Ability Matters, and Heterogeneity Can Be Good: The Effect of Heterogeneity on the Performance of Tournament Participants

Rudi Stracke
University of St. Gallen
IZA, Bonn
CEPR, London

Uwe Sunde
University of St. Gallen

August 18, 2011

Abstract

This paper analyzes the effect of heterogeneity between tournament participants on their performance in a static and in a dynamic tournament model. We show that it is important to separately consider changes in the average level of ability and variations in the degree of heterogeneity. When keeping the average ability constant, we find that the degree of heterogeneity has a negative effect on performance in the static tournament. The opposite holds, however, for the dynamic tournament model, where the degree of heterogeneity has a strictly positive effect. We also present experimental evidence that is in line with these theoretical predictions.

JEL-classification: M52, J33

Keywords: Tournaments, Heterogeneity, Incentive Provision, Ability, Experiment

*Contact: Swiss Institute for Empirical Economic Research (SEW), Varnhüelstrasse 14, CH-9000 St. Gallen.
E-mail: rudi.stracke@unisg.ch, uwe.sunde@unisg.ch
1 Introduction

Tournaments constitute an important element within the field of Personnel Economics ever since Lazear and Rosen (1981) showed that rank-order tournaments are optimal labor contracts under certain conditions. In particular, if workers’ individual effort is not verifiable (e.g., because it is observed with some noise), it might not be possible or optimal to implement piece rates or other pay-for-performance remuneration schemes (Malcomson 1984). As long as an ordinal ranking of workers’ performance is still possible, Lazear and Rosen’s results show that rank-order tournaments for discrete prizes or bonus payments can be used in such settings as a compensation scheme to provide workers with efficient incentives for effort provision. Tournaments are not only used to model the competition for bonus payments within a company, however. Internal labor markets are often modeled as promotion tournaments along the lines of Rosen (1986). Surveys of the respective literature are provided by McLaughlin (1988) and Prendergast (1999).

One issue that received comparably little attention in the tournament literature is the effect of heterogeneity between participants on their equilibrium performance in dynamic settings, such as promotion tournaments with multiple stages. In reality, workers typically differ in their ability, which implies that virtually all tournaments in involve heterogeneous participants. Moreover, empirical evidence by Gibbs and Hendricks (2004) shows that promotion tournaments are an important means for the provision of incentives in organizations. Still, the incentive properties of dynamic tournaments with heterogeneous participants are largely unexplored. It remains an open question whether or not the result by Lazear and Rosen (1981) that effort provision and performance decrease with the degree of heterogeneity in static tournaments carries over to dynamic settings.

This paper takes a closer look at the effects of heterogeneity of tournament participants on overall tournament performance. We consider both a static tournament model and a dynamic model with multiple stages, applying the same tournament setup as in Rosen (1986). The theoretical analysis of these two models delivers three main results: First, we find that the average ability level of tournament participants has a strong impact on their performance, independent of the tournament format. Second, our results show that the incentive effect of heterogeneity on the overall performance of tournament participants depends on the structure of the tournament. To isolate the effect of heterogeneity on incentives, we compare homogeneous and heterogeneous situations, where the average ability level of participants is the same. This allows us to separate effects of changes in ability from the effect of variations in the degree of heterogeneity on incentives. The findings show that the incentive effect of heterogeneity on performance is negative in static tournaments, which is in line with the common
perception in the literature. However, the opposite holds in the dynamic specification, where heterogeneity has a strictly positive incentive effect on the overall performance of tournament participants. The reason is that heterogeneity increases the value of winning in early stages for strong agents, as they anticipate that it will be easier to win another time in later stages of the tournament due to the presence of weak agents. Therefore, the performance of strong agents is higher in early stages. Third, the comparison of both, the direct (absolute) ability effect and the incentive effect of heterogeneity through relative ability reveals that the effect of variation in heterogeneity through changes in the average level of absolute abilities dominates the corresponding effect through incentives, if ability and heterogeneity are changed simultaneously.

In the second part of the paper, we provide some experimental evidence regarding the theoretical predictions. The findings provide empirical support for the main qualitative predictions: First, we find that the level of average ability of tournament participants has a strong influence on performance. Second, the incentive effect of heterogeneity on performance is negative in the static tournament treatments, while the effect is positive in dynamic tournaments. The (negative or positive) effect of heterogeneity is much more pronounced than one would expect from the theory. While the theoretical results predict a 4% reduction of overall performance due to heterogeneity for the static tournament, we observe a reduction of more than 15%; similarly, overall performance should be approximately 1.5% higher in the dynamic tournament, but we observe an increase of almost 20% in response to a higher degree of heterogeneity. As a consequence, we find little evidence for the third theoretical result that the effect of ability on overall performance dominates the effect of heterogeneity. Instead, the experimental analysis suggests that both effects are equally important under the parametric setup of the experiment.

Our results have several interesting implications for the performance of corporate tournaments. We show that the effect of heterogeneity between participants of a tournament can be affected by the structure of the tournament. Consequently, the specific tournament format plays a role as to whether or not it makes sense for a tournament designing principal to separate or pool different types. In addition, the results also show that the average ability of the workforce participating in a tournament can be as important as negative (or positive) effects of heterogeneity. According to our findings, an increase of the average ability does always have a strictly positive effect, independent of the tournament structure. This suggests that the ability of employees might be more relevant for hiring decisions than potential concerns for the homogeneity among participants of corporate bonus or promotion tournaments for

1Average ability and the degree of heterogeneity are changed simultaneously if, for example, a participant i of the tournament is replaced with somebody who has either a higher or a lower ability than i.
reasons of incentive provision.

This paper complements the existing theoretical and empirical literature on tournaments in several ways. We provide a systematic comparison of the effects of heterogeneity in the two most prominent tournament models in the Personnel Economics literature, the static one-shot tournament along the lines of Lazear and Rosen (1981), and the dynamic multi-stage tournament as suggested by Rosen (1986). Existing theoretical comparisons between static and dynamic tournament models either assume homogeneity of participants (Gradstein and Konrad 1999), or consider the case of a perfectly discriminating all-pay auction (Moldovanu and Sela 2006). In the latter case, it is assumed that signals on the relative performance of tournament participants are always correct and fully informative, an assumption that is likely to be violated in reality. The results of this paper also complement earlier studies which suggest that the tournament designing principal has an incentive to induce self-sorting of worker types by ability into different tournaments (O’Keeffe, Viscusi, and Zeckhauser 1984, Bhattacharya and Guasch 1988), or alternatively, if types are observable, to handicap stronger workers (Lazear and Rosen 1981, Gürtler and Kräkel 2010).

Second, this paper is related to existing experimental work on behavior in tournaments. The papers most closely related are the ones by Sheremeta (2010) and Altmann, Falk, and Wibral (2011), who compare static one-stage and dynamic two-stage tournaments with homogeneous participants. We complement their work and additionally consider settings with heterogeneous agents. Further, our experimental analysis is related to research by Bull, Schotter, and Weigelt (1987), Orrison, Schotter, and Weigelt (2004), and Harbring and Lünser (2008), who analyze the behavior in static tournaments. These studies consider homogeneous and heterogeneous treatments, but do not provide a systematic assessment of the strength of the effect of heterogeneity on performance, since average ability of participants is not held constant across treatments. Finally, the paper is also related to the empirical literature that has investigated the performance effects of heterogeneity. In this strand of the literature, field data from sports (Abrevaya 2002, Sunde 2009, Brown 2010) and corporations (Knoeber and Thurman 1994, Eriksson 1999) has been used to test the implications of heterogeneity that follow from static one-stage tournament models. We provide this empirical literature with a new testable hypothesis for dynamic tournaments with multiple stages, which are quite common both in corporate and sport tournaments.

The remainder of the paper is structured as follows. Section 2 presents a theoretical analysis of equilibrium behavior in static one-stage and dynamic two-stage tournaments. Section 3 presents experimental evidence of tests of the main theoretical predictions, and section 4 concludes.
2 Theoretical Analysis

We consider two different tournament models. In both models, we allow for ability differences between tournament participants, which we will refer to as “workers” subsequently. The baseline specification is a static one-shot tournament, in which two workers compete for some exogenously given prize $P^2$. The prize can be understood as a performance reward for a worker, who receives some bonus payment or a promotion to a better paid position. The second model is a straightforward dynamic extension of the one-shot tournament in the spirit of Rosen (1986). By adding a qualification stage to the static tournament, one can analyze a dynamic tournament with two stages. In the first stage of this tournament, four workers compete in two separate pairwise interactions for a promotion to stage 2. The two losers of the first stage are eliminated from the competition, while the two winning workers are promoted. They encounter each other in stage 2, where they compete for some exogenously given prize $P$, as in the static tournament model $^3$.

The remainder of this section first derives equilibrium solutions for homogeneous and heterogeneous specifications of both tournament models, which allow us to describe optimal behavior of workers in the respective setting. Then, we analyze the effect of heterogeneity on measures of interest for a tournament designing principal in both the static and the dynamic tournament model. The analysis focuses on two central questions: Should a principal separate strong and weak workers from each other, given that both types are employed in his company? And second, should hiring decisions of new workers be influenced by concerns for homogeneity of the workforce? At the end of this theoretical section, we discuss the implications and the robustness of our results for the optimal design of tournaments.

