Sorting and communication in weak-link group contests

Philip Brookins* John P. Lightle† Dmitry Ryvkin‡

Abstract

We experimentally study the effects of sorting and communication in contests between groups of heterogeneous players whose within-group efforts are perfect complements. Contrary to the common wisdom that competitive balance bolsters performance in contests, in this setting theory predicts that aggregate output increases in the variation in abilities between groups, i.e., it is maximized by the most unbalanced sorting of players. However, the data does not support this prediction. In the absence of communication, we find no effect of sorting on aggregate output, while in the presence of within-group communication aggregate output is 33% higher under the balanced sorting as compared to the unbalanced sorting. This reversal of the prediction is in line with the competitive balance heuristic. The results have implications for the design of optimal groups in organizations using relative performance pay.

Keywords: group contest, sorting, complementarity, heterogeneous players, experiment

*Corresponding author: Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113, Bonn, Germany; e-mail: brookins@coll.mpg.de.

†Department of Economics, Virginia Commonwealth University, Richmond, VA 23284-4000, USA; jplightle@vcu.edu.

‡Department of Economics, Florida State University, Tallahassee, FL 32306-2180, USA; dryvkin@fsu.edu.

1 Introduction

Many organizations use relative performance pay schemes, or *contests*, to incentivize high effort from workers. Under such schemes, workers exert costly effort to increase their chances of winning a reward, e.g., a bonus or promotion. Production often takes place in teams where it is difficult to distinguish between individual inputs of team members; hence, contest incentives are applied to teams and the whole team is rewarded in the case of success.[1] For example, in an attempt to increase sales of U.S. beef, E-Mart Everyday, a chain of Korean grocery stores, awarded gift cards to the staff from the stores with the highest sales. As a result, the percentage of sales of U.S. beef sold by E-Mart Everyday increased six-fold (from 4 to 25%) and remained at high levels afterwards.[2] Similarly, Adventist Health System, a healthcare organization based in Florida, held a contest between several hospitals' inpatient and emergency department units with the goal of improving patients' satisfaction. Staff members at top-performing hospitals received free registration to an upcoming conference with a monetary value of US$1,300.[3]

Within virtually any organization, workers differ in their ability to perform tasks. Hence, organizations using team production and contest incentives may be interested in the following question: When workers are heterogeneous in their ability, and the objective is to maximize aggregate output, how should workers be *sorted* by ability into competing groups? For example, should a firm use a "balanced" sorting, in which high-ability and low-ability workers are mixed together so that aggregate ability does not vary much across groups, or should an "unbalanced" sorting be used, where groups are stratified by ability? As it turns out, the answer to this question is not unique and depends on the degree of complementarity of individual efforts in team production.

In this paper, we experimentally explore how the sorting of heterogeneous players affects aggregate output in a group contest when the total output of each group is characterized by the perfect complements, or "weak-link," production technology. The success of many organizations often depends on the productivity of the least able worker; therefore, using a weak-link production technology is natural and widely applicable. For an example of such strong complementarity, consider the operation performed by Transportation and Security Administration (TSA) workers within U.S. airports. The TSA workers oper-

[1] Since the late 1980s, the number of organizations utilizing teamwork has increased substantially (Lazear and Shaw, 2007). From 1987 to 1999, the percentage of firms with at least 20% of personnel assigned to teams increased from 37 to 61% (Lawler, Mohrman and Benson, 2001; Lawler, Mohrman and Ledford, 1995), and so has the use of team contests (Chen and Lim, 2013).

[2] Source: `http://www.agweb.com/article/sales_competition_boosts_us_beef_at_korean_grocery_chain_NAA_News_Release/`.

[3] Source: `http://www.adventisthealthsystem.com/page.php?section=news&page=article&id=1393`.

ate in tandem, each with their own specialization: Checking identification and boarding passes, administering random screenings, examining luggage, operations oversight, etc. If any one of these workers performs inefficiently, the entire operation is impeded.[4] For another example, consider the operation performed by emergency department employees at a given hospital. Emergency medical technicians, triage nurses, general physicians and surgeons each play a highly specialized role in the caring for patients.

Laboratory experiments are utilized extensively in personnel economics as an alternative methodology to analyzing naturally occurring data. Field data, even if available at the employee level, only contain information on production outcomes (e.g., a measure of output), but not on the underlying effort and other factors, such as ability and luck. Moreover, assignment of workers to groups in the field is endogenous, and hence it would be difficult to isolate the effects of sorting. A laboratory experiment allows for a much tighter control over workers' incentives, the team production technology, abilities, winner determination process and group assignment.

We consider two sortings of heterogeneous players in our experiment: (i) the *unbalanced* sorting which maximizes the variance in ability across groups, and (ii) the *balanced* sorting which minimizes the variance. As we show in Section 3, when group production follows weak-link aggregation and the contest is modeled assuming an imperfectly discriminating lottery,[5] the unique optimal sorting of players is the unbalanced sorting. For the parameters of the experiment, total predicted output in the case of the unbalanced sorting exceeds the output under any other sorting by at least 25%. This suggests that managers and supervisors can benefit from the careful assignment of personnel into competing groups, especially considering the low-cost nature of making such a decision.

In a group contest with weak-link aggregation, equilibrium efforts within each group are equalized. There are multiple equilibria (Lee, 2012), but in this paper, for reasons that are explained in detail in Section 3, we focus on the unique coalition-proof equilibrium which is effectively the equilibrium in a contest between the lowest-ability individuals from each group. In the experiment, we use contests between two groups of two players each[6]. In our setting, the coordination game involves heterogeneous players and is embedded

[4]While the main purpose of the TSA is to protect the nation's transportation systems, the organization's *Vision* places a great deal of importance on the efficiency of transportation security.

[5]Under the lottery, the group with the highest output does not win the contest with certainty but its probability of winning is proportional to the share of its output in the total output of all groups. By using a noisy winner determination process, we model situations in which it may be difficult or impossible for the principal to measure output precisely, or output is a combination of effort and idiosyncratic luck. This is in contrast to perfectly discriminating contests where participants with the highest output win with certainty (e.g., Baye, Kovenock and De Vries, 1996).

[6]Two-player groups are the most likely to succeed at coordinating on the coalition-proof Pareto-dominant equilibrium.

into the strategic contest environment, and we would like to make coordination as easy as possible so that subjects could focus more on the contest component of the incentives. We assign subjects one of four types – A, B, C, or D – that differ in their cost of effort from lowest to highest cost (i.e., highest to lowest ability). Under the *balanced* sorting, types A and D compete against types B and C, whereas under the *unbalanced* sorting types A and B compete against types C and D. Thus, in the coalition-proof equilibrium, under the balanced sorting the group contest is effectively reduced to an individual contest between types C and D, while under unbalanced sorting it is reduced to a contest between types B and D. Because total equilibrium effort is higher in the latter case, the unbalanced sorting yields higher predicted aggregate output than the balanced sorting.

In addition to sorting, we explore the effect that within-group communication has on individuals' effort decisions. Exploring how communication affects the behavior of individuals in group contests is not only interesting because of its natural occurrence in organizations, but also because it may improve coordination between group members and help them play more strategically. Players in a group contest with weak-link aggregation face a somewhat similar setting, in the following sense. Consider a group contest between groups 1 and 2, with two players each. For a given output E_2 of group 2, consider a reduced game involving effort choices (e_{11}, e_{12}) of players 11 and 12 of group 1. This is a coordination game with multiple Pareto-ranked equilibria of the form (x_1^*, x_1^*), where the set of possible equilibria dependson E_2 (cf. Lee, 2012). The main difference with the simple coordination setting is that in group contests a group's success not only depends on the coordination of efforts within a group, but also on the competition between groups, i.e., a group's coordination problem is embedded into another strategic environment. One important consequence of this is that the Pareto ranking of equilibria in the reduced game is no longer unidirectional, i.e., increasing group output does not necessarily lead to higher payoffs.

The balanced sorting minimizes the variance in ability between groups; however, it leads to a large variance in ability within each group, which creates an interesting within-group dilemma: When players within a group have misaligned incentives, how will communication affect their effort decisions? Will the high-ability player be able to convince her teammate to choose higher effort which benefits the high-ability player but hurts the low-ability player? Or will the low-ability player convince her teammate to choose a lower effort which is more in line with the low-ability player's best response? Or will they arrive at some sort of compromise? At the same time, in the case of unbalanced sorting the difference between players within each group is smaller, and it may be easier for the players to coordinate, both with and without communication.

In the absence of communication, total output under the balanced and unbalanced sortings are indistinguishable. This is achieved because (A,B) groups (i.e., the groups comprised of player types A and B) under the unbalanced sorting produce more than (A,D) groups under the balanced sorting, but (C,D) groups under the unbalanced sorting produce less than (B,C) groups under the balanced sorting, and the two differences nearly compensate each other. However, in the presence of communication we find that total output is significantly (33%) higher under the balanced sorting as compared to the unbalanced sorting, which is a reversal of the theoretical predictions. We find that (B,C) groups under balanced sorting increase their output in the presence of communication, while (C,D) groups under the unbalanced sorting have a surprisingly strong decrease in output when they can communicate.

Our results highlight important differences between the settings with symmetric and asymmetric groups competing in a contest with weak-link aggregation technology. We show that it is not always the case that all group types increase their output in the presence of within-group communication. Furthermore, our results imply that the "common wisdom" – that competitive balance is important in sustaining competition and high effort provision in group contests – is more robust than suggested by theory. Simply put, competitive balance does not appear to be beneficial without communication, but in the more ecologically valid setting which allows for communication the positive role of competitive balance is restored.

The rest of the paper is organized as follows. We begin by briefly discussing the most relevant literature in Section 2. Section 3 presents the theoretical model and predictions that serve as the basis for the experimental design presented in Section 4. Section 5 presents the experimental results and Section 6 contains a discussion and concluding remarks.

2 Related literature

In addition to the expansive literature on individual contests (for reviews see, e.g., Lazear, 1999; Connelly et al., 2014; Dechenaux, Kovenock and Sheremeta, 2015), there is a growing experimental literature that focuses exclusively on group contests (for a review see Sheremeta, 2017). While most of these examine the case of symmetric players and perfectly substitutable effort within groups (see, for example, Nalbantian and Schotter, 1997; Abbink et al., 2010; Ahn, Isaac and Salmon, 2011; Cason, Sheremeta and Zhang, 2012, 2015), a few experimental studies have considered group contests with heterogeneous players. Sheremeta (2011) compares performance under three types of group effort aggregation technologies – weak-link, best-shot, and perfect substitutes – in a setting with two competing groups of three players (one high-valuation player and two low-valuation players in each group) and linear costs of effort. Brookins, Lightle and Ryvkin (2015*a*) is the only existing experimental study of the effects of sorting heterogeneous players in group contests. For perfectly substitutable within-group effort, the authors find, in agreement with theory, that balanced sorting produces a higher aggregate output. One common feature of most group contest experiments, similar to individual contests, is overinvestment of effort as compared to the risk-neutral Nash equilibrium predictions (Sheremeta, 2013). In the presence of heterogeneous players, low-ability (low-valuation or high-cost) players tend to overinvest more.

Our experiment builds on the existing theoretical analysis of group contests. Lee (2012) characterizes the set of equilibria present in group contests with weak-link aggregation and linear costs of effort, but does not consider the question of sorting which is relevant to our study. Rather, the optimal sorting of heterogeneous players in group contests was first explored by Ryvkin (2011). Provided the cost function of effort is not "too steep," Ryvkin (2011) theoretically shows that when within-group efforts are perfectly substitutable and the probability of winning is given by the lottery contest success function (CSF) (Tullock, 1980), the optimal sorting is the one that minimizes the variance in ability across groups, i.e., the most "balanced" sorting. Extending these results to group contests with arbitrary levels of within-group complementarity, Brookins, Lightle and Ryvkin (2015*b*) show that the optimal sorting may be either balanced or unbalanced, and depends on both the degree of complementarity of efforts within groups and steepness of the effort cost function.

