Collaboration and Free-riding in Team Contests*

Mürüvvet Büyükboyacı† Andrea Robbett‡

May 10, 2016

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.

Keywords: Contest, Weakest-link technology, Comparative advantage, Coordination, Communication, Experiment, Team performance
JEL classification codes: C91, C92, C71, C72, D72, H41

---

*The research was funded by European Commission under grant 2015ABH67840007. All errors are ours.

†Corresponding author: Department of Economics, Middle East Technical University, Ankara 06800, Turkey. E-mail: muruvvet@metu.edu.tr; Tel: +90 312 210 2057; Fax: +90 312 210 7964.

‡Department of Economics, Middlebury College.

1 Introduction

Many firms employ team-based incentives in production and the use of such incentives can affect worker productivity in different ways (Hamilton, Nickerson, and Owan, 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 contests, with experimental evidence regularly finding that team effort is higher than predicted, especially when teams are engaged in a competition with another team (Dechenaux, Kovenock, and Sheremeta, 2015). 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 the work. 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

---

[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.

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, Huang, and Sefton (2015). 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, teammates are no longer forced to catch an equal number of each colored ball, 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 the potential for coordination failure. As a second treatment dimension, we vary whether teammates are able to communicate.

The experiment allows us to test the following hypotheses: (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 strong positive effect in collaborative teams. 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 their production strategy 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. For instance, when directly comparing chosen effort in individual contests with team contests, both Abbink, Brandts, Herrmann, and Orzen (2010) and Ahn, Isaac, and Salmon (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, Sonnemans, and van Winden (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. Babcock, Bedard, Charness, Hartman, and Royer (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.[2] Sutter and Strassmair (2009) find that intra-group communication further reduces free-riding in group contests.

One explanation for these findings comes from social identity theory (Sherif et al. 1961, Brewer 1979, 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, Rigotti, and Rustichini, 2007; Li, Dogan, and Haruvy, 2011; Eckel and Grossman, 2005; Chen and Li, 2009; Goette, Huffman, and Meier, 2006).

Sheremeta (2011) studies the performance of 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. While that experiment studies individuals engaged in an identical task with heterogeneous values for the outcome, we consider individuals engaged in multiple tasks, with heterogeneous skills at each, while holding constant their overall costs and benefits. However, our input-substitutable

---

[2] For instance, in an early study on group contests, Nalbantian and Schotter (1997) find that inter-group 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.

treatment shares features of the weakest link game if teammates always specialize in separate tasks. 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 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), 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 conceptual framework for our experiment. In *Section 3*, we explain experimental design and the 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, are presented in an *Appendix*.

2 Conceptual Framework

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 more time he must devote to Task B.

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, Huang, and Sefton, 2015), in order to promote subject understanding and to generate team-level dominant strategies. In other words, individuals choosing how to allocate time across tasks need only form beliefs about their teammates' strategies, and not beliefs about the other team's strategies. In this section, we describe the predictions for the linear contest success function, including for the specific parametrization of our experiment. We note, however, that the gist of our predictions can be generalized to other contest success functions, such as the Tullock contest.

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:

$$\text{Probability of Success } (x_i, x_k) = \tfrac{50+x_i-x_k}{100}$$

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 i's payoff is given by:

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

From the first order condition, $V/100 = f/\alpha_i + f/\beta_i$, we observe 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 $\beta_j = 25$. For both individuals, the marginal cost of producing a unit of output is equal to 1.2 ($= 24/30 + 24/60 = 24/100 + 24/25$) and the marginal benefit is equal to a 1/100 increase in probability of winning the prize of 200, or 2. Therefore, participants in the individual contest of our experiment are expected to work fully. 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 $X_m = min\{a_i, b_i\} + min\{a_j, b_j\}$ and the winning team shares the prize, V. An agent producing x_i units, paired with a teammate producing x_j, against an opposing team producing a of X_k units has utility:

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

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, Huang, and Sefton, 2015), which generates the same prediction. In our team contest, individual i will devote all of his time to the tasks and produce the maximum amount whenever $V/200 > f/\alpha_i + f/\beta_i$.

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 social dilemma: they maximize their own payoffs by working fully in an individual contest but fully freeride in the team contest, which is socially inefficient. This is the situation faced by the individuals 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. To the extent that individuals do work, they are expected 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 he chooses to work.