2.1 Static and Dynamic Tournament Models

Both types of tournament models describe a situation in which a principal awards some valuable prize to the best worker, i.e., to the worker who produces the highest amount of output in a given time frame. We define the individual output produced by a type $i$ worker as $y_i(a_i, x_i) = a_i x_i$, where output is the product of ability $a_i$ and effort $x_i$. Given individual outputs of two workers $i$ and $j$, the probability $^4$.
that the prize is awarded to worker $i$ equals

$$p_i = \frac{[y_i(a_i, x_i)]^r}{[y_i(a_i, x_i)]^r + [y_j(a_i, x_i)]^r}.$$  

This formulation is similar to the one used by Rosen (1986) and implies that the principal cannot always perfectly observe which worker produced more, i.e. the monitoring technology is affected by some random component. The parameter $r$ reflects the degree of this randomness: When $r$ approaches infinity, the probability of the worker with the higher output winning converges to 1, implying that the principal can perfectly observe which of the two workers produced more output. For all strictly positive and finite values of $r$, the monitoring technology implies that the probability to win is greater than 0.5 for the agent whose contribution to aggregate output is higher. Consequently, the winning probability is strictly increasing in the individual output $y_i(a_i, x_i)$, and strictly decreasing in the output $y_j(a_j, x_j)$ produced by the opponent $j$ for all values of $r > 0$.

In both theoretical models considered below, we use this monitoring technology for reasons of analytical tractability.

### 2.1.1 Model 1: Static Tournament

We start with the static baseline model, where two risk neutral workers compete with each other for some prize $P$. For simplicity, it is assumed that workers receive no fixed wages. Workers can be of two different types: They are either “strong” (type $S$), or “weak” (type $W$). Types may differ with respect to their productive ability $a_i$ ($a_S \geq a_W$) or their disutility of labor (or effort costs) $c_i$ ($c_S \leq c_W$), or both. Compared to weak workers, strong workers either have a higher productivity or a lower disutility of labor, or both. Workers are assumed to know their type and the types of their competitors.

The two type assumption allows for three different settings: Either, both workers are strong ($SS$) or weak ($WW$), or workers are of different types ($SW$), i.e., we have to consider two homogeneous and one heterogeneous tournament settings. It suffices to solve the general case where workers are allowed to be of different types, however, because one can derive the respective expressions for the homogeneous settings by simply imposing the restriction that type specific parameters are equal. Therefore, we start by considering a situation where one worker is of type $S$, while his opponent is of type $W$. Formally, the

---

5 In the literature, this monitoring technology is usually referred to as the ratio contest success function à la Tullock. Another prominent functional form is the additive noise difference specification. See Hirshleifer (1989) for a comparison of the properties of different contest success functions.

6 A third possibility would be to model heterogeneity in terms of heterogeneous valuations of prizes. Using this specification would leave the main results unaffected. Details are available upon request.
optimization problems can be described as follows:

\[
\text{Type } S: \quad \max_{x_S} \Pi_S = \frac{a_S^r x_S^r}{a_S^r x_S^r + a_W^r x_W^r} P - c_S x_S
\]

\[
\text{Type } W: \quad \max_{x_W} \Pi_W = \frac{a_W^r x_W^r}{a_S^r x_S^r + a_W^r x_W^r} P - c_W x_W,
\]

where each worker maximizes his expected payoff \( \Pi_i \) by choosing effort \( x_i \). Workers face a trade-off with respect to effort provision: On the one hand, effort increases individual output \( y_i(a_i, x_i) = a_i x_i \) and therefore the probability to win the tournament. At the same time, however, each worker bears marginal costs \( c_i \) for each unit of effort provided, no matter whether he wins the tournament or not.\(^7\)

In equilibrium, workers choose their level of effort provision optimally such that marginal costs equal marginal benefits. Note, however, that first-order conditions are necessary and sufficient for optimal behavior only if the strategic advantage of strong workers is not too high for the given precision of the monitoring technology \( r \). For the necessary conditions to be sufficient, heterogeneity between workers must not exceed a certain threshold if the monitoring technology is relatively precise, otherwise pure strategy equilibria do not exist, as was shown by Nti (1999). Apart from the fact that only mixed-strategy equilibrium in such a scenario, little is known about the properties of equilibria in tournaments where this restriction is violated, which is why we restrict attention to equilibria in pure strategies throughout the paper.\(^8\) To ensure the existence of equilibria in pure strategies, we pose a parametric restriction on heterogeneity. For notational clarity, denote the relative ability advantage of strong workers in terms of ability and effort costs by

\[
\phi = \left( \frac{a_S c_W}{a_W c_S} \right)^r,
\]

where \( \phi \geq 1 \), since by assumptions workers of type \( S \) have a higher productive ability as well as a lower dis-utility of labor (or effort) than workers of type \( W \). Essentially, \( \phi \) measures the degree of heterogeneity in the tournament: If \( \phi = 1 \), both worker types are identical and the tournament is homogeneous, while high values of \( \phi \) indicate that types differ substantially. We use this measure of heterogeneity to ensure that first-order conditions characterize optimal behavior (and hence the existence of pure strategy

\(^7\)The assumption of constant marginal costs implies a substantial simplification in terms of analytical tractability, but is not central for the main results of this paper.

\(^8\)The design of the experiments presented below ensure that this condition holds. To our knowledge, the only paper which addresses mixed-strategy equilibria in a Tullock contest is the one by Baye, Kovenock, and de Vries (1994).
equilibria), which is the case if and only if the relation
\[ r \leq 1 + \frac{1}{\phi} \]  
(4)
is satisfied.

**Assumption 1.** Relation (4) is always satisfied, which implies that the degree of heterogeneity between workers (measured by $\phi$) is not too high for the given precision $r$ of the monitoring technology.

Under Assumption 1, the definition of $\phi$, and the two first-order optimality conditions which follow from the optimization problem described above, equilibrium efforts are given by
\[ x^*_S(SW) = r \left( \frac{1}{c_S} \right) \frac{\phi}{[1 + \phi]^2} P \quad \text{and} \quad x^*_W(SW) = r \left( \frac{1}{c_W} \right) \frac{\phi}{[1 + \phi]^2} P. \]  
(5)
Inserting equilibrium efforts $x^*_S(S, W)$ and $x^*_W(S, W)$ in (1) and (2) determines the corresponding equilibrium payoffs
\[ \Pi^*_S(SW) = \frac{\phi^2 + (1 - r)\phi P}{[1 + \phi]^2} \quad \text{and} \quad \Pi^*_W(SW) = \frac{1 + (1 - r)\phi}{[1 + \phi]^2} P, \]  
(6)
which solves the heterogeneous interaction (SW). The expressions in (5) and (6) can then be used directly to characterize equilibrium behavior and outcomes in each of the two homogeneous settings, SS and WW. Recall that $\phi = 1$ by definition in homogeneous specifications. Imposing this assumption on (5), we obtain equilibrium efforts
\[ x^*_S(SS) = r \left( \frac{1}{4c_S} \right) P \quad \text{and} \quad x^*_W(WW) = r \left( \frac{1}{4c_W} \right) P, \]  
(7)
which, when inserted into the formal maximization problems, imply that workers in the homogeneous interactions can expect equilibrium payoffs of
\[ \Pi^*_S(SS) = \frac{2 - r}{4} P \quad \text{and} \quad \Pi^*_W(WW) = \frac{2 - r}{4} P. \]  
(8)
Under Assumption 1, these are strictly positive. 

\footnote{A proof for this claim is provided by Nti (1999).}
2.1.2 Model 2: Dynamic Tournament

The static baseline model can be extended to a dynamic tournament model along the lines of Rosen (1986) by adding a qualification stage to each of the three specifications of the static baseline model, as illustrated in Figure 1. In the case of the homogeneous setting with strong workers only (SS), this implies adding two pairwise stage 1 interactions with two strong workers each; this dynamic setting is denoted SSSS. Similarly, setting WWWW is the dynamic extension of the static model with two weak workers (WW). Analogously, one can add a qualification stage to the static model with heterogenous workers (SW), where two strong and two weak workers compete with each other in stage 1; in what follows, we will refer to this dynamic setting as SSWW. A common feature of these dynamic tournaments is that two workers compete for the right to participate in stage 2 in two separated stage 1 interactions. One worker from each interaction qualifies for stage 2, where the two stage 1 winners compete for prize $P$ as in the static model considered in the previous section; the workers who lost in stage 1 are eliminated from the competition.

All three settings SSSS, SSWW and WWWW are solved via backwards induction due to the dynamic structure of the tournament. Equilibrium efforts in the pairwise interactions on stage 2 are already known from the analysis of the previous section and are given by the respective expressions in (5).
and (7). On stage 1, the optimization problems differ across specifications, and we start by analyzing setting \textit{SSSS}. Note that the two stage 1 interactions are fully symmetric, since two workers of the same type compete in each of the two interactions for the right to participate in stage 2. Participation in stage 2 is valuable for workers, because they have a chance to win the prize \( P \) only if they reach stage 2. This \textit{continuation value} is given by the payoff that a strong worker can expect in equilibrium if he competes with a strong worker for a prize \( P \). Using the results about the expected equilibrium payoff from the previous section in equation (8), the continuation value is given by \( \Pi^*_S(SS) = \frac{2-r}{4} P \).

