Do Competitive Incentives in Team Production Lead to More Effort or More Sabotage? *

E. Glenn Dutcher[†], Regine Oexl[‡], Dmitry Ryvkin[§], Timothy C. Salmon[¶]

This version: August 23, 2019

Abstract

There is an important tension in the incentive systems of many firms. While firms want to incentivize team production by providing rewards for overall output, they also want to incentivize individual effort with relative performance-based bonuses. When production relies on team cooperation though, these competitive incentives may harm production as team members may refrain from helping each other and could even begin sabotaging their colleagues in order to get ahead. We examine these issues via theory and experiments in the presence of individual heterogeneity to try to determine the manner in which individual effort provision as well as help/sabotage behavior may change as competitive incentives increase and as team composition changes. Our results help understand better how to balance out individual versus team rewards and how firms might want to structure teams when employees have heterogeneous abilities.

Keywords: contest, help and sabotage, team composition, incentive structure
JEL classification codes: C92, D01

*Financial support from the Austrian Science Fund (FWF, P27912-G27 and SFB F63) is gratefully acknowledged. We are grateful for valuable comments from the conference participants at the 2018 ESA World Conference in Berlin, Germany; 2018 GSDS Conference on Decision Sciences, Konstanz, Germany; 2018 THEEM in Kreuzlingen, Switzerland; 2018 ASFEE in Nice, France; 2018 LAWEBESS in Santiago, Chile; 2019 Coller Conference on Behavioral Economics in Tel Aviv, Israel; Texas Experimental association conference at Baylor University, Waco, USA, as well as seminar participants at the University of Alaska, USA, and Ohio University, USA.

[†]Ohio University
[‡]University of Innsbruck
[§]Florida State University
[¶]Southern Methodist University

1 Introduction

Many tasks are best handled by a team.[1] The success of a team often relies on cooperation among team members such as when one member of the team helps another by sharing knowledge or takes on part of a teammate's task (Lazear and Shaw, 2007).[2] Such cooperation is often fostered by team incentives, with some portion of pay based on collective output.[3] While team-based incentives may be intended to promote cooperation, they may also dampen individual effort due to free-riding concerns. The alternative approach to motivating employees, advocated by the likes of Jack Welch, argues that competition among employees and appropriate rank-based rewards are the best way to motivate workers and drive the organization toward constant improvement.[4] Competitive mechanisms, however, may diminish the willingness of individuals to help others, and may even incentivize intra-team sabotage, leading to potentially lower total output. The question of whether rank-based reward mechanisms lead to increased or decreased production in a team setting have been much debated in many large organizations over the past few decades, and while the ranking mechanisms may be on the decline, which approach corporations should adopt has yet to be settled. For example, while Microsoft recently decided to scrap their rank-order system, Yahoo! announced they were implementing one.[5] Given the relevant implications of these compensation issues, a more complete understanding of how they actually affect employee behavior is warranted.

Evaluating these claims and determining the effectiveness of rank-based mechanisms is vital to unlocking team dynamics and ultimately in resolving questions regarding how employers might want to assemble their workers into teams. The dual incentive problem that firms face in trying to incentivize team production – incentivizing individual effort while also trying to encourage cooperation – becomes even more complicated once

[1]The number of organizations utilizing teamwork has been growing since the 1980s (Lazear and Shaw, 2007). Between 1987 and 1999, the percentage of firms with at least 20% of employees working in teams increased from 37 to 61% (Lawler, Mohrman and Benson, 2001).

[2]Ultimately, efficient utilization of complementarities defines also the boundaries of firms (Alchian and Demsetz, 1972).

[3]Lawler, Mohrman and Benson (2001) report some level of "gainsharing"in 53% of surveyed firms.

[4]"Jack Welch: 'Rank-and-Yank'? That's Not How It's Done,"The Wall Street Journal, November 14, 2013, https://www.wsj.com/articles/8216rankandyank8217-that8217s-not-how-it8217s-done-1384473281.

[5]"Microsoft Ditches the Stack Ranking System. Yahoo! Lays off 600 because of It," InfoQ, November 16, 2013, https://www.infoq.com/news/2013/11/stack-ranking-microsoft-yahoo; "'Because Marissa Said So' – Yahoo's Bristle at Mayer's QPR Ranking System and 'Silent Layoffs,"' All Things D, November 8, 2013, http://allthingsd.com/20131108/because-marissa-said-so-yahoos-bristle-at-mayers-new-qpr-ranking-system-and-silent-layoffs/; "Microsoft axes its controversial employee-ranking system," The Verge, November 12, 2013, https://www.theverge.com/2013/11/12/5094864/microsoft-kills-stack-ranking-internal-structure.

the realistic assumption of worker heterogeneity is considered. This is because workers of varying ability may respond to these rank-based mechanisms differently and their response may also differ based on the composition of their team. This makes determining the effectiveness of the payment scheme more difficult and compounds the problem by introducing a new puzzle regarding the optimal way to compose teams. The issue is no longer one related only to incentives, but now involves a question of whether it is best to construct homogeneous or heterogeneous teams where the answer very likely depends on the nature of the rank-based reward system.

Our goal in this paper is to assess the validity of the competing claims regarding the effect of competitive incentives on individual effort and helping behavior in a team production setting with heterogeneous agents and then determine what those results suggest about the optimal structure of teams. Exploring these questions using field data would be difficult because such data rarely contains information on effort and abilities, and is almost guaranteed to omit information on behaviors involving help and sabotage. Further, the endogeneity of team construction and design of the compensation mechanism would make it difficult to identify causal relationships. We will, therefore, investigate these questions through the use of controlled laboratory experiments guided by a theoretical model of the underlying incentives.

To help us understand the basic questions better, it is worth discussing the competing schools of thought in more detail to understand the essential elements of their claims and what evidence might exist for or against these claims. Fundamentally, the two sides of this debate make conflicting claims regarding how workers might react to different incentive schemes in regard to both their individual effort and their tendencies to help fellow co-workers. Those who believe strongly in the importance of competitive incentives are implicitly claiming that such incentives lead to large increases in individual effort, but do not significantly reduce helping behaviors or at least do not lead to such a large reduction that overall productivity is harmed. The group who believes that competitive incentives damage teamwork and impair overall productivity are essentially claiming the opposite which is that while the competitive incentives might or might not increase individual effort, those incentives will ruin the willingness of co-workers to help each other and could even lead to acts of sabotage.[6] This viewpoint will usually come with an optimistic view of the ability of people to cooperate with each other as well.

One of the reasons that neither viewpoint has achieved dominance is that there is

[6] See, e.g., "Companies Revisit 'Rank And Yank' of 1980s," NPR, December 2, 2013, https://www.npr.org/2013/12/02/248151316/companies-revisit-1980s-rank-and-yank, and "Why Stack Ranking Is a Terrible Way To Motivate Employees," Business Insider, November 15, 2013, https://www.businessinsider.com/stack-ranking-employees-is-a-bad-idea-2013-11.

substantial prior evidence favoring the core arguments from both. For example, there is a long literature examining behavior in tournaments and contests, and the most common result found in that literature is that competitive incentives drive individuals to exert substantially more effort than predicted in a standard model (see, e.g., a review by Dechenaux, Kovenock and Sheremeta, 2015). These results are quite strong and one might think that they lend credence to the use of competitive incentives in the field. On the other hand, there is also a very long literature showing that individuals are much better at cooperating than one would expect given the predictions from a standard model (see, e.g., Ledyard, 1995). This literature showing that individuals will often contribute much more than expected in a public goods setting suggests that in corporate team production settings, teammates may also be able to solve the cooperation problems and, therefore, generate high output without the need for the competitive incentives. These base studies on behavior in contests and public goods environments do not, however, provide clear answers to the question of how the competitive incentives affect cooperative behavior because in most of these prior studies the two issues are examined separately.

There are a few prior studies which examine more directly the effect of competition on cooperation. Buser and Dreber (2016) examine the issue in a setting where participants either compete for a prize or engage in piece-rate work prior to playing a public goods game. That study finds that people do tend to cooperate less after they have competed with each other. These findings do not seem conclusive, though, due to the fact that the study finds similar results when the prize is allocated purely randomly rather than through a competition. This suggests the lack of cooperation in the public goods game may simply be due to unequal endowments rather than the experience of competition, as has been found in other studies (Heap, Ramalingam and Stoddard, 2016). Danilov, Harbring and Irlenbusch (2019) present an experiment more similar to ours, as they address directly a subset of the questions in which we are interested, with a similar experimental design. That study investigates a setting of team production where teammates can help and sabotage each other to examine the effects of introducing a prize for the worker who contributes the most individual effort. The study finds results which largely comport with standard theoretical predictions showing that increasing the size of the prize increases individual effort and decreases the willingness of workers to help each other. Our interest is in taking the examination of these issues further by examining teams with various compositions of heterogeneous workers and trying to identify systematic deviations from the standard theoretical predictions that will provide a deeper understanding of this behavior.

To begin to understand how heterogeneity complicates these issues, consider a heterogeneous team composed of one member whose ability is far superior to that of her teammates. Relative to a purely team-based incentive scheme, the introduction of a competitive bonus may not actually induce the better player to exert much higher effort if she expects to easily win the competition anyway. Even though the incentive may not induce higher effort, because she knows she faces very little competition for this prize, she may still help her less productive co-workers to keep team production high and secure higher team-level payments. Additionally, if her less productive teammates expect little chance of winning the bonus, it might also be the case that they too will not increase their effort, but this also implies that their willingness to help others does not diminish. If a manager wishes to increase the competitiveness of the environment, she may reassign workers so that the team includes multiple strong members. With this new configuration, introducing tournament incentives could yield a very different impact. The effort of the high-ability team members could increase substantially as each strives to win the prize, but they may no longer be willing to spend the time to help others as it could improve the competitiveness of others which indirectly decreases their own chance of winning. At the extreme, when all members of a team are of similar abilities, the addition of a competitive prize may actually lead to team members sabotaging each other. Of course the firm does not want to encourage such behavior, but more important to the firm is the total effect. Reductions in help resulting from more competitive settings may be optimal if they are more than offset by increases in effort.

What these few examples make clear is that when teams are comprised of workers pf heterogeneous ability, their potential responses to the introduction of a competitive element are quite complicated and are potentially driven by a range of conflicting motives. Depending on how different motivations balance out, it may lead to firm managers preferring to try to form relatively homogenous teams, as this may maximize the effect of bonuses on individual effort. The firm may also wish to form heterogeneous teams, as this may better preserve the willingness of teammates to cooperate with each other. It is also possible that the optimal team configuration could depend on the level of competitive incentives.

To investigate these issues, we start by presenting a theoretical model of decision making for workers in a team where output is rewarded by a combination of team-based and competitive incentives. Workers are heterogeneous in their productivity (in the experiment, we restrict heterogeneity to having only two types) and are able to exert individual effort as well as effort towards helping or sabotaging other team members. Our main interest is in understanding if the competitiveness of the setting leads to levels of effort and help/sabotage that differ from money-maximizing behavior and if such behavior supports either of the two competing schools of thought on corporate compensation. The theoretical predictions provide a baseline for money-maximizing behavior, while the actual choices are the result of an experiment which isolates the relevant areas of the broader debate and allows causal identification. Specifically, in the experiment, we examine how behavior changes as we vary the proportion of high and low ability workers on a team and how behavior changes as we increase the strength of the competitive incentives.

In general, we find a mix of evidence for and against both the competitive and the cooperative (or noncompetitive) schools of thought. We find that moderate competitive incentives drive effort above the equilibrium prediction, but this fails at high levels of incentives where effort drops below the corresponding equilibrium. Individuals do cooperate well in the no competition case by providing effort above the individually optimal levels, but their helping behavior is lower than predicted. On the other hand, when competitive incentives rise to the level where sabotage would be predicted and extreme sabotage might be feared based on the competitive school of thought, we find that people sabotage much less than expected. This generates a complex set of findings which may help further explain why neither payment scheme is universally adopted in the field. Understanding the strengths and weaknesses of both may help understand where one might or might not consider using competitive incentives in a team production setting.

The rest of the paper is organized as follows. Section 2 presents the theoretical model while Section 3 presents the experimental design and exploratory hypotheses for behavioral deviations that may be expected. The results are presented in Section 4, and Section 5 concludes.

2 Model

In this section we present a model that provides a set of predictions regarding how individuals will behave when competitive incentives are introduced into a team production setting. Consistent with our issues of interest, the model allows for heterogeneity in ability of team members and for team members to choose to devote their energy toward individual effort, helping another teammate or sabotaging another teammate. There are many different assumptions one can make in constructing such a model that will affect its predictions regarding effort, help and sabotage levels, and how helping behavior occurs between agents of various types.

Our goal is not to produce a general model which is calibrated on any specific setting. Rather, what we need from the model is a flexible and straightforward method of providing a set of baseline predictions regarding behavior in an environment which is amenable to conducting experiments. The model we present was constructed with this goal in mind, noting that our interest in the end will be mostly in examining the data for systematic patterns regarding how individuals alter their behavior as we increase the relative magnitude of competitive incentives and change the ability composition of the team.

Our model is a variation of several existing models of help and sabotage in teams employing homogeneous (Garvey and Swan, 1992; Danilov, Harbring and Irlenbusch, 2019) and heterogeneous (Kräkel, 2005; Gürtler and Münster, 2013) agents.[7]

Consider a team consisting of $n \geq 2$ risk-neutral agents indexed by $i \in \{1, \ldots, n\}$ and characterized by (possibly heterogeneous) ability parameters $\gamma_i > 0$. Each agent i chooses effort $x_i \in \mathbb{R}_+$ associated with a strictly convex, increasing cost function $c(x_i)$. In addition, agent i chooses, for every agent $j \neq i$ in the team, the level of effort-modifying activity $k_{ij} \in \mathbb{R}$, where $k_{ij} > (<) 0$ corresponds to agent i helping (sabotaging) agent j. Help and sabotage are associated with a strictly convex cost function $s(k_{ij})$, which is increasing (decreasing) for $k_{ij} > (<) 0$. The output of agent i is given by

$$y_i = \gamma_i \max\{0, x_i + \sum_{j \neq i} k_{ji}\}, \tag{1}$$

Equation (1) ensures that output cannot be negative for any levels of sabotage. In the experiment, we choose parameters so that in equilibrium the constraints $x_i \geq 0$ and $y_i \geq 0$ are not binding.

Team incentives without competition

Every team member receives a piece rate r per unit of total team output $Y = \sum_{i=1}^{n} y_i$. For simplicity, suppose that effort and effort-modifying activities have the same cost, and both cost functions are quadratic: $c(x_i) = \frac{1}{2\alpha} x_i^2$ and $s(k_{ij}) = \frac{1}{2\alpha} k_{ij}^2$.[8] Note that k_{ij} can be positive or negative, but $s(k_{ij})$ is increasing in $|k_{ij}|$. This gives agent i's utility (payoff) in the form

$$\pi_i = r \sum_{j=1}^{n} y_j - \frac{x_i^2}{2\alpha} - \sum_{j \neq i} \frac{k_{ij}^2}{2\alpha}.$$

[7] The main difference between these model and ours is in that they model the tournament component of the incentives à la Lazear and Rosen (1981) whereas we employ a lottery contest success function of Tullock (1980). Ultimately, both are a form of a noisy winner determination process. One advantage of our model is that it allows for a flexible closed-form solution for heterogeneous agents.

[8] Parameter $\alpha > 0$ can be subsumed in γ_i and is redundant for modeling, but it will be helpful in calibrating the experiment.

Maximizing π_i with respect to x_i and k_{ij} (for each $j \neq i$), obtain

$$x_i^*(0) = r\alpha\gamma_i, \quad k_{ij}^*(0) = r\alpha\gamma_j.$$

These levels of effort and effort-modifying activities constitute the unique Nash equilibrium (NE) in dominant strategies.