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 $f/\beta_j < V/200$,[3] then the teammates find themselves in 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 β_j and j is willing to work on Task B up to α_i. In this case, any outcome such that $a_i = b_j < \min\{\alpha_i, \beta_j\}$ can be sustained as an 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 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.[4] 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). If Type 2 (agent j) were motivated to continue working after

---

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

[4] 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. 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.

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.[5] 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. In half of the sessions, the individual contests came first and in half the team contests came first. 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 Gächter, Huang, and Sefton (2015). 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

---

[5] The instructions in a session were read in three groups: 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.

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 evenly spaced increments, so that someone spending x seconds on a task at which she has skill level α is always expected to catch $x\alpha/90$ balls. Finally, we eliminate the cost per click that was used by Gächter, Huang, and Sefton (2015) 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 seconds) 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. 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 = 3f/60 = f/20 and for an individual of Type 2, the total costs are f/100 + f/25 = 5f/100 = f/20. Likewise, the maximum output that each team member could produce independently is equal to 20.

In the contest stages, participants could choose how they wished to allocate their 90 seconds 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 Figure ??. At the beginning of the period, the screen is blank except for three buttons (Red, Blue, Neither), any of which could be pressed at anytime. If the participant pressed the "Play 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 second period (or .267 tokens per second spent on a task). They did not bear any cost when they hit neither button and the tasks were not displayed. The consequences of these costs in each treatment were described in Section 2.