Pre-play communication and group decision-making can help players behave more strategically in complex environments, as shown in experiments on signaling games (e.g., Cooper and Kagel, 2005) or contests (Sheremeta and Zhang, 2010). For simple coordination games with multiple Pareto-ranked equilibria, many studies have found that the presence of communication can help players overcome coordination failure and coordinate on more efficient equilibria (e.g., Cooper et al., 1992; Blume and Ortmann, 2007; Brandts and Cooper, 2007; Devetag and Ortmann, 2007). For group contests with weak-link aggregation, Cason, Sheremeta and Zhang (2012) explore how communication affects effort decisions when players are symmetric. In the presence of within-group communication, the authors find significantly higher group output as well as improved coordination within groups. However, as a result of the increased between-group competition, individual payoffs decreased, on average, due to substantial overinvestment as compared to equilibrium effort levels. Similar results are obtained by Cason, Sheremeta and Zhang (2015) who allow symmetric players to endogenously choose whether or not to communicate with each other. Most subjects choose to communicate, which also leads to higher group output and lower individual payoffs, on average, due to substantial overprovision of effort. While Cason, Sheremeta and Zhang (2012, 2015) use group contests with the Tullock CSF, Sutter and Strassmair (2009) obtained similar results in a group tournament setting with a Lazear and Rosen (1981) style CSF and perfectly substitutable within-group effort.

The experiment we present in this paper contributes further to the finding in the literature that communication in group contests can lead to unanticipated results. In contrast to Cason, Sheremeta and Zhang (2012, 2015), the focus of the present paper is on sorting; therefore, we consider a weak-link group contest of *heterogeneous* players, with heterogeneity both within and between groups, which has not yet been explored in the literature. The effect of communication in this setting, if any, may vary depending on the sorting.

3 Model and theoretical predictions

We consider a contest between two groups, indexed by $i = 1, 2$, of two risk-neutral players, indexed by $ij = i1, i2$. Players simultaneously and independently exert efforts $e_{ij} \geq 0$ and incur costs $c_{ij} e_{ij}^{\gamma}$, $\gamma > 1$, where $c_{ij} > 0$ are the players' heterogeneous cost parameters. The output, E_i, of group i is given by the weak-link technology: $E_i = \min\{e_{i1}, e_{i2}\}$. The probability that group i wins the contest is given by $p_i = \frac{E_i}{E_1+E_2}$.[7] Each member of the winning group receives a prize $V > 0$, while the members of the other group receive nothing.[8] Thus, the expected payoff of player ij in group i is given by

$$\pi_{ij} = V \frac{E_i}{E_1 + E_2} - c_{ij} e_{ij}^{\gamma}. \tag{1}$$

It is obvious that in equilibrium efforts within each group will be equalized, i.e., $e_{i1} = e_{i2}$ for $i = 1, 2$. As discussed by Lee (2012), there may be a continuum of equilibria in a group contest.[9] Here, we focus our attention on the unique coalition-proof equilibrium, which is Pareto optimal within each group and determined by the competition of the high-cost players in each group in an effective individual contest. We focus on the coalition-proof equilibrium for several reasons. First, we use a fixed matching protocol in

[7]In the event that $E_1 + E_2 = 0$, we assume $p_i = 1/2$.

[8]We assume that the prize is a group-specific public-good (see, e.g., Baik, Kim and Na, 2001). Alternative prize distribution rules have also been explored, such as Nitzan and Ueda (2011) and Nitzan and Ueda (2014).

[9]In Lee (2012), effort costs are linear. Although we assume a convex cost of effort, the equilibrium analysis remains the same.

our experimental design, which facilitates coordination and convergence to more efficient effort choices, from a group's perspective. Second, in two of our treatments we allow subjects to communicate, which has also been shown to improve coordination between group members. The coalition-proof equilibrium is also the equilibrium in which group members are exerting the highest levels of effort. Considering most experimental studies on contests report rampant overbidding (Sheremeta, 2013), we expect to observe behavior to be closest to this equilibrium.

Let $x_i = e_{i1} = e_{i2}$ denote the equilibrium effort in each group. Without loss of generality, suppose $c_{i1} < c_{i2}$ for $i = 1, 2$, i.e., players $i2$ are the high-cost players. In the coalition-proof equilibrium the payoffs of players 12 and 22 are

$$\pi_{12} = V\frac{x_1}{x_1 + x_2} - c_{12}x_1^\gamma, \qquad \pi_{22} = V\frac{x_2}{x_1 + x_2} - c_{22}x_2^\gamma. \tag{2}$$

The following proposition characterizes the equilibrium.

Proposition 1 *In the individual contest with payoffs defined by (2), there exists a unique coalition-proof equilibrium in the group contest with effort levels given by*

$$e_{11}^* = e_{12}^* = \left(\frac{V(\frac{c_{12}}{c_{22}})^{1/\gamma}}{\gamma c_{12}(1 + (\frac{c_{12}}{c_{22}})^{1/\gamma})^2}\right)^{1/\gamma}, \quad e_{21}^* = e_{22}^* = \left(\frac{V(\frac{c_{12}}{c_{22}})^{1/\gamma}}{\gamma c_{22}(1 + (\frac{c_{12}}{c_{22}})^{1/\gamma})^2}\right)^{1/\gamma}. \tag{3}$$

The equilibrium efforts (3) are obtained by solving the system of first-order conditions corresponding to the payoffs (2). The proof of Proposition 1 follows directly from the analysis of Lee (2012), with efforts and group output redefined as $y_{ij} = e_{ij}^\gamma$ and $E_i = \min\{y_{i1}^{1/\gamma}, y_{i2}^{1/\gamma}\}$, respectively.

In the experiment, we use the following parameterization of the model. Two groups consisting of two players compete for a prize of $V = 1000$, which is awarded to each member of the winning group. The exponent on the individual cost of effort is $\gamma = 1.2$, i.e., a total cost of effort for player ij of $c_{ij}e_{ij}^{1.2}$. Players are heterogeneous in their cost parameters c_{ij} and are assigned to Types A, B, C and D, with associated cost parameters $c_A = 5, c_B = 7, c_C = 13$ and $c_D = 15$, respectively. This implies that Type A players are the lowest-cost (highest-ability) players, followed by Type B players, and so on. In the *unbalanced* treatments, an (A,B) group (a group comprised of Types A and B) competes against a (C,D) group, i.e., $c_{11} = c_A, c_{12} = c_B, c_{21} = c_C$ and $c_{22} = c_D$. In the *balanced* treatments, an (A,D) group competes against a (B,C) group, i.e., $c_{11} = c_A, c_{12} = c_D, c_{21} = c_B$ and $c_{22} = c_C$. For convenience, we use the term *cost follower* (*cost leader*) to refer to the member of a group with the *highest* (*lowest*) cost of effort.

	Balanced				Unbalanced			
	Group 1		Group 2		Group 1		Group 2	
Type	A	D	B	C	A	B	C	D
Individual effort	8.9	8.9	10.1	10.1	15.6	15.6	8.2	8.2
Individual payoff	801.0	662.6	818.0	722.2	918.9	865.0	582.8	557.7
Group output	8.9		10.1		15.6		8.2	
Group payoff	1,463.6		1,540.2		1,783.9		1,140.5	
Winning probability	0.4702		0.5298		0.6536		0.3464	
Total output	19.0				23.8			

Table 1: Predicted equilibrium effort levels, group output, payoffs, and winning probabilities.

Given the parameters of the experiment, the predicted equilibrium individual effort, individual payoff, group output, aggregate group payoff, group winning probability, and total contest output can be computed using Proposition 1. The results are shown in Table 1 and lead to the following hypotheses.

Hypothesis 1 *Within-group efforts are equalized.*

Hypothesis 2 *Total output is higher under the unbalanced sorting than under the balanced sorting.*

Hypothesis 3 *(a) Type A's effort and type B's effort are higher under the unbalanced sorting than under the balanced sorting.*
(b) Type C's effort and type D's effort are higher under the balanced sorting than under the unbalanced sorting.

Hypothesis 1 follows directly from Proposition 1, and is independent of sorting. Recall that, following Proposition 1, equilibrium efforts are determined by the effective individual contest between cost followers in each group. Thus, under the unbalanced sorting the group contest reduces to competition between Types B and D whereas under the balanced sorting it reduces to competition between Types C and D. The average ability, and hence total output, is higher in the former case. For the parameters of the experiment, this leads to a 25.3% higher total output under the unbalanced sorting than under the balanced sorting (Hypothesis 2). Parts (a) and (b) of Hypothesis 3 are a direct consequence of sorting. Types A and B (C and D) are better off in the unbalanced (balanced) sorting due to the benefit associated with the increased ability of the other member of their group relative to the balanced (unbalanced) sorting.

In addition to sorting, we also explore the effects of pre-play communication. In the communication treatments, subjects are allowed to chat with the other member of their group before making individual effort decisions. Agreements made, if any, between the two members are non-binding and can, therefore, be considered "cheap talk." While the presence of cheap talk does not change the equilibrium predictions in Table 1, there exists experimental evidence that cheap-talk may have significant effects on behavior. For example, in simple weak-link games, it helps players coordinate better on Pareto-dominant equilibria (e.g., Cooper et al., 1992). Communication may also help players make better strategic decisions in complex environments such as signaling games (Cooper and Kagel, 2005) and beauty-contests (Kocher and Sutter, 2005). However, the experimental evidence on the effects of communication and group decision-making in contests is somewhat contradictory. On the one hand, Sheremeta and Zhang (2010) find that communicating teams making a joint decision choose lower contest expenditures than individuals. Given that individuals typically overexpend effort in such contests, the results suggest that teams behave more strategically. On the other hand, for group contests with weak-link aggregation Cason, Sheremeta and Zhang (2012) and Cason, Sheremeta and Zhang (2015) find that within-group communication leads to higher effort, i.e., it moves players further away from equilibrium.

In all the settings discussed in the previous paragraph, players are symmetric. In our setting, however, heterogeneity exists within and between groups, which may lead to several interesting alternative hypotheses. First, it is not clear whether communication will improve coordination within groups. Cost leaders may be trying to convince cost followers to exert higher effort, and vice versa, with various degrees of success, depending on sorting. Second, communication may or may not help players make better strategic decisions. Strong heterogeneity may weaken group identity and lead to more individualistic behavior. At the same time, communication and exchange of ideas may facilitate the adoption of strategies that are closer to best response.

4 Experimental design

Our experiment is designed to test how both sorting and communication affect investment in a group contest where individual efforts within a group are perfect complements. We, therefore, employ a 2×2 between-subject design with four treatments outlined in Table 2, where subjects are sorted into groups using either a balanced or an unbalanced sorting, and within-group communication may or may not be available.

All sessions were conducted at a standard laboratory at a large public university, with

	Communication Condition	
Sorting Condition	No Communication	Communication
Balanced	BAL-NOCOMM (2 sessions, 44 subjects)	BAL-COMM (2 sessions, 40 subjects)
Unbalanced	UNBAL-NOCOMM (2 sessions, 44 subjects)	UNBAL-COMM (2 sessions, 44 subjects)

Table 2: Summary of experimental treatments.

the experiment programmed in z-Tree (Fischbacher, 2007) and 172 subjects (60.4% of them female) recruited using ORSEE (Greiner, 2004) from a pool of pre-registered FSU undergraduate students. Subjects earned \$21.82 on average, including a \$10.00 show-up fee, for a session that lasted approximately one hour, on average. The experiment consisted of three parts. At the beginning of each part, printed instructions for that part were distributed and read aloud. The instructions for the group contest part of the experiment can be found in Appendix A.[10]

In the first part of the experiment, we elicited subjects' risk, ambiguity and loss aversion in a series of three decision tasks similar in spirit to the list elicitation method of Holt and Laury (2002).[11] In the second part of the experiment, all subjects performed a number counting task (Ryvkin and Semykina, 2013) for three minutes. On a computer screen, subjects were shown a string containing 20 random letters and numbers and were asked to count how many numbers appeared in the string. They entered their answer into the computer, after which a new string would appear. The instructions stated that subjects who correctly completed more of the number counting tasks would earn an advantage over other subjects in the subsequent parts of the experiment, but no further details were provided at that point.