Team incentives with competition for a bonus

We will now introduce an intra-team contest. We assume that there is a manager who imperfectly observes individual output levels y_i and rewards the agent whose output is perceived as the highest with a bonus $V \geq 0$. We model the winner determination process using the Tullock/lottery contest success function (CSF) whereby the probability for agent i's output to be perceived as the highest is $\frac{y_i}{\sum_{j=1}^n y_j}$. In this setting, agent i's expected payoff function is

$$\pi_i = \frac{V y_i}{\sum_{j=1}^n y_j} + r\sum_{j=1}^n y_j - \frac{x_i^2}{2\alpha} - \sum_{j\neq i} \frac{k_{ij}^2}{2\alpha}. \tag{2}$$

In order to find the equilibrium, consider the system of first-order conditions for effort and help/sabotage levels, assuming interior solutions:[9]

$$\frac{V\gamma_i \sum_{m\neq i} y_m}{(\sum_{m=1}^n y_m)^2} + r\gamma_i = \frac{x_i}{\alpha}, \quad x_i \geq 0; \tag{3}$$

$$-\frac{V y_i \gamma_j}{(\sum_{m=1}^n y_m)^2} + r\gamma_j = \frac{k_{ij}}{\alpha}, \quad j \neq i. \tag{4}$$

For brevity, let $Y = \sum_{i=1}^n y_i$. Equations (3) and (4) can be manipulated to obtain a closed-form solution. Expressing x_i from (3) and k_{ji} from (4), obtain individual outputs,

$$\begin{aligned} y_i &= \gamma_i \left(x_i + \sum_{j\neq i} k_{ji} \right) \\ &= \alpha\gamma_i \left(\frac{V(Y-y_i)}{Y^2} + r\gamma_i - \frac{V(Y-y_i)}{Y^2} + (n-1)r\gamma_i \right) = nr\alpha\gamma_i^2, \end{aligned} \tag{5}$$

and aggregate team output: $Y^* = \alpha rn \sum_{i=1}^n \gamma_i^2$. Plugging this expression and (5) into (3)

[9] In the experiment, we choose parameters so that the interior solution to first-order conditions indeed provides best responses in each case.

and (4), obtain

$$x_i^*(V) = \gamma_i \left[r\alpha + \frac{V \sum_{m \neq i} \gamma_m^2}{rn(\sum_{m=1}^n \gamma_m^2)^2} \right], \quad k_{ij}^*(V) = \gamma_j \left[r\alpha - \frac{V \gamma_i^2}{rn(\sum_{m=1}^n \gamma_m^2)^2} \right]. \tag{6}$$

As seen from (6), while equilibrium effort increases with the introduction of the bonus, help decreases so as to exactly offset the impact of the increase in effort on aggregate output; the latter is independent of the bonus. For a sufficiently large V, help becomes negative, i.e., it turns into sabotage. Note that, other things being equal, this turning point is lower for higher-ability agents.

3 Experimental Design and Predictions

3.1 Overview

Sessions were conducted in November 2017 and May 2018 at the Innsbruck Econ Lab. The experiment was computerized via z-Tree (Fischbacher, 2007), and recruitment of subjects took place via the recruitment system *hroot* (Bock, Baetge and Nicklisch, 2014). Once all subjects were checked in and seated at a computerized workstation, instructions were handed out and were read out loud.[10] After all questions were answered, the experiment began.

3.2 Treatments

In the experiment, we utilized two ability levels; high (H) and low (L). The ability of each subject was exogenously assigned at the beginning, and fixed throughout the session. Additionally, subjects were anonymously assigned into a fixed team with three others to make a team of four. In order to fully understand the effects of team composition, we utilized a between subjects design, where subjects were assigned into one of five potential team composition treatments – *HHHH*, *HHHL*, *HHLL*, *HLLL* or *LLLL*. In a given session, the group composition was fixed and each subject knew their own ability along with the ability of their three teammates.

The main part of the experiment consisted of three 8-round blocks for a total of 24 rounds. The first block included only a team incentive: all team members received the same payoff, proportional to the total team output; there was no contest incentive in this block. The second and third blocks added on the contest incentives, and we varied the size

[10] All instructions were neutrally framed. Instructions for one of the treatments are included in the Appendix.

of the prize across these blocks, with one block using a relatively small prize, a low powered incentive, and the other a rather large prize, to represent a high powered incentive. The reason for using within-subject variation in the contest prize – and beginning all sessions with the no-contest baseline – is that we wish to understand how behavior changes with the *introduction* of contest incentives. Starting from a situation without contest incentives and then adding them in represents the change we are considering when a corporation implements a different system. To control for order effects in examining the different sizes of contest prizes, we varied the order of the prizes. In 'order 1', the low powered incentive was introduced in block 2 and the high powered in block 3; in 'order 2', these were reversed with the high powered incentive in block 2 and the low powered incentive in block 3.[11] New instructions were handed out, and read out loud prior to each block, to introduce and explain the changes to the incentive scheme. It was common knowledge that one round per block was chosen randomly for payment.

In each round, all subjects made four simultaneous choices. They had to choose how many points to allocate to their own effort and how many to allocate to modifying their three other teammates' effort. Individual effort could be any integer from 0 to 150, while modifications ranged from -150 to 150.[12] Each choice entailed a cost, which was presented to subjects in a table in their instruction packet.[13] After all subjects made their choices, they were shown a results screen. On the results screen, subjects were reminded of their own choices and were shown the average help/sabotage in their team directed at all members, their own total effort (which is a combination of their own effort and effort modifying choices of their team members), their output after accounting for their ability, and the total team output. They were also reminded of the cost of each decision, the group payment from the team output, whether they won the prize or not (in blocks 2 and 3), and their total payoff in that round should it be chosen for payment. Following the main game, subjects' risk preferences were elicited using the 'bomb' risk elicitation task (Crosetto and Filippin, 2013).

[11]The experiment was designed to obtain two sessions of each team composition with those two sessions using the different orderings. In the end, we conducted 11 sessions because one session configuration was accidentally used in two sessions.

[12]All amounts in the main part of the experiment were denominated in tokens. At the end of the session payoffs were translated into Euros at the exchange rate 100 tokens = €6.

[13]In order to ensure numeracy was not a concern, subjects had access to an on-screen calculator which calculated hypothetical payoffs given their own choices and hypothetical choices they entered for other members in their group. Sample screenshots are included in the Appendix.

3.3 Parameters and Equilibrium Predictions

The goal of our experiments is to examine how behavior changes as we increase tournament incentives and change group composition in a team production environment. We have constructed a set of parameters which are intended to allow us to do just that. First, we vary the size of the prize, V, between the values 0, 100 and 500 to represent no tournament incentives, moderate and then high incentives. Second, we vary the ability of our workers by setting $\gamma_L = 1$ for low ability and $\gamma_H = 2$ for high ability workers. Third, we vary the team composition by considering all possible configurations of L and H types, in four-person teams, as previously described. We set the piece rate to $r = 0.25$ and $\alpha = 39.0625$, which results in the cost functions $c(x_i) = 0.0128x_i^2$ and $s(k_{ij}) = 0.0128k_{ij}^2$. With these parameters, equations (5) and (6) lead to a set of equilibrium point predictions of behavior as displayed in Table 1.

There are a few key elements to note about these predictions. First, total group output, Y, increases with the number of H types in the group, yet, given a fixed group composition, it does not vary with V. Importantly, individual effort, x_i, does depend on V, as does help/sabotage, k_{ij}. These values move in opposite directions, and the parameters are chosen so that they exactly offset each other in determining total output under our production function. This symmetric offset makes it easy to examine the exact channel through which V affects output and allows us to cleanly test the two schools of thought. Relatedly, these parameters also allow us to cleanly examine behavioral effects from different group compositions. That is, with a group of eight workers – four H types and four L types – divided into two groups of any configuration, the total output is predicted to be constant, 781.25, for any value of V. This point will be relevant when we later discuss what our results suggest about optimal group composition as, in equilibrium, all group constructions are equivalent to the firm. We may, however, find that they are not behaviorally equivalent.

A second element to note is that helping behavior is predicted to be directed mostly towards H types rather than L types, as both H and L types help H types more than L types in equilibrium. Prior studies examining helping behavior have found the opposite (Hamilton, Nickerson and Owan, 2003; Brandts et al., 2016). In these studies, high ability workers tend to help the low ability ones. The difference is due to the specification of help in our model, where a transfer of helping effort k_{ij} from agent i to agent j is augmented by j's productivity parameter γ_j. This is certainly one of many possible ways help can operate. Think, for example, of a team of lawyers, in which senior lawyers are responsible for strategic decisions while junior lawyers help them with mundane tasks, such as summarizing precedents. In this case, the junior lawyers helping the senior ones is the most efficient way to increase productivity of the firm. One should not mistake the prediction our model makes for a general claim – we take no stance on what the "correct" relationship is, we rather want to understand how observed behavior changes, as V increases, relative to some predicted level of cooperative behavior. Our environment was constructed to induce levels of help that were high enough to allow us to observe how they might change. Their relative sizes between types are of no importance to our research questions.

Table 1: Equilibrium predictions

	LLLL	*HLLL*	*HHLL*	*HHHL*	*HHHH*
			$\boldsymbol{V} = 0$		
x_L	9.77	9.77	9.77	9.77	
x_H		19.53	19.53	19.53	19.53
k_{LL}	9.77	9.77	9.77		
k_{LH}		19.53	19.53	19.53	
k_{HL}		9.77	9.77	9.77	
k_{HH}			19.53	19.53	19.53
			$\boldsymbol{V} = 100$		
x_L	28.52	22.01	18.77	16.87	
x_H		31.78	31.53	30.18	28.91
k_{LL}	3.52	7.72	8.77		
k_{LH}		15.45	17.53	18.35	
k_{HL}		1.60	5.77	7.40	
k_{HH}			11.53	14.80	16.41
			$\boldsymbol{V} = 500$		
x_L	103.52	70.99	54.77	45.27	
x_H		80.76	79.53	72.79	66.41
k_{LL}	-21.48	-0.44	4.77		
k_{LH}		-0.88	9.53	13.61	
k_{HL}		-31.05	-10.23	-2.07	
k_{HH}			-20.47	-4.14	3.91
y_L	39.06	39.06	39.06	39.06	
y_H		156.25	156.25	156.25	156.25
Y	156.25	273.44	390.63	507.81	625.00

3.4 Behavioral Predictions

Based on the values in Table 1 – or equations (5) and (6) – we can construct a basic set of testable predictions. Even though such tests could be useful in generating a first estimate of behavior in this setting, it is highly unlikely that we will find consistent support for the point predictions. Our real interest is not in simply demonstrating whether the model works, but rather in using these theoretical benchmarks to test for systematic behavioral deviations from the theory, and testing if these deviations support one of the competing schools of thought described above. Our behavioral hypotheses are built around the core idea that the two schools of thought differ in the perceived benefits and drawbacks of competition. Those arguing in favor of competition claim that benefits will be manifested via increased effort. Likewise, those arguing against competitive incentives hold that competition will lead to lower levels of help and increases in sabotage. Important to our study are the two ways that influence the competitiveness of the setting: The size of the prize and the composition of the group. As the size of the prize increases, we assume the competition increases. We also assume that group composition matters and the competitiveness of the setting is increasing in the number of H types. The first set of hypotheses concerns the effort decisions of the subjects.

Hypothesis 1 (Effort 1) *Observed effort from H and L types will exceed the equilibrium predictions for all values of V.*

This prediction is consistent with the large prior experimental literature on contests, showing that effort is usually provided in excess of the equilibrium predictions when $V > 0$,[14] and is in line with findings from similar revenue sharing schemes when $V = 0$.[15] Our greater interest though is in how effort changes as competition increases; that is, as V increases, and as group composition changes. The next two hypotheses concern these two comparative statics. Both statements are testing the credibility of the school of thought that claims that individuals tend to behave as though the contest places more competitive pressure on their behavior than expected. If this holds true in our environment, then, as competitive pressure rises – due to the presence of more strong competitors, or due to the prize increasing in value – we would expect that individuals respond with greater deviations from the equilibrium predictions. The effects of the additional competitive pressure due to an increase in V are clear. Similarly, as competitive pressure increases from additional H types in the group, the predicted effect on H types is also clear, as they should feel additional pressure to compete and hence increase individual effort. The effect on L types is less clear, as they could also respond to the competitive pressure by increasing effort, or, realizing that their chances of winning are diminished with each additional H type added to their group, they could respond by decreasing their effort. We keep the hypothesis for the L types in line with the hypothesis for the H types – an

[14]Overbidding is widely documented in experiments on lottery contests, see Sheremeta (2013) for a review. At the same time, bidding at or below equilibrium predictions is typical for experiments utilizing the Lazear-Rosen tournament framework (see, e.g., Bull, Schotter and Weigelt, 1987; Dutcher et al., 2015). However, the latter environment is not directly comparable to our setting.

[15]See, e.g., the review by Ledyard (1995).

increase in effort with an increasing number of H types – which is consistent with the notion that the competitiveness of the setting leads to an additional (positive) behavioral response, but recognize the opposite may feasibly occur.

Hypothesis 2 (Effort 2) *The deviation of the observed effort from the equilibrium prediction, $x_{i,V} - x^*_{i,V}$, will be increasing in the value of V, for both H and L types.*

Hypothesis 3 (Effort 3) *The deviation of the observed effort from the equilibrium prediction, $x_{i,V} - x^*_{i,V}$, will be increasing in the number of H types in the group, for both H and L types.*

The next set of hypotheses concerns help/sabotage behavior of the subjects. Many prior examinations of social dilemmas indicate that people will often cooperate at a higher rate than predicted by standard models. As our basic prediction we use the one derived from this literature, claiming that people are more cooperative than expected, and suppose that this holds for all V.

Hypothesis 4 (Help 1) *Observed help from H and L types towards both H and L types will exceed the equilibrium predictions for all values of V.*

How this behavior might shift based on changes in group composition and V is less clear; prior results regarding public goods games provide no strong basis for predictions. We can, however, form some predictions for how behavior might deviate from the standard model, if we presume that, similar to the "noncompetitive "school of thought, individuals are driven by competitive pressures. In this case, these competitive pressures should generally drive people to reduce the amount of help directed to others who might be rivals for the prize. In our setting, this lower level of help could also imply more sabotage than predicted. As before, we expect competitive pressure to go up with the size of the prize and the number of H types in the group.

Hypothesis 5 (Help 2) *The difference between the observed amount of help and the equilibrium prediction, $k_{ij,V} - k^*_{ij,V}$, will be decreasing in the value of V, for both H and L types.*

Hypothesis 6 (Help 3) *The difference between the observed amount of help and the equilibrium prediction, $k_{ij,V} - k^*_{ij,V}$, will be decreasing in the number of H types in the group, for both H and L types.*

In line with the noncompetitive school of thought, the interpretation of *Help 2* and *Help 3* is that observed help will be declining relative to predicted help; we do not intend for them to imply that observed help will converge to the predicted value from above as observed help could certainly drop below predicted levels and continue dropping below as competitive pressure increases.

From a managerial perspective, the main variable of interest is usually total output. *If* effort is increasing faster than predicted *and* help is declining faster than predicted, the total impact of these two effects on output is heavily dependent on the relative importance of the two in a given production function of the firm. It is important to note the importance of the form of the production function in this overall effect because, while we expect our behavioral results on effort and help/sabotage should extend to other settings, their combined impact on total production could be different. We will provide results on the impact of these behaviors on total output given our specified production function, but we do so with the caveat that the overall effect might change with different production functions.

One can specify hypotheses regarding how total output changes as a result of these behavioral effects to be consistent with the view of those who favor competitive incentive schemes or those who think they are harmful. In line with our prior hypotheses following the competitive school of thought, we will state a set of hypotheses consistent with the notion that the positive effects of the competition will outweigh the negative impacts, but of course the hypotheses could be stated either way.

Hypothesis 7 (Total Output 1) *Total output will be greater than the equilibrium predictions for all values of V.*

Hypothesis 8 (Total Output 2) *The difference between the observed total output and the equilibrium prediction, $Y - Y^*$, will be increasing in the value of V.*

Hypothesis 9 (Total Output 3) *The difference between the observed total output and the equilibrium prediction, $Y - Y^*$, will be increasing in the number of H types in the group.*

The results of testing these hypotheses should be useful in helping to understand the optimal group structure in situations where there is heterogeneity in ability and the opportunity for help and sabotage. These results should also indicate the degree to which adding competitive incentives to a team production environment may or may not be helpful in increasing overall production.

4 Results

Overall, 264 subjects participated in one of 11 sessions. Including the €9 participation payment, average earnings were €24.58 for an experiment that lasted about 90 minutes. Table 2 provides a breakdown of the number of subjects and groups in each session. With one exception, there were 48 individuals in each treatment. The only exception is the *HHHL* treatment, which includes 72 individuals resulting from one extra session of this treatment. Thus, there are 12 groups (non-interacting clusters) in each treatment except *HHHL* that contains 18 groups, leading to a total of 66 groups. Each subject made choices in 24 decision rounds, eight in each of the payment schemes. This leads to a total of 6,336 individual decision rounds.