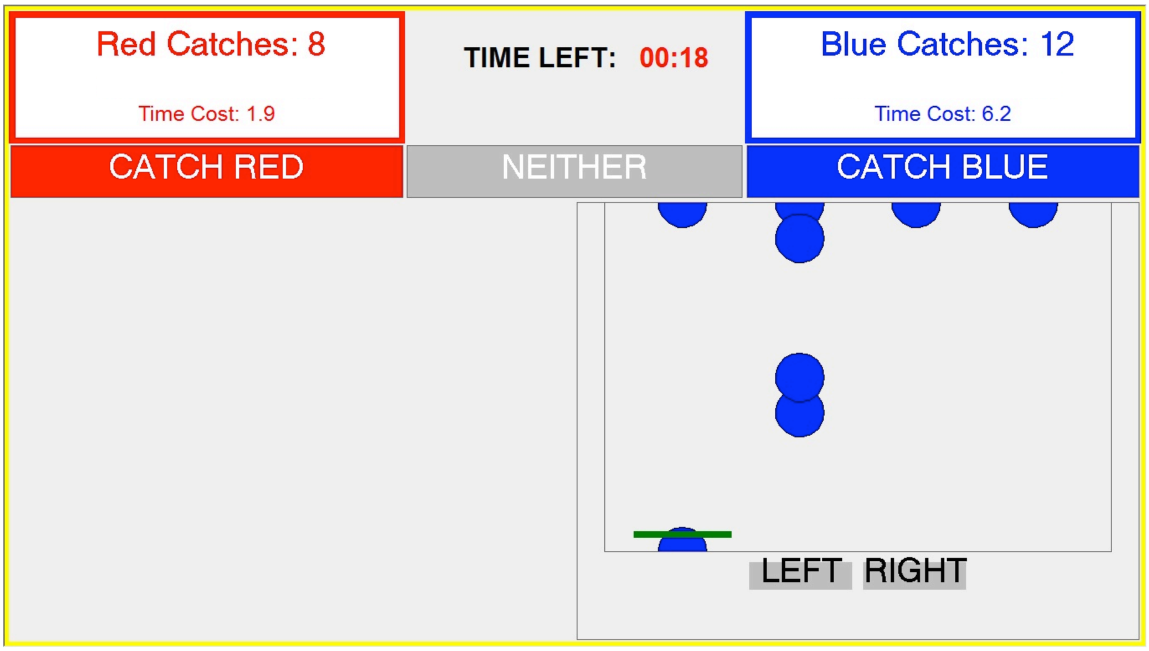

Figure 1: Contest Screen

In the experiment, we implemented a 3 × 2 design. Specifically, we varied the type of contest (i.e., individual contest, input-substitutable team contest, or output-substitutable team contest) and whether communication between teammates was possible. In the team contests, we always paired subjects with different productivity levels in Task A and Task B, to create complementarity. All subjects participated in a set of individual contests *and* a set of team contests, so that we can make within-subject comparisons between individual and team contests. However, we varied across subjects whether the team contests experienced were: output-substitutable without communication, output-substitutable with communication, input-substitutable without communication or input-substitutable with communication. The within-subject design for the individual contest allowed us to understand how the same subject allocates his time when he is competing for himself and when he is competing for his team. In order to check whether there is a role of learning due to repetition we also varied the order of individual contests and team contests.

In the individual contests, subjects compete with other individuals of the same productivity type for a prize of 200 tokens. They were matched with the same opponent for all 6 individual contest periods. If a subject won (lost) the contest, s/he earned 200 (0) points. A subject's points from the game were determined according to the minimum of red and blue balls s/he caught. The likelihood of winning the prize was determined by the subject's points and his/her competitor's points. We used the same linear contest function as Gill and Prowse (2012) and Gächter, Huang, and Sefton (2015): the probability that a subject earning x_i points wins a contest against an opponent earning x_j points is: $(x_i - x_j + 50)/100$. The computer then randomly determined the winner according to this probability. Regardless of whether they won the contest, the incurred time cost was subtracted from the subject's earnings. At the end of each period, subjects saw how many of each colored ball they collected, their points (output) during the contest, and whether they won the contest or not. They did not receive feedback on the points scored by their opponent.

In the team contest, subjects were matched with the same partner for all 6 periods, such that Type 1 individuals were matched with Type 2 individuals. They competed against the same team for these 6 periods. The team's points depended on the treatment assigned to them, either output-substitutable or input-substitutable. 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 teammates' 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). The winning team shares the 200 token prize and the contest success function is the same as in the individual contest: The probability that a team earning X_i points wins a contest against an opposing team earning X_j points is: $(X_i - X_j + 50)/100$.

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 team contest, subjects saw how many of each colored ball they collected (as in the individual contest) as well as how many of each colored ball their teammate collected. As in the individual contest, they did not receive feedback on the performance of the opposing team. Half of the teams 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 seconds. 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. There were sixteen sessions and each lasted 60 minutes. Each subject participated in only one session. 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.

4 Results

In this section, we examine the overall productivity of individuals and teams in each type of contest, before turning to the question of how participants of different skill-types produce and allocate their time to different tasks.

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 expectation. 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 **??**, we summarize the performances in the first part of the experiment, when subjects are paid according to a piece rate and complete only one task at a time. 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 1: Performances in the Piece-Rate

Types	Piece-Rate	Task A	Task B
Type 1	In the Experiment	29.3(0.1)	59.9(0.4)
(30/60)	*Expected*	*30*	*60*
Type 2	In the Experiment	97.1(0.4)	24.8(0.1)
(100/25)	*Expected*	*100*	*25*

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.[6] Overall, the output per teammate in the team contests is significantly greater than the output produced by individuals in the individual contest. Taking the individual (in individual contests) or team (in team contests) as the unit of observation, we find that the difference is significant at all conventional levels (15.34 vs. 19.56, $Z = 3.74$, $p < .01$). Table **??** further breaks down these results by production technology and communication. First, from the first row of the table, we observe 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 ($Z = 0.281$, $p = .779$), individuals

---

[6] 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.

in the input-substitutable teams are significantly more productive than either individuals or teams where collaboration is not possible.

Moving to the next two rows of Table **??**, we see that input-substitutable teams are more productive both without communication (row 2) and, even more so, when teammates can chat (row 3). While teammates in the output-substitutable teams are similarly productive with and without communication, the availability of chat has a significant positive effect on productivity in the teams where collaboration is possible.

Table 2: Output by Treatment

	Individual	Output Subs.	Input Subs.	z (Output = Input)
Overall	15.34	16.15	22.97	4.85***
No Chat	-	15.98	21.16	2.99***
Chat	-	16.33	24.82	3.87***
z (Chat = No Chat)	-	.289	2.05**	-

Notes: Each individual (column 1) or each team (columns 2 and 3) is considered one observation.