In the third part of the experiment, subjects competed in a group contest between two groups of two players. Subjects were assigned types based on their relative performance in the number counting task. Specifically, the top 25% of performers were assigned Type A, the second 25% were assigned Type B, the third 25% were assigned Type C, and the bottom 25% were assigned Type D, with ties broken randomly. Subjects kept their type throughout this part of the experiment. We chose the number counting task as a means for type assignment because performance on it is determined primarily by effort and should not be correlated with cognitive or mathematical ability; hence, it should not

[10] The rest of the instructions are straightforward and available from the authors upon request.

[11] Various versions of these measures are collected routinely in contest experiments; the results, however, are mixed (Sheremeta, 2013). We found no systematic effects of any of these measures on subjects' behavior. Detailed results are available upon request.

affect subjects' ability to make optimal decisions in the experiments. The goal of using the task for the assignment of types was to reinforce the idea that the types are "deserved" or "earned" and thus minimize other-regarding behavior (see, e.g., Hoffman et al., 1994).[12] The subjects participated in a total of 20 rounds of the group contest game, split into two sub-parts of 10 rounds each, with a short break in between. At the beginning of the first sub-part, subjects were told that it would last for 10 rounds, but they were not informed that the second sub-part would follow. At the beginning of the second sub-part, subjects were again informed that it would be 10 more rounds in exactly the same environment.[13]

Before the first round in all treatments, subjects were paired into groups and matched with an opposing group according to their types and whether the sorting was balanced or unbalanced. This matching remained fixed throughout this part of the experiment. The repeated interaction induced by the fixed matching allowed for a more natural setting in which groups receive feedback and have the capacity to learn both about how to play the game and also about their opponents' strategy. This design gives the subjects the best chance to coordinate on a Pareto-dominant equilibrium of the reduced game and to approach the coalition-proof equilibrium in the group contest game. For notational convenience, we will use "match" to refer to the composition of the two repeatedly competing groups.

In every round of the group contest game, subjects chose a level of investment toward a group project. They were given an endowment of 400 points and selected a level of investment which was an integer between 0 and 50. The level of investment is equivalent to effort in our model, and the cost of effort depended on a subject's type and the level of investment exactly as described in Section 3. Subjects were provided with a table which listed the costs associated with each level of investment for each type; thus, costs of investment for all types were common knowledge (see Table 7 in Appendix A). Group output was determined by the weak-link technology, i.e., by the minimum of the two efforts chosen in a group. The winning group was determined randomly according to the lottery contest success function described in Section 3. If a group won the contest, both members received a prize of 1000 points; otherwise, they received zero points. Thus, round payoff for subject ij was $1400 - c_{ij}g(e_{ij})$ if her group won and $400 - c_{ij}g(e_{ij})$ if her group lost.

[12]Using the fact that there is within-type variation in the performance on the task, we ran regressions of subjects' effort on their performance on the task for each type and treatment condition. We found no systematic correlations between task performance and effort chosen in the experiment. Out of 16 regressions, only 3 produced a significant correlation, negative in two cases and positive in one.

[13]Each time, the number of rounds was fixed and known to subjects; therefore, the equilibrium predictions of the one-shot game described in Section 3 hold for the subgame-perfect equilibrium of the super-game.

In the treatments with communication, subjects were given one minute in each round to communicate with the other member of their group using a computerized chat interface. Subjects could not communicate with members of the other group. Communication occurred immediately prior to the subjects' choosing their investments. Subjects were permitted to discuss anything provided they did not use profane, threatening, or self-identifying language. In the treatments without communication, in each round subjects went directly to the stage where investment was chosen.

The following feedback was provided to subjects after each round: The subject's own investment, the investment of their partner, their group's output, the group output of the other group, their group's probability of winning given the two outputs, and the realized outcome of the contest. Four randomly chosen rounds out of the 20, with two rounds selected from each sub-part, counted toward final payment at the exchange rate of \$1 = 200 points.[14]

After the first 10 rounds, there was a short break, and then the group contest restarted and continued for 10 more rounds. Types, group compositions, and opponents remained the same. Having a restart in the middle allows us to observe learning not only within but also between 10-round super-games. It also serves as an additional robustness check for the treatment effects.

At the conclusion of the experiment, subjects were paid anonymously by check the sum of their earnings from all parts.

5 Results

5.1 Preliminary analysis

Figure 1 shows average total output by round for each treatment. Without communication, average total output in UNBAL-NOCOMM (37.9) appears to be higher than in BAL-NOCOMM (33.5), as predicted by theory. The difference, however, is not statistically significant ($p = 0.502$).[15] In contrast, in the presence of communication average total output in BAL-COMM (47.5) is 25.4% higher than in UNBAL-COMM (37.8), and

[14]Paying subjects for a subset of rounds is common in the literature. While Azrieli, Chambers and Healy (2015) show that paying for one randomly selected round is the only incentive-compatible scheme that induces subjects to behave in each round as if it were a one-shot game, they discuss paying for several random rounds as a reasonable practical compromise.

[15]Here and below, unless specified otherwise, the p-values are two-sided and obtained by running OLS regressions on the appropriate set of dummy variables and their interactions, with match-level clustered standard errors, and conducting pairwise comparisons of coefficients using Wald tests. Clustering at the match level allows for arbitrary correlations among subjects' effort choices within and between groups and across time in each match, and thus constitutes a conservative approach.

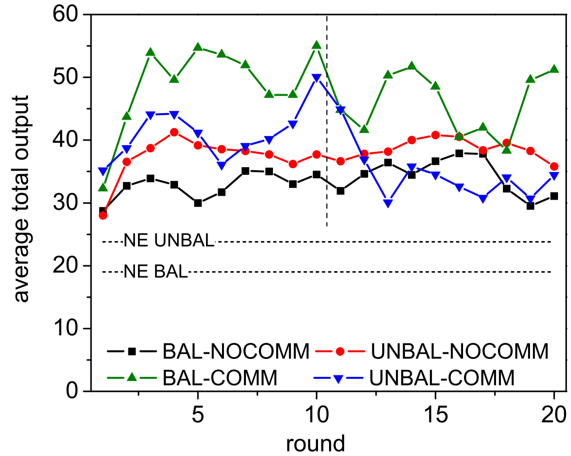

Figure 1: Average total output by round, by treatment.

this difference is statistically significant ($p = 0.079$). Interestingly, in the presence of communication the observed effect of sorting is a reversal of Hypothesis 2. Holding the sorting fixed and comparing output with and without communication, we find that average total output in BAL-COMM is significantly higher than in BAL-NOCOMM ($p = 0.033$), but the difference between UNBAL-COMM and UNBAL-NOCOMM is not statistically significant ($p = 0.988$).

It is also of interest to explore how well individuals coordinate efforts within each group, and examine how coordination evolves over time. As a measure of coordination between group members, we define *average wasted effort* of group i as the average of the difference between individual effort and group output, i.e., $Waste_i = (e_{i1} + e_{i2})/2 - \min\{e_{i1}, e_{i2}\}$ (Riechmann and Weimann, 2008).[16] Figure 2 shows average wasted effort (left panel) and the percentage of groups with nonzero waste (right panel) by round for each treatment. Recall that subjects went through two identical 10-round sub-parts, with a break in between. For convenience, we will refer to the first 10 rounds as Half 1 and the second 10 rounds as Half 2. From Figure 2, it is evident that both average wasted effort and the percentage of groups with nonzero waste are decreasing over time in Half 1 but mostly stabilizing in Half 2 in all treatments.[17] For all treatments, average wasted effort and the percentage of groups with nonzero waste are significantly lower in Half 2 as compared to Half 1, suggesting that group members coordinate on effort significantly

[16] Riechmann and Weimann (2008) define wasted effort of individual ij as $Waste_{ij} = e_{ij} - \min\{e_{i1}, e_{i2}\}$, and use the average of individual waste within a group to measure coordination success at the group-level.

[17] Because of the possibility of strategic signalling of efforts by group members, it is reasonable to relax the definition of coordination to include groups with positive, but small, waste. We analyzed the percentage of groups with waste greater than 1, 2 and 3. There still exists a negative time trend in Half 2 when looking at the percentage of groups with waste above 1 and 2, but this trend stabilizes when looking at the percentage of groups with waste above 3, for all treatments.

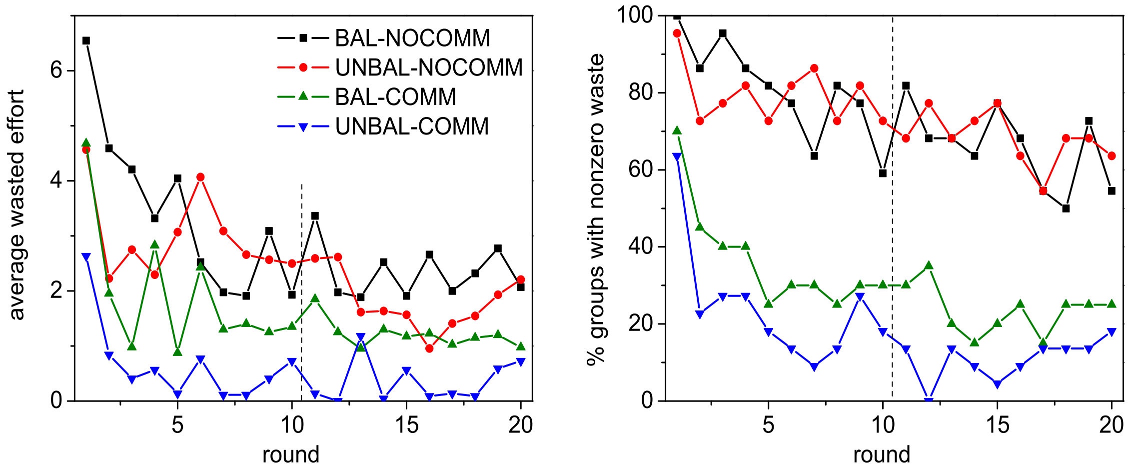

Figure 2: Average wasted effort (left panel) and percentage of groups with nonzero waste (right panel) by round, by treatment.

better in Half 2 relative to Half 1.[18]

Even though waste is significantly lower in Half 2 as compared to Half 1, average waste in Half 2 is still significantly above zero in all treatments ($p < 0.05$ for NOCOMM treatments and $p < 0.01$ for COMM treatments); thus, coordination remains imperfect and Hypothesis 1 is not supported by the data at the aggregate level. Nevertheless, coordination is successful 57.4% of the time across all treatments, with the highest success rate in UNBAL-COMM (89.1%), followed by BAL-COMM (76.5%), UNBAL-NOCOMM (34.1%), and BAL-NOCOMM (31.8%). As expected, communication facilitates coordination for both sortings. Moreover, coordination is better in UNBAL-COMM than in BAL-COMM ($p = 0.097$), i.e., with communication it is easier to coordinate efforts when players within groups are less heterogeneous. Somewhat surprisingly, the same is not true without communication, as there is no difference in average waste between treatments UNBAL-NOCOMM and BAL-NOCOMM ($p = 0.947$).

We also tested for differences in average waste between treatments separately in Half 2. The only statistically significant effect we find is that of communication under the unbalanced sorting ($p = 0.000$). Thus, after the initial learning period communication remains effective at reducing waste only when the players in each group are not too heterogeneous. For strongly heterogeneous groups under the balanced sorting, learning in Half 1 leads to the same average waste in Half 2 with and without communication.

[18] For differences in average waste between Half 1 and Half 2, the corresponding p-values for treatments BAL-NOCOMM, BAL-COMM, UNBAL-NOCOMM and UNBAL-COMM are 0.025, 0.000, 0.001 and 0.024, respectively. For differences in the percentage of groups with nonzero waste between Half 1 and Half 2, the p-values are 0.001, 0.017, 0.062 and 0.001, respectively, and are based on a logistic regression.

Result 1 *(a) In all treatments, average wasted effort and the percentage of groups with nonzero waste are significantly lower in Half 2 as compared to Half 1.*
(b) Average wasted effort in Half 2 is significantly greater than zero in all treatments, i.e., miscoordination of effort between group members persists at least in some groups.
(c) Communication helps improve long-run coordination under the unbalanced sorting but not under the balanced sorting.