Table 2: Overview of treatments, sessions and the number of observations

Treatment	# Sessions	# Subjects	# Rounds	# Groups	# Obs.
LLLL	2	48	24	12	1,152
HLLL	2	48	24	12	1,152
HHLL	2	48	24	12	1,152
HHHL	3	72	24	18	1,728
HHHH	2	48	24	12	1,152
Total	11	264		66	6,336

4.1 Overview

We begin with an overview of the data. Table 3 presents the equilibrium predictions and observed averages in each treatment and reward scheme. While these summary statistics are useful for gaining an initial impression of the results, we will not conduct an exhaustive sequence of tests comparing the observed averages to theoretical predictions. As discussed above, our approach is instead to look for patterns in *deviations* from equilibrium predictions resulting from the introduction of competition and changes in the competitive environment, including group composition.[16]

From a brief examination of the summary statistics in Table 3, it appears that in the $V = 0$ and $V = 100$ pay schemes, average effort (x_i) is generally close to or above the predicted value, while in the $V = 500$ pay scheme, it is generally close to or below the

[16]For interested readers, Table 8 in Appendix C presents the observed averages with robust standard errors that can be used to perform comparisons to theory. Also included in the Appendix are Figures 7-10 which provide a graphical summary of the data.

predicted value. Help (k_{ij}) is generally close to or below the predicted value in the $V = 0$ and $V = 100$ pay schemes. In the $V = 500$ case, when predicted help is very low (including sabotage), observed help appears generally above the prediction while when the prediction for help is higher, help appears lower than predicted leading to no general relationship. Similarly, we find that total group output sometimes exceeds predicted output while at other times it is lower than the predicted level. As we examine each of our hypotheses, we will attempt to better define these apparent relationships.

Table 3: Equilibrium predictions (Eq.) and observed averages (Obs.)

	LLLL		*HLLL*		*HHLL*		*HHHL*		*HHHH*	
	Eq.	Obs.	*Eq.*	Obs.	*Eq.*	Obs.	*Eq.*	Obs.	*Eq.*	Obs.
					$\boldsymbol{V} = 0$					
x_L	*9.77*	23.64	*9.77*	24.46	*9.77*	21.01	*9.77*	30.69		
x_H			*19.53*	25.88	*19.53*	37.23	*19.53*	31.36	*19.53*	35.54
k_{LL}	*9.77*	8.26	*9.77*	8.93	*9.77*	9.28				
k_{LH}			*19.53*	14.77	*19.53*	17.20	*19.53*	7.14		
k_{HL}			*9.77*	6.00	*9.77*	9.94	*9.77*	7.34		
k_{HH}					*19.53*	20.21	*19.53*	11.83	*19.53*	14.36
y_L	*39.06*	49.36	*39.06*	48.31	*39.06*	50.24	*39.06*	53.18		
y_H			*156.25*	140.38	*156.25*	184.00	*156.25*	137.54	*156.25*	151.26
Y	*156.25*	197.46	*273.44*	285.30	*390.63*	468.48	*507.81*	465.79	*625.00*	605.04
					$\boldsymbol{V} = 100$					
x_L	*28.52*	33.28	*22.01*	31.64	*18.77*	31.31	*16.87*	39.69		
x_H			*31.78*	35.78	*31.53*	47.67	*30.18*	44.74	*28.91*	39.05
k_{LL}	*3.52*	0.66	*7.72*	2.54	*8.77*	4.65				
k_{LH}			*15.45*	4.29	*17.53*	8.61	*18.35*	0.68		
k_{HL}			*1.60*	-3.20	*5.77*	2.32	*7.40*	1.91		
k_{HH}					*11.53*	5.00	*14.80*	3.55	*16.41*	7.10
y_L	*39.06*	36.00	*39.06*	34.44	*39.06*	41.79	*39.06*	48.58		
y_H			*156.25*	97.25	*156.25*	139.98	*156.25*	120.31	*156.25*	120.80
Y	*156.25*	144.00	*273.44*	200.58	*390.63*	363.53	*507.81*	409.5	*625.00*	483.21
					$\boldsymbol{V} = 500$					
x_L	*103.52*	57.81	*70.99*	50.44	*54.77*	48.89	*45.27*	45.47		
x_H			*80.76*	60.60	*79.53*	69.57	*72.79*	57.44	*66.41*	63.40
k_{LL}	*-21.48*	-5.46	*-0.44*	-3.80	*4.77*	0.81				
k_{LH}			*-0.88*	-3.51	*9.53*	4.59	*13.61*	-7.17		
k_{HL}			*-31.05*	-11.48	*-10.23*	-1.94	*-2.07*	-2.00		
k_{HH}					*-20.47*	-2.42	*-4.14*	-1.59	*3.91*	2.35
y_L	*39.06*	46.77	*39.06*	39.83	*39.06*	51.39	*39.06*	47.64		
y_H			*156.25*	101.58	*156.25*	155.76	*156.25*	115.31	*156.25*	144.58
Y	*156.25*	187.09	*273.44*	221.08	*390.63*	414.29	*507.81*	393.58	*625.00*	578.33

4.2 Testing the hypotheses

In order to better understand underlying behavior in our setting, we turn our attention to the empirical tests of our hypotheses. We will establish statistical significance of our results as they pertain to specific hypotheses using regression analysis. Given that the primary interest of the paper is to understand the behavioral response to competitive pressure, the dependent variable measures the difference between an individual's observed and predicted choice. The main explanatory variables will be the sources of competitive pressure previously described.[17]

4.2.1 Effort

Table 4: Deviation of effort from eq. prediction; testing hypothesis *Effort 1* in columns (1) and (2), and hypotheses *Effort 2* and *Effort 3* in columns (3) and (4)

	(1)	(2)	(3)	(4)
	L	H	L	H
	Dep. variable: $x_{i,V} - x^*_{i,V}$			
$V = 100$	-4.50^{***}	-0.41	-4.50^{***}	-0.41
	(1.49)	(1.86)	(1.49)	(1.86)
$V = 500$	-38.99^{***}	-23.32^{***}	-38.99^{***}	-23.32^{***}
	(3.68)	(2.98)	(3.68)	(2.98)
Number H types			7.72^{***}	1.75
			(2.19)	(1.90)
Constant	14.61^{***}	12.78^{***}	6.16^{***}	7.53
	(1.50)	(1.63)	(2.28)	(5.91)
Observations	3,024	3,312	3,024	3,312
Number of indiv.	126	138	126	138
Number of groups	54	54	54	54

Robust standard errors clustered on group level in parentheses; *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$. 'Number H types' runs from 0 to 4.

We begin by testing the first hypothesis regarding aggregate effort. To robustly test the hypothesis that effort is larger than predicted for all values of V, Table 4 reports the results of a random effects GLS regressions with standard errors clustered at the group

[17]We have conducted additional robustness checks on these results to examine ordering effects, the explanatory power of risk aversion, and learning effects, see Appendix A. These extensions generally provide a qualitatively similar view of the results though some additional nuances to the behavior do come through in these additional tests.

Figure 1: Av. effort depending on V

(a) H types (b) L types

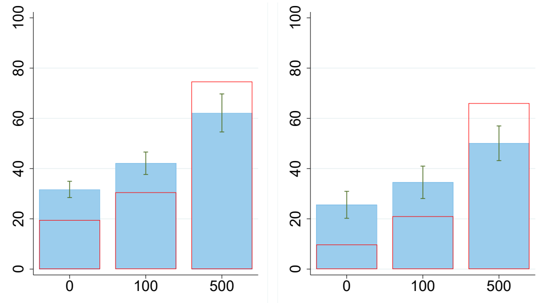
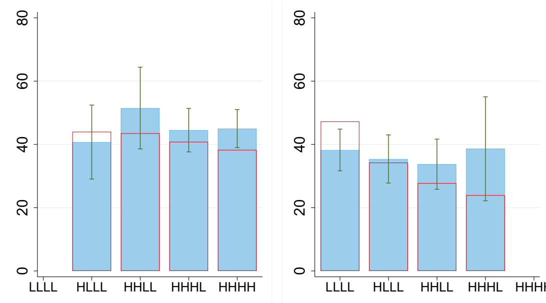


Figure 2: Av. effort dep. on # of H types

(a) H types (b) L types



Predicted effort levels are shown as empty boxes and average observed effort levels as filled boxes, with the error bars showing the 95% confidence interval. In Figure 1 (a) and (b), we treat each group in one block as one independent observation, i.e. we have one observation per block for each group. In Figure 2 (a) and (b), we treat each single group as one independent observation. For groups containing heterogeneous types, we have the same group once in the graphs for H types and once in the graphs for L types.

level. Columns (1) and (2) report results for the L and the H types respectively.[18] The dependent variable is the difference between an individual's observed effort choice and the theoretically predicted effort in a given period. Positive numbers imply effort is higher than predicted. We include data from all periods and include binary variables for $V = 100$ and $V = 500$ which implies the reference group is when $V = 0$.

The statistically significant and positive signs on the constant terms in columns (1) and (2) imply that observed effort is higher than the predicted value when $V = 0$ for both types. In the $V = 0$ case, effort above the level predicted suggests individuals may be resolving some of the collective action problems as effort above the self-interested level could be consistent with individuals working towards the good of their group. Post-estimation Wald tests show that effort is also above equilibrium for both types when $V = 100$ ($p < 0.01$ for both types), which could either be due to individuals valuing the good of their group or from competitive desire to obtain the prize. Effort is below equilibrium at $V = 500$ ($p < 0.01$ for both types) suggesting that at this level of incentives, subjects are unwilling to contribute effort even up to the point of individual self-interest. Consequently, $H1$ is supported for $V = 0$ and $V = 100$ for both types, but not for $V = 500$ for either type. This leads to our first result:

Result 1 (Effort 1) *$H1$ is partially rejected: For both types, observed effort is significantly higher than predicted when $V = 0$ and $V = 100$, while it is significantly lower than*

[18]See Tables 9, 10, and 11 in Appendix C for robustness checks controlling for order effects, risk aversion and learning, respectively.

predicted for $V = 500$.

Result 1 characterizes an average deviation of effort from the prediction, but does not uncover underlying behavioral responses to competition, which is our main interest. Specifically, looking only at the magnitude of the response ignores the behavioral responses to an increasingly competitive setting. To better understand these behavioral effects, we now test our second and third hypothesis regarding deviations from the predicted values for effort, as the competitiveness of the setting changes, i.e., as V changes, and as the number of H types changes. Panels (a) and (b) in Figure 1 show that the competitive pressure of the prize does not appear to have a significant impact on deviations in effort, as deviations do not appear to increase as V increases, which contradicts our second hypothesis. On the contrary, panels (a) and (b) in Figure 2 seem to show support for the third hypothesis as deviations from the predictions appear to be positive and increasing in the number of H types.

Columns (3) and (4) of Table 4 display the results of a random effects GLS regression, again for the L and the H types, respectively. The dependent variable is, again, the deviation in observed effort from equilibrium. The first explanatory variables are once again binary variables equal to one when $V = 100$ and zero otherwise, and similar for $V = 500$; these are meant to test our second hypothesis, *Effort 2*. The third explanatory variable accounts for the number of H types in the group (ranging from 0 to 4). It captures the responses to an increase in the competitiveness due to group composition, as outlined in our third hypothesis, *Effort 3*. Columns (3) and (4) display results for the L and the H types respectively.

The results indicate that moving from $V = 0$ to $V = 100$ leads to a reduction in deviations of effort for the L types, and does not change the deviations for the H types. When moving from $V = 0$ to $V = 500$, the results indicate that deviations in effort are even more negative, not positive. Likewise, when moving from $V = 100$ to $V = 500$ post-estimation Wald tests indicate lower levels of deviations ($p < 0.01$ for both types). These findings contradict our second hypothesis and lead to our second result:

Result 2 (Effort 2) *$H2$ is rejected: For both types, deviations of observed effort from the equilibrium prediction are not increasing in V.*

Columns (3) and (4) of Table 4 also indicate that the number of H types in a group positively influences the deviations in effort for the L types, but not for the H types. For the H types, the effect is no different than zero, which is in support of the standard theoretical prediction. This leads to our third result:

Result 3 (Effort 3) *$H3$ is partially rejected:*

(a) the deviation of the observed effort from the equilibrium prediction for H types, $x_H - x_H^*$, is not increasing in the number of H types in the group;

(b) the deviation of the observed effort from the equilibrium prediction for L types, $x_L - x_L^*$, is increasing in the number of H types in the group.

Overall, we find mixed support for the notion that an increase in the competitiveness of the setting leads to an increase in effort above the standard theoretical prediction. The one exception is the L types' response to an increase in the number of H types as statistically they appear to be increasing their effort relative to the equilibrium prediction as the number of H types increases. Examining panels (a) and (b) in Figure 2 makes it clear that this finding is due to predicted effort declining while observed effort remains about the same. This suggests that the L types may simply be nonresponsive to the configuration of their group as they supply on average the same amount of effort regardless of the group configuration. The overall conclusion from our examination of the effort decisions is that we do not find much support for a behavioral effect claimed by the competitive school of thought, which predicts that increasing competitive pressure either through increasing the size of the prize or through increasing the number of H types drives up chosen effort to some hypercompetitive level above the equilibrium prediction.

4.2.2 Help and sabotage

Having established how competitive pressure alters effort provision, we now analyze how these same pressures affect help and sabotage behavior; a test of our fourth, fifth and sixth hypothesis. We begin by testing our fourth hypothesis on aggregate help. Columns (1) and (2) of Table 5 display the results of random effects GLS regressions.[19] The dependent variable is the difference between actual and predicted help, where a negative deviation could imply less help than predicted (or more sabotage than predicted). Once again, we include two binary variables for $V = 100$ and $V = 500$ implying our reference category is $V = 0$.

When $V = 0$, deviations in help are negative, not positive, for both types, which may indicate that any sort of cooperative preferences are manifested through effort allocations (see *Result 1*). Furthermore, post-estimation Wald tests indicate that deviations in help are also negative when $V = 100$ ($p < 0.01$ for both types), but are not different from zero when $V = 500$ for the L types ($p = 0.587$), and are greater than zero for the H types

[19] See Tables 12, 13, and 14 in Appendix C for robustness checks controlling for order effects, risk aversion and learning, respectively.

Table 5: Deviations of help and sabotage from eq. predictions; testing hypothesis *Help 1* in column (1) and (2) and hypotheses *Help 2* and *Help 3* in columns (3) and (4)

	(1)	(2)	(3)	(4)
	L	H	L	H
	Dep. var.: $k_{ij,V} - k^*_{ij,V}$			
$V = 100$	-3.77^{***}	-3.69^{***}	-3.77^{***}	-3.69^{***}
	(0.87)	(0.94)	(0.87)	(0.94)
$V = 500$	4.65^{**}	8.23^{***}	4.65^{**}	8.23^{***}
	(2.05)	(1.57)	(2.05)	(1.57)
Number H types			-6.09^{***}	-3.37^{***}
			(2.25)	(0.87)
Constant	-3.28^{**}	-4.39^{***}	3.38^{**}	5.70^{**}
	(1.59)	(0.71)	(1.53)	(2.52)
Observations	3,024	3,312	3,024	3,312
Number of indiv.	126	138	126	138
Number of groups	54	54	54	54

Robust standard errors clustered on group level in parentheses; *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$. 'Number H types' runs from 0 to 4.

($p = 0.017$). Thus, in only one out of six instances do we find support for the fourth hypothesis, stated in our fourth result.

Result 4 (Help 1) *$H4$ is partially rejected: Observed help from H and L types is lower than predicted when $V = 0$ and $V = 100$; for $V = 500$, it is equal to the predicted value for the L types and greater than the predicted value for the H types.*

As with effort, the magnitudes of deviations for each V serve as useful baselines, but do not address responses to the competitiveness of the setting. Figure 3 shows how helping behavior changes as V increases and as the number of H types increase. Panels (a) and (b) indicate some support for hypothesis *Help 2*, as the deviation from the prediction decreases when moving from $V = 0$ to $V = 100$, but not when going from $V = 100$ to $V = 500$ (or from $V = 0$ to $V = 500$). Panels (a) and (b) provide an indication that help may be dropping relative to the predicted level as the number of H types in the group increases, as predicted by hypothesis *Help 3*.