Figure **??** graphs the average output per person over the course of the contest series and further confirms the results of Table **??**: 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. Further, we see no evidence of a "decay" in teammates' efforts, as is often observed in public goods games where participants face a similar social dilemma (Ledyard, 1995). In fact, team output significantly *increases* over time in both input-substitutable treatments and in the output-substitutable treatment without communication.

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), seems to suggest that free-riding may not be much of a factor in our team contests.

This is confirmed by Figure **??**, which presents the amount of free-riding in each treatment, where free-riding is the defined as the percentage of the 90-second contest the individual spent on neither of the two tasks. (The allocation of time across Task A and Task B will be more fully

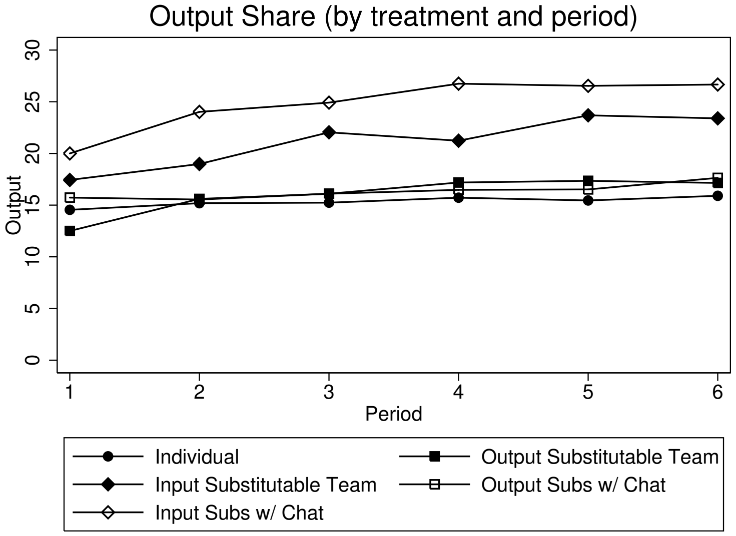

Figure 2: Output over Time by Treatment

explored in the following subsection.) Taking each individual as the unit of observation, there are no significant pair-wise differences across any of the treatments ($p > .30$ in all cases). Qualitatively, the level of free-riding is actually highest in the individual contest (where no free-riding incentives exist) and lowest 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 individuals 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. The estimates are reported in Table **??** and provide several overall insights regarding each type of individual's performance in each task. Although we do not expect any effect of type on the number of catches in Task A and in Task B in the individual contests or in the output-substitutable

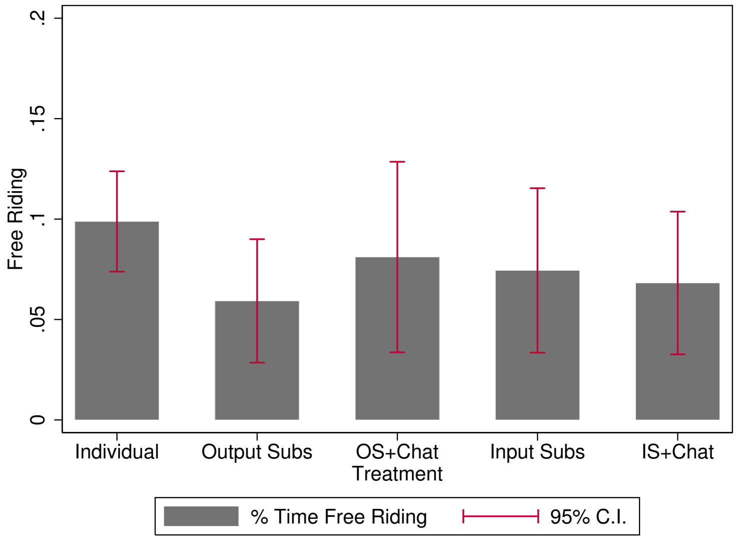

Figure 3: Percentage of 90-second period spent on neither task. The individual is the unit of observation.

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). This result indicates that subjects perform more on their stronger task even when efficient production requires that they perform each equally. Moving to the estimates for the input-substitutable contests, we observe that the coefficient on the Type 2 indicator variable becomes more extreme, indicating that subjects in input-substitutable teams move in the direction of specializing in their stronger task. Finally, we observe that gender, contest order, and contest period have very little effect on performance. We therefore aggregate the data in the analysis to follow.