Consequently, the two workers in each of the homogeneous stage 1 interactions compete for a prize of \( \Pi^*_S(SS) \). Recall that the equilibrium efforts for two strong workers who compete for a prize \( P \) are given in (7); replacing \( P \) by the expression for \( \Pi^*_S(SS) \), we obtain:

\[
\begin{align*}
x^*_s(SSSS) &= r \left( \frac{1}{4c_S} \right) \frac{2-r}{4} P, \\
\end{align*}
\]

where the superscript 1 indicates that effort is provided in stage 1 of setting \textit{SSSS}. Note that an analogous line of argument applies to setting \textit{WWWW}, where weak workers compete in two separate stage 1 interactions for the value of participation in stage 2, which is given by \( \Pi^*_W(WW) = \frac{2-r}{4} P \) according to equation (8). Consequently, equilibrium effort in stage 1 by weak workers equals

\[
\begin{align*}
x^*_w(WWWW) &= r \left( \frac{1}{4c_W} \right) \frac{2-r}{4} P.
\end{align*}
\]

Finally, we analyze the slightly more complicated setting \textit{SSWW} with heterogeneous workers. Note that both stage 1 pairings are between workers of the same type, i.e., strong workers compete in one, while weak workers compete in the other interaction. Therefore, the value of participation in stage 2 depends on the type of a worker. We already saw in the previous section that strong workers can expect a payoff that amounts to \( \Pi^*_S(SW) = \frac{\phi^2 + (1-r)\phi}{[1+\phi]^2} P \) in a competition with a weak worker for a prize \( P \). Similarly, a weak worker can expect a payoff that amounts to \( \Pi^*_W(SW) = \frac{1+(1-r)\phi}{[1+\phi]^2} P \) in equilibrium. Consequently, the two strong workers compete for a prize \( \Pi^*_S(SW) \) in stage 1, while the prize amounts to \( \Pi^*_W(SW) \) in the interaction with weak workers. From the analysis of the static tournament model, we know that this implies equilibrium efforts

\[
\begin{align*}
x^*_s(SSWW) &= r \left( \frac{1}{4c_S} \right) \frac{\phi^2 + (1-r)\phi}{[1+\phi]^2} P, \\
x^*_w(SSWW) &= r \left( \frac{1}{4c_S} \right) \frac{1+(1-r)\phi}{[1+\phi]^2} P.
\end{align*}
\]

This completes the solution of the static and dynamic tournament models. These solutions constitute the incentive compatibility constraints for a tournament designing principal. Expected equilibrium
payoffs are strictly positive for both types and both tournament models under Assumption 1.

2.2 Optimal Tournament Design: The Principal’s Perspective

This section analyzes how ability and heterogeneity between workers affect output as the central measure of interest of a tournament designing principal. Both the direction and the magnitude of these effects are important to answer the two central questions of our analysis, namely whether or not strong and weak workers should be separated by the principal, and to what extent concerns for the homogeneity of the workforce should affect hiring decisions. We assume that the principal’s objective is to maximize profits of the company and abstract from other objective functions that a principal might have. The principal has prior information about the type of each worker. Since total wage costs as well as the price for the output good are assumed to be given exogeneously, the principal’s problem reduces to a maximization of total output (denoted $Y$ subsequently) produced by all employees. Following the literature on tournament design, we abstract from ability specific tasks or complementarities between output by individual workers and consider the simple case where total output $Y$ equals the sum of individual outputs of all $S$ and $W$ type workers. Recall that individual output is given by the product of ability and effort of a worker, i.e. $y_i(a_i, x_i) = a_i x_i$. Then, total output is formally defined by the relation

$$Y = K \cdot y_S(a_S, x_S) + M \cdot y_W(a_W, x_W),$$

(12)

where $K$ and $M$ are the numbers of $S$ and $W$ type workers the principal employs. Total output production for all specifications of the static tournament can then be computed using the expressions for equilibrium efforts given in (5) and (7) as

$$Y(SS) = r \left( \frac{a_S}{c_S} \right) \frac{1}{2} P, \quad Y(SW) = r \left( \frac{a_S}{c_S} + \frac{a_W}{c_W} \right) \frac{\phi}{1 + \phi}^2 P, \quad \text{and} \quad Y(WW) = r \left( \frac{a_W}{c_W} \right) \frac{1}{2} P.$$

(13)

It is slightly more complicated to compute total output levels for the dynamic tournament specifications. Note, however, that total output in stage 2 is already known, since the stage 2 interaction is completely identical to the respective static tournament setting. When adding output produced in both stage 1

---

10We refrain from modeling the outside options of workers or details of the hiring process. Instead, we assume that workers receive some wage payment by the principal in addition to the tournament compensation. Non-negative expected payoffs are sufficient for participation in the bonus or promotion tournament.

11Subsequent results do not depend on this specific functional form assumption. All that is needed is a complementarity between ability and effort.
interactions, we obtain

\[
Y(\text{SSSS}) = r \left( \frac{a_s}{c_s} \right) \frac{4 - r}{4} P \quad \text{and} \quad Y(\text{WWWW}) = r \left( \frac{a_w}{c_w} \right) \frac{4 - r}{4} P
\]

(14)

for the homogeneous specifications, while total output in the heterogeneous case amounts to

\[
Y(\text{SSWW}) = r \left[ \left( \frac{a_s}{c_s} \right) \frac{\phi^2 + (3 - r)\phi}{2[1 + \phi]^2} + \left( \frac{a_w}{c_w} \right) \frac{1 + (3 - r)\phi}{2[1 + \phi]^2} \right] P.
\]

(15)

We are now in the position to compare total output across different specifications for a particular model. When comparing total output levels of the homogeneous and heterogeneous specifications, the following relations hold for both the static and the dynamic tournament model:

(i) Total output in the homogeneous setting with strong workers is always higher than output in the heterogeneous setting with equal shares of strong and weak workers, i.e.

\[
Y(\text{SS}) \geq Y(\text{SW}) \quad \text{and} \quad Y(\text{SSSS}) \geq Y(\text{SSWW}).
\]

(ii) Total output in the homogeneous setting with weak workers is always lower than output in the heterogeneous setting with equal shares of strong and weak workers, i.e.

\[
Y(\text{SW}) \geq Y(\text{WW}) \quad \text{and} \quad Y(\text{SSWW}) \geq Y(\text{WWWW}).
\]

(iii) Total output in the homogeneous setting with strong workers is always higher than output in the homogeneous setting with weak workers, i.e.

\[
Y(\text{SS}) \geq Y(\text{WW}) \quad \text{and} \quad Y(\text{SSSS}) \geq Y(\text{WWWW}).
\]

The first statement is in line with the standard perception that heterogeneity is associated with lower effort provision by workers and therefore lower total output. The comparison of homogeneous and heterogeneous settings in (ii) tells a very different story, however, indicating that output is always higher in the heterogeneous settings \(\text{SW}\) and \(\text{SSWW}\) compared to the homogeneous settings with weak workers. Finally, relation (iii) states that total output is not the same in two different homogeneous settings, illustrating that the common distinction between homogeneous and heterogeneous tournament settings is sometimes misleading. The intuition for these relations becomes obvious once one separately considers changes of the average ability level of all workers who participate in a certain tournament, and changes in terms of relative abilities of different worker types: In (iii), for example, settings \(\text{SS}/\text{SSSS}\) and \(\text{WW}/\text{WWWW}\) differ only in terms of the average ability level; relative abilities, which are measured by the degree of heterogeneity \(\phi\), are identical. Strong workers are more productive and face a lower disutility of working by definition, such that average ability and therefore total output is higher in the

\[12\] Relations (i) to (iii) below hold with strict equality whenever strong and weak worker types differ, i.e. if \(\phi > 1\).
situation where strong workers compete with each other. Consequently, the average level of ability has a strictly positive effect on total output when keeping heterogeneity (or relative abilities) constant. Since this is a general result, we summarize this finding in the following Proposition:

**Proposition 1 (Ability Effect).** When holding the degree of heterogeneity, as measured by relative abilities $\phi$, constant, total output is strictly increasing in the (absolute) ability of each worker in both the static and the dynamic tournament model.

*Proof.* See Appendix.

Note that any ceteris paribus increase in the absolute ability level of any worker type increases the average level of abilities. This fact can explain the seemingly contradictory findings in (i) and (ii). Note that settings $SS/SSSS$ and $SW/SSSW$, as well as settings $WW/WWWW$ and $SW/SSWW$ differ in two dimensions, namely in terms of both the average ability level and the degree of heterogeneity $\phi$, measured by relative abilities of different worker types. The average ability level is higher in the homogeneous setting with strong workers only as compared to the heterogeneous situation with equal numbers of strong and weak workers (see comparison in (i)). The opposite holds for the comparison in (ii), where the average level of abilities is higher in the heterogeneous specification with both strong and weak workers than in the homogeneous one with weak workers only. Consequently, the two comparisons of homogeneous and heterogeneous tournaments cannot be used to determine the effect of heterogeneity (or relative abilities) on total output, since the average level of abilities changes at the same time. To measure the incentive effect of heterogeneity on total output, however, one has to compare a heterogeneous tournament setting with a homogeneous situation where workers have the same average level of ability. In other words, for a meaningful comparison, average ability must be kept constant to isolate the effect of heterogeneity on incentives. Essentially, we use the concept of a mean preserving spread in this comparison, since average ability (the mean) is held constant, while relative ability differences and therefore heterogeneity are increased. Using this approach, the effect of heterogeneity on total output crucially depends on the tournament format:

**Proposition 2 (Incentive Effect).** When holding the average ability level constant, total output is

(a) *decreasing* in the degree of heterogeneity between workers (as measured by relative abilities $\phi$) in the static tournament model.