Columns (3) and (4) of Table 5 present the results of random effects GLS regressions for the L and H types respectively, in which the dependent variable is the deviation in help/sabotage from its equilibrium level and the independent variables are the same ones

Figure 3: Av. help depending on V

(a) H types (b) L types

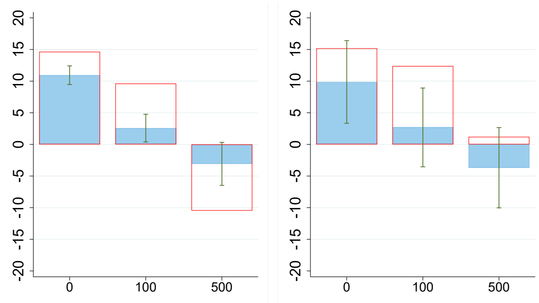
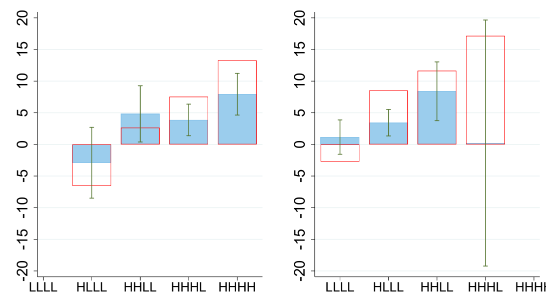


Figure 4: Av. help dep. on # of H types

(a) H types (b) L types



Predicted help levels are shown as empty boxes and average observed help levels as filled boxes, with the error bars showing the 95% confidence interval. In Figure 3 (a) and (b), we treat each group in one block as one independent observation, i.e. we have one observation per block for each group. In Figure 4 (a) and (b), we treat each single group as one independent observation. For groups containing heterogeneous types, we have the same group once in the graphs for H types and once in the graphs for L types.

as used in Table 4. Once again, of particular interest are the variables capturing changes in the prize structure and the number of H types in a group.

The results indicate that when the prize increases from $V = 0$ to $V = 100$, deviations in help for both types decrease, as predicted by *Help 2*. However, when the prize increases from $V = 0$ to $V = 500$, deviations in help increase for both types. Likewise, post-estimation tests indicate that when the prize increases from $V = 100$ to $V = 500$, deviations increase ($p < 0.1$ for both types), as summarized in our next result.

Result 5 (Help 2) *$H5$ is partly rejected: For both types, the difference between observed and predicted help is decreasing in V when going from $V = 0$ to $V = 100$, while it is increasing in V when going from $V = 100$ to $V = 500$ and when going from $V = 0$ to $V = 500$.*

Note that this result is driven by a decrease in help when $V = 100$, *and* less sabotage in the case of $V = 500$. That is, a higher prize leads to lower levels of help on average when help is predicted to be positive, but when the prediction is to sabotage, individuals refrain from sabotaging more than predicted. This is especially relevant in the $V = 500$ case for the H types where the highest level of sabotage is predicted, cf. Figure 3 (a).

That is not to say that there is no response to increases in competitiveness; We find that as the number of H types increases, the deviations from the prediction of help decrease. This is true for both types and leads to our next result.

Result 6 (Help 3) *H6 is not rejected: For both types, the difference between the observed and predicted help is decreasing in the number of H types in the group.*

Overall, the results on *Help 2* and *Help 3* mostly support the idea that competitiveness leads to a behavioral "anti-help" mindset. However, we also observe the opposite effect – aversion to sabotage – when sabotage is predicted. High levels of sabotage are predicted, on average, for the H types when $V = 500$, and in the $HLLL$ setting for the H types and the $LLLL$ setting for the L types. In all of these cases individuals are largely refraining from sabotaging each other, cf. Figures 3 and 4. Thus, even though competitiveness may lead to lower levels of help, this aversion implies that the main drawback from competitiveness, i.e., sabotage, may be more limited than predicted thus blunting the worst of the impacts predicted by the noncompetitive school of thought. These results suggest that increased competitiveness may not lead to the "toxic" work environments that are often predicted by the detractors of high-powered incentives.

4.2.3 Total output

The established mutually offsetting results on deviations in effort and help could lead to higher or lower total group output, since it is the sum of all effort and mutual help/sabotage within a group. Our hypotheses predict total output to be higher than predicted (*Total Output 1*) and for competitive pressures to lead to further increases in total output (*Total Output 2* and *Total Output 3*). Figure 5 compares the predicted to the average group output by group composition and size of the prize. We observe that total output is more often equal to or lower than the predicted value.

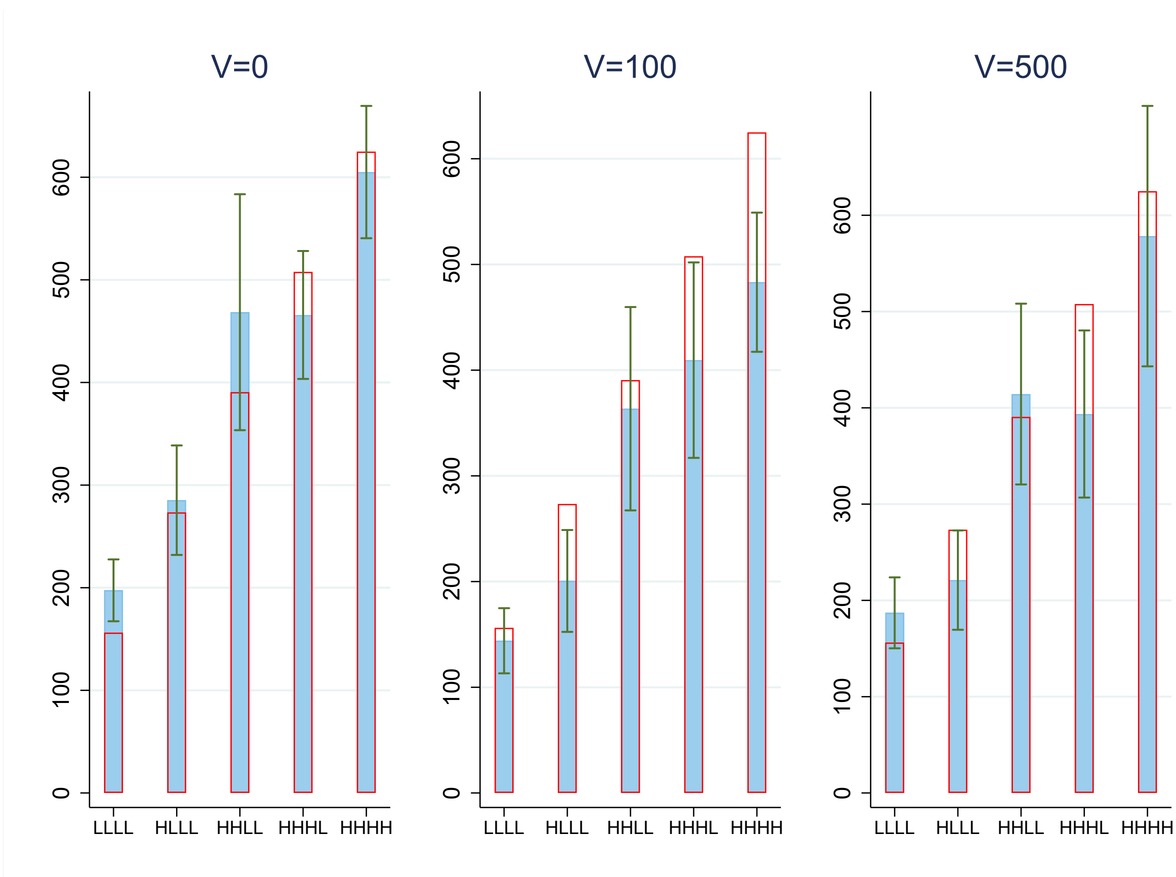
Column (1) of Table 6 displays the results of a random effects GLS regression, where the dependent variable is the difference in actual versus predicted total group output in every round.[20] The two independent variables are binary variables for $V = 100$ and $V = 500$. There are 66 groups, and output was observed 24 times per group, which leads to a total of 1,584 observations.

The statistically insignificant estimate of the constant term implies that when $V = 0$, output is no different than predicted. Post-estimation Wald tests indicate that total output is lower than predicted when $V = 100$ ($p < 0.01$) and when $V = 500$ ($p = 0.041$). None of these findings are in line with our first prediction on total output.

Result 7 (Total Output 1) *H7 is rejected: Total output is not greater than the equilibrium predictions for any value of V.*

[20] Tables 15 and 16 in Appendix C control for order effects and learning.

Figure 5: Average group output



Predicted levels are shown as empty boxes and average observed levels as filled boxes, with the error bars representing the 95% confidence intervals. We treat each group in one block as one independent observation, i.e., we have one observation per block in each group.

Table 6: Deviations from group output; testing hypotheses *Total Output 1*, *Total Output 2*, and *Total Output 3*

	(1)	(2)
	H7	H8 & H9
	Dep. var.: $Y - Y^*$	
$V = 100$	-81.71^{***}	-81.71^{***}
	(13.38)	(13.39)
$V = 500$	-47.96^{***}	-47.96^{***}
	(17.84)	(17.85)
Number H types		-24.01^{***}
		(7.86)
Constant	8.72	58.93^{***}
	(15.04)	(18.14)
Observations	1,584	1,584
Number of groups	66	66

Robust standard errors clustered on group level in parentheses; *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$. 'Number H types' runs from 0 to 4.

Although the aggregate output somewhat informs on the total effect of the reward scheme, it does not capture the behavioral effects stemming from the change in competitiveness of the setting. However, no general trends appear to be present in Figure 5, which suggests that aggregate output may not be as influenced by these behavioral effects. For a formal test of Hypotheses 8 and 9, column (2) of Table 6 includes a variable accounting for the prize structure and the number of H types in the group.

When testing the eighth hypothesis (*Total Output 2*), the negative and significant estimates of the coefficients on $V = 100$ and $V = 500$ imply that deviations from predictions in both of these settings are below the deviations when $V = 0$ (Table 6, column (2)). A post-estimation Wald test confirms that the deviation is smaller in magnitude when $V = 500$ as compared to when $V = 100$ ($p = 0.018$). Thus, the hypothesis is supported only partially, and overall its support is not very convincing given the large negative deviation still present at $V = 500$ relative to $V = 0$.

Result 8 (Total Output 2) *H8 is partially rejected: The difference between observed and predicted total output is initially decreasing when going from $V = 0$ to $V = 100$, but then it is increasing when going from $V = 100$ to $V = 500$; still, the difference between observed and predicted total output is decreasing when going from $V = 0$ to $V = 500$.*

Turning to the final hypothesis, the significant and negative estimate of the coefficient on the number of H types implies that, as the competitiveness of the setting increases due to changes in the composition of the group, total deviations in output decrease. This is against our hypothesis derived from the competitive school of thought (and hence would be in line with the noncompetitive school of thought).

Result 9 (Total Output 3) *H9 is rejected: The difference between observed total output and the equilibrium prediction is decreasing in the number of H types in the group.*

Our results show that total output declines when shifting from $V = 0$ to $V = 100$ and then increases when shifting from $V = 100$ to $V = 500$; however, the output at $V = 500$ is still lower than at $V = 0$. What this shows is that, given our production function, the trade-off between effort and help as the prize increases does not combine into higher overall output. In the case of the shift from $V = 0$ to $V = 100$, we find that effort increases above the prediction but helping behavior diminishes enough to counter that positive outcome. In the shift from $V = 100$ to $V = 500$, we find the opposite to be the case as effort drops below the equilibrium prediction but we get less sabotage and, therefore, more help than predicted. This balances out to less of a hit on overall output but it does not lead to a better outcome than at $V = 0$. For a production function that relied less on helping behavior, it is possible that a different result could have emerged for the $V = 100$ case and for a production function which relied less on individual effort, the $V = 500$ case could have yielded a different outcome. The main take-away from our current analysis is that increasing the value of the prize employees compete for seems to have some unexpected effects as the prize moves from zero to moderate and then high levels. Understanding these effects would be important when a firm is choosing how much to reward individual effort in a group production setting.[21]

4.3 Organization of employees into teams

In addition to incentive mechanisms, a manager also has a choice on how to organize her fixed set of employees into teams. Since we studied different team structures, we can also examine how the manager's decision to allocate individuals into teams affects output. In this section, we empirically analyze how organizing a fixed group of employees into different team compositions may affect output for different values of V. The hypothetical

[21] An additional aspect one might be concerned about with regard to the incentives is the inequality induced, given that the H types are predicted to win much more often than the L types. We actually find that the L types win more than expected, indicating that while they still end up receiving the bonus payment much less often than the H types, the induced inequality is not as bad as predicted.

exercise here is to consider how a manager should allocate individuals to teams when she needs two teams of four members each and she has eight employees, four high ability and four low, to allocate between both teams. This implies that the manager can organize her employees into two Homogeneous teams (*HHHH* and *LLLL*), two Balanced teams (*HHLL* and *HHLL*), or two Asymmetric teams (*HHHL* and *HLLL*).[22] As stated in Section 3.3, our model predicts that total output of each team composition is constant, for any value of V. Yet, if behavioral effects are present, team structures may matter.

Figure 6 shows the predicted and observed outcomes, and Table 7 displays the results of random effects GLS regressions. The dependent variable is the deviation in total group output (the same variable is used in Table 6).[23]. The independent variables are binary variables equal to one if the group is classified as Homogeneous (*HHHH* and *LLLL*), and equal to one if the group is classified as Asymmetric (*HHHL* and *HLLL*), which implies the reference group is the Balanced group (*HHLL* and *HHLL*). The first column displays results for $V = 0$, the second for $V = 100$, and the third for $V = 500$.[24]

We find that in the Balanced case, the observed total output does not differ from the predicted output for any value of V. The coefficient on *Homogeneous* tests whether output is different between that case and Balanced, and again we find no difference. We do, however, find that for the Asymmetric case, output is significantly less than in the Balanced case in the $V = 0$ and $V = 500$ cases. We can also use linear combinations of the constant and the coefficients for those two conditions to test whether the output in Homogeneous and Asymmetric groups is different from the predicted level. We find that for Homogeneous groups, when $V = 100$, output is less than the predicted level ($p < 0.01$), but is not significantly different from the predicted level for $V = 0$ ($p = 0.52$) and $V = 500$ ($p = 0.80$). For Asymmetric groups, the observed output is not significantly different from the predicted level for $V = 0$ ($p = 0.31$); however, output is less than predicted for $V = 100$ ($p < 0.01$) and $V = 500$ ($p < 0.01$). This leads to the following result.

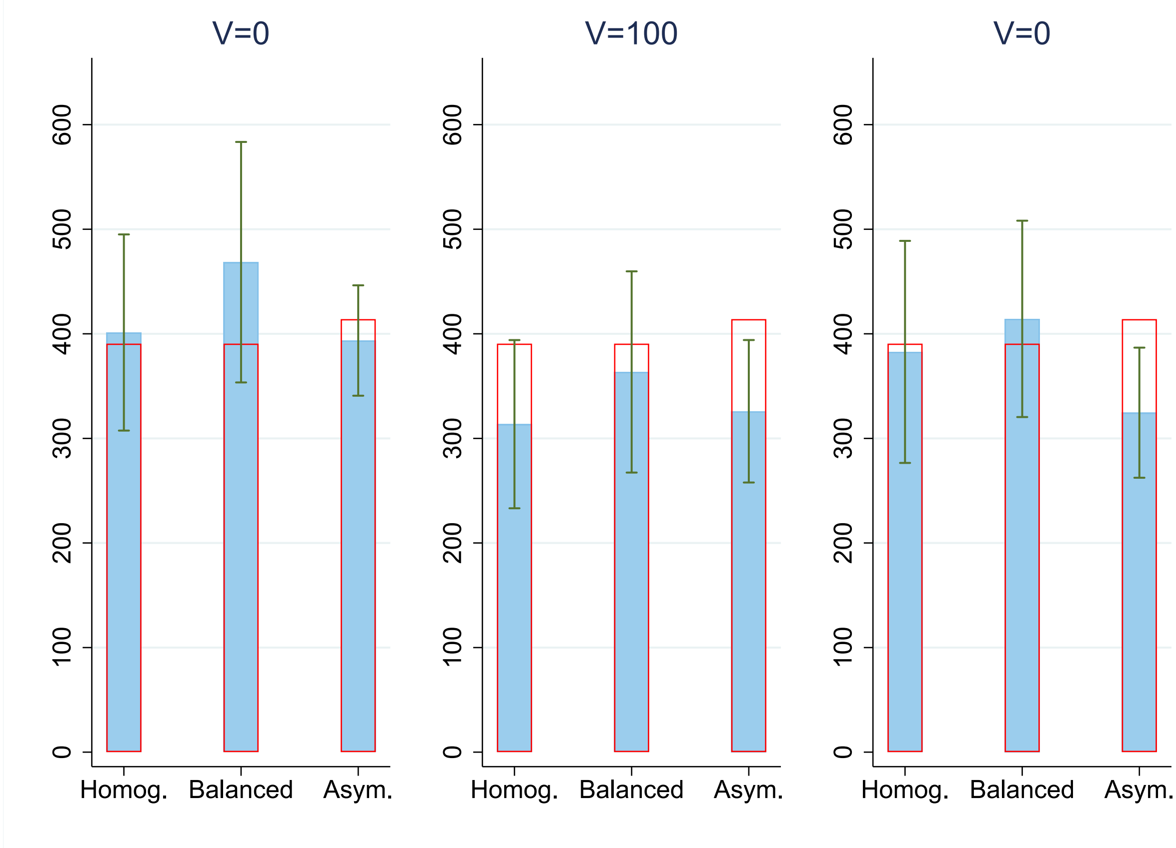
Result 10 (Group composition) *Homogeneous and Asymmetric group compositions*

[22]If the assumption that a manager has equal H and L types is relaxed, the only aspect that changes is how teams are formed. Importantly, if eight employees are present, a manager must have at least two of each types to have a choice on group composition. For instance, if a manager has two L types and six H types, the manager can either form one team of *HHHH* and one team of *HHLL* or two teams of *HHHL*. The empirical results based on the assumption of an equal number of each type are already informative on general behavior and thus we leave out these additional analysis for succinctness.