In view of the results of this section, for the remainder of the analysis we will focus on Half 2 observations. Our goal is to identify robust treatment effects, and we expect play in Half 2 to be a better representation of how experienced subjects behave in the group contest environment. By Half 2, subjects significantly improve their within-group coordination in all treatments, especially in the presence of communication. The improved coordination in Half 2 may be due to experience gained through repeated play under fixed grouping and matching, or it may indicate subjects understand the game better, which is a reasonable assumption as wasted effort may serve as a proxy for subjects' understanding of the environment.

5.2 Average total output by treatment

Table 3 reports summary statistics of individual effort, group output, total output, waste, group winning probability, individual payoff, and group payoff in Half 2. In the absence of communication, we do not find a statistically significant difference in average total output between BAL-NOCOMM (34.3) and UNBAL-NOCOMM (38.6) ($p = 0.557$). However, in the presence of communication, average total output is significantly (33%) higher in BAL-COMM (45.9) than in UNBAL-COMM (34.5) ($p = 0.046$). Holding the sorting fixed and comparing the COMM and NOCOMM conditions, we find that average total output is significantly (33.8%) higher in BAL-COMM (45.9) than in BAL-NOCOMM (34.3) (p=0.098); however, there is no difference between UNBAL-COMM (34.5) and UNBAL-NOCOMM (38.6) ($p = 0.505$).

Result 2 *(a) In the absence of communication, there is no statistically significant difference in average total output between the balanced and unbalanced sorting.*
(b) In the presence of communication, average total output is 33% higher under the balanced sorting than under the unbalanced sorting.
(c) Under the balanced sorting, average total output is 33.8% higher in the presence of communication as compared to no communication.
(d) Under the unbalanced sorting, there is no statistically significant difference in average total output between the communication and no communication conditions.

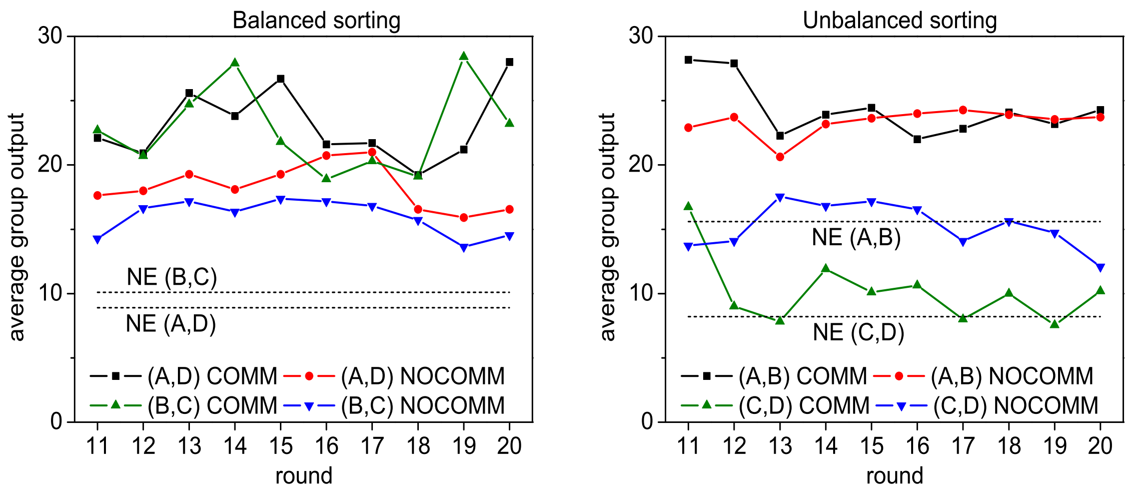

Figure 3: Average group output by treatment and group composition: (A,D) and (B,C) groups under the balanced sorting (left panel), and (A,B) and (C,D) groups under the unbalanced sorting (right panel).

Result 2 (a) and (b) are both contrary to theory, with the latter being in an interesting aggregate effect resulting in a reversal of the equilibrium predictions. Results 2 (c) and (d) combined show that a significant aggregate effect of communication is only present under the balanced sorting. Recall that previous studies of weak-link group contests with symmetric players find that communication significantly increases output relative to no communication (Cason, Sheremeta and Zhang, 2012, 2015). What is driving the difference between sortings when there is within-group communication? Why does communication only have an effect under the balanced sorting? Does communication affect each group composition in the same way? In the next section, we address these questions by disaggregating the data and analyzing output at the group level.

5.3 Average group output by treatment

In this section, we explore how each group type contributes to total output in order to gain insight into which groups are driving the aggregate treatment effects. Figure 3 shows average group output by treatment for each group composition: (A,D) and (B,C) groups under the balanced sorting (left panel), and (A,B) and (C,D) groups under the unbalanced sorting (right panel). For the balanced sorting, we do not find a statistically significant difference in the average group output of (A,D) groups between BAL-NOCOMM (18.3) and BAL-COMM (23.1) ($p = 0.227$), but we do find that the average group output of (B,C) groups is significantly higher in BAL-COMM (22.8) than in BAL-NOCOMM (16.0) ($p = 0.080$). For the unbalanced sorting, we do not find a statistically significant

	BAL				UNBAL			
	Group 1		Group 2		Group 1		Group 2	
Type	A	D	B	C	A	B	C	D
Individual effort								
Predicted	8.9	8.9	10.1	10.1	15.6	15.6	8.2	8.2
NOCOMM	22.0	19.5	18.5	18.0	25.2	24.4	18.2	16.5
	(3.1)	(2.6)	(3.2)	(3.0)	(3.3)	(3.1)	(2.9)	(2.3)
COMM	26.3	23.8	23.5	22.9	24.6	24.3	10.8	10.8
	(2.5)	(3.1)	(2.5)	(2.6)	(1.9)	(2.0)	(1.9)	(2.2)
Overbid NOCOMM	13.1	10.6	8.4	7.9	9.6	8.8	10.0	8.3
Overbid COMM	17.4	14.9	13.4	12.8	9.0	8.7	2.6	2.6
Waste NOCOMM	2.4		2.3		1.5		2.1	
	(0.5)		(0.7)		(0.3)		(0.5)	
Waste COMM	2.0		0.5		0.1		0.6	
	(1.2)		(0.2)		(0.1)		(0.3)	
Individual payoff								
Predicted	801.0	662.6	818.0	722.2	918.9	865.0	582.8	557.7
NOCOMM	790.3	450.4	557.9	360.9	862.6	777.2	248.5	238.7
	(80.2)	(114.0)	(73.3)	(102.1)	(37.1)	(43.8)	(93.9)	(86.2)
COMM	650.9	215.2	570.4	312.9	944.6	853.2	364.6	324.6
	(69.6)	(95.6)	(77.2)	(92.8)	(58.3)	(65.2)	(56.2)	(72.8)
Group output								
Predicted	8.9		10.1		15.6		8.2	
NOCOMM	18.3		16.0		23.4		15.2	
	(2.5)		(3.0)		(3.4)		(2.4)	
COMM	23.1		22.8		24.3		10.2	
	(3.2)		(2.6)		(2.0)		(1.9)	
Group payoff								
Predicted	1,463.6		1,540.2		1,783.9		1,140.5	
NOCOMM	1,240.7		918.8		1,639.8		487.2	
	(190.6)		(172.9)		(80.0)		(178.7)	
COMM	866.1		883.3		1,797.8		689.2	
	(142.3)		(166.8)		(123.2)		(128.1)	
Winning probability								
Predicted	0.4702		0.5298		0.6536		0.3464	
NOCOMM	0.6000		0.4000		0.7091		0.2909	
	(0.0688)		(0.0688)		(0.0343)		(0.0343)	
COMM	0.5100		0.4900		0.7818		0.2182	
	(0.0722)		(0.0722)		(0.0483)		(0.0483)	
Total output								
Predicted	19.0				23.8			
NOCOMM	34.3				38.6			
	(5.3)				(5.4)			
COMM	45.9				34.5			
	(4.7)				(3.3)			

Table 3: Theoretical predictions and summary statistics based on Half 2 data (rounds 11-20), with match-level clustered standard errors in parentheses.

difference in the average group output of (A,B) groups between UNBAL-NOCOMM (23.4) and UNBAL-COMM (24.3) ($p = 0.804$), but we do find that the average group output of (C,D) groups is significantly higher in UNBAL-NOCOMM (15.2) than in UNBAL-COMM (10.2) ($p = 0.094$).

Result 3 *(a) Under the balanced sorting, average group output of (A,D) groups is not statistically distinguishable between communication conditions; however, average group output of (B,C) groups is siginficantly higher in the presence of within-group communication, relative to no communication.*
(b) Under the unbalanced sorting, average group output of (A,B) groups is not statistically distinguishable between communication conditions; however, average group output of (C,D) groups is significantly lower in the presence of within-group communication, relative to no communication.

Result 3 suggests that the difference in average total output between BAL-COMM and UNBAL-COMM is due to the decrease in the average group output of the underdog (C,D) groups in UNBAL-COMM, as well as to the increase in the average group output of (B,C) groups in BAL-COMM, relative to no communication. Note also that the output of (C,D) groups in UNBAL-NOCOMM is only less than 2 output points away from the equilibrium prediction, suggesting communication enables the underdog groups to play closely to the Nash predictions.[19]

Recall that we do not find a statistically significant difference in total output between the two sortings in the absence of communication. This is achieved because (A,B) groups in UNBAL-NOCOMM produce more than (A,D) groups in BAL-NOCOMM, but (C,D) groups in UNBAL-NOCOMM produce slightly less than (B,C) groups in BAL-NOCOMM, and the two differences nearly compensate each other. While each of these comparative statics is consistent with theory, the reason why there is no large enough difference between total outputs is in that, contrary to theory, (A,D) groups produce more than (B,C) groups in BAL-NOCOMM. This is driven primarily by the D types in the (A,D) groups who overbid by more than any other cost follower compared to equilibrium predictions.

Our results regarding the effects of communication on behavior in weak-link group contests differ from the results of studies in a similar environment for symmetric players that observe higher average group output in the presence of within-group communication (Cason, Sheremeta and Zhang, 2012, 2015). Our findings suggest that the direction of the effect of within-group communication depends on the heterogeneity within and between

[19] Of course, the (C,D) groups are still bidding substantially more than their best response to the behavior of (A,B) groups.

groups. While under the balanced sorting the effect of communication is similar to the symmetric case, under the unbalanced sorting it is not. The difference arises primarily due to the more strategic behavior of the underdog (C,D) groups.

5.4 Overbidding

One of the most robust and widespread findings in the experimental literature on contests is the overexpending of effort relative to the risk-neutral Nash equilibrium predictions (for a review, see Sheremeta, 2013).[20] Comparing effort levels reported in Table 3, we find that observed effort exceeds the predictions for each player type in all treatments. The predicted average effort across types is 9.5 and 11.9 under the balanced and unbalanced sorting, respectively. Under the balanced sorting and in the absence (respectively, presence) of communication, average observed effort across types is 19.5 (respectively, 24.1), which exceeds the prediction by 105.3% (respectively, 153.7%). Similarly, under the unbalanced sorting and in the absence (respectively, presence) of communication, average observed effort across types is 21.1 (respectively, 17.6), which exceeds the prediction by 77.3% (respectively, 47.9%). Thus, we find that overbidding, while present in both sortings, is more severe under the balanced sorting relative to the unbalanced sorting. We verified this result by running a regression, with match-level clustered standard errors, of the variable for overbidding, defined as a difference between observed effort and the equilibrium prediction, on the treatment dummy BAL equal to unity for the balanced sorting. The coefficient estimate on BAL is positive (4.7) and significant (p=0.055).

Interestingly, overbidding appears to be less pronounced in BAL-NOCOMM than in BAL-COMM, suggesting communication may have an enhancing effect on the extent of players' overbidding under the balanced sorting. In contrast, it appears that communication has a mitigating effect on overbidding under the unbalanced sorting, as the extent of players' overbidding is lower in UNBAL-COMM than in UNBAL-NOCOMM. In either case, however, we do not find a statistically significant difference.