(b) *increasing* in the degree of heterogeneity between workers (as measured by relative abilities $\phi$) in the dynamic tournament model.

*Proof.* See Appendix.

---

13It is worth noting in this context that on the individual level the effect of heterogeneity is detrimental for individual effort provision in static tournaments since $x^*_W(SW) < x^*_W(WW)$ and $x^*_S(SW) < x^*_S(SS)$. Much of the previous literature has focused on this relation without considering the effects on total tournament performance.
Proposition 2 shows that the incentive effect of heterogeneity between workers can have a positive or a negative effect on performance. In case of the static tournament model, we find support for the general perception in the tournament literature that heterogeneity reduces incentives for effort provision, such that total output decreases. Surprisingly, however, the opposite holds for the dynamic tournament, where total output increases in the degree of heterogeneity. Different effects are at work here, which we will analyze separately below. We start with the static tournament model, since it also captures the effect of heterogeneity in stage 2 of the dynamic model.

A higher degree of heterogeneity, or a higher value for $\phi$, implies that weak workers reduce effort and therefore the production of output independent of the effort provided by the opponent, i.e., the best response function of weak workers is lower for all effort levels. On the other hand, strong workers also reduce their equilibrium effort to account for the fact that it is easier to win against a relatively weaker opponent. Both effects unambiguously reduce effort provision, and therefore output. There is an opposing effect, however, in the dynamic tournament setting. The fact that workers of both types provide less effort in stage 2 makes it more attractive to reach stage 2, since the promotion to the top level position becomes cheaper. This holds particularly for strong workers, whose winning probability is increased by a higher relative ability. Consequently, the continuation value of strong workers increases, such that they provide more effort and produce more output in stage 1. Weak workers provide slightly less effort in stage 1, as their continuation value decreases, which means that the effect of heterogeneity on individual output differs across worker types. The overall effect of heterogeneity on total output is unambiguously positive in the dynamic tournament model, however.

It is important to highlight that the effect of heterogeneity is rather small in both the static and the dynamic tournament for modest degrees of heterogeneity. Figure 2 shows the percentage change in total output due to changes of the degree of heterogeneity for both tournament formats. One can see, for example, that degrees of heterogeneity in the range of a 10 to 20% difference in abilities between types have almost no detectable effect on total output. For the same average level of abilities, total output in static tournaments is between 1 and 2% lower in a heterogeneous specification where strong workers are 10-20% stronger than weak ones; in dynamic tournaments, output is 0-0.5% higher in the heterogeneous specification for this degree of heterogeneity. Apart from that, the figure also shows that the strength of the heterogeneity effect depends on the precision of the monitoring technology. In line with economic intuition, the positive or negative effect of heterogeneity is stronger when the signal that the principal receives becomes more precise (in the sense of a higher discriminatory power $r$ of the contest success function), since this discourages the weak worker type, whose chances to win are decreased.

While the effect of heterogeneity on total output is small, changes in the average ability of workers have a large effect on total output: A 1% increase (reduction) of the average ability level increases
Figure 2: The incentive effect of heterogeneity on total output

(reduces) total output by 1%. Therefore, the effect of a change in average ability dominates the effect of heterogeneity on total output if both average ability and the degree of heterogeneity are changed simultaneously. Comparison (ii) between settings $\text{WW}$ and $\text{SW}$ above illustrates this fact for the static tournament: The positive effect on total output of an increase in the average ability level according to Proposition 1 dominates the negative effect of higher heterogeneity (Proposition 2), i.e., when moving from a homogeneous situation with two weak workers to a heterogeneous setting with one worker of each type, total output does always increase, irrespective of how small is the difference in ability. The average level of ability is also more important than the degree of heterogeneity in the dynamic tournament specification: When moving from a homogeneous situation with four strong workers to a heterogeneous setting with two workers of each type, total output does always decrease. The negative effect on total output caused by the decrease of the average ability level according to Proposition 1 dominates the positive effect of a higher degree of heterogeneity (Proposition 2). We summarize these findings in Proposition 3:

**Proposition 3 (Strength of Ability and Incentive Effects).** If the average ability level and the degree of heterogeneity are changed simultaneously, the effect of the change in absolute average ability on total output is always stronger than the corresponding incentive effect of changes in heterogeneity that works through relative ability. This relation holds in static and dynamic tournaments.

*Proof.* See Appendix.
This finding is particularly important in reality for hiring or firing decisions. The replacement of workers usually implies a simultaneous change of ability and heterogeneity, because it is rather unlikely that a newly hired worker has exactly the same ability level as workers already employed by the company, or as a worker who was recently fired. In such a setting, Proposition 3 tells us that the principal should focus entirely on the absolute ability of this worker, and neglect potential effects on the degree of heterogeneity.

2.3 Discussion

The previous analysis of static and dynamic tournaments illustrates that the distinction between the average level of abilities and the degree of heterogeneity in terms of relative abilities is important. We find that individual abilities of a worker determine his general willingness to provide effort in a tournament where ability and effort are complements. Heterogeneity between participants may have a positive or a negative effect on incentives and therefore the production of output, depending on the tournament format: heterogeneity can in fact be good. As a general result, however, changes in the level of average abilities have a stronger impact on effort and total output than corresponding changes in the degree of heterogeneity. Therefore, the theoretical analysis suggests that it is the average ability of a workforce rather than its degree of homogeneity which is crucial for a firm that makes use of tournaments as a compensation scheme. In other words, a principal should be interested mainly in the average ability of his workforce rather than its homogeneity. Whether strong and weak workers are separated into different tournaments does not matter much. It can even be in the interest of the principal not to separate different worker types in dynamic tournaments with multiple stages. While these results suggest that larger heterogeneity can be beneficial for output in the typical static and dynamic tournament settings where ability and effort are complements due to the dominant effect of ability, the results do not imply that greater heterogeneity is always good, however. Under alternative production technologies, e.g., when ability and effort are substitutes, or when firms are primarily interested in balanced competition for different reasons than total output maximization, the results might differ.\[14\]

Two questions remain: First, it is not clear how general the previous results are, since the theoretical analysis restricted attention to two special cases, namely to static tournaments with two, and to dynamic tournaments with four workers. Second, relatively little is known about the behavior of participants in heterogeneous tournaments, and it cannot be taken for granted that individuals react to heterogeneity in the same way as theory predicts. In fact, some existing evidence suggests that behav-

\[14\] See, e.g., Gürtler and Kräkel (2011) for a description of alternative tournament settings in which firms might have an incentive to hire low-ability workers.
ior in tournaments with heterogeneous agents strongly deviates from Nash equilibrium predictions.\footnote{See for example Bull, Schotter, and Weigelt (1987), or Harbring and Lünser (2008).}

Therefore, we briefly discuss the robustness of our results to changes in the theoretical setup, before we test the relevance of our theoretical results for actual behavior in the next section.

**Robustness**  The previous theoretical analysis is restricted along two important dimensions: First, we only consider static tournaments with two, and dynamic tournaments with four workers. Second, we assume that there are always equal shares of strong and weak workers in heterogeneous situations. A more general model with both arbitrary numbers of workers and varying shares of worker types can only be analyzed if the generality of the model is restricted in other dimensions, such as the potential for variations in the precision of the monitoring technology (parameter $r$). However, none of the two restrictions mentioned above affects the main result that the average level of ability has a sizeable effect on total output, while the positive or negative effect of heterogeneity on incentives for effort provision is comparably weak, as we show next.

Two complications arise when one increases the number of workers in the static tournament model and allows for variations in the share of strong and weak workers. First, a closed-form analytical solution is only available for a specific precision of the monitoring technology (lottery contest with $r = 1$) if more than two workers interact.\footnote{See Stein (2002) for a solution of such a model.} Second, if at least two $S$ type workers jointly compete with some workers of type $W$, the latter ones might optimally drop out, i.e., their relative costs of participation in the tournament may exceed potential gains, such that they optimally produce nothing. This will, however, only occur in cases where the strategic disadvantage of weak workers is either extremely high, or alternatively, if each weak worker faces a large number of strong opponents, i.e., if weak workers are very rare. Still, the main results are not affected even in these extreme cases, as the analysis of the case with $r = 1$ for situations with four or eight agents and different shares of strong and weak workers shows.\footnote{Results are available upon request.} Absolute ability does still have a positive effect on total output that is much stronger than the negative effect of heterogeneity. Surprisingly, the relative strength of the effect of heterogeneity on total output is completely independent of the number of competing workers, as long as equal shares of both types compete with each other, as we show in part B of the Appendix. We will come back to this finding in the discussion of the experimental setup.