[23]We use the deviation between observed and predicted output here despite the fact that total predicted output is common across all group configurations, to filter out the level effect, and to be able to use the standard significance test on the constant term to test whether observed output deviates from predicted output in the Balanced case.

[24]Again, Tables 17 and 18 in Appendix C control for order effects and learning.

Figure 6: Average group output for Balanced, Asymmetric and Homogeneous groups



Predicted levels are shown as empty boxes and average observed levels as filled boxes, with the error bars representing the 95% confidence intervals. We treat each group in one block as one independent observation, i.e., we have one observation per block in each group.

Table 7: Output of each group if the manager uses Homogeneous groups (*HHHH* and *LLLL*), Balanced heterogeneous groups (*HHLL* and *HHLL*) or Asymmetric heterogeneous groups (*HHHL* and *HLLL*).

	(1)	(2)	(3)
	$V = 0$	$V = 100$	$V = 500$
	Dep. var.: $Y - Y^*$		
Homogeneous	-67.22	-49.92	-31.57
	(53.23)	(47.09)	(52.07)
Asymmetric	-98.32*	-61.03	-113.14**
	(54.44)	(50.27)	(48.99)
Constant	77.85	-27.10	23.66
	(50.49)	(42.25)	(41.23)
Observations	528	528	528
Number of groups	66	66	66

Robust standard errors clustered on group level in parentheses; *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

can lead to output lower than theoretically predicted while the Balanced group composition leads to output that is not significantly different from the predicted level.

These results suggest that the Balanced group configuration is at least weakly better than the other two. While the Homogeneous case is generally not significantly different from the Balanced case, it is somewhat lower on average and low enough that it can be distinguished from the theoretical prediction in the $V = 100$ case. The Asymmetric case generally leads to lower output than in the Balanced case and output that is significantly less than the predicted level. The indication is that balanced teams may lead to overall better in-group interactions in terms of allowing incentives to encourage reasonable effort while not impairing helping behavior as much as the other configurations.[25]

5 Conclusions

In this paper, our goal was to empirically examine if competitive pressures in a group production setting led to consistent behavioral responses not predicted by standard theory. Our hypotheses were built around two schools of thought: the "competitive" school,

[25]It may also be of interest whether some group compositions produce more volatility in output than others. We examined the variance of aggregate output across treatments and found that output variance is somewhat lower in Balanced groups as compared to Homogeneous and Asymmetric, but the differences are not statistically significant.

which argues competitive pressures lead to additional increases in effort, and the "non-competitive" school, which argues that competitive pressures lead to toxic environments, (dramatically) lower levels of help, and higher levels of sabotage.

In general, we find fairly mixed support for both schools of thought. While moderate incentives do seem to stimulate effort levels above theoretical predictions, high incentives have the opposite impact. In regard to helping behavior, we find that even with no competitive incentives, our subjects provided less help than predicted, but as incentives increase, we do not find that individuals resort to sabotage as much as expected. These results help us to understand why there is still an active debate as to whether high competitive incentives are good or bad for team production. Our results do not suggest that a general conclusion to this question will exist. That makes it very important to understand how different production technologies in regard to the importance of teamwork in a firm will lead to different conclusions on this question.

We get a similarly mixed set of results regarding how changing the composition of a team – in terms of having more or less high ability types in there – affects behavior. We generally find that neither high nor low ability types change their effort much, as we change the composition of the team. This suggests that people are less sensitive to team composition than expected. In regard to helping behavior, the prediction in our environment is that team members should take advantage of the efficiency enhancing nature of helping high productivity team members, and so, as their number increases, helping behavior should rise. We find that it generally does, though not as much as predicted. There may be a behavioral effect here of the increased competition from too many high productivity types harming helping behavior.

By combining the insights from these two sets of results, we find that teams that are a balanced mix of high and low ability workers may perform best and that moderate rather than high powered incentives may be better if a firm wants to adopt an incentive mechanism for individual effort. We do not, however, find results suggesting that competitive incentives are all that valuable to team production. This might be considered a counter-intuitive result and so it is important to understand the reason for this finding. In most similar public goods style frameworks, the default prediction for individual effort is 0 due to the production technology. Our environment, however, does involve providing incentives for individuals to provide effort, even with $V = 0$, by paying them based on the overall production level of the team such that positive individual effort is the equilibrium prediction. Thus, our $V = 0$ case should not be interpreted as a case with "no incentives". It is, however, a case where one might be concerned about the possibility of individuals free-riding on the effort of others and providing less help and or effort than is ideal. As expected, we find that individuals provide effort in excess of the Nash prediction, so people do not generally fully free ride on the effort of others, though they do seem to help less than predicted. On balance, this leads to total output being in line with the prediction. As we ramp up competitive incentives, we initially get an increase in individual effort which comes at too great a cost in helping behavior. As competitive incentives rise further, the lift to individual effort trails off, and while help doesn't decline as much as expected, it certainly does not offset the lack of lift to the effort levels. While we do not observe the worst predictions from the "noncompetitive" viewpoint, we also do not observe the best predictions from the "competitive" viewpoint. On balance, competitive incentives neither excessively hurt nor help. The group incentives appear sufficient for encouraging team production and we find little need to augment those incentives with the competitive incentives. As we note though, different production functions which give rise to different values of help and individual effort could yield different conclusions. Thus our results regarding the underlying effects on effort and helping behavior should be considered in light of any alternative weighting for those elements of team production.

References

Alchian, Armen A., and Harold Demsetz. 1972. "Production, information costs, and economic organization." *The American Economic Review*, 62(5): 777–795.

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

Brandts, Jordi, David J Cooper, Enrique Fatas, and Shi Qi. 2016. "Stand by Me - Experiments on Help and Commitment in Coordination Games." *Management Science*, (forthcoming): 1–21.

Bull, Clive, Andrew Schotter, and Keith Weigelt. 1987. "Tournaments and Piece Rates: An Experimental Study." *Journal of Political Economy*, 95(1): 1–33.

Buser, Thomas, and Anna Dreber. 2016. "The flipside of comparative payment schemes." *Management Science*, 62(9): 2626–2638.

Crosetto, Paolo, and Antonio Filippin. 2013. "The "bomb" risk elicitation task." *Journal of Risk and Uncertainty*, 47(1): 31–65.

Danilov, Anastasia, Christine Harbring, and Bernd Irlenbusch. 2019. "Helping under a combination of team and tournament incentives." *Journal of Economic Behavior & Organization*, 162: 120–135.

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.

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.

Garvey, Gerald T., and Peter L. Swan. 1992. "Managerial objectives, capital structure, and the provision of worker incentives." *Journal of Labor Economics*, 10(4): 357–379.

Gürtler, Oliver, and Johannes Münster. 2013. "Rational self-sabotage." *Mathematical Social Sciences*, 65(1): 1–4.

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

Heap, Shaun P Hargreaves, Abhijit Ramalingam, and Brock V Stoddard. 2016. "Endowment inequality in public goods games: A re-examination." *Economics Letters*, 146: 4–7.

Kräkel, Matthias. 2005. "Helping and sabotaging in tournaments." *International Game Theory Review*, 7(02): 211–228.

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.

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." *European Journal of Political Economy.*

Ledyard, John O. 1995. "Public Goods: A Survey of Experimental Research." 111–194.

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

Tullock, Gordon. 1980. "Toward a Theory of the Rent Seeking Society." *in book*, 97–112.

A Robustness checks

In the following, we will summarize our main findings from the robustness checks.[26] In separate analysis, we have controlled for i) order effects (Robustness check 1), i.e., we have run separate regressions for when the order of V was 0, 100, 500 (order 1), and when the order of V was 0, 500, 100 (order 2); ii) risk aversion (Robustness check 2), including an item that runs from -100 to 0, with higher numbers indicating higher risk aversion, and iii) learning (Robustness check 3), i.e., we have run separate regressions for periods 1-4, 9-12, 17-20 (first half of periods) and regressions for periods 5-8, 13-16, 21-24 (second half of periods). Results are displayed in Tables 9 - 18 in Appendix C.

While controlling for order and learning mainly leaves qualitative results unchanged[27], controlling for risk aversion alters some of our result, which suggests that these preferences are also important in our setting and could be potentially used as an explanation of some of our results. When controlling for risk aversion, for some levels of risk aversion, effort is *not different* than the predicted value when $V = 0$ and $V = 100$ (and lower for some levels of risk aversion), instead of being higher, which implies that risk preferences may explain the observed overcontribution outlined in *Result 1*. Also, when controlling for risk aversion, for some levels of risk aversion, help/sabotage is *not different* than the predicted value when $V = 0$, instead of being lower, and it is *higher* than the predicted value for the L types for $V = 500$, instead of being equal (for some values of risk aversion, interpretation changes again). It seems that, similar to the effort choices, risk preferences play an important role also in help/sabotage decisions, and may explain the observed undercontribution outlined in *Result 4*.

Summary:

Result 1 Additional effects when including risk aversion as a control; no changes when controlling for order / learning.

Result 2 No changes.

Result 3 Small changes to order. No changes when including risk aversion and learning.

Result 4 Additional effects when including risk aversion as a control; no changes when controlling for order / learning.

[26]Given the length, we have placed the extensive description of these additional results in an extra subsection, Appendix B.

[27]The only exception is Result 3; here, the deviation of observed effort from the equilibrium prediction for the H types is *increasing* in the number of H types in the group in order 2.

Result 5 No major changes.

Result 6 No changes.

Result 7 No changes.

Result 8 No major changes.

Result 9 No changes.

B Details on the Robusness Checks

In the following paragraphs, we restated the respective result, and the corresponding results from the robustness checks. Where the robustness check regression differs from the original regression, we have noticed this in red! "–" stands for 'qualitatively not different for the stated result'.

Result 1: [Effort 1]H_1 is partly confirmed: For both types, observed effort is higher than predicted when $V = 0$ and $V = 100$, while it is lower than predicted for $V = 500$.

RC1 (Order, Table 9, columns (1) & (2) and (5) & (6)) H_1 is partly confirmed: For both types, observed effort is higher than predicted when $V = 0$ and $V = 100$, while for $V = 500$ it is lower than predicted for the L types and for the H types, even though for the H types in order 1 the difference is not significant.

RC2 (Risk Av., Table 10, columns (1) & (2)) H_1 is not confirmed: For both types, observed effort is equal to the predicted value when $V = 0$ and $V = 100$, for some levels of risk aversion, and lower for other levels of risk aversion. Observed effort is lower than predicted for $V = 500$ for all levels of risk aversion.

RC3 (First half / second half, Table 11, columns (1) & (2) and (5) & (6)) –

Result 2: [Effort 2]H_2 is not confirmed: For both types, deviations of observed effort from the equilibrium prediction are not increasing in V.

RC1 (Order, Table 9, columns (3) & (4) and (7) & (8)) –

RC2 (Risk Av., Table 10, columns (3) & (4)) –

RC3 (First half / second half, , Table 11, columns (3) & (4) and (7) & (8)) –

Result 3: [Effort 3]H_3 is partly confirmed: (a) the deviation of the observed effort from the equilibrium prediction for H types, $x_H - x_H^*$, is not increasing in the number of H types in the group; (b) the deviation of the observed effort from the equilibrium prediction for L types, $x_L - x_L^*$, is increasing in the number of H types in the group.

RC1 (Order, Table 9, columns (3) & (4) and (7) & (8)) H_3 is partly confirmed: (a) the deviation of the observed effort from the equilibrium prediction for H types, $x_H - x_H^*$, is not increasing in the number of H types in the group in Order 1, while it is *increasing* in Order 2; (b) –

RC2 (Risk Av., Table 10, columns (3) & (4)) H_3 is partly confirmed: (a) – (b) the deviation of the observed effort from the equilibrium prediction for L types, $x_L - x_L^*$, is increasing in the number of H types in the group for some levels of risk aversion.

RC3 (First half / second half, Table 11, columns (3) & (4) and (7) & (8)) –

Result 4: [Help 1]H_4 is partially confirmed: Observed help from H and L types is lower than predicted when $V = 0$ and $V = 100$; for $V = 500$, it is equal to the predicted value for the L types and greater than the predicted value for the H types.

RC1 (Order, Table 12, columns (1) & (2) and (5) & (6)) H_4 is partially confirmed: Observed help from H and L types is lower than predicted when $V = 0$ and $V = 100$, even though for $V = 0$ the differences are not significant; for $V = 500$, it is equal to the predicted value for the L types and greater than the predicted value for the H types.

RC2 (Risk Av., Table 13, columns (1) & (2)) H_4 is partially confirmed: For some levels of risk aversion, observed help from L types is equal to the predicted value when $V = 0$ and $V = 100$; for others, it is lower. For the H types, for some levels of risk aversion, observed help is lower than predicted when $V = 0$ and $V = 100$, even though for $V = 0$ the difference is not significant. For $V = 500$, it is greater than the predicted value for the L types and greater than the predicted value for the H types, for some levels of risk aversion.

RC3 (First half / second half, Table 14, columns (1) & (2) and (5) & (6)) H_4 is partially confirmed: Observed help from H and L types is lower than predicted when $V = 0$ and $V = 100$; for $V = 500$, it is equal to the predicted value for the L types and greater than the predicted value for the H types, even though for the H types in the second half of the periods the difference is not significant (coefficient $= 2.28$, $p = 0.20$).

Result 5: [Help 2]H_5 is partly confirmed: For both types, the difference between observed and predicted help is decreasing in V when going from $V = 0$ to $V = 100$, while it is increasing in V when going from $V = 100$ to $V = 500$ and increasing when going from $V = 0$ to $V = 500$.

RC1 (Order, Table 12, columns (3) & (4) and (7) & (8)) H_5 is partly confirmed: For both types, the difference between observed and predicted help is decreasing in V when going from $V = 0$ to $V = 100$, while it is increasing in V when going from $V = 100$ to $V = 500$ and increasing when going from $V = 0$ to $V = 500$, even though the latter difference is not significant for the L types in order 1.

RC2 (Risk Av., Table 13, columns (3) & (4)) H_5 is partly confirmed: For both types, the difference between observed and predicted help is decreasing in V when going from $V = 0$ to $V = 100$, while it is increasing in V when going from $V = 100$ to $V = 500$ and increasing when going from $V = 0$ to $V = 500$, for some levels of risk aversion.

RC3 (First half / second half, Table 14, columns (3) & (4) and (7) & (8)) –

Result 6: [Help 3]H_6 is confirmed: For both types, the difference between the observed and predicted help is decreasing in the number of H types in the group.

RC1 (Order, Table 12, columns (3) & (4) and (7) & (8)) –

RC2 (Risk Av., Table 13, columns (3) & (4)) –

RC3 (First half / second half, Table 14, columns (3) & (4) and (7) & (8)) –

Result 7: [Total Output 1]H_7 is not confirmed: Total output is not greater than the equilibrium predictions for any value of V.

RC1 (Order, Table 15, columns (1) and (3)) –

RC3 (First half / second half, Table 16, columns (1) and (3)) –

Result 8: [Total Output 2]H_8 is partially confirmed: The difference between observed and predicted total output is initially decreasing when going from $V = 0$ to $V = 100$, but then it is increasing when going from $V = 100$ to $V = 500$; still, the difference between observed and predicted total output is decreasing when going from $V = 0$ to $V = 500$.

