collective contests for commons and club goods. J. Public Econ. 93 (1), 48–55.

Olson, M., Zeckhauser, R., 1966. An economic theory of alliances. Rev. Econ. Stat. 48 (3), 266–279.