Finally, consider dynamic promotion tournaments with more than four workers, or with different shares of strong and weak workers. Simply adding stages to the tournament while maintaining the assumption of equal shares of workers of both types implies no qualitative changes as there are simply more pairwise interactions between workers of the same type before the final stage is reached.\footnote{A detailed analysis is available upon request.} One
The result may change, however, if the shares of strong and weak workers are varied, or if workers are
seeded differently in stage 1. Then, heterogeneity can have a negative effect in certain situations, for
example if there are three strong and only one weak worker in the tournament. Yet, it is important
to stress that the dynamic nature of multi-stage tournaments always reduces the (already relatively
weak) negative effect of heterogeneity in comparable static tournaments. The reason is again that the
competition for promotions in later stages of the tournament becomes cheaper for strong workers due
to the existence of weak workers, which induces the strong workers to increase effort and consequently
output production in earlier stages. This effect exists in all dynamic tournaments, but its strength
varies across different specifications, such that it overcompensates the negative effect of heterogeneity
on total output in later stages in some, but not in all cases.

In summary, the result that heterogeneity can have a positive incentive effect on performance
appears to be robust, contrary to the perceived wisdom of a negative effect of heterogeneity on perfor-
mance. The next section investigates the empirical relevance of this result.

3 Experimental Evidence

The theoretical analysis in section 2 provides several testable hypotheses. We use laboratory ex-
periments to test the theoretical predictions, because the use of experimental methods has the clear
advantage that all relevant parameters, in particular in terms of ability, heterogeneity, and the struc-
ture of the tournament, are fixed by the experimental design. This allows us to test the theoretical
predictions in a direct and controlled way. An investigation using an empirical approach with data
from personnel files of companies, like Eriksson (1999), or using data from sports tournaments, would
be more difficult because reliable information about absolute levels of ability, an essential component
of all our theoretical predictions, is typically not available.

3.1 Experimental Design

Following the theoretical model, we assume that agents can be of two different types: Either, they are
of the strong type $S$, or of the weak type $W$. The cost of effort (or disutility of labor) parameter for
weak agents is equal to $c_W = 1.50$ as compared to $c_S = 1.00$ for strong ones. By assumption, productive
ability of both agent types is normalized to one ($a_W = a_S = 1$), such that total output equals total
effort. For the remainder of this section, we will use the term total output.

\footnote{Seedings are considered in detail by Höchtl, Kerschbamer, Stracke, and Sunde (2011). Using the solution of the theoretical model where both stage 1 interactions are heterogeneous (setting $SWSW$) from this paper, one can show that the effect of heterogeneity on total output is weakly negative for low degrees of heterogeneity, but strongly positive if heterogeneity is high. Details are available upon request.}

\footnote{For instance, even very rich sports data typically only provide ordinal rankings of ability. See Abrevaya (2002), Brown (2010), or Sunde (2009).}
Table 1: Theoretical Equilibrium Predictions of Total Output

<table>
<thead>
<tr>
<th></th>
<th>Static Tournament</th>
<th>Dynamic Tournament</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td><strong>Total Output</strong></td>
<td><strong>Total Output</strong></td>
</tr>
<tr>
<td>SSSS\textsubscript{1}</td>
<td>180</td>
<td>SSSS\textsubscript{2}</td>
</tr>
<tr>
<td>SSWW\textsubscript{1}</td>
<td>144</td>
<td>SSWW\textsubscript{2}</td>
</tr>
<tr>
<td>WWWW\textsubscript{1}</td>
<td>120</td>
<td>WWWW\textsubscript{2}</td>
</tr>
</tbody>
</table>

mean(SSSS\textsubscript{1}, WWWW\textsubscript{1}) = 150

Note: Equilibrium predictions for total output, i.e. the sum of individuals efforts, in the specific treatment with a single prize of value 240. Note that total effort provision corresponds to total output given the linear production technology with ability normalized to 1 and heterogeneity affecting effort costs. See text for details.

Apart from \( a_i \) and \( c_i \), there are two additional free parameters in the theoretical model, namely \( P \), which is the value of the bonus payment, or the value of the promotion to the top level position, respectively, and \( r \), which measures the precision of the monitoring technology used by the principal. We set \( P = 240 \), and \( r = 1 \). The choice of \( r = 1 \) has several advantages: First, this case is easy to explain to experimental subjects, which might be the main reason for its popularity in the experimental literature on tournaments.\(^{21}\) Second, as already mentioned previously, this specification allows us to analytically solve a static tournament model with more than two agents, even if agents are heterogeneous. Therefore, we can consider tournaments with four agents for both one-stage and two-stage tournaments, which facilitates comparison, since the tournaments are completely identical in all but one dimension: The one-stage tournament is static, while the two-stage tournament has a dynamic dimension. A solution to the static tournament with four workers is provided in Appendix B, where we also show that the strength of the effect of the degree of heterogeneity on total output is independent of the number of participating workers.\(^{22}\)

Overall, we consider six different treatments, three treatments for the static tournament and three treatments for the dynamic tournament. For both tournament formats, we have one treatment with strong agents only, denoted SSSS\textsubscript{i}, where \( i = 1 \) (\( i = 2 \)) in the tournament with one (two) stages. All agents are of type W in the second homogeneous treatment WWWW\textsubscript{i}, while equal shares are strong and weak in the heterogeneous treatment SSWW\textsubscript{i}. A list of these six treatments and the corresponding theoretical equilibrium predictions for total output in each treatment is provided in Table\(^{[1]}\). This experimental design allows us to test three different hypotheses which directly follow from the theoretical analysis. First, Proposition \(^{[1]}\) implies that total output is increasing in the absolute ability of each participating worker in both static and dynamic tournaments. Therefore, we should observe that total

\(^{21}\)See for example Sheremeta (2011), Sheremeta (2010), and the references provided therein.

\(^{22}\)See also Stein (2002) for details.
output is higher in the treatment with four strong agents than in the treatment with four weak ones; theory predicts that output equals 180 units in $SSSS_i$ for both one-stage and two-stage tournaments, as compared to 120 units in the $WWW_i$ treatments.

**Hypothesis 1 (Ability Effect).** Total output is increasing in the ability of workers in static and in dynamic tournaments. Therefore:

- (a) Total output in $SSSS_1 >$ Total output in $WWW_1$
- (b) Total output in $SSSS_2 >$ Total output in $WWW_2$

Following the order of Propositions in the theoretical analysis, Hypothesis 2 addresses the effect of heterogeneity on total output. According to Proposition 2, heterogeneity reduces total output in static tournaments, while heterogeneity has a positive incentive effect through changes in relative ability on total output in dynamic tournaments. Recall that it is essential that the average level of ability remains constant when comparing homogeneous and heterogeneous specifications. This is not the case in any pairwise comparison of our experimental treatments, which were designed to keep types exactly comparable across the different tournament settings. However, one can easily construct a respective contrast by a simple thought experiment. Suppose that the principal employs eight workers, where four are strong and four are weak. Then, there are (at least) two design options: He can either separate types, which implies that two tournaments with four players each are homogeneous ($SSSS_i$ and $WWW_i$), or he mixes types and designs two heterogeneous tournaments ($2 \times SSWW_i$). Absolute and average abilities are identical in both options. Therefore, the comparison of these two design options allows us to isolate the effect of heterogeneity on total output, while keeping absolute ability constant. In what follows, we will therefore use the average value of total output in treatments $SSSS_i$ and $WWW_i$ and compare this value with total output in the heterogeneous treatment $SSWW_i$, which analogously ensures that absolute ability is unchanged. As Table 1 shows, the average of total output in the homogeneous treatments equals 150 units in both the one-stage and the two-stage tournament. Heterogeneity reduces total output to 144 units in the static tournament model, while total output increases to a value of 152 units due to heterogeneity in the dynamic tournament model with two stages.

**Hypothesis 2 (Incentive Effect).** The incentive effect of heterogeneity on total output depends on the tournament format: The effect is negative in static, and positive in dynamic tournaments, i.e.:

- (a) mean(Total output in $SSSS_1$ and $WWW_1$) > Total output in $SSWW_1$
- (b) mean(Total output in $SSSS_2$ and $WWW_2$) < Total output in $SSWW_2$

Proposition 3 above provides us with another testable hypothesis. According to this Proposition, the effect of changes of the average level of ability on total output is stronger than the corresponding effect of variations in the degree of heterogeneity, if ability and heterogeneity are changed simultaneously. In
terms of our experimental treatments, this implies that we have to compare treatments in which ability and heterogeneity effects work in opposite directions. For the static tournament, this is the case if we compare treatments $SSWW_1$ and $WWWW_1$: Average ability is higher in $SSWW_1$, which should increase total output. At the same time, however, heterogeneity is also higher in $SSWW_1$ than in the homogenous treatment $WWWW_1$, which tends to reduce total output. Theory predicts that the ability effect dominates, since total output amounts to 144 units in $SSWW_1$ and to 120 units in $WWWW_1$, respectively (see Table 1). In the dynamic two-stage tournament setting, we have to compare total output of treatments $SSSS_2$ and $SSWW_2$, since a higher level of average ability positively affects total output in $SSSS_2$, while heterogeneity tends to increase total output in $SSWW_2$. Theory predicts that total output amounts to 180 units in $SSSS_2$ as compared to 152 units in $SSWW_2$ (see Table 1).