RC1 (Order, Table 15, columns (2) and (4)) H_8 is partially confirmed: The difference between observed and predicted total output is initially decreasing when going from $V = 0$ to $V = 100$, but then it is increasing when going from $V = 100$ to $V = 500$, even though the difference is not significant for order 1 (difference = 3.94, Wald test, $p = 0.84$); still, the difference between observed and predicted total output is decreasing when going from $V = 0$ to $V = 500$, even though, again, the difference is not significant for order 1.

RC3 (First half / second half, Table 16, columns (2) and (4)) H_8 is partially confirmed: The difference between observed and predicted total output is initially decreasing when going from $V = 0$ to $V = 100$, but then it is increasing when going from $V = 100$ to $V = 500$; still, the difference between observed and predicted total output is decreasing when going from $V = 0$ to $V = 500$, even though the difference is not significant for order 1.

Result 9: [Total Output 3]H_9 is not confirmed: The difference between observed total output and the equilibrium prediction is decreasing in the number of H types in the group.

RC1 (Order, Table 15, columns (2) and (4)) –

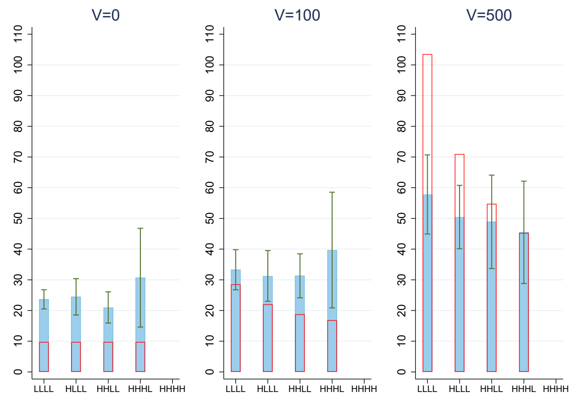
RC3 (First half / second half, Table 16, columns (2) and (4)) –

Group composition: Table 17: signs stay as they were; for V = 0 and Order 2, group output is significantly larger than predicted in the symmetric heterogeneous groups (is in line with our result anyways).

Group composition: Table 18: signs stay as they were; for V = 0 and 'second half periods', our result anyways).

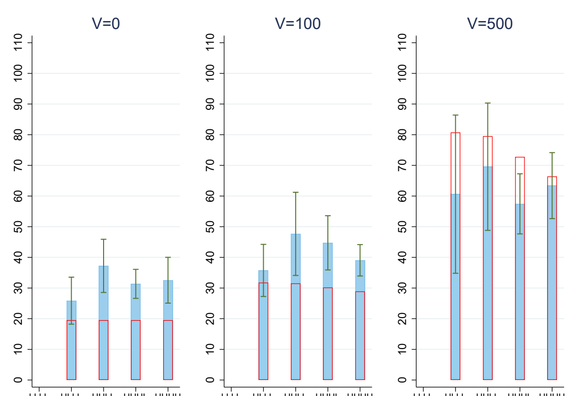
C More Figures and Results

Figure 7: Average effort, L-types



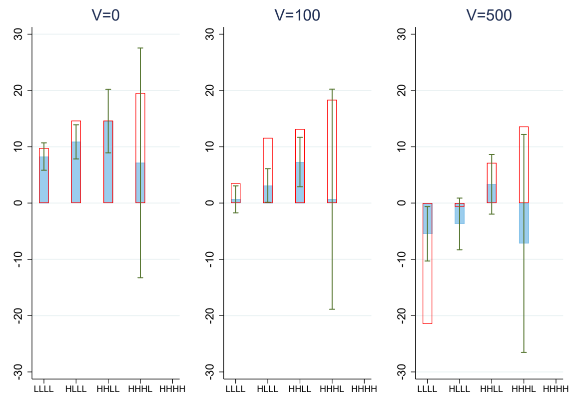
Predicted levels are shown as empty boxes and actual observed levels as filled boxes, with the standard error bars of the 95% confidence interval; we treat each group in one block as one independent observation, i.e. we have per group 1 observation per block – if groups are of heterogeneous types, we have the same group once in the graphs for H-types and once in the graphs for the L-types.

Figure 8: Average effort, H-types



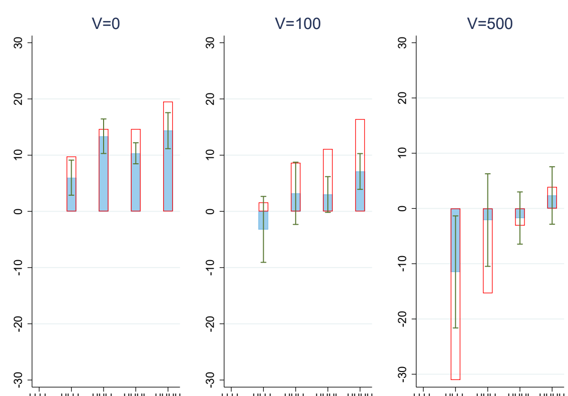
Predicted levels are shown as empty boxes and actual observed levels as filled boxes, with the standard error bars of the 95% confidence interval; we treat each group in one block as one independent observation, i.e. we have per group 1 observation per block – if groups are of heterogeneous types, we have the same group once in the graphs for H-types and once in the graphs for the L-types.

Figure 9: Average help, L-types



Predicted levels are shown as empty boxes and actual observed levels as filled boxes, with the standard error bars of the 95% confidence interval; we treat each group in one block as one independent observation, i.e. we have per group 1 observation per block – if groups are of heterogeneous types, we have the same group once in the graphs for H-types and once in the graphs for the L-types.

Figure 10: Average help, H-types



Predicted levels are shown as empty boxes and actual observed levels as filled boxes, with the standard error bars of the 95% confidence interval; we treat each group in one block as one independent observation, i.e. we have per group 1 observation per block – if groups are of heterogeneous types, we have the same group once in the graphs for H-types and once in the graphs for the L-types.

Table 8: Overview of observed averages and clustered standard errors

	LLLL	*HLLL*	*HHLL*	*HHHL*	HHHH
			$\boldsymbol{V} = 0$		
x_L	23.64	24.46	21.01	30.69	
	(1.42)	(2.69)	(2.31)	(7.63)	
x_H		25.88	37.23	31.36	32.54
		(3.48)	(3.94)	(2.24)	(3.39)
k_{LL}	8.26	8.93	9.28		
	(1.11)	(1.10)	(1.72)		
k_{LH}		14.77	17.20	7.14	
		(2.29)	(3.53)	(9.67)	
k_{HL}		6.00	9.94	7.34	
		(1.42)	(1.22)	(0.67)	
k_{HH}			20.21	11.83	14.36
			(3.63)	(1.25)	(1.46)
y_L	49.36	48.31	50.24	53.18	
	(3.42)	(4.81)	(4.48)	(7.54)	
y_H		140.38	184.00	137.54	151.26
		(13.10)	(25.58)	(11.08)	(7.32)
Y	197.46	285.30	468.48	465.79	605.04
	(13.70)	(24.25)	(52.24)	(29.54)	(29.29)
			$\boldsymbol{V} = 100$		
x_L	33.28	31.26	31.31	39.69	
	(2.97)	(3.75)	(3.26)	(8.93)	
x_H		35.74	47.67	44.74	39.05
		(3.86)	(6.16)	(4.19)	(2.32)
k_{LL}	0.66	2.54	4.65		
	(1.09)	(1.28)	(1.11)		
k_{LH}		4.29	8.61	0.68	
		(1.79)	(2.61)	(9.27)	
k_{HL}		-3.20	2.32	1.91	
		(2.66)	(2.25)	(1.23)	
k_{HH}			5.00	3.55	7.10
			(3.65)	(1.71)	(1.45)
y_L	36.00	34.44	41.79	48.58	
	(3.50)	(5.08)	(5.42)	(10.23)	
y_H		97.25	139.98	120.31	120.80
		(11.67)	(21.09)	(13.95)	(7.48)
Y	144.00	200.58	363.53	409.50	483.21
	(14.00)	(21.89)	(43.70)	(43.83)	(29.91)
			$\boldsymbol{V} = 500$		
x_L	57.81	50.44	48.89	45.47	
	(5.85)	(4.69)	(6.91)	(7.90)	
x_H		60.60	69.57	57.44	63.40
		(11.72)	(9.43)	(4.64)	(4.89)
k_{LL}	-5.46	-3.80	0.81		
	(2.20)	(2.04)	(1.07)		
k_{LH}		-3.51	4.59	-7.17	
		(2.39)	(3.30)	(9.18)	
k_{HL}		-11.48	-1.94	-2.00	
		(4.61)	(3.31)	(1.96)	
k_{HH}			-2.42	-1.59	2.35
			(5.27)	(2.42)	(2.36)
y_L	46.77	39.83	51.39	47.64	
	(4.18)	(4.68)	(8.56)	(8.73)	
y_H		101.58	155.76	115.31	144.58
		(21.04)	(22.77)	(13.48)	(15.36)
Y	187.09	221.08	414.29	393.58	578.33
	(16.74)	(23.44)	(42.66)	(41.13)	(61.45)
numb. indiv.	48	48	48	72	48
numb. groups	12	12	12	18	12
numb. rounds	8	8	8	8	8

Stand. errors are clustered on the group level and displayed in parenthesis.

Table 9: Effort – Robustness check 1: controlling for order effects

	(1) L	(2) H	(3) L	(4) H	(5) L	(6) H	(7) L	(8) H
	Order 1				Order 2			
	Dep. variable: $x_{i,V} - x^*_{i,V}$							
V = 100	-3.08^{*}	3.13	-3.08^{*}	3.13	-6.07^{**}	-5.02^{*}	-6.07^{**}	-5.02^{*}
	(1.65)	(2.11)	(1.65)	(2.11)	(2.57)	(3.04)	(2.57)	(3.04)
V = 500	-38.83^{***}	-19.20^{***}	-38.83^{***}	-19.20^{***}	-39.15^{***}	-28.68^{***}	-39.15^{***}	-28.68^{***}
	(4.81)	(4.22)	(4.81)	(4.22)	(5.74)	(4.13)	(5.74)	(4.13)
Number H types			5.28^{**}	-0.43			10.76^{**}	3.93^{*}
			(2.36)	(2.62)			(4.24)	(2.29)
Constant	16.16^{***}	14.64^{***}	9.92^{***}	15.93^{*}	12.90^{***}	10.37^{***}	2.15	-1.43
	(1.78)	(1.97)	(2.98)	(8.40)	(2.38)	(2.71)	(3.68)	(6.87)
Observations	1,584	1,872	1,584	1,872	1,440	1,440	1,440	1,440
Number of indiv.	66	78	66	78	60	60	60	60
Number of groups	30	30	30	30	24	24	24	24

Robust standard errors clustered on group level in parentheses; *** = p<0.01 ** = p<0.05 * = p<0.1. 'V = 100' is a dummy equal to 1 if V = 100 and zero otherwise; similarly for 'V = 500'. 'Number H types' runs from 0 to 4.

Table 10: Effort – Robustness check 2: controlling for risk aversion

	(1) L	(2) H	(3) L	(4) H
	Dep. var.: $(x_{i,V} - x^*_{i,V})$			
V = 100	-4.50^{***}	-0.41	-4.50^{***}	-0.41
	(1.49)	(1.86)	(1.50)	(1.86)
V = 500	-38.99^{***}	-23.32^{***}	-38.99^{***}	-23.32^{***}
	(3.68)	(2.98)	(3.68)	(2.98)
Number H types			7.71^{***}	1.87
			(2.00)	(1.94)
Risk Aversion	-0.39^{*}	-0.29^{***}	-0.38^{**}	-0.29^{***}
	(0.20)	(0.10)	(0.17)	(0.10)
Constant	-1.44	0.32	-9.85	-5.43
	(7.96)	(4.62)	(7.88)	(8.16)
Observations	3,024	3,312	3,024	3,312
Number of indiv.	126	138	126	138
Number of groups	54	54	54	54

Robust standard errors clustered on group level in parentheses; *** = p<0.01 ** = p<0.05 * = p<0.1. ; risk aversion runs from -100 to 0 with higher numbers indicating a higher risk aversion. 'V = 100' is a dummy equal to 1 if V = 100 and zero otherwise; similarly for 'V = 500'. 'Number H types' runs from 0 to 4.

Table 11: Effort – Robustness check 3: controlling for learning

	(1) L	(2) H	(3) L	(4) H	(5) L	(6) H	(7) L	(8) H
	First half of the periods				Second half of the periods			
	Dep. variable: $x_{i,V} - x^*_{i,V}$							
V = 100	-5.98^{***}	0.02	-5.98^{***}	0.02	-3.03^{**}	-0.85	-3.03^{**}	-0.85
	(2.01)	(2.17)	(2.01)	(2.17)	(1.47)	(2.05)	(1.47)	(2.05)
V = 500	-41.19^{***}	-22.74^{***}	-41.19^{***}	-22.74^{***}	-36.79^{***}	-23.90^{***}	-36.79^{***}	-23.90^{***}
	(4.16)	(3.10)	(4.16)	(3.10)	(3.58)	(3.30)	(3.58)	(3.30)
Number H types			8.73^{***}	2.71			6.70^{***}	0.79
			(2.17)	(1.95)			(2.31)	(1.95)
Constant	16.61^{***}	13.01^{***}	7.04^{***}	4.87	12.61^{***}	12.56^{***}	5.27^{**}	10.19
	(1.43)	(1.61)	(2.53)	(5.76)	(1.75)	(1.96)	(2.35)	(6.33)
Observations	1,512	1,656	1,512	1,656	1,512	1,656	1,512	1,656
Number of indiv.	126	138	126	138	126	138	126	138
Number of groups	54	54	54	54	54	54	54	54

Robust standard errors clustered on group level in parentheses;*** = p<0.01 ** = p<0.05 * = p<0.1. ‘V = 100’ is a dummy equal to 1 if V = 100 and zero otherwise; similarly for ‘V = 500’. ‘Number H types’ runs from 0 to 4.

Table 12: Help/sabotage – Robustness check 1: controlling for order effects

	(1) L	(2) H	(3) L	(4) H	(5) L	(6) H	(7) L	(8) H
	Order 1				Order 2			
	Dep. variable: $k_{ij,V} - k^*_{ij,V}$							
V = 100	-3.52^{***}	-2.24^{**}	-3.52^{***}	-2.24^{**}	-4.05^{***}	-5.59^{***}	-4.05^{***}	-5.59^{***}
	(1.15)	(1.00)	(1.15)	(1.00)	(1.34)	(1.65)	(1.34)	(1.65)
V = 500	2.35	8.31^{***}	2.35	8.31^{***}	7.18^{**}	8.14^{***}	7.18^{**}	8.14^{***}
	(2.79)	(2.08)	(2.80)	(2.09)	(2.96)	(2.43)	(2.96)	(2.44)
Number H types			-2.74^{**}	-4.01^{***}			-11.06^{**}	-2.72^{*}
			(1.10)	(1.03)			(5.03)	(1.41)
Constant	-1.09	-4.92^{***}	2.15^{**}	7.11^{**}	-5.70^{*}	-3.71^{***}	5.36	4.46
	(1.20)	(0.92)	(1.09)	(3.07)	(3.11)	(1.13)	(3.35)	(4.00)
Observations	1,584	1,872	1,584	1,872	1,440	1,440	1,440	1,440
Number of indiv.	66	78	66	78	60	60	60	60
Number of groups	30	30	30	30	24	24	24	24

Robust standard errors clustered on group level in parentheses; *** = p<0.01 ** = p<0.05 * = p<0.1. ‘V = 100’ is a dummy equal to 1 if V = 100 and zero otherwise; similarly for ‘V = 500’. ‘Number H types’ runs from 0 to 4.

Table 13: Help/sabotage – Robustness check 2: controlling for risk aversion

	(1) L	(2) H	(3) L	(4) H
	Dep. var.: $(k_{ij,V} - k^*_{ij,V})$			
V = 100	-3.77^{***}	-3.69^{***}	-3.77^{***}	-3.69^{***}
	(0.87)	(0.94)	(0.87)	(0.94)
V = 500	4.65^{**}	8.23^{***}	4.65^{**}	8.23^{***}
	(2.05)	(1.57)	(2.05)	(1.57)
Number H types			-6.08^{***}	-3.39^{***}
			(2.09)	(0.87)
Risk Aversion	-0.27	-0.06	-0.27	-0.07
	(0.26)	(0.05)	(0.24)	(0.04)
Constant	7.83	-1.77	14.47	8.64^{***}
	(9.52)	(2.17)	(10.72)	(3.34)
Observations	3,024	3,312	3,024	3,312
Number of indiv.	126	138	126	138
Number of groups	54	54	54	54

Robust standard errors clustered on group level in parentheses; *** = p<0.01 ** = p<0.05 * = p<0.1; risk aversion runs from -100 to 0 with higher numbers indicating a higher risk aversion. 'V = 100' is a dummy equal to 1 if V = 100 and zero otherwise; similarly for 'V = 500'. 'Number H types' runs from 0 to 4.