The number of units of each task produced by each type is presented in Figure **??**, 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 Figure **??**. 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

Table 3: Task Performance by Contest Type

	Individual Contest		Input-Subs. Team Contest		Output-Subs. Team Contest	
	Catches A	Catches B	Catches A	Catches B	Catches A	Catches B
Type 2	8.3***(5.5)	-5.5***(-6.1)	42.4***(9.9)	-34.4***(-10.0)	11.3***(3.0)	-6.8***(4.0)
Period	-0.5(-1.5)	-0.3(-1.9)	-0.2(0.5)	0.4(1.3)	-0.8*(-2.1)	-0.1(-0.4)
Ind. First	3.7*(2.1)	1.8(1.5)	-7.2(-1.7)	-3.8(-1.3)	1.3(0.5)	-0.1(-0.1)
Male	-0.6(-0.4)	0.2(0.3)	-3.6(-1.0)	3.5(1.3)	-1.4(-0.6)	-0.5(-0.4)
Comm			-1.9(-0.8)	2.9(0.6)	-0.1(-0.1)	-1.0(-0.6)
Comm x Type			15.9**(2.5)	-4.9(-1.0)	-7.0(-1.7)	2.1(1.0)
Constant	18.2***(8.1)	22.1***(16.5)	14.3***(4.3)	40.9***(10.3)	22.5***(6.7)	23.5***(9.3)
Subjects	192		96		96	
Observations	1152		576		576	

Standard errors are clustered at the subject level
Dependent Variables: Catches in Task A and Task B
Z-values reported in parentheses.
***=statistically significant at 1% level **=statistically significant at 5% level

can reject the hypothesis that the production in Task A is equal to production in Task B at all conventional levels ($Z = 6.8$ for Type 1 and $Z = 6.7$ for Type 2, with the individual as the unit of observation). 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 ($Z = 2.27$, $p = .023$).

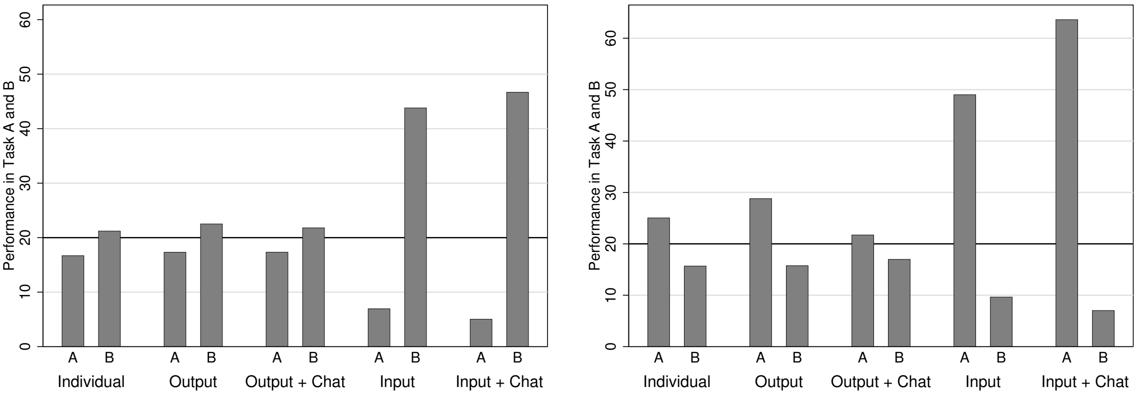

Figure 4: Performance of Task A and Task B for Type 1 (Left) and Type 2 (Right) by Treatment
Note: The vertical line shows the output-maximizing performance in the Individual and Output-Substitutable Team Contests.