**Hypothesis 3 (Relative Strength of Ability and Incentive Effect).** *When the effects of changes in ability and heterogeneity work in opposite directions, the effect of absolute ability on total output dominates the effect of heterogeneity, i.e.,*

(a) Total output in $WWWW_1 <$ Total output in $SSWW_1$

(b) Total output in $SSSS_2 >$ Total output in $SSWW_2$

The theoretical predictions in Table 1 show that there is one additional advantage of considering a static tournament with four rather than two workers in the experimental implementation. According to the theoretical predictions for this specification, total output should be identical in both tournament formats when the worker pool is homogeneous. The design in terms of a static or dynamic tournament should not matter for performance if workers are homogeneous. This theoretical result is well-known and goes back to Gradstein and Konrad (1999). Yet, the table also shows that this result does not hold when workers are of different types: Total output is predicted to be higher in the dynamic setting $SSWW_2$ than in the static specification $SSWW_1$ (152 compared to 144 units, respectively). This constitutes an additional testable hypothesis that serves as a robustness check for our theoretical prediction that the effect of heterogeneity depends on the tournament format.

**Hypothesis 4 (Heterogeneity and Tournament Format).** *Total output is the same in static and dynamic tournament specifications if workers are homogeneous, while total output differs across the two tournament formats if workers are heterogeneous, i.e.,*

(a) Total output in $SSSS_1 = $ Total output in $SSSS_2$

(b) Total output in $WWWW_1 = $ Total output in $WWWW_2$

(c) Total output in $SSWW_1 < $ Total output in $SSWW_2$

23Experimental tests of this hypothesis are also provided by Sheremeta (2010).
3.2 Experimental Implementation

In the experimental sessions, we adopted a between-subject design, such that experimental subjects encountered only one of the six treatments. Each participant played the same tournament 30 times. We use the experimental currency “Taler”, where 200 Taler equal 1.00 Euro. As mentioned previously, we define \( P = 240 \) such that subjects compete for a single prize of 240 Taler in each interaction. Effort provision was implemented in terms of investments into a lottery: Participants were told that they could buy a discrete number of balls in each interaction. The chosen value for \( P \) ensures that equilibrium efforts in all stages and both tournaments are positive integers, which implies that the discrete grid has no consequences for the equilibrium strategies; the equilibrium in pure strategies is unique. The balls purchased by the subjects as well as those purchased by their respective opponents were then said to be placed in the same ballot box, out of which one ball was randomly drawn. This setting reflects the experimental implementation of the monitoring technology with precision \( r = 1 \) from the theoretical set-up. Players had to buy (and pay for) their desired number of balls before they knew whether or not they won the prize in a given tournament. Therefore, each participant received an endowment of 240 Taler in each round to avoid limited liability problems. This endowment could be used to buy balls. In multi-stage treatments, a subject that reached stage 2 could use whatever remained of the endowment to buy balls on the stage 2 interaction. The part of the endowment that a participant did not use to buy balls was added to the payoffs for that round. Since the endowment was as high as the prize that could be won, agents were not budget-constrained at any time. Experimental subjects were told that the endowment could only be used in a given round, transfers across decision rounds were not possible. Therefore, the strategic interaction was the same in each of the 30 decision rounds. Random matching in each round ensured that the same participants did not interact repeatedly. Matching groups corresponded to the entire session. After each decision round, participants were informed about their own decision, the decision(s) of their immediate opponent(s), and about their own payoff. This setting allows for an investigation of whether players learn when completing the task repeatedly. To avoid income effects, however, the participants were told that only four decision rounds (out of 30) would be chosen randomly and paid out at the end of the experiment.

The protocol of an experimental session was as follows for all treatments: First, the participants received some general information about the experimental session. Then, they were given instructions for the respective main treatment (one-stage or two-stage tournament) with four players, which is described above. After each participant confirmed that he/she had understood the instructions on the computer screen, subjects were informed about their type, i.e., about their individual cost parameter \( (c_S = 1.00 \text{ or } c_W = 1.50) \); the assignment of types was random. Subsequently, participants had to answer

\[24\] A translated version of the instructions is provided in Appendix C. The original instructions (which are in German) are available upon request.
a set of control questions to ensure that they had fully understood the instructions. Only once the control questions were answered correctly did the first decision round start. We ran a total of 15 computerized sessions with 20 participants each: Two sessions for each treatment of the static, and three sessions for each treatment of the dynamic tournament. The experiments were implemented using the software z-Tree (Fischbacher, 2007). All 300 participants were students from the University of Innsbruck, which were recruited using ORSEE (Greiner 2004). Each session lasted approximately 1.5 hours, and participants earned between 10-20 Euros (approximately 15 Euros on average).

### 3.3 Experimental Results

Table 2 provides session and first round means of total output for each of the six different treatments. Before assessing the empirical validity of Hypotheses 1–4, we compare the session means to the respective theoretical predictions, which are also shown in Table 2. This comparison reveals a high degree of over-provision of effort by experimental subjects: Observed total output over all decision rounds is substantially higher than theory predicts. Such substantial over-provision is, however, not uncommon in tournament experiments. Sheremeta (2010) reports very similar degrees of over-provision for homogeneous one-stage and two-stage tournaments in treatments which are almost identical to the ones implemented here, and presents evidence that the size of the endowment, which is equal to the prize in both his and our treatments, is responsible for this result. Note, however, that both the endowment and the prize for the winner of the tournament are identical in all treatments, such that differences between treatments, on which Hypotheses 1–4 rely, cannot be attributed to the endowment.

One potential reason for the over-provision of effort is that experimental subjects might have difficulties in determining optimal or equilibrium effort levels. Hence, one might conjecture that experimental subjects “learn” over time and reduce their effort provision correspondingly. Figure 3 shows that this is indeed the case. The degree of over-provision is lower in later than in earlier decision periods, especially for the dynamic tournaments. However, even in later decision rounds total output is substantially higher than theory predicts. Therefore, it seems that learning reduces over-provision only partly. Importantly, learning has virtually no influence on the qualitative relations between different treatments.

---

25 Each experimental session consisted of several parts. At the beginning of each session, subjects were told that they would get 3 Euro show-up fee, and that that the experiment consists of three parts. Part 1 is the main treatment which is described above. Risk preferences were elicited in part 2, and finally distributional preferences in part 3. Subjects received instructions for each part only right before the start of the respective part.

26 See also Sheremeta (2011). He shows that a reduction of the endowment causes a proportional reduction of total effort as long as the endowment is not binding for equilibrium effort levels.

27 This would only be an issue if the endowment is binding in some and not in other treatments. However, the share of experimental subjects who spend their whole endowment does not systematically differ between treatments. If we exclude all observations in which the endowment is binding, for example, total output is somewhat lower in all treatments, but the qualitative findings remain unchanged. Details available upon request.

28 Bull, Schotter, and Weigelt (1987), for example, find that average effort in simple two-person tournaments converges to equilibrium predictions in homogeneous, but not in heterogeneous settings.
Table 2: Total Output

<table>
<thead>
<tr>
<th></th>
<th>Static Tournament</th>
<th></th>
<th>Dynamic Tournament</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>homogeneous</td>
<td>heterogeneous</td>
<td>homogeneous</td>
<td>heterogeneous</td>
</tr>
<tr>
<td>Treatment</td>
<td>Data</td>
<td>Theory</td>
<td>Treatment</td>
<td>Data</td>
</tr>
<tr>
<td></td>
<td>session</td>
<td>1st round</td>
<td>session</td>
<td>1st round</td>
</tr>
<tr>
<td>SSSS&lt;sub&gt;1&lt;/sub&gt;</td>
<td>308.69</td>
<td>325.60</td>
<td>180</td>
<td>SSW&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td>(41.35)</td>
<td>(47.62)</td>
<td></td>
<td>(26.49)</td>
</tr>
<tr>
<td>WWWW&lt;sub&gt;1&lt;/sub&gt;</td>
<td>215.73</td>
<td>190.50</td>
<td>120</td>
<td>WWWW</td>
</tr>
<tr>
<td></td>
<td>(5.40)</td>
<td>(22.03)</td>
<td></td>
<td>(21.22)</td>
</tr>
<tr>
<td>∅</td>
<td>262.21</td>
<td>258.05</td>
<td>150</td>
<td>∅</td>
</tr>
</tbody>
</table>

Note: The numbers in the column “Data” denote total average output observed in all rounds and the first round of the experimental sessions, respectively. Total output is the sum of individuals outputs (in experimental currency, Taler); standard errors in parantheses, based on 2 (10) independent observations for session (1st round) means of the static tournament, and 3 (15) independent observations for session (1st round) means of the dynamic tournament. The column “Theory” provides the theoretical equilibrium prediction for total output production.
treatments, as suggested by a closer look at Figure 3 and the comparison of session and first round means for total output in Table 2.