Analyzing overbidding by player type, we find substantial overbidding by all types in all treatments, with the exception of Type C and D players in UNBAL-COMM. We verified this result by running a regression, without an intercept and with match-level clustered standard errors, of the variable for overbidding on dummies for each player type, treatment dummies, and the necessary interactions, and then conducting Wald

[20]Traditionally, overbidding is defined in the literature as the difference between observed effort and equilibrium predictions. Alternatively, it can be defined as the difference between observed effort and the player's best response to the opponents' behavior. Given that, on average, players overbid relative to equilibrium, overbidding would be even higher under this alternative definition.

tests for the equivalence of overbidding to zero by type and treatment. The p-values satisfy $p < 0.05$, almost all satisfying $p < 0.01$, except for Type C and D in UNBAL-COMM ($p = 0.177$ and $p = 0.239$, respectively). This is somewhat surprising given that prior experiments report overbidding by all types even in heterogeneous group contests (Sheremeta, 2011; Brookins, Lightle and Ryvkin, 2015*a*).

It also appears from Table 3 that, on average, cost leaders overbid more than cost followers (the only exception is Type C and D players in UNBAL-COMM who overbid by the same amount); however, we only find significant differences in UNBAL-NOCOMM (p=0.086) and in BAL-COMM (p=0.000).

Result 4 *In all treatments and for all types of players, there is excessive overbidding of effort as compared to the risk-neutral Nash equilibrium predictions, with the exception of Type C and Type D players in the unbalanced sorting in the presence of within-group communication.*

Interpreting overbidding as a deviation from strategic behavior, we conclude that our findings provide an interesting mixture in terms of agreement with the previous literature. The behavior of the favorite (A,B) groups under the unbalanced sorting and of both group types under the balanced sorting is more consistent with literature on symmetric weak-link group contests that finds that in the presence of communication players care more about winning and are less strategic. On the other hand, the behavior of the underdog (C,D) groups is in agreement with the literature that finds that groups behave more strategically in the presence of communication.

We conjecture that due to the prevailing common wisdom of competitive balance the members of (A,D) and (B,C) groups under the balanced sorting view themselves as evenly matched even though the incentive structure implies otherwise. As a result, for these groups communication ramps up competition similar to the symmetric case. Type D players are the ones who should resist these groups following down this path, but they either do not even try or give in to the pressure from A types. In contrast, the members of (A,B), and especially (C,D) groups under the unbalanced sorting clearly understand their favorite and underdog roles, and this understanding is enhanced by communication. In Section 5.8 below, we analyze the content of chat messages to explore this further.

5.5 Average individual effort by treatment and type

For all pairwise treatment comparisons, we do not find any statistically significant difference in average effort for Types A and B. In the presence of communication, the average effort of Types C and D is significantly higher in BAL-COMM (22.9 and 23.8, respectively)

	Balanced		Unbalanced	
	(A,D)	(B,C)	(A,B)	(C,D)
NOCOMM	2.51	0.41	0.79	1.71
	(0.65)	(0.97)	(0.51)	(1.00)
COMM	2.48	0.62	0.24	0.03
	(2.57)	0.45	(0.16)	(0.57)

Table 4: Average difference in effort between cost leaders and cost followers by group type for all treatments, with match-level adjusted standard errors in parentheses.

than in UNBAL-COMM (10.8 and 10.8; $p = 0.000$ and $p = 0.001$, respectively); however, these results do not hold in the absence of communication. Thus, we find evidence in support of part (b) of Hypothesis 3 only in the presence of communication.

Result 5 *(a) Average effort of Types A and B is indistinguishable between the balanced and unbalanced sorting, regardless of whether or not group members can communicate.*
(b) In the presence of within-group communication, average effort of Types C and D is significantly higher under the balanced sorting than under the unbalanced sorting.

Note that both Type C and Type D players are the weak links in their groups under the balanced sorting, and the fact that communication affects them differently depending on the sorting they are in is consistent with the conjecture formulated in the previous section.

In Section 5.1, we discussed that although the frequency of perfect coordination on effort between groups members, i.e., zero wasted effort, was high for all treatments, there existed miscoordination. Which player types contributed the most towards waste? Table 4 shows the average difference in effort, $e_{CL} - e_{CF}$, between cost leaders (CL) and cost followers (CF), for each treatment and group type. As expected, in all treatments cost leaders contributed more towards wasted effort than cost followers as the difference between efforts is positive for each group type. However, it is significantly different from zero only for (C,D) groups in UNBAL-NOCOMM ($p = 0.086$) and for (A,D) groups in BAL-NOCOMM ($p = 0.000$). We also did not find a statistically significant difference between the NOCOMM and COMM condition for each group type.

Result 6 *In all treatments, cost leaders contribute at least as much as cost followers towards wasted group effort, and contribute significantly more than cost followers in (A,D) groups in BAL-NOCOMM and in (C,D) groups in UNBAL-NOCOMM.*

One possible explanation for this behavior in treatments without communication is that cost leaders may be trying to signal to cost followers that they want to exert a higher effort in the next round. Cost followers may also be reducing effort to signal to cost leaders that they prefer a lower effort in the next round. We explore this issue further in Section 5.7.2, where we use a dynamic regression model explicitly controlling for lagged effort choices and outcomes.

In treatments with communication, however, it is puzzling that average wasted effort exceeds zero 10.9% of the time in UNBAL-COMM and 23.5% of the time in BAL-COMM. One possible explanation is the existence of a conflict between group members. If such conflicts arise at all, they are most likely to emerge in the strongly heterogeneous (A,D) groups. This is indirectly confirmed by the observation that the average $e_{CL} - e_{CF}$ in (A,D) groups is almost the same with and without communication, but there is much more noise in this measure in the presence of communication (cf. Table 4). The increase in noise suggests that while some groups become more successful at coordinating effort, other groups become less successful, as compared to the NOCOMM case. In Section 5.8, we explore this by closely analyzing the content of the chat messages.

5.6 Average winning probabilities

Average winning probability for all group compositions and treatments are reported in Table 3. Under the balanced sorting, (A,D) groups win the contest 60% and 51% of the time in BAL-NOCOMM and BAL-COMM, respectively, and although both of these estimates are above the equilibrium prediction of 47.02%, only in BAL-NOCOMM is the difference statistically significant ($p = 0.058$). This is consistent with our findings that the level of overbidding by Type A and D players is higher relative to Type B and C players in the absence of communication, but with communication the two group types overbid to the same extent, and hence the winning probabilities in BAL-COMM are very close to the theoretical predictions.

Under the unbalanced sorting, the winning probability of (A,B) groups is 70.91% and 78.18% in UNBAL-NOCOMM and UNBAL-COMM, respectively, both of which exceed the theoretically predicted 65.36%, but in this case the significant difference is only observed in UNBAL-COMM ($p = 0.009$). In this treatment, Type A and B players have overbid more than Type C and D players, who actually have the lowest degree of overbidding, in relative terms, compared to any other group composition in any treatment. In UNBAL-NOCOMM, the two group types overbid by the same amount, in relative terms, and hence the winning probabilities are close to the theoretically predicted ones.

Result 7 *Under the balanced sorting, communication shifts winning probabilities closer to the balanced theoretical predictions; however, under the unbalanced sorting communication* *shifts the winning probabilities further away from the unbalanced theoretical predictions.*

Result 7 is consistent with the conjecture formulated at the end of Section 5.4. If indeed under the balanced sorting players feel entitled to winning the contest approximately 50% of the time (which happens to be close to the equilibrium prediction), communication helps them achieve that goal, albeit via substantial overbidding by all types. Under the unbalanced sorting, given the disparity in effort costs, it may seem counterintuitive that the favorite group should only win about 2/3 of the time (as predicted), and hence the groups move to an even more extreme disparity in winning probabilities in the presence of communication.

5.7 Dynamics

In this section, we explore the dynamics in subjects' decisions. Subjects' feedback at the end of each round included their group's output, the output of the opponent group, and the outcome (i.e., whether their group won or lost).

5.7.1 Dynamics at the group level

We start with the analysis of the dynamics of output at the group level. It is not clear *a priori* how the presence of communication or the type of sorting will affect the reactions of groups to the events of the previous round. On the one hand, communication allows subjects to better coordinate their reaction to changes in the environment (e.g., to a change in the behavior of the other group or in the frequency of winning), and hence it may reduce the persistence in own output but strengthen the response to the behavior of the opponent group. On the other hand, communication makes it easier for groups to establish a "type" early on and stick to it regardless of idiosyncratic changes in the environment, resulting in higher persistence of own output and weaker reactions to the opponents' behavior. The same two countervailing factors apply to the effect of sorting. Under the unbalanced sorting, it is easier for groups to coordinate and hence to react to changes in the environment, but it is also easier for them to establish a type.

Table 5 reports OLS regressions, with standard errors clustered at the match level, of group output on various controls including lagged choices and outcomes as well as dummies for each type of group composition. The models also include *Round* to capture time trend, but the coefficient estimate on it is not statistically significant in any of our specifications.

Model (1) is a regression of group output on dummies representing each type of group composition (with the (C,D) groups in the unbalanced sorting as the baseline) and the

Output	(1)	(2)	(3)
(A,D) Group	3.0545	1.1592	1.0556
	(3.374)	(1.407)	(1.464)
(A,D) Group×COMM	4.7800	5.5745	5.5665
	(3.903)	(3.768)	(3.742)
(B,C) Group	0.7272	-0.0231	-0.4236
	(3.713)	(1.129)	(1.168)
(B,C) Group×COMM	6.7972*	6.2858*	6.4538*
	(3.397)	(3.352)	(3.316)
(A,B) Group	8.1091***	2.1170*	1.2799
	(2.350)	(1.081)	(1.036)
(A,B) Group×COMM	0.9545	10.9670***	10.8329***
	(3.829)	(3.312)	(3.185)
(C,D) Group×COMM	-5.0545*	1.7489	1.5862
	(2.953)	(2.676)	(2.019)
Output_lag		0.8550***	0.8882***
		(0.058)	(0.056)
Output_lag×BAL		-0.2362**	-0.2513***
		(0.088)	(0.091)
Output_lag×COMM		-0.4975***	-0.4960***
		(0.128)	(0.127)
Output_lag×BAL×COMM		0.4745***	0.4721***
		(0.166)	(0.166)
Other_Output_lag		0.0618	0.0598
		(0.0424)	(0.048)
Other_Output_lag×BAL		0.2374***	0.2480***
		(0.069)	(0.072)
Other_Output_lag×COMM		0.1006	0.1031
		(0.081)	(0.083)
Other_Output_lag×BAL×COMM		-0.3143**	-0.3224**
		(0.123)	(0.123)
Win_lag			-0.0926
			(0.493)
Lost_Money_lag			-2.3830*
			(1.269)
Round	-0.1352	0.0239	0.0258
	(0.125)	(0.0824)	(0.083)
Constant	17.3405	0.0880	0.2452
	(2.864)	(1.350)	(1.449)
Observations	860	860	860
Groups	86	86	86
R-squared	0.170	0.590	0.593

Table 5: OLS regression results, match-level clustered standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

interactions of group types with the *COMM* dummy, which provides a replication of Result 3 presented in Section 5.3.

Model (2) in Table 5 additionally controls for lagged own and other group's output (*Output_lag* and *Other_Output_lag*) and their interactions with the *BAL* and *COMM* treatment dummies. Looking at the coefficients on *Output_lag* and its interactions, we observe that group output is fairly persistent from one round to the next. However, the evidence on the effects of sorting and communication on output persistence is mixed. The persistence is the highest in UNBAL-NOCOMM and the lowest in UNBAL-COMM. In the absence of communication, the persistence is lower under the balanced sorting ($p = 0.011$), which is consistent with groups being better at establishing a fixed type when sorting is unbalanced. However, in the presence of communication, the persistence is lower under the unbalanced sorting ($p = 0.096$), which would be consistent with more coordinated reactions to changes in the environment. Looking at the effect of communication, there is less persistence in output when communication is present under the unbalanced sorting ($p = 0.000$), which is also consistent with the more effective coordination explanation, but under the balanced sorting there is no effect of communication on persistence ($p = 0.829$).