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 Figure **??**. 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 point is higher than the gain obtained through each point and thus payoff-maximizing subjects in the output-substitutable team contest should not perform at all in either task. Nevertheless, their performance in team contests with or without communication is very similar to their performance in the individual contest. Similar to our observation for subjects in the individual contest, Type 1s and Type 2s overproduce on their stronger task and underproduce on their weaker task in output-substitutable team contests. As in the individual contest, the difference in Task A and B production is significant at all reasonable levels for both types, both with and without communication. 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 ($Z = 6.03$). While communication appears to have little effect on the behavior of Type 1 individuals, we note that there is some evidence of communication helping Type 2 individuals close the gap between Task A production and Task B production.[7]

The final bars in Figure **??** 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 1s performed Task B more than Task A, while Type 2s performed Task A more than Task B. In the input-substitutable treatments, we expect the team to produce 60 if teammates specialize in the task they do best, plus an additional 8 units if Type 2s 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 by only Type 2s and Type 1s should produce 0 units in Task A. However, as we can see from Figure **??** this is not the case: with or without communication, Type 1s do perform Task A instead of exclusively focusing on Task B and catch fewer than the 60 balls in Task B. On the other hand, Type 2s 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 Figure **??**, without communication, Type 2s underperform on the task

---

[7] Taking each individual as the unit of observation, the difference in the production gap for Type 2 individuals with and without communication is significant at the $p = .078$ level in a one-sided test.

they are better at, but, with communication they collect more balls than the maximum of their Type 1 partners. In particular, communication seems to have a role for subsidies by Type 2s.

Finally, we look at the amount of time that each type spent on Task A or Task B in each of the treatments, shown Figure **??**. For the individual contests, we expect Type 1 individuals to spend 67% of their time (60 seconds) on the task they are worse at (Task A) and 33% of their time (30 seconds) on the task that they are better at (Task B). We expect Type 2 individuals to spend 80% of their time (72 seconds) on the task they are worse at (Task B) and 20% of their time (18 seconds) on the task they are better at (Task A). The figure shows the percentage of working time (i.e., time not spent free-riding) 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. While, as predicted, both types do spend significantly more time on their weaker task ($Z = 6.11$ for Type 1 and $Z = 7.3$ for Type 2), they still allocate too much of their working time to their stronger task.[8]

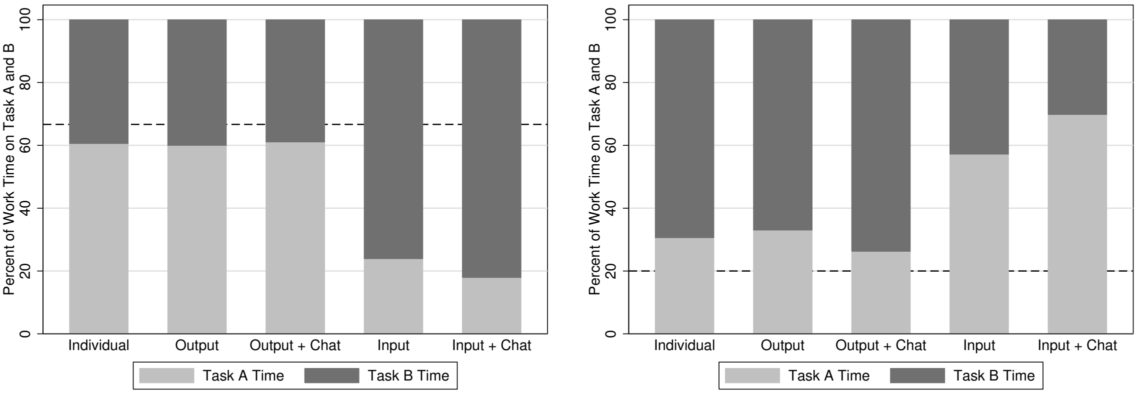

Figure 5: Percent of Work Time Spent on Task A and Task B for Type 1 (Left) and Type 2 (Right) by Treatment

Note: The vertical line shows the output-maximizing allocation to Task A in the Individual and Output-Substitutable Team Contests.

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 divide their time identically to participants in the individual contests: Type 1 teammates should spend 66.7% of their time on their

---

[8] 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 each individual as the unit of observation.

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 Figure **??**, 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 ($p = .5$ or greater 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%, $Z = .22$ and $p = .826$). However, with communication, they succeed in allocating *more* time to their weaker task (74.1%, $Z = 2.004$ and $p = .045$), 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 will 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 Figure **??**, this prediction is largely confirmed. Type 1 individuals 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.[9] Likewise, Type 2 individuals spend only 42.3% (no communication) and 31% (with communication)[10] of their time on their weaker task, down from 67% to 74% in the individual and output-substitutable contests.[11] 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 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

---

[9] 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.