Table 14: Help/sabotage – Robustness check 3: controlling for learning

	(1) L	(2) H	(3) L	(4) H	(5) L	(6) H	(7) L	(8) H
	First half of the periods				Second half of the periods			
	Dep. variable: $k_{ij,V} - k^*_{ij,V}$							
V = 100	-3.25^{***}	-2.20^{**}	-3.25^{***}	-2.20^{**}	-4.30^{***}	-5.19^{***}	-4.30^{***}	-5.19^{***}
	(0.76)	(0.94)	(0.76)	(0.94)	(1.15)	(1.11)	(1.15)	(1.11)
V = 500	5.88^{***}	10.84^{***}	5.88^{***}	10.84^{***}	3.41^{*}	5.63^{***}	3.41^{*}	5.63^{***}
	(2.16)	(1.50)	(2.16)	(1.50)	(2.05)	(1.80)	(2.05)	(1.80)
Number H types			-6.63^{***}	-3.20^{***}			-5.55^{**}	-3.53^{***}
			(2.23)	(0.86)			(2.29)	(0.90)
Constant	-3.65^{**}	-5.44^{***}	3.61^{**}	4.17^{*}	-2.92^{*}	-3.35^{***}	3.16^{*}	7.24^{***}
	(1.55)	(0.71)	(1.49)	(2.42)	(1.69)	(0.82)	(1.68)	(2.71)
Observations	1,512	1,656	1,512	1,656	1,512	1,656	1,512	1,656
Number of indiv.	126	138	126	138	126	138	126	138
Number of groups	54	54	54	54	54	54	54	54

Robust standard errors clustered on group level in parentheses;*** = p<0.01,** = p<0.05,* = p<0.1.; 'V = 100' is a dummy equal to 1 if V = 100 and zero otherwise; similarly for 'V = 500'. 'Number H types' runs from 0 to 4.

Table 15: Total output – Robustness check 1: controlling for order effects

	(1)	(2)	(3)	(4)
	$H7$	$H8\&H9$	$H7$	$H8\&H9$
	Order 1		Order 2	
	Dep. variable: $Y - Y^*$			
V = 100	-47.08^{***}	-47.08^{***}	-123.26^{***}	-123.26^{***}
	(17.22)	(17.23)	(18.60)	(18.61)
V = 500	-43.14	-43.14	-53.75^{**}	-53.75^{**}
	(26.58)	(26.60)	(23.37)	(23.38)
Number H types		-21.65^{**}		-29.41^{**}
		(10.38)		(12.71)
Constant	22.12	69.04^{**}	-7.37	51.45^{**}
	(20.15)	(26.86)	(22.63)	(24.37)
Observations	864	864	720	720
Number of groups	36	36	30	30

Robust standard errors clustered on group level in parentheses; *** = p<0.01,** = p<0.05,* = p<0.1.; 'V = 100' is a dummy equal to 1 if V = 100 and zero otherwise; similarly for 'V = 500'. 'Number H types' runs from 0 to 4.

Table 16: Total output – Robustness check 2: controlling for learning

	(1)	(2)	(3)	(4)
	$H7$	$H8\&H9$	$H7$	$H8\&H9$
	First half of periods		Second half of periods	
	Dep. variable: $Y - Y^*$			
V = 100	-59.15^{***}	-59.15^{***}	-104.27^{***}	-104.27^{***}
	(15.14)	(15.15)	(16.37)	(16.38)
V = 500	-21.31	-21.31	-74.62^{***}	-74.62^{***}
	(20.35)	(20.37)	(19.66)	(19.67)
Number H types		-20.38^{**}		-27.64^{***}
		(8.78)		(7.64)
Constant	-3.51	39.11^{**}	20.94	78.74^{***}
	(14.49)	(17.19)	(18.23)	(21.54)
Observations	792	792	792	792
Number of groups	66	66	66	66

Robust standard errors clustered on group level in parentheses; *** = p<0.01,** = p<0.05,* = p<0.1.; 'V = 100' is a dummy equal to 1 if V = 100 and zero otherwise; similarly for 'V = 500'. 'Number H types' runs from 0 to 4.

Table 17: Group composition – Robustness check 1: controlling for order effects

	(1) V=0	(2) V=100 Order 1	(3) V=500	(4) V=0	(5) V=100 Order 2	(6) V=500
	Dep. var.: $Y - Y^*$					
Homogeneous	−50.41	−70.76	−106.11	−84.04	−29.09	42.96
	(93.98)	(73.63)	(78.69)	(51.35)	(53.68)	(62.36)
Asymmetric	−50.52	−35.04	−118.59*	−163.17***	−121.63**	−129.95**
	(93.62)	(73.35)	(70.78)	(57.09)	(58.75)	(59.37)
Constant	64.18	16.14	73.64	91.52**	−70.34	−26.32
	(91.24)	(67.87)	(62.74)	(44.60)	(44.91)	(46.11)
Observations	288	288	288	240	240	240
Number of groups	36	36	36	30	30	30

Robust standard errors clustered on group level in parentheses; *** = p<0.01, ** = p<0.05, * = p<0.1.

Table 18: Group composition – Robustness check 2: controlling for learning

	(1) V=0	(2) V=100 First half of periods	(3) V=500	(4) V=0	(5) V=100 Second half of periods	(6) V=500
	Dep. var.: $Y - Y^*$					
Homogeneous	−34.80	−29.12	−30.05	−99.65	−70.72	−33.10
	(46.87)	(50.45)	(53.40)	(67.35)	(45.97)	(55.42)
Asymmetric	−74.34*	−49.74	−131.70***	−122.29*	−72.32	−94.58*
	(44.86)	(54.89)	(49.11)	(69.66)	(48.29)	(52.79)
Constant	42.93	−29.46	45.97	112.77 *	−24.73	1.35
	(40.74)	(46.77)	(39.57)	(65.18)	(38.98)	(45.65)
Observations	264	264	264	264	264	264
Number of groups	66	66	66	66	66	66

Robust standard errors clustered on group level in parentheses; *** = p<0.01, ** = p<0.05, * = p<0.1.

D Instructions

Part 1

Dear participants,

welcome to today's experiment.

Please read the instructions for the experiment carefully. For a better understanding, in the following we will only use male designations. Those should be understood gender neutral. All statements in the instructions are true, and all participants receive exactly the same instructions. The experiment as well as the data analysis is anonymous.

We ask you to not talk to other participants and to use only the resources and devices that are provided by the conductors of the experiment. Please switch off all electronic devices. In addition, at the computer you are only allowed to use features that are necessary for the experiment. If you do not comply with these rules, you won't be paid in this experiment and you are not allowed to participate in any further experiments.

Your earnings in the experiment depend on your decisions and potentially the decisions of others. The currency used in the experiment is Tokens. Tokens will be converted to Euros at a rate of 100 Tokens to 6 Euro. You have already received a Euro 9.00 participation fee. Your earnings from the experiment will be incorporated into your participation fee. At the end of today's experiment, you will be paid in private and in cash.

The experiment will last around 90 Minutes. It consists of two parts; both parts are completely independent from each other. That is, your payment for part x only depends on decisions that you take in part x, and does not depend on decisions you take in the other part of the experiment.

At the beginning of each part you receive the corresponding instructions. We will read the instructions out loud and will give you time for questions. If you have a question, please raise your hand. Your question will then be answered privately. Thank you for your attention and for participating in this experiment.

Part 2

The first part of the experiment consists of 24 periods, divided in three blocks of 8 periods; Blocks 1, 2 and 3. In each period you will be asked to make a set of decisions. At the beginning of each block, you will receive a new set of instructions. At the end of Part 1, we will randomly choose one period from each block to determine your earnings from Part 1. Because you do not know which periods will be chosen when you are making your decisions, you should make decisions in each period as if it were to be paid.

Remember that you were given 9 Euro show up fee at the beginning of the Experiment. Any gains or losses incurred in this part of the experiment will be offset against this amount.

Block 1

Matching

In Block 1, you will be matched with three other participants to make a group of four. You will stay in the same group for all 8 periods. To ensure anonymity, you and the three others in your group will be labeled by the computer program as member A, B, C and D. Each group member's label will be the same for all 8 periods.

Types

At the beginning of Block 1, each participant will be randomly assigned a type. You will be either an H or L type. You will keep this assignment for all 8 periods as well. You will know what type you are, and you will know the types of the other members in your group. More on the role of the types in a moment.

Decisions – Overview

Each round, you and your other group members will be working to generate a group output. The group output will determine how much each of you will earn. "Working" means that each of you will choose how many effort points to invest into an **individual output**. The sum of all individual outputs will determine your group output.

For your individual output you choose effort points on the range of 0 to 150 (a whole number). Each point you choose will be costly to you. You can find a table of costs on the separate sheet. These costs are denoted in Token. Notice that the first point will cost you 0.01 Token, 10 points cost you 1.28 Token, and 100 points 128 Token. This indicates that your per point-cost of effort is increasing with the total effort. These effort point costs will be subtracted from your earnings in each round.

If you are an L type, then your **total individual output** will be equal to your total effort points. If you are an H type, your total individual output will be equal to twice your total effort points.

All group members will choose simultaneously their own effort points.

The number of Token each group member receives from the total group output is equal to the total group output multiplied by 0.25. That is, for each point you and your group member's generate towards total group output, you and each of your group members receive 0.25 Token. For instance, if the total group output was 162 points, then you and every member of your group would receive 162*0.25=40.5 Token. To determine your net earnings you would then have to subtract off the cost of your chosen effort. For instance, if your effort was 23 points, your cost would be 6.72 Token. This would result in total earnings of 40.5-6.72 = 33.78 Token. Please turn to your screen and I will go through an example of how these decisions look like on the screen.

[read out loud] On the screen, you see a brief reminder of your task and a box where you will be able to type in the number of effort points you wish to choose for your individual output. You can choose any number of points between 0 and 150. Please type in 20. On the bottom of the screen, there is a calculator to calculate the costs. This calculator automatically updates the costs of your choices when you press the "calculate" button. These costs are the same as those in the Cost Table. If you have entered a choice of 20, and you press the "calculate" button, you will notice that it shows you the cost is 5.12 Token, the number that corresponds to a cost of a choice of 20 on your Cost Table. Notice that what you earn from your 20 points of effort you chose for your individual output is $20 * 0.25 = 5$ *Token if you an L type and* $2 * 20 * 0.25 = 10$ *Token if you are an H type. Please turn your attention back to the instructions and we will describe the next task in the experiment.*

Alterations to Efforts of Other Group Members

In addition to making your own effort choice, in each round you will also be able to affect the effort of your group members. On the screen you can choose additional effort points towards increasing or decreasing the effort of your group members. You will be able to modify the effort of others by increasing or decreasing their effort by up to 150 effort points per group member. Each of these efforts again means costs to you as shown in the table.

This means you will have a total of 4 decisions to make per period. You will choose how many effort points to exert towards your own effort. Then you decide for each of your group members regarding whether and how much you want to alter their effort. Regardless of your type, each effort point you choose towards raising or lowering the effort of others changes their effort by one point.

When determining each group members total individual output, we will first add their own effort points with all of the effort points others have chosen to increase or decrease the effort. If that individual is an L type, their total individual output is the same as their total effort. If that individual is an H type, their total individual output will be twice this sum. Note that this means that you can alter the total individual output of an H type by two points per 1 point of effort you chose. Each point of effort you choose to alter t the individual output of an L type alters their total effort by only 1 point. Similarly, your group member's choices affect your total individual output.

For instance, if you chose 10 effort points for your effort and your group members modified your effort by 5, -2, and 19 points, your modified individual effort would be 10+5-2+19=32. If you are an L type, your total individual output would be 32. If you are an H type, your total individual output would be 32*2 = 64. Similar calculations also hold for your group members who are H or L types. It is possible that your total individual output will be negative. In this case, the computer will assign you a total individual output of zero so that you will never have a negative total individual output.

Each effort point is costly, independently on whether you chose it for your own effort or to affecting the efforts of others. All of your efforts determine your total cost (in Token). For instance, if you altered (increased or decreased) the effort of each team member by 10 points and chose an individual effort of 10 points, the total cost to you would be 1.28+1.28+1.28+1.28=5.12 Token: the cost of an effort level of 10 is equal to 1.28 (as in the cost Table).

It is important to note that effort costs are treated separately for each decision. If for instance, consider the following example: you chose to reduce the effort of one group member by 20 (by choosing -20), leave the one of another group member unchanged (by choosing 0) and increased the one of the third group member by 10 (by choosing 10). In addition, you chose an own effort of 10 points. Then you would still be expending a total of 40 effort points (20+10+10). But, according to the cost table, your cost for these decisions would be 5.12+0+1.28+1.28=7.68 Token, which are the costs associated with effort points of 20, 0, 10 and 10. Please turn to your screen, on which I will walk you through such a decision situation.

On your screen, you are given a brief reminder of your task and an input box for each of your group members. The points chosen for each of these input boxes determine the amount you wish to alter each group member's effort. You can choose any number between -150 and 150 (negative 150 and positive 150) points. Notice beside each input box is each group members' label (A, B, C or D) and their type (H or L).

Please type in -18 in the first box, 10 in the second and 23 in the third. On the bottom of the screen, there is a calculator for you to use and calculate costs. This calculator automatically updates with the costs of your choices when you press the "calculate" button. These costs are taken from the cost table you've been given. If you have chosen -18, 10 and 23 points and you press the "calculate" button, you see the following cost for these choices: 12.2 Token (4.15+1.28 +6.77). These are the same as the sum of costs of alterations of 18, 10 and 23 on your cost table.

You will also notice that the on-screen calculator asks you to enter a hypothetical amount you believe your group members will contribute to the group output. This is purely for you to be able to understand how payoffs work. Anything entered here has no impact on your actual payoff or the decisions of others. Below this, you can see how the group output and total payoffs change as you change your choices of your effort points and modify the efforts of others. Please turn your attention back to the instructions and we will go through the feedback you will receive after a round is over.

Feedback

After each round, you will be shown a results screen which will show you your four decisions. In addition, you see the following information: your total individual effort and total individual output after the modifications from your group members, the average modification of efforts from your group, the total group output, the costs associated with your choices, and your payoff, should that round be chosen for payment.

Payoff Example

We will now explain you by means of an example how your payoff is calculated. Let's assume that the total group output from your group was 162 points. The gain you would receive from this total group output is 162*0.25=40.5 Token. Let's also assume you chose 20 points for your own effort and chose to alter the efforts of your group members by -18, 10 and 23 points, respectively. This would result in a total cost of 5.12+4.15+1.28+6.77 = 17.32 Token. Thus, in this example, the Token gained in this period would be your gains minus your total costs: 40.5-17.32=23.18.

Summary

At the beginning of the first block, you will be randomly assigned a type, L or H, and will be grouped with 3 other people to make a group of 4. You will keep this type and this group for the entire 8 periods of block 1. Each period you must choose the number of points for your own individual output and modifications to each of your group member's individual outputs. The costs of these decisions will be deducted from your gains from the total group output.

Are there any questions?

If not, please turn to your screen. There you will be shown your type and the types of your other 3 group members. After reviewing this information, please click on the "continue" button, and the first round of Block 1 will begin. If you are finished making decisions on a screen, you must click on the "continue" button to advance. The program only advances if everyone has clicked on the continue button for a given portion so, please pay attention to the screen and click the "Continue" button if you are finished making decisions on that screen.

Block 2

Block 2 is similar to Block 1 except for one change. You and your group members will still take decisions for 8 periods. You have the same four choices of your own effort and modifications to your group members' efforts as in Block 1. The costs and gains from the group output are as previously defined. Also, you are in the same group as before, and your types are the same as in Block 1.

In Block 2, however, there will be a bonus awarded to one of the group members, to encourage higher effort. The bonus will be awarded using a lottery, where your probability of winning is increasing in your total individual output, and is decreasing in the total individual output of the others.