The effect of the opponent group output in the previous round is only statistically significant in treatments BAL-NOCOMM ($p = 0.000$) and UNBAL-COMM ($p = 0.022$). In the absence of communication, there is less of a reaction to opponent group output in the unbalanced sorting ($p = 0.001$), which is again consistent with subjects being better at establishing fixed types under the unbalanced sorting. In the presence of communication, however, the difference between the two sortings is no longer significant ($p = 0.452$). Holding fixed the communication condition, there is less of a reaction to opponent group output in the balanced sorting when communication is present ($p = 0.027$), but under the unbalanced sorting there is no effect of communication on how groups respond to lagged opponent output ($p = 0.219$).

Model (3) additionally includes the dummy *Win_lag* equal to one if the group won the contest in the previous round, and the dummy *Lost_Money_lag* equal to one if both group members experienced negative earnings in the previous round. An increase (respectively, decrease) in group output in response to losing (respectively, winning) the contest in the previous round would be consistent with the basic directional and reinforcement learning dynamics in contest experiments (e.g., Dutcher et al., 2015). However, we observe that group output is not affected by the contest outcome in the previous round as the coefficient estimate on *Win_lag* is not statistically significant.[21] We do observe a negative

[21]We also tried including the interactions of *Wibn_lag* with the treatment dummies but found no differences in its effect across treatments.

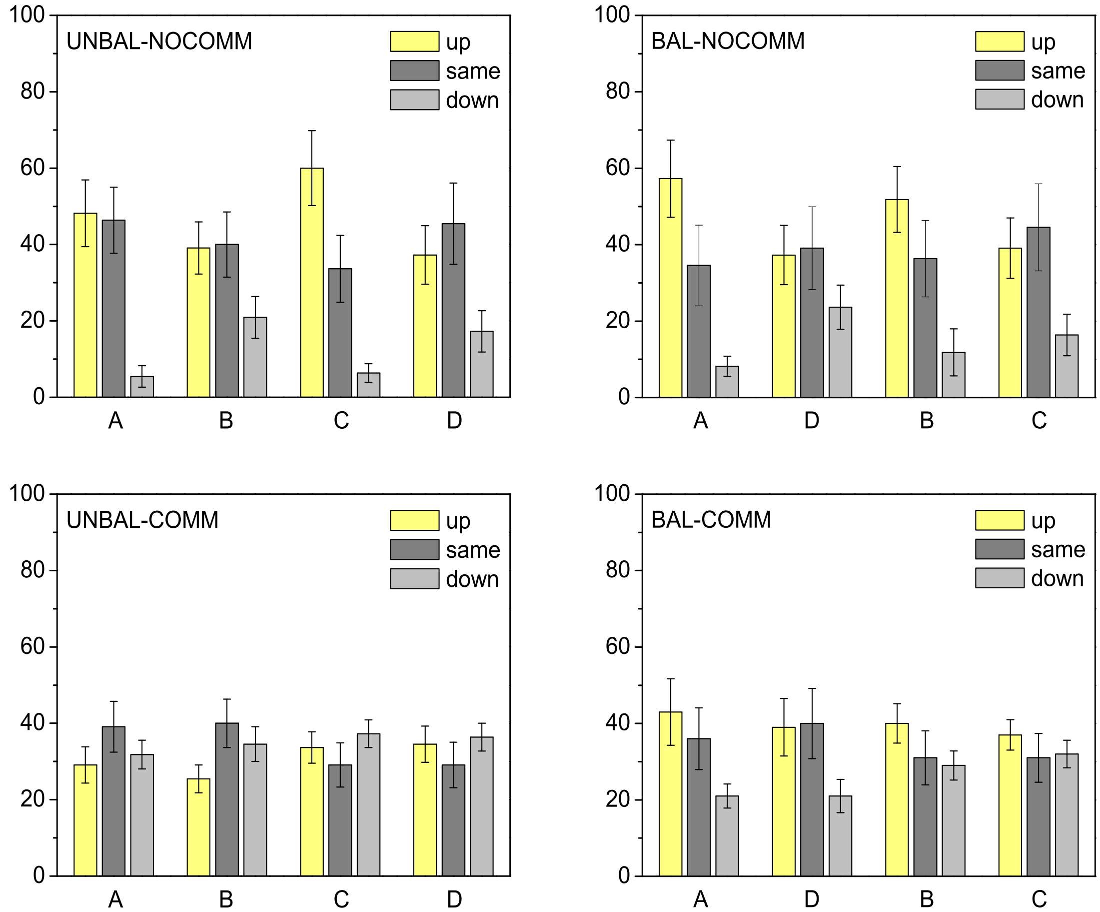

Figure 4: The percentage of instances when individual effort in period t went up, stayed the same or went down, as compared to the output of the individual's group in period $t-1$, averaged by treatment and type. The error bars show one match-level clustered standard error above and below the averages.

and significant effect on group output when both members from the group experienced negative earnings in the previous round.[22]

5.7.2 Dynamics at the individual level

We now turn to the analysis of dynamics at the individual level. First, we explore how within-group coordination evolves over time, i.e., how subjects adjust their individual effort in response to instances of miscoordination in the previous round. *Ceteris paribus*,

[22] We also ran additional regressions where *Lost_Money_lag* was defined alternatively as the indicator for at least one of the group members losing money and at least one of the group members earning the amount below their endowment of 400 points in the previous round, and found no significant effects in either case. This suggests that subjects view zero, and not the periodic endowment, as their reference point, which is consistent with the results of Brookins, Lightle and Ryvkin (2015*a*).

we expect subjects who were (were not) the weak link in their group in round $t-1$ to be more likely to increase (respectively, decrease) their effort in round t. This is precisely what we find in the absence of communication.[23] With communication, it is puzzling why miscoordination would arise in the first place. The most obvious explanation is that subjects cannot agree on any level of output and continue to choose different efforts. Thus, with communication, the presence of miscoordination is an indicator of entrenched disagreement, and hence subjects are less likely to respond to waste in the previous round by adjusting their effort as described above. Indeed, in UNBAL-COMM the adjustments are still present for both types in (A,B) groups, but only for Type C in (C,D) groups, while in BAL-COMM the adjustments are observed only for Type A in (A,D) groups and for Type C in (B,C) groups.

A different way of looking at individual effort dynamics is to explore how individual effort is adjusted relative to the *group output* in the previous round. In the weak-link aggregation environment, a subject cannot increase her group's output or chances of winning by unilaterally increasing effort. Having observed their group output in period $t-1$, a subject can exert effort that is higher than, the same as, or lower than that output. Exerting a higher effort can be viewed as a (possibly costly) signal to their partner that they would like to increase output. Similarly, a lower effort may signal a willingness to decrease output or, if the subject generated waste in the previous round, a desire to reduce the waste and agree to a lower level of output. In the presence of communication, when subjects can explicitly discuss and agree on output, costly signaling and strategic teaching through effort may no longer be necessary, although whether and to what extent they are reduced is an empirical question because communication is still cheap talk.

Figure 4 shows the percentages of instances when individual effort in period t went up, stayed the same or went down, as compared to the output of the individual's group in period $t-1$, averaged for each treatment and type. As seen from the figure, in the absence of communication (the top panels) cost leaders tend to move up more often, and to move down less often, than cost followers for both sortings and all group types. These differences are statistically significant for both (A,B) and (C,D) groups in UNBAL-NOCOMM (Type A is 15% less likely to go down than Type B, $p = 0.021$; Type C is 23% more likely to go up than Type D, $p = 0.088$) and for (A,D) groups in BAL-NOCOMM (Type A is 20% more likely to go up and 15% less likely to go down than Type D, $p = 0.006$ and $p = 0.040$, respectively). These results are consistent with the presence of

[23] The effects are statistically significant for both sortings and all subject types. We also ran the same regressions additionally controlling for gender and the measures of risk aversion, ambiguity aversion and loss aversion elicited in Part 1 of the experiment and found no systematic effects. Estimation details are available from the authors upon request.

strategic teaching. Alternatively, cost leaders may be more likely to go up and less likely to go down, as compared to cost followers, simply because effort is marginally cheaper for them. In contrast, in the presence of communication (the lower panels), there appear to be no differences between cost leaders and followers in any of the frequencies (the only statistically significant difference is observed for (B,C) groups in BAL-COMM where Type B is less likely to move down, $p = 0.082$, but the size of the effect is only 3 percentage points).

5.8 Chat analysis

One of our main findings is that, contrary to the equilibrium predictions, the balanced sorting produces significantly higher total output relative to the unbalanced sorting when group members are allowed to communicate with one another prior to making individual effort decisions. Moreover, we find that the behavior of (B,C) groups in the balanced sorting and (C,D) groups in the unbalanced sorting differs significantly between communication conditions, which is the driving force behind our main result. We now take a detailed look at the content of chat messages between group members. Chat content analysis may provide additional insights into subjects' reasoning.[24]

We use standard message categorization methods which have been widely used to analyze the content of chat messages (see, e.g., Cooper and Kühn, 2014). Two coders independently coded all conversations between group members; thus, the unit of observation for chat messages is at the group level. Table 6 briefly describes the categories that were coded and reports the average frequency coded broken down by group type and treatment. Categories were chosen prior to being coded, and their inclusion was based on anything we expected a group may talk about. If a conversation between group members contained relevant messages that fit the description of a particular category, a code of 1 was given; otherwise, a code of 0 was given. Due to the subjective nature of codings, and the difficulty of making such judgments, we rely on Cohen's κ-coefficient of intercoder reliability (Cohen, 1960) to distinguish cases of agreement between coders from pure chance: the higher the value of κ, the more reliable the coding. The κ-coefficient for each category is reported in Table 6.

As seen from Table 6, many of the message categories were coded with $\kappa < 0.3$, which is typically considered "unreliable" (Landis and Koch, 1977). For this reason, we will focus on the categories whose κ is in the range of "fair" ($0.3 \leq \kappa \leq 0.4$) and higher.

[24] Although in the previous sections we restricted the analysis to Half 2 rounds, we use the coded chat messages from all 20 rounds for this section. This is a reasonable approach because we expect many important messages between group members to occur early on.

	Percent Observed Balanced		Percent Observed Unbalanced		
Message Description	(A,D)	(B,C)	(A,B)	(C,D)	κ
Make decision by discussing previous round	52.29	63.07	57.11	55.05	0.31*
Make decision by discussing multiple, past rounds	14.00	15.08	14.91	10.32	0.26
Discuss changing strategy	4.00	4.77	8.72	10.78	0.26
Strategic thinking about other group	18.86	33.42	36.47	25.35	0.61***
Agreed on effort level	72.57	75.38	82.34	79.36	0.78***
Disagree on effort level	0.86	1.76	1.15	1.38	0.28
Discussed coordinating on same effort level	74.29	80.86	86.24	84.40	0.74***
Discussed coordinating on same/similar effort level as other group	1.71	5.28	2.75	3.44	0.23
First proposal by Type A	50.57	–	37.61	–	0.80***
First proposal by Type B	–	54.02	52.75	–	0.87***
First proposal by Type C	–	37.69	–	43.81	0.86***
First proposal by Type D	40.57	–	–	45.41	0.84***
No clear first proposal	7.71	6.78	9.40	9.86	0.79***
First proposal accepted	61.43	57.79	71.56	65.37	0.74***
Appeal to give up or concede	1.14	2.01	0.23	4.85	0.22
Appeal to give up or concede	1.14	2.01	0.23	4.85	0.22
Conversational leader Type A	10.29	–	3.90	–	0.20
Conversational leader Type B	–	5.78	5.05	–	0.20
Conversational leader Type C	–	4.02	–	9.17	0.30*
Conversational leader Type D	4.57	–	–	5.05	0.25
No clear conversational leader	85.14	90.20	91.06	85.78	0.18
Group members compared individual payoffs	1.14	4.52	6.88	8.76	0.12
Appeal to win or beat other group	6.00	5.78	7.34	3.67	0.17
Any mention of fairness	1.43	1.26	0.92	2.98	0.14
Did either group member feel forced to choose a given level of effort	0.86	0.75	0.00	2.29	0.12

Table 6: Chat message descriptions with average frequency coded by group type and sorting including Cohen's κ-coefficient of intercoder reliability. * $0.3 \leq \kappa \leq 0.4$, ** $0.4 < \kappa \leq 0.6$, *** $\kappa > 0.6$.