[10] The amount of time spent on the weaker task is significantly lower with communication than without communication at $p = .03$.

[11] 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.

which they are most productive. To better understand how participants used the communication period in each treatment, we classify the content of their chats according to several categories: specifically, whether they discussed each teammate working separately to catch an equal number of each colored ball, discussed each teammate specializing in one task, discussed free-riding, or did not discuss the game at all. The percentage of chats corresponding to each category is presented in Table **??**. For each of the first two categories, we further consider the percent of conversations within this category in which participants set a target number of catches.

Table 4: Chat Content

Message Type	Output Substitutable	Input Substitutable
Suggest individuals catch an equal number of red and blue	54.17%	6.25 %
Given equal, set a target	38.46%	33.3 %
Suggest one person catch red and the other blue	6.25%	77.08%
Given specialization, set a target	0%	35.14 %
Suggest free-riding	4.16%	6.25%
Other	16.67%	10.42%
Did not talk	18.75%	0%

First, we note that participants in the output-substitutable teams most frequently discuss 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, over three-quarters of participants in the input-substitutable teams discuss a strategy of specialization in which one member completes each task, with over a third also establishing a target number of balls for the teammates to catch. Finally, nearly a fifth of the chat opportunities are not utilized by participants in the output-substitutable treatment, where, given that agents produce, the production strategy is unique and straightforward. In contrast, teammates in the input-substitutable treatment, where there exist a multiplicity of equilibria, always use the chat period to discuss the game.

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, Sonnemans, and van Winden (2001), chosen effort experiments finding below-expectation free-riding in team contests (Abbink et al., 2010; Ahn, Isaac, and Salmon, 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 contest.

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.

References

[1] Abbink, K., Brandts, J., Herrmann, B., and Orzen, H. (2010). “Intergroup Conflict and Intragroup Punishment in an Experimental Contest Game.” *The American Economic Review,* 100(1): 420-447.

[2] Ahn, T.K., Isaac, R.M., and Salmon, T. (2011). “Rent seeking in groups.” *International Journal of Industrial Organization*, 29: 116-125.

[3] Alchian, A. A. and Demsetz, H. (1972). “Production, Information Costs, and Economic Organization.” *The American Economic Review*, 62: 777-795.

[4] Babcock, P., Bedard, K., Charness, G., Hartman, J., and Royer, H. (2015). “Letting Down the Team? Social Effects of Team Incentives.” *Journal of the European Economic Association*, 13(5): 841-870.

[5] Bracha, A., Fershtman, C. (2013). “Competitive incentives: Working Harder or Working Smarter?” *Management Science*, 59(4): 771-781.

[6] Brewer, M. B. (1979). “In-group bias in the minimal intergroup situation: A cognitive motivational analysis.” *Psychological Bulletin*, 86(2): 307-324.

[7] Charness, G., Rigotti, L., and Rustichini, A. (2007). “Individual behavior and group membership.” *The American Economic Review*, 97(4): 1340-1352.

[8] Che, Y. and Gale, I. (2000). “Difference-Form Contests and Robustness of All-Pay Auctions.” *Games and Economic Behavior,* 30: 22-43.

[9] Chen, R., Chen, Y. (2011). “The Potential of Social Identity for Equilibrium Selection.” *American Economic Review*, 101(6): 2562-2589.

[10] Cooper, D. J., Sutter, M., (2013). “Role Selection and Team Performance.” IZA Discussion Paper No. 5892.

[11] Dechenaux, Emmanuel, Dan Kovenock, and Roman Sheremeta (2015). “A Survey of Experimental Research on Contests, All-Pay Auctions and Tournaments.” *Experimental Economics*, 18(4): 609-669.

[12] Eckel and Grossman (2005). “Managing diversity by creating team identity.” Journal of Economic Behavior and Organization, 58(3): 371-392.

[13] Epstein, G.S., Mealem, Y. (2009). “Group specific public goods, orchestration of interest groups with free riding.” *Public Choice*, 139 (3): 357-369.

[14] Esteban, J., Ray, D. (2001). “Collective action and the group paradox,” *American Political Science Review*, 95 (3): 663-672.