If you win the bonus, you get 100 [500] Token. Only one group member can win the bonus per period. Your probability of winning is determined as

$$\text{Chance of winning} = \frac{\text{Your total individual output (TIO)}}{\text{TIO of A + TIO of B + TIO of C + TIO of D}}$$

As an example, suppose that your total individual output was 30 points and that the other members of your group had total individual outputs of 21, 52 and 9 points. Your chance of winning is thus 30/(30+21+52+9)=0.27, or 27% (rounded). Likewise, the chance of each of your group members to win the bonus is 19%, 46% and 8% for the group members who had 21, 52 and 9 points respectively. It is easy to see that increases in total individual output lead to a greater chance in winning the prize.

To see how the likelihoods work, imagine the percentages represent the number of balls each group member has in a common container. If someone randomly selected one of these balls to determine the winner, the chance of winning can now be thought of as the likelihood your own ball is drawn. Group member C, who has 46 balls, has a much higher chance of their ball being drawn than group member D, who only has 8 balls.

Let's go through another example. If your group members' total indiv. outcomes had remained the same, but you had a total individual output of 40 (instead of 30), your chance of winning would increase from 27% to 33% (40/(40+21+52+9)=0.33 or 33%). Since your chance of winning went up, your group members' chances of winning went down to 17%, 43% and 7% (for the group members who had total individual outputs of 21, 52 and 9 points respectively).

Likewise, the chances will also change if your total individual output had remained the same, but the total individual output of one of your group members had changed (because they chose a different effort or their total individual output was modified by you or other group members).

For this example, assume that your total individual output was again 30 and the total individual output of two of your group members was still 21 and 52, but the fourth group member had an increase in his total individual output to 18 (instead of 9). Now, instead of you having a 27% chance of winning you would have a chance of winning of 23% (30/(30+21+52+18) = 0.23 or 23%). Similarly, your group members would have a 16%, 40% and 14% chance of winning respectively. Notice that the group member whose total

individual output is higher, now has a larger chance of winning, while all other group members have a smaller chance of winning. Similarly, your chances of winning the bonus can increase if the total individual output of another member of your group goes down.

Continue with the same example. **Suppose** your total individual output is 30, the total indiv. outputs of two other group members are 21 and 9, but the total individual output of the group member who previously had 52 decreases to 30. Then your chance of winning is 33% (30/(30+21+9+30)=0.33 or 33%).To assist you in your decision, the calculator on the screen will now also show you how your chance of winning changes as you change your choices.

End of a Period

At the end of each period, you will see the same information as in Block 1, except now, you will also be told if you won the bonus or not.

Do you have questions?

Block 3

The instructions for Block 3 are very similar to those from Block 2. Specifically, you and your group members will still make decisions for 8 periods where you have the same four choices: your own effort and the modifications to your group members' efforts. The costs and the and gains from the group output are as previously defined. You are in the same group as before and the types are the same as before. The only change is that in Block 3, the size of the bonus has increased to 500 [100] Token. Everything else stays as in Block 2. Do you have questions?

Part 3

On your computer screen you will see a square composed of 100 numbered boxes, like shown below.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Behind one of these boxes hides a mine; all the other 99 boxes are free from mines. You do not know where this mine lies. You only know that the mine can be in any place with equal probability. Your task is to decide how many boxes to collect. Boxes will be collected in numerical order, starting with number 1. So you will be asked to choose a number between 1 and 100. At the end of the experiment we will randomly determine the number of the box containing the mine. If you happen to have harvested the box where the mine is located – i.e. if your chosen number is greater than or equal to the drawn number – you will earn zero. If the mine is located in a box that you did not harvest – i.e. if your chosen number is smaller than the drawn number – you will earn an amount equivalent to the number you have chosen.

Table 19: Cost Table

Your choice of effort	costs	Your choice of effort	costs	Your choice of effort	costs
0	0.00	51	33.29	101	130.57
1	0.01	52	34.61	102	133.17
2	0.05	53	35.96	103	135.80
3	0.12	54	37.32	104	138.44
4	0.20	55	38.72	105	141.12
5	0.32	56	40.14	106	143.82
6	0.46	57	41.59	107	146.55
7	0.63	58	43.06	108	149.30
8	0.82	59	44.56	109	152.08
9	1.04	60	46.08	110	154.88
10	1.28	61	47.63	111	157.71
11	1.55	62	49.20	112	160.56
12	1.84	63	50.80	113	163.44
13	2.16	64	52.43	114	166.35
14	2.51	65	54.08	115	169.28
15	2.88	66	55.76	116	172.24
16	3.28	67	57.46	117	175.22
17	3.70	68	59.19	118	178.23
18	4.15	69	60.94	119	181.26
19	4.62	70	62.72	120	184.32
20	5.12	71	64.52	121	187.40
21	5.64	72	66.36	122	190.52
22	6.20	73	68.21	123	193.65
23	6.77	74	70.09	124	196.81
24	7.37	75	72.00	125	200.00
25	8.00	76	73.93	126	203.21
26	8.65	77	75.89	127	206.45
27	9.33	78	77.88	128	209.72
28	10.04	79	79.88	129	213.00
29	10.76	80	81.92	130	216.32
30	11.52	81	83.98	131	219.66
31	12.30	82	86.07	132	223.03
32	13.11	83	88.18	133	226.42
33	13.94	84	90.32	134	229.84
34	14.80	85	92.48	135	233.28
35	15.68	86	94.67	136	236.75
36	16.59	87	96.88	137	240.24
37	17.52	88	99.12	138	243.76
38	18.48	89	101.39	139	247.31
39	19.47	90	103.68	140	250.88
40	20.48	91	106.00	141	254.48
41	21.52	92	108.34	142	258.10
42	22.58	93	110.71	143	261.75
43	23.67	94	113.10	144	265.42
44	24.78	95	115.52	145	269.12
45	25.92	96	117.96	146	272.84
46	27.08	97	120.44	147	276.60
47	28.28	98	122.93	148	280.37
48	29.49	99	125.45	149	284.17
49	30.73	100	128.00	150	288.00
50	32.00				

E Screenshots

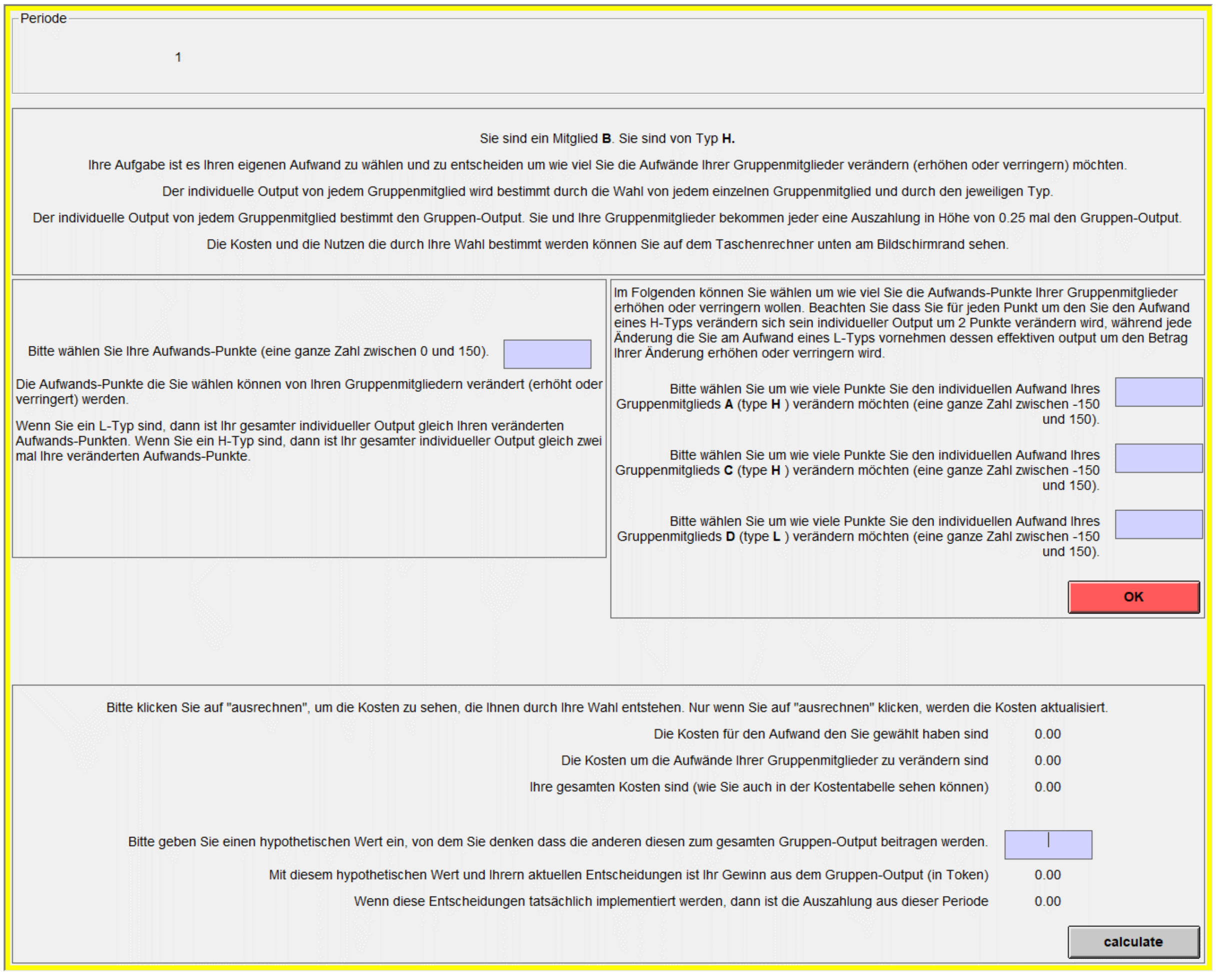


Figure 11: Full screen

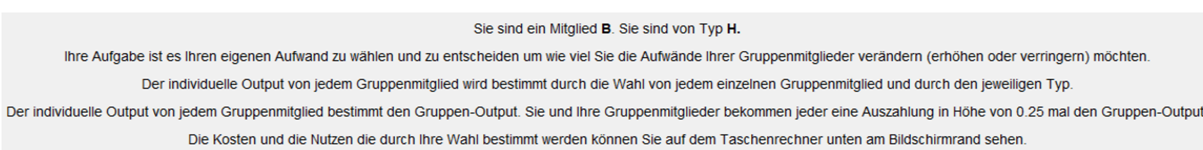


Figure 12: Full screen

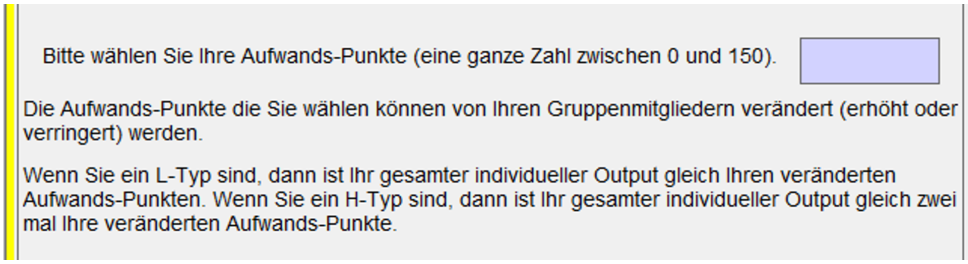


Figure 13: Full screen

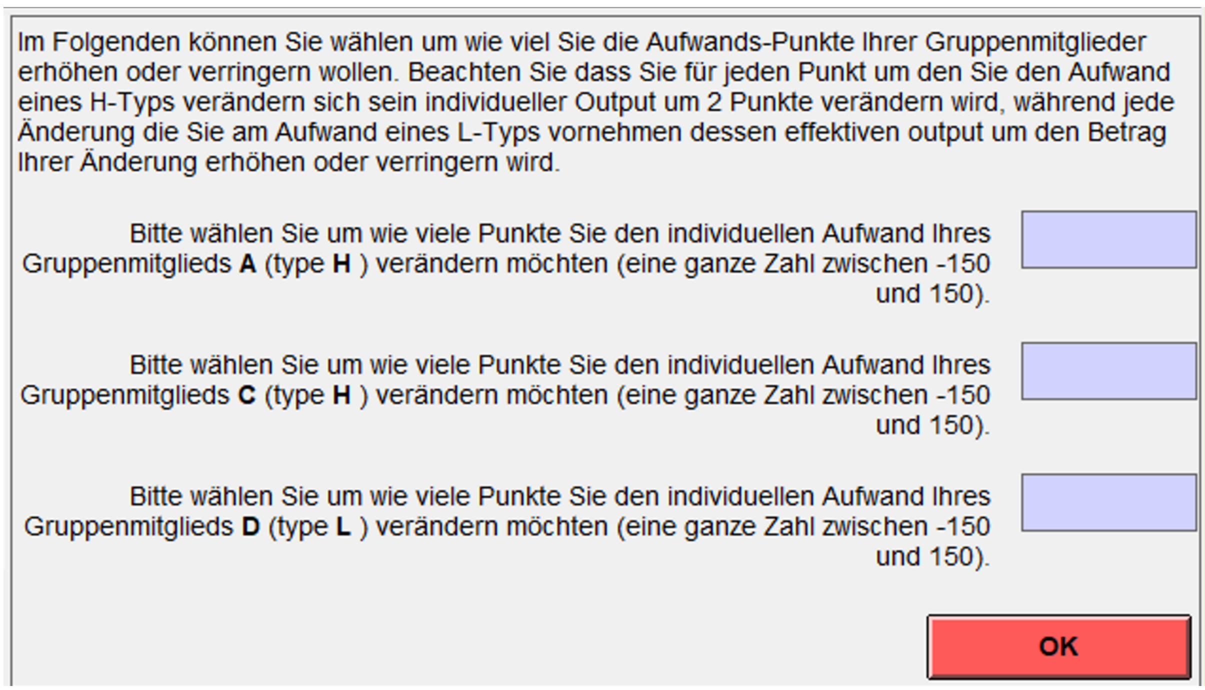


Figure 14: Full screen

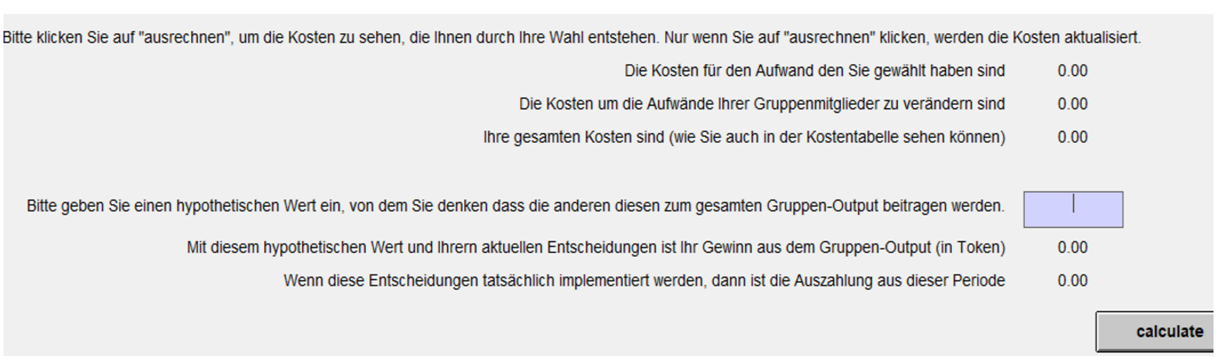


Figure 15: Full screen

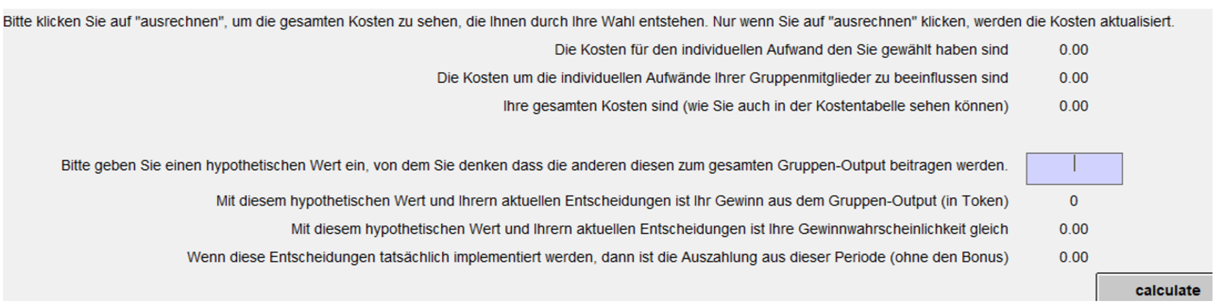


Figure 16: Full screen