The category "Make decision by discussing previous round" captures explicit discussions of responses to the events of the previous round, i.e., to the feedback subjects receive. Such discussions are present in between 52 and 63% of conversations, depending on group type, and are indicative of directional and reinforcement learning type reasoning. In contrast, the category "Strategic thinking about other groups" captures explicit discussions of the behavior of the opponent group, such as attempts to forecast what the opponent group will do or reasoning from their point of view. Such discussions are only present in between 18 and 36% of conversations.

We find that communication significantly reduces wasted effort by each group type, with the exception of (A,D) groups in balanced, suggesting groups improve coordination.[25] We find additional support for this finding in the coded chat messages as group members frequently (about 81% of conversations) discussed coordinating on the same level of effort, which is captured by the chat message category "Discussed coordinating on same effort level."

The category "Agreed on effort level" was coded in about 77% of conversations between group members, which suggests group members frequently came to an agreement on a particular level of effort regardless of the within-group heterogeneity present. Thus, for cost followers Type C and D in the balanced sorting, we do not find evidence of conflict or resistance in expending high effort that the cost followers would prefer to choose. This further supports the conjecture that under the balanced sorting, the (A,D) and (B,C) groups believe they are evenly matched, and thus we do not see cost followers resisting competition by disagreeing and choosing lower effort.

We also explore which player types make the first proposal regarding what effort to choose, and whether or not this proposal was accepted by their partners. This is captured through the categories "First proposal by Type X" (X=A,B,C,D) and "First proposal accepted." Under the balanced sorting, we see that cost leaders make the first proposal more frequently in (B,C) groups (about 54% by Type B as compared to 38% by Type C, $p = 0.011$), but not in (A,D) groups (about 50% by Type A compared to 40% by Type D, $p = 0.344$). In contrast, cost followers make the first proposal more frequently under the unbalanced sorting in (A,B) groups (about 38% by Type A compared to 53% by Type B, $p = 0.032$), but not in (C,D) groups (44% by Type C compared to 45% by Type D). The first proposal was accepted about 61% of the time under the balanced sorting and

[25]To verify differences in average wasted effort by group types between communication conditions, we ran a regression, without an intercept and with match-level clustered standard errors, of wasted effort on dummies for each group type and the necessary interactions with the treatment dummies, then conducted pairwise tests for the equivalence between coefficients using Wald tests. All differences were significant at the 1% or 5% confidence level.

about 72% of the time under the unbalanced sorting. One would expect to see more first proposals accepted under the unbalanced sorting due to the low heterogeneity between group members in that sorting compared to the balanced sorting; however, the difference is not statistically significant ($p = 0.247$).

One interesting result we found for the unbalanced sorting was that relative to no communication, within-group communication makes (C,D) groups choose effort more strategically, resulting in group output closer to the equilibrium predictions. We conjectured in Section 5.4 that when communication is present the salience of being the underdog group is enhanced and group members discuss conceding or giving up, which would be captured by group members agreeing on a low effort or no effort at all. We cannot confirm that the underdog (C,D) groups behave more strategically simply due to a higher frequency of conversations involving the excplicit mention of "giving up" or "conceding", as the category for this type of conversation "Appeal to give up or concede" has a κ shy of fair intercoder agreement. However, we do find that the frequency in which Type C players are coded as the "Conversational leader," a category coded 1 if they appeared to lead the conversation, 0 otherwise, is more than twice as high when paired with Type D players under the unbalanced sorting as compared to being matched with Type B players under the balanced sorting (about 4% compared to 9%, $p = 0.023$).[26] While this is anecdotal, we reproduce a conversation which nicely highlights conversational leadership by a Type C participant:

Type D: *okay keep 0?*

Type D: *or bid higher?*

Type C: *they are bidding pretty high... i think 0 is our best bet*

Type D: *they went down this time*

Type D: *okay zero then*

Type C: *Because if not we will lose money...we atleast want more than 400?*

Type D: *good idea!*

Type C: *kk*

[26]More specifically, for the category "Conversational leader Type C" coders were instructed to code a 1 if the messages from Type C involved rationalizing what decisions to make, convincing or persuading the other group member, or persistence in suggesting what decisions to make, and code a 0 otherwise.

As can be seen from the conversation, it is clear that Type C is able to convince the Type D player to follow their lead and choose zero effort, which is strategically better than competing by exerting effort above equilibrium. Indeed, the opponent group in this situation had overbid substantially (by about 41%) in the previous round and it is therefore a best-response of this group, taking the previous round's bid as given, to bid below equilibrium, albeit not as low as zero.

We conclude that subjects spend most of their conversations on relatively unsophisticated discussions related to coordinating on the same level of effort within group and reinforcement learning type reactions, as opposed to more strategic discussions of best-responding to the other group. Even the underdog (C,D) groups reduce their effort not so much as a result of strategic response to the behavior of (A,B) groups as simply to avoid losses.

6 Discussion and conclusions

When a firm utilizes a production technology with perfect complementarity among workers' effort, a manager is faced with a difficult but important task: The manager must assign individuals into groups who will not only work well together, but will be productive. The groups must coordinate their efforts and motivate each other to maintain a high level of performance. In this study, we consider a firm which has chosen to incentivize high effort by utilizing a group contest and compare two extremes in how a manager may choose to sort workers into groups. Because we believe workers typically have the opportunity to communicate when working toward a common goal, we also examine the role that communication plays in assisting coordination and motivation in these group contests.

According to the unique coalition-proof equilibrium of the weak-link group contest game, sorting workers in an unbalanced way – creating a single group of the most talented workers – should result in the highest total output. The main intuition behind this result is that in equilibrium each group's output is determined by the least talented worker, and by pooling the low performing workers together a manager can create an environment where fewer groups are "low performing." Even in this unbalanced environment, the inherent probabilistic structure of the contest promotes beneficial competition.

While the equilibrium comparative statics are not supported in the absence of communication, our results in the presence of communication are in stark contrast to the predictions. The unbalanced sorting significantly underperforms relative to balanced when workers can communicate. A closer look at the data reveals that two idiosyncrasies of communication are primarily responsible for this result. The first one is not very surprising: under the balanced sorting, the strongly heterogeneous (A,D) groups do not reduce their output when communication is possible, while the less heterogeneous (B,C) groups experience sizable gains in output. The second finding is more unexpected: under the unbalanced sorting, the underdog (C,D) groups produce 33% less output when they can communicate, even though communication helps the coordination process by reducing wasted effort by 71%. This finding is in contrast to previous studies that find that communication always increases output in a similar environment.

In order to better understand these idiosyncrasies, we categorized and analyzed the messages sent by subjects during the experiment. Why do (B,C) groups have higher output in the presence of communication? For one, we noticed a high frequency of agreements between group members in regards to effort decisions across group types. The fact that Type B profits more than Type C from a given level of effort was not a concern. Even when the difference in talent is extreme, i.e., in (A,D) groups, the less able worker Type D is still unable to come to an agreement on lower effort with Type A. Rather, we find that roughly 80% of groups discussed coordinating on an effort, and most of these successfully coordinated. Our chat analysis provides an insight into why the results diverge from theory: despite the prediction that a group will coordinate on the effort preferred by the weaker worker, we find that under the balanced sorting the stronger worker (Type B) is more likely to be the first to propose an effort in (B,C) groups, whereas both types A and D are equally likely to propose an effort in (A,D) groups. On the other hand, under the unbalanced sorting, it is the weaker worker (Type B) who is more likely to propose an effort in (A,B) groups and Type D is as likely to make a first proposal as Type C in (C,D) groups. Note that it is the overbidding by (B,C) groups and the relatively reserved bidding by (C,D) groups that drives the reversal of the theoretical predictions in the presence of communication. The chat analysis suggests that the result is due to the unexpected within-group leadership potential of high-ability workers under the balanced sorting.

The fact that output of (C,D) groups declines when communication is possible is surprising and more difficult to explain. One would expect that these groups are more likely to discuss conceding the contest than any other group type, but we do not observe many such conversations. It might be the case that when a group is a clear underdog, the avenue of communication only serves to further discourage workers from attempting to win a contest. Our chat analysis reveals that Type C workers from the underdog (C,D) are the only player type to show "converstational leadership," in the sense that they were able to lead the group to a decision on effort based on strategic thinking. Upon closer inspection, we find evidence that it is these instances of leadership that lead the underdog group to more strategic behavior.

Our experiment shows that communication has an important impact on behavior in group contest games, and its impact may differ depending on the configuration of players' heterogeneity. When managers implement group contests to improve productivity, they must consider the extent to which employees can motivate or discourage other workers in their group. When choosing how to sort heterogeneous workers into groups, we show that the theoretical predictions of optimality based on the behavior of self-interested actors are actually reversed in the presence of free-form communication between group members. One interpretation is that the benefits of having a mentor in a workplace might extend to more than the fact that mentors can teach new skills to newer employees. Specifically, having a system where potential mentors are assigned to work one-per-group is preferable to one where these experienced workers are combined to form a single group because it raises the effort level of less talented workers even when, as in our experiment, it is not possible for the workers to learn skills they did not already possess.[27]

More generally, our results demonstrate the robustness of the optimality of competitive balance in group contests. The positive effect of competitive balance on output will be further enhanced in settings where the leadership roles of high-ability workers are well-defined and/or there exists a possibility for help in the form of transferred effort or skills.

References

Abbink, Klaus, Jordi Brandts, Benedikt Herrmann, and Henrik Orzen. 2010. "Intergroup conflict and intra-group punishment in an experimental contest game." *American Economic Review*, 420–447.

Ahn, TK, R Mark Isaac, and Timothy C Salmon. 2011. "Rent seeking in groups." *International Journal of Industrial Organization*, 29(1): 116–125.

Azrieli, Yaron, Christopher P Chambers, and Paul J Healy. 2015. "Incentives in experiments: A theoretical analysis." `http://healy.econ.ohio-state.edu/papers/Azrieli_Chambers_Healy-CombinedDecisions.pdf`.

[27] It can be argued that one of the advantages of a balanced sorting is the ability of more productive group leaders to help others in their group. However, such help may be impossible in production processes with highly specialized workers.

Baik, Kyung Hwan, In-Gyu Kim, and Sunghyun Na. 2001. "Bidding for a group-specific public-good prize." *Journal of Public Economics*, 82(3): 415–429.

Baye, Michael R., Dan Kovenock, and Casper G. De Vries. 1996. "The all-pay auction with complete information." *Economic Theory*, 8(2): 291–305.

Blume, Andreas, and Andreas Ortmann. 2007. "The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria." *Journal of Economic theory*, 132(1): 274–290.

Brandts, Jordi, and David J Cooper. 2007. "It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure." *Journal of the European Economic Association*, 5(6): 1223–1268.

Brookins, Philip, John Lightle, and Dmitry Ryvkin. 2015*a*. "An experimental study of sorting in group contests." *Labour Economics*, 35: 16–25.

Brookins, Philip, John Lightle, and Dmitry Ryvkin. 2015*b*. "Optimal sorting in group contests with complementarities." *Journal of Economic Behavior and Organization*, 112(1): 311–323.

Cason, Timothy N., Roman M. Sheremeta, and Jingjing Zhang. 2012. "Communication and efficiency in competitive coordination games." *Games and Economic Behavior*, 76(1): 26–43.

Cason, Timothy N., Roman M. Sheremeta, and Jingjing Zhang. 2015. "Asymmetric and endogenous communication in competition between groups." *Available at SSRN.*

Chen, Hua, and Noah Lim. 2013. "Should managers use team-based contests?" *Management Science*, 59(12): 2823–2836.

Cohen, Jacob. 1960. "A coefficient of agreement for nominal scales." *Educational and psychological measurement*, 20(1): 37–46.

Connelly, Brian L., Laszlo Tihanyi, T. Russell Crook, and K. Ashley Gangloff. 2014. "Tournament Theory: Thirty Years of Contests and Competitions." *Journal of Management*, 40(1): 16–47.

Cooper, David J., and John H. Kagel. 2005. "Are two heads better than one? Team versus individual play in signaling games." *American Economic Review*, 477–509.