Hypothesis 1 suggests that total output should be higher if four strong workers compete for a prize than if all workers are weak, independent of the tournament format. Table 2 reveals that the experimental results are qualitatively in line with this theoretical prediction, no matter whether session averages or first round means of total output are considered: In the static tournament, session (first round) averages of total output amount to 308.69 (325.60) units if all workers are strong, as compared to 215.73 (190.50) units in setting \( \text{WWW}_1 \). Similarly, the session (first round) average of total output in the dynamic tournament equals 304.51 (406.93) units in setting \( \text{SSSS}_2 \), while weak workers produce 201.88 (231.40) units on average. To determine the statistical significance of this result, we use t-tests and separately consider session and first round means. We find that the difference between total output levels is significantly different (greater) than zero in both tournament formats; the respective p-values of a two-sided test are 0.152 (0.044) for session and 0.002 (0.001) for first round means of the static (dynamic) tournament, respectively.

We summarize this finding as follows:

**Result 1.** Ability has a strictly positive effect on total output in both the static and the dynamic tournament treatments with homogeneous participants. The strength of this effect is consistent with the theoretical prediction in both tournament formats.

Hypothesis 2 makes different predictions for static and dynamic tournaments. According to this hypothesis, heterogeneity is expected to have a negative effect on total output in the static tournament (Hypothesis 2a), compared to a positive effect in dynamic tournaments (Hypothesis 2b). Table 2 shows that we observe this pattern in the experimental data: Total output is lower in the heterogeneous than in the homogeneous specifications for the static tournament, no matter whether session (220.33 versus 262.21) or first round means (252.40 and 258.05) are examined. The opposite holds for the dynamic tournament model, where session averages of total output amount to 303.75 units in the heterogeneous, compared to 253.20 units in the homogeneous setting; the corresponding values on the first decision round are 381.13 (heterogeneous) and 319.17 (homogeneous), respectively.

Even though we focus on aggregate outcomes rather than individual effort decisions in this paper, it is interesting to note that strong participants are responsible for the direction of the incentive effect of heterogeneity: In the dynamic tournament, production by strong agents is increasing with heterogeneity, particularly in stage 1 (as theory suggests); the opposite holds in the static tournament, where strong workers produce less in the homogeneous than in the heterogeneous situation. Thus, the

\(^{29}\) We test and reject the null hypothesis \( H_0: \) Total output \( \text{WWW}_i = \text{Total output \text{SSSS}}_i \) for \( i = 1, 2 \). We use the parametric t-test rather than the non-parametric Mann-Whitney U-test throughout the paper for consistency reasons. A non-parametric three sample mean test, which we would need to test Hypothesis 2 (see below), is not available. P-values of the Mann-Whitney U-Test for Hypothesis 1 and 3 are very similar to p-values of the t-test, however, and available upon request.
Figure 3: The effect of experience on total output
reaction of strong agents to heterogeneity crucially depends on the tournament structure. Interestingly, this is not the case for weak agents. Independent of the tournament format, both absolute and relative over-provision by weak agents are higher in heterogeneous than in homogeneous settings. This suggests that weak types try to compensate for their strategic disadvantage in any case.

To examine the statistical significance of both parts of Hypothesis 2, we employ a three sample t-test. Testing indicates that the negative incentive effect of heterogeneity in the static tournament is insignificant (p-value > 0.10 both for session and first round means), while the positive effect of heterogeneity on total output is significant in the dynamic tournament; the corresponding p-values are 0.065 for session and 0.05 for first round means. Therefore, even though the incentive effect of heterogeneity is insignificant in static tournaments, it is fair to say that the experimental results are qualitatively in line with the theoretical predictions. In summary, we view this evidence to be consistent with Hypothesis 2.

**Result 2.** The direction of the incentive effect of heterogeneity is in line with the theoretical model. We find that heterogeneity has a

(a) **negative** effect on total output in the static tournament treatments.

(b) **positive** effect on total output in the dynamic tournament treatments.

The data only provide weak evidence for Hypothesis 3, which is on the relative strength of the effect of heterogeneity and ability on total output. According to the theoretical model, the effect of changes in the degree of heterogeneity on total output (be it positive or negative) is always weaker than the corresponding effect of variations in the level of average ability if the degree of heterogeneity and the average ability level are changed jointly. That is, theory predicts for the static tournament that total output is higher in treatment \(SSW_W_1\) than in \(WWW_W_1\), since the negative effect of heterogeneity on total output is dominated by the positive effect of higher average ability. For the dynamic tournament, total output in treatment \(SSS_S_2\) is predicted to be higher than in \(SSW_W_2\), since the negative effect due to the reduction of average ability is more pronounced than the positive effect of an increase of the degree of heterogeneity when moving from a situation with only strong workers to a setting where equal shares are strong and weak. Figure 3 shows that the ability and the incentive effect of heterogeneity are

---

30 Further details on individual effort decisions are available upon request.

31 Note that this response of weak agents to a strategic disadvantage has previously been documented for static tournaments with heterogeneous participants (Bull, Schotter, and Weigelt 1987, van Dijk, Somemans, and van Winden 2001, Harbring and Lünser 2008).

32 Formally, we test the hypotheses \(\mu(SSS_S_1) + \mu(WWW_W_1) = 2\mu(SSW_W_1)\) and \(\mu(SSS_S_2) + \mu(WWW_W_2) = 2\mu(SSW_W_2)\), where \(\mu(X)\) is the average total output in setting \(X \in \{SSS_S, WWW_W, SSW_W\}\), \(i = 1, 2\). The corresponding test statistic is

\[
T = \frac{\mu(SSS_S_i) + \mu(WWW_W_i) - 2\mu(SSW_W_i)}{\sqrt{\frac{\sigma(SSS_S_i)^2}{n(SSS_S_i)} + \frac{\sigma(WWW_W_i)^2}{n(WWW_W_i)} + \frac{4\sigma(SSW_W_i)^2}{n(SSW_W_i)}}}
\]

with \(n(SSS_S_i) + n(WWW_W_i) + n(SSW_W_i) - 3\) degrees of freedom.
equally strong both for the static and the dynamic tournament, since there is hardly any difference between the total output produced in treatments $SSW_1$ and $WWW_1$, or $SSS_2$ and $SSW_2$, respectively. When considering the session means, we find that total output equals 215.73 units in treatment $WWW_1$ and 220.33 units in $SSW_1$, which is qualitatively in line with the theoretical prediction. However, the difference is statistically insignificant (p-value > 0.10) and much lower than expected (2% rather than 20%)\textsuperscript{33} The pattern is almost the same in the dynamic tournament treatments, where the session means of total output amount to 304.51 units in setting $SSS_2$, compared to 303.75 units in $SSW_2$ (difference equals 0.2% in the data, compared to 18% predicted by theory). Again, the differences between treatments are insignificant for session and first round means (p-value > 0.10 in both cases).\textsuperscript{33}

These results indicate that either the effect of changes in the level of average ability is much weaker, or that the effect of variations in the degree of heterogeneity is much stronger than theory predicts. Experimental results suggest the latter explanation, since the strength of the pure ability effect is in line with theory, while the incentive effect of heterogeneity is much stronger than expected: Theory predicts that output should be approximately 33% lower in treatment $WWW_i$ than in treatment $SSS_i$, independent of the tournament format ($i = 1, 2$). In fact, session means of total output are 31% (33%) lower in the static (dynamic) tournament with only weak workers than in the corresponding treatment with only strong workers. The isolated incentive effect of heterogeneity, however, is much more pronounced than expected: While theory predicts that total output decreases by 4% in the static tournament, the session means show a reduction of more than 15%. Similarly, the session means of total output increase by almost 20% in the dynamic tournament, while the increase should be slightly more than 1% according to the theoretical model.\textsuperscript{35}

\textbf{Result 3.} \textit{When the ability effect and the incentive effect of heterogeneity work in opposite directions, the effects are approximately equally strong and offset each other in both the static and the dynamic tournament treatments; the incentive effect of heterogeneity is stronger than predicted.}

Finally, we briefly consider Hypothesis 4, which makes predictions about the relation between static and dynamic tournaments: Total output should be the same if workers are homogeneous, while this measure is predicted to differ across the two tournament formats if workers are heterogeneous. This pattern emerges from the results in Table 2 when considering session means: Total output is similar when only strong or only weak workers compete with each other (308.69 vs. 304.51, and 215.73 vs. 201.88, respectively), while the difference between the two tournament models is remarkable in the heterogeneous settings (220.33 compared to 303.75). Based on two-sided t-tests, we cannot reject the

\textsuperscript{33} Even the difference of total output in the first decision round, which is somewhat higher, is statistically insignificant. We test whether $H_0 : \text{Total output in } SSW_1 - \text{Total output in } WWW_1 < 0$. $H_0$ cannot be rejected.

\textsuperscript{34} We test whether $H_0 : \text{Total output in } SSS_2 - \text{Total output in } SSW_2 < 0$. $H_0$ cannot be rejected.

\textsuperscript{35} Both for the static and the dynamic tournament, we compare both the theoretical precition and the session means of $SSW_i$, and mean($SSS_i, WWW_i$) for $i = 1, 2$, respectively.
null hypothesis of equal means in either comparison of the homogeneous settings, $SSSS_1$ and $SSSS_2$, and $WWW_1$ and $WWW_2$, respectively. However, the null of equality of total output being equal in settings $SSWW_1$ and $SSWW_2$ can be rejected with a p-value of 0.051).