[15] Fischbacher, U. (2007) “z-tree: Zurich toolbox for ready-made economic experiments.” *Experimental Economics*, 10: 171-178.

[16] Gächter, S., Huang, L. and Sefton, M. (2015). “Combining 'Real Effort' with Induced Effort Costs: The Ball Catching Task.” *Experimental Economics*, forthcoming.

[17] Gill, D. and V. Prowse (2007). “A Structural Analysis of Disappointment Aversion in a Real Effort Competition.” *American Economic Review*, 102(1): 469 - 503.

[18] Goette, L., Huffman, D., and Meier, S. (2012a). “The Impact of Social Ties on Group Interactions : Evidence from Minimal Groups and Randomly Assigned Real Groups.” *American Economic Journal: Microeconomics*, 4(1): 101-115.

[19] Goette, L., Huffman, D., Meier, S., and Sutter, M. (2012b). “Competition Between Organizational Groups: Its Impact on Altruistic and Antisocial Motivations.” *Management Science*, 58(5): 948-960.

[20] Gunnthorsdottir, A. and Rapoport, A. (2006).“Embedding social dilemmas in intergroup competition reduces free-riding.” *Organizational Behavior and Human Decision Processes,* 101:184-199.

[21] Hamilton, B.H., Nickerson, J.A., and Owan, H. (2003), “Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation.” *Journal of Political Economy*, 111: 465-497.

[22] Holmstrom, B. (1982). “Moral hazard in teams.” *Bell Journal of Economics*, 13(2): 324-340.

[23] Katz, E., Nitzan, S., and Rosenberg, J. (1990). “Rent-seeking for pure public goods.” *Public Choice* 65(1): 49-60.

[24] Lazear, E. (1998). “Personnel Economics for Managers.” New York: John Wiley & Sons.

[25] Lazear, E. (2000). “Personnel Economics and Economic Approaches to Incentives.” *HKCER Letters* 61: 1- 8.

[26] Lee, S. (1995). “Endogenous sharing rules in collective-group rent-seeking.” *Public Choice*, 85(1-2): 31-44.

[27] Lee, D. (2012). “Weakest-link contests with group-specific public good prizes.” *European Journal of Political Economy*, 28: 238-248.

[28] Li, S. X., Dogan, K., and Haruvy, E. (2011). “Group identity in markets.” *International Journal of Industrial Organization*, 29: 104-115.

[29] Nalbantian, H. and Schotter, A. (1997). “Productivity under Group Incentives: An Experimental Study.” *The American Economic Review*, 87 (3): 314-341.

[30] Nitzan, S. (1991). "Collective Rent Dissipation." *The Economic Journal*, 101(409): 1522-1534.

[31] Nitzan, S., Ueda, K. (2009). "Collective contests for commons and club goods." *Journal of Public Economics*, 93 (1): 48-55.

[32] Olson, M., and Zeckhauser, R. (1966). "An economic theory of alliances." *The Review of Economics and Statistics*, 48(3): 266-279.

[33] Ryvkin, D. (2011). "The Optimal Sorting of Players in Contests between Groups." *Games and Economic Behavior*, 73: 564-572.

[34] Sheremeta, R. (2011). "Perfect Substitutes, Best shots and Weakest Link Contests between Groups." *The Korean Economic Review*, 27(1): 5-32.

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

[36] Sutter, M. (2009). "Individual behavior and group membership: Comment." *The American Economic Review*, 99(5): 2247-2257.

[37] Sutter and Strassmair (2009). "Communication, cooperation, and collusion in team tournaments – An experimental study." *Games and Economic Behavior*, 66: 506-525.

[38] Tajfel, H. and Turner, J. (1979). "An integrative theory of intergroup conflict." In: Austin, William, Worchel, Stephen (Eds.), The Social Psychology of Intergroup Relations, Brooks/Cole Publ., Monterey, CA, pp. 33-48.

[39] van Dijk, F. Sonnemans, J. and van Winden, F. (2001). "Incentive Systems in a Real Effort Experiment." *European Economic Review*, 45: 187-214.

[40] Winter, E. (2014). "Feeling Smart: Why our emotions are more rational than we think." New York: Public Affairs.