Cooper, David J, and Kai-Uwe Kühn. 2014. "Communication, Renegotiation, and the Scope for Collusion." *American Economic Journal: Microeconomics*, 6(2): 247–278.

Cooper, Russell, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross. 1992. "Communication in coordination games." *Quarterly Journal of Economics*, 739–771.

Dechenaux, Emmanuel, Dan Kovenock, and Roman M. Sheremeta. 2015. "A survey of experimental research on contests, all-pay auctions and tournaments." *Experimental Economics*, 18(4): 609–669.

Devetag, Giovanna, and Andreas Ortmann. 2007. "When and why? A critical survey on coordination failure in the laboratory." *Experimental economics*, 10(3): 331–344.

Dutcher, E. Glenn, Loukas Balafoutas, Florian Lindner, Dmitry Ryvkin, and Matthias Sutter. 2015. "Strive to be first or avoid being last: An experiment on relative performance incentives." *Games and Economic Behavior*, 94: 39–56.

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental economics*, 10(2): 171–178.

Greiner, Ben. 2004. "The online recruitment system orsee 2.0-a guide for the organization of experiments in economics." *University of Cologne, Working paper series in economics*, 10(23): 63–104.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. 1994. "Preferences, property rights, and anonymity in bargaining games." *Games and Economic Behavior*, 7(3): 346–380.

Holt, Charles A., and Susan K. Laury. 2002. "Risk aversion and incentive effects." *American Economic Review*, 92(5): 1644–1655.

Kocher, Martin G, and Matthias Sutter. 2005. "The Decision Maker Matters: Individual Versus Group Behaviour in Experimental Beauty-Contest Games*." *Economic Journal*, 115(500): 200–223.

Landis, J Richard, and Gary G Koch. 1977. "An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers." *Biometrics*, 363–374.

Lawler, Edward E, Susan Albers Mohrman, and George Benson. 2001. *Organizing for high performance: Employee involvement, TQM, reengineering, and knowledge management in the Fortune 1000: The CEO report.* Jossey-Bass.

Lawler, Edward E, Susan Albers Mohrman, and Gerald E Ledford. 1995. *Creating high performance organizations: Practices and results of employee involvement and total quality management in Fortune 1000 companies.* Jossey-Bass San Francisco, CA.

Lazear, Edward P. 1999. "Personnel Economics: Past Lessons and Future Directions Presidential Address to the Society of Labor Economists, San Francisco, May 1, 1998." *Journal of Labor Economics*, 17(2): 199–236.

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

Lazear, Edward P., and Sherwin Rosen. 1981. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy*, 89(5): 841–864.

Lee, Dongryul. 2012. "Weakest-link contests with group-specific public good prizes." *European Journal of Political Economy*, 28(2): 238–248.

Nalbantian, Haig R, and Andrew Schotter. 1997. "Productivity under group incentives: An experimental study." *American Economic Review*, 314–341.

Nitzan, Shmuel, and Kaoru Ueda. 2011. "Prize sharing in collective contests." *European Economic Review*, 55(5): 678–687.

Nitzan, Shmuel, and Kaoru Ueda. 2014. "Intra-group heterogeneity in collective contests." *Social Choice and Welfare*, 43(1): 219–238.

Riechmann, Thomas, and Joachim Weimann. 2008. "Competition as a coordination device: Experimental evidence from a minimum effort coordination game." *European Journal of Political Economy*, 24(2): 437–454.

Ryvkin, Dmitry. 2011. "The optimal sorting of players in contests between groups." *Games and Economic Behavior*, 73(2): 564–572.

Ryvkin, Dmitry, and Anastasia Semykina. 2013. "The chicken or the egg: An experimental study of democracy survival, income, and inequality." *working paper.*

Sheremeta, Roman M. 2011. "Perfect-substitutes, best-shot, and weakest-link contests between groups." *Korean Economic Review*, 27: 5–32.

Sheremeta, Roman M. 2013. "Overbidding and heterogeneous behavior in contest experiments." *Journal of Economic Surveys*, 27(3): 491–514.

Sheremeta, Roman M. 2017. "Behavior in group contests: A review of experimental research." *Journal of Economic Surveys.*

Sheremeta, Roman M., and Jingjing Zhang. 2010. "Can groups solve the problem of over-bidding in contests?" *Social Choice and Welfare*, 35(2): 175–197.

Sutter, Matthias, and Christina Strassmair. 2009. "Communication, cooperation and collusion in team tournamentsan experimental study." *Games and Economic Behavior*, 66(1): 506–525.

Tullock, Gordon. 1980. "Efficient rent seeking." *Toward a Theory of the Rent-Seeking Society*, , ed. James M. Buchanan, Robert D. Tollison and Gordon Tullock, 97–112. College Station:Texas A&M University Press.

A Experimental instructions

The following instructions are preceded by an introduction, instructions for the risk, ambiguity, and loss aversion task and instructions for the number counting task (e.g., Ryvkin and Semykina, 2013) used to assign the types. Those instructions are straightforward and are available from the authors upon request. The instructions below are for the first part of the Balanced treatment.[28] The differences with the Unbalanced treatment are highlighted in footnotes 29 and 30 which are not part of the original instructions. Treatments in which subjects were allowed to communicate had an additional section titled "*Communication*."

Instructions for Part 3

All amounts in this part of the experiment are expressed in *points*. The exchange rate is 200 points = $1.00.

Types
You will be assigned one of the four types – A, B, C or D – depending on your ranking in the number addition task you performed in the previous part of the experiment. Specifically, the top 25% of performers will be assigned type A, the second 25% - type B, the third 25% - type C, and the bottom 25% - type D. Ties will be broken randomly. Types will remain fixed until the end of this part.

Decision rounds
This part of the experiment consists of a sequence of decision rounds, all in the same basic setting.

Groups
In the beginning of this part of the experiment, you will be randomly divided into groups of 2 participants referred to as *your group*. Groups will be arranged by types in the following manner:[29]

(A,D) groups: type A will always be in a group with type D
(B,C) groups: type B will always be in a group with type C

[28]Recall, subjects participated in 2 parts of each treatment, each consisting of 10 round each. After completion of the first 10 rounds, there was a break and instructions were provided for the next part of the experiment. Subjects were told that the next part of the experiment would be identical to the previous part.

[29]In the Unbalanced treatment, the following two lines are replaced with "(A,B) groups: type A will always be in a group with type B" and "(C,D) groups: type C will always be in a group with type D."

After each round, groups will be randomly re-shuffled, i.e., you will be randomly re-matched with another participant of the corresponding type. In any given round, any participant with the right type has equal chances to be matched with you.

Matching
In the beginning of this part of the experiment, your group will be randomly matched with a group of the other type; that is,[30]

> **an (A,D) group will always be matched with a (B,C) group**.

The two groups matched together will stay matched together throughout this part of the experiment. In other words, the same (A,D) group will always stay matched with the same (B,C) group, in every round.

Endowment and investment
In each round, you will be given an endowment of 400 points. You can invest any integer number **between 0 and 50** (0 and 50 inclusive) into a group project.

Group output and project success
Your group will generate a certain level of *group output*. Your *group output* will be calculated as the **minimum** investment of the two group members. In other words, the amount of output for your group will be a number equal to whichever individual investment is the smallest between the two individuals in a group. If the two investments are equal, then your group output will be equal to the investment of one of the individuals.

Group output = Minimum of {*(your investment) and (the other member's investment)*}.

Your group project can be either successful or unsuccessful. If your group project is **successful**, both you and the other member of your group will receive **1000 points of revenue** that round. If your group project is **unsuccessful**, you both will **not receive any revenue** that round.

Likelihood of success
Each round, only one of the two group projects in the match will be successful. The probability (likelihood) that your group's project is successful is determined as follows. First, the group output of both groups is computed. Then, your group's probability of

[30] In the Unbalanced treatment, the following line is replaced with "an (A,B) group will always be matched with a (C,D) group."

success is computed as the *share of your group's output in the total output from both groups.*

The numbers in the following example are hypothetical and in no way suggest what you should do in the actual experiment. **For example**:

- Suppose in one group (group 1), one of the members chose an investment of 10 and the other chose an investment of 25. Then this group as a **group ouput of 10** because 10 is the minimum (smallest number) out of 10 and 25.
- If the other group (group 2) had individual investments of 20 and 20, then the other group's **output is 20**, because 20 is the minimum of 20 and 20.
- Then total output from both groups is **10 + 20 = 30**, and the probabilities of success for each group are given below:

group	group output	probability of success
1	10	10/30 = 33.33%
2	20	20/30 = 66.67%

Are there any questions?

Cost of investment

Your cost of investment depends on your type and can be found in the table provided at the end of these instructions. To determine your costs of investment, locate the investment of your choice in the "Investment" column and find the corresponding number in the "Cost" column for your type. For example, if your type is B, the cost of investment of 20 is 255.

Note that for any given investment level,

type A has the lowest cost of investment
type B has the second lowest cost of investment
type C has the third lowest (second highest) cost of investment
type D has the highest cost of investment

Feel free to look through the costs table for all four types to compare the costs of investments for different types.

Are there any questions?

Payoff in a given round

Once the probability of project success of your group and the other group is calculated, the computer randomly chooses one group whose project is successful, *in accordance with the probabilities calculated*. That is, for example, if one group's probability of success is 20%, and the other group's probability of success is 80%, the first group's project is four times less likely to be successful than the second group's project.

Your **individual payoff** in a given round is determined as follows:

If your group project is successful:	If your group project is unsuccessful:
+400 (endowment)	+400 (endowment)
+1000 (revenue)	+0 (no revenue)
(your cost of investment)	(your cost of investment)
1400 (your cost of investment)	400 (your cost of investment)

Are there any questions?

Communication In each round, you will be able to communicate with the other member of your group using a computerized chat interface prior to making your individual investment decisions.l You can type messages in toe txt box at the botoom of the screen and submit them by pressing ENTER. Submitted messages will appear in the window above and will only be visible to the member of your group. Your message will be identifited by your type.

You will have 60 seconds to chat in each round. The remaining time will be shown at the bottom of the screen.

Please refrain from using profane or threatening language, or identifying yourself in any way.

Are there any questions?

How your earnings from this part of the experiment are determined.
You will go through 10 decision rounds. After that, **two** of these rounds will be chosen randomly (with all rounds being equally likely to be chosen) to base your earnings on. At the end of the experiment, you will be informed about the rounds chosen, your earnings in those rounds, and the total earnings from this part.

Are there any questions?

You will now start the actual decision rounds. Please do not communicate with other participants or look at their monitors. If you have a question or problem, from this point on please raise your hand and one of us will assist you in private. Please remember to click CONTINUE to proceed.

	type A	type B	type C	type D
investment	cost	cost	cost	cost
0	0	0	0	0
1	5	7	13	15
2	11	16	30	34
3	19	26	49	56
4	26	37	69	79
5	34	48	90	103
6	43	60	112	129
7	52	72	134	155
8	61	85	158	182
9	70	98	182	209
10	79	111	206	238
11	89	124	231	267
12	99	138	256	296
13	109	152	282	326
14	119	166	309	356
15	129	180	335	387
16	139	195	362	418
17	150	210	389	449
18	160	225	417	481
19	171	240	445	514
20	182	255	473	546
21	193	270	502	579
22	204	286	531	612
23	215	301	560	646
24	227	317	589	680
25	238	333	619	714
26	249	349	649	748
27	261	365	679	783
28	273	382	709	818
29	284	398	739	853
30	296	415	770	888
31	308	431	801	924
32	320	448	832	960
33	332	465	863	996
34	344	482	895	1032
35	356	499	926	1069
36	369	516	958	1106
37	381	533	990	1143
38	393	551	1023	1180
39	406	568	1055	1217
40	418	586	1087	1255
41	431	603	1120	1293
42	443	621	1153	1330
43	456	639	1186	1369
44	469	657	1219	1407
45	482	674	1253	1445
46	495	692	1286	1484
47	508	711	1320	1523
48	521	729	1353	1562
49	534	747	1387	1601
50	547	765	1421	1640

Table 7: Cost of effort as a function of chosen investment and type.