When comparing the first round rather than the session means of the heterogeneous treatments, the results are qualitatively unchanged. Total output equals 252.40 units in the static, compared to 381.13 units in the dynamic tournament. This difference is statistically significant (p-value < 0.01). However, first round means of total output are much higher in the homogeneous treatments of dynamic tournaments (325.60 vs. 406.93, and 190.50 vs. 231.40 for $SSSS_i$ and $WWW_i$ ($i \in \{1, 2\}$, respectively), even though the differences are statistically insignificant (p-value > 0.10). While our findings for the homogeneous treatments differ from previous results in the literature when considering session means, it is interesting to note that our results are comparable with previous findings if we use the first decision period only. For instance, Sheremeta (2010) suggests that effort provision is higher in the two-stage tournament with homogeneous participants due to joy of winning. His experimental design allows for learning, but he uses a mixture of between and within subject comparison, while we employ a between-subject design. Altmann, Falk, and Wibral (2011) also find that effort provision is higher in dynamic tournaments when considering one-shot interactions, and a between-subject comparison. The remarkable difference between session and first round means in dynamic tournaments (which we do not observe for static tournaments), suggests that learning patterns may differ between static and dynamic tournaments. This conjecture receives some support when considering Figure 3. There is a downward trend of total output in dynamic tournaments, i.e., experimental subjects seem to realize in later rounds that their initial effort provision was too high. This is different in static tournaments, where total output starts at comparably lower initial levels. Note that the differential learning trends in static and dynamic tournaments are of no consequences for the our main results with respect to Hypotheses 1–3, since we compare different specifications of the same tournament format, and learning trends seem to be very similar for different treatments within a certain tournament class.

**Result 4.** Total output does not differ significantly between static and dynamic tournament treatments if participants are homogeneous. In line with the theoretical predictions, however, there is a difference between the two tournament formats in the heterogeneous treatments; this difference is more pronounced than predicted by theory.

Overall, Result 4 provides additional support for the theoretical results with respect to the effects of ability and heterogeneity. The data match all qualitative relations not only within, but even across different tournament formats, in particular when using session means.

---

36 This p-value refers to a t-test of equality of session means for total output across the two settings.
37 The investigation of the dynamic patterns is an interesting topic for future work that directly compares different tournament formats.
4 Conclusion

This paper analyzes the effect of variations in the degree of heterogeneity between workers on their performance in corporate tournaments. The analysis considers the two most prominent tournament formats in the Personnel Economics literature, namely static one-shot tournaments, in which participants compete for a bonus payment, and dynamic multi-stage tournaments, which are often used to model promotion tournaments in companies with several hierarchy levels. The theoretical results show that the ability of workers who participate in the tournament has a strong impact on the overall level of effort provided, and therefore also on the total amount of output produced. The results also suggest that the common perception that heterogeneity between participants is detrimental for incentives and performance in tournaments is correct in static one-stage tournaments. In dynamic multi-stage elimination tournaments, however, the effect of heterogeneity on incentives for effort provision is strictly positive, i.e., incentives for effort provision are higher in tournaments with heterogeneous participants than in tournaments with homogeneous agents if the level of average ability is held constant. This is because the negative incentive effect for weak workers is more than compensated for by a higher value of promotions for strong workers, who anticipate that their chances for promotions in the future are higher if there is a chance that they have to compete with a weak opponent in later stages. This possibility strongly increases the value of a promotion today, and consequently incentives to provide effort in early stages of the tournament. From the theoretical analysis, it also follows that the effect of heterogeneity is rather weak in both tournament models, and in particular much weaker than the effect of changes of the average ability level of tournament participants on performance, which is measured by output throughout the paper. The second part of the paper presents evidence from laboratory experiments that is largely consistent with the theoretical predictions. The experimental findings suggest that ability of workers has a strong impact on the performance in both tournament formats. Further, we find that incentives and consequently the overall performance are lower in static tournaments with heterogeneous subjects than in comparable treatments with homogeneous participants. The experimental results suggest, however, that the negative effect of heterogeneity in static tournaments is stronger than predicted by theory. The pattern is similar in our dynamic tournament treatments, where the direction of the heterogeneity effect is in line with the theoretical prediction, but the effect is again stronger than predicted. We find that heterogeneity is associated with a strongly increased effort provision in dynamic tournaments. Concerning the relative strength of ability and heterogeneity effects, we find that the empirical effects of heterogeneity and of changes of the average ability level of tournament participants are of similar size, which is in contrast to the theoretical prediction. However, this suggests that that the influence of the tournament structure on incentives in case of heterogeneity might be much more important in reality than it is in theory.
The results have important implications for both practical applications and future research. The theoretical analysis and the experimental results show that the ability of workers can be very important for performance, potentially more important than a homogeneous workforce, even if tournament compensation schemes are used for the provision of incentives. Second, heterogeneity is not necessarily bad per se in a corporate tournament setting. Whether or not corporations should assign similarly productive employees or different types to the same tournament crucially depends on its structure. An important topic for future is the surprisingly strong effect of heterogeneity between participants on behavior in the experimental treatments, both in the static tournament, where the effect is negative, and in the dynamic tournament, where the effect is positive.
References


ALTMANN, S., A. FALK, AND M. WIBRAL (2011): “Promotions and Incentives: The Case of Multi-


BROWN, J. (2010): “Quitters Never Win: The (Adverse) Incentive Effects of Competing with Super-

Study,” Journal of Political Economy, 95(1), 1–33.


Journal, 109(458), 536–545.


GÜRTLER, O., AND M. KRÄKEL (2010): “Optimal Tournament Contracts for Heterogeneous Workers,”


Appendix

A Proofs

Proof of Proposition 1: In the static tournament model, we consider three different specifications, namely $SS$, $WW$, and $SW$. Inspection of equation (13) immediately reveals that the respective expressions are strictly increasing in the ability measure $(\frac{a_S}{c_S})$, or $(\frac{a_W}{c_W})$, or both, respectively, when keeping $\phi$ constant. The same holds for all three settings of the dynamic tournament model, namely $SSSS$, $WWWW$, and $SSWW$, as inspection of equations (14) and (15) immediately reveals.

Proof of Proposition 2: We separately consider parts (a) and (b) of the Proposition:

(a) From (13), total output in the heterogeneous specification $SW$ is given by

$$Y(SW) = r \left( \frac{a_S}{c_S} + \frac{a_W}{c_W} \right) \frac{\phi}{[1 + \phi]^2} P.$$

Recall that $\phi = \left( \frac{a_SC_W}{a_Wc_S} \right)^r$ measures heterogeneity in terms of relative ability. Due to the assumption that workers of type $S$ have a higher productive ability and/or a lower dis-utility of labor than these of type $W$, $\phi > 1$ holds, which implies that the degree of heterogeneity increases with $\phi$. Next, we determine total output in a homogeneous setting where workers have the same average ability level. Workers in the heterogeneous setting have an average ability level of

$$\bar{a}/\bar{c} = \frac{1}{2} \left( \frac{a_S}{c_S} + \frac{a_W}{c_W} \right).$$

Then, it follows from (13) that total output in a homogeneous tournament with two workers with ability $\bar{a}/\bar{c}$ amounts to

$$Y(hom) = r \left( \frac{a_S}{c_S} + \frac{a_W}{c_W} \right) \frac{1}{4} P.$$

To prove that heterogeneity has a negative effect on total output, we have to show that $Y(hom) > Y(SW)$ does always hold, i.e.

$$r \left( \frac{a_S}{c_S} + \frac{a_W}{c_W} \right) \frac{1}{4} P > r \left( \frac{a_S}{c_S} + \frac{a_W}{c_W} \right) \frac{\phi}{[1 + \phi]^2} P.$$

$$\iff \frac{1}{4} > \frac{\phi}{[1 + \phi]^2}$$

$$\iff \phi^2 + 2\phi + 1 > 4\phi$$

$$\iff (\phi - 1)^2 > 0$$
Since $\phi > 1$ by construction, this relation is always satisfied. Moreover, the strength of the detrimental effect of heterogeneity is increasing in $\phi$, i.e., the higher $\phi$, the larger is the difference between the homogeneous and the heterogeneous setting. Therefore, total output is decreasing in the degree of heterogeneity between workers, which completes the proof.

(b) Total output in the heterogeneous specification $Y^{(SS\overline{WW})}$ is given by

$$Y^{(SS\overline{WW})} = r \left[ \left( \frac{a_S}{c_S} \right) \cdot \frac{\phi^2 + (3 - r)\phi}{2[1 + \phi]^2} + \left( \frac{a_W}{c_W} \right) \cdot \frac{1 + (3 - r)\phi}{2[1 + \phi]^2} \right] P.$$ 

Recall that $\phi = \left( \frac{a_{SW}}{a_{WS}} \right)^r$ measures heterogeneity in terms of relative ability. Due to the assumption that workers of type $S$ have both a higher productive ability and a lower disutility of labor than these of type $W$, $\phi > 1$ holds. This implies that the degree of heterogeneity increases with $\phi$.

Next, we have to determine total output in a homogeneous setting where workers have the same average ability level. Workers in the heterogeneous setting have an average ability level of

$$\bar{a} = \frac{1}{2} \left( \frac{a_S}{c_S} + \frac{a_W}{c_W} \right).$$

Then, it follows from (14) that total output in a homogeneous tournament with four workers with ability $\bar{a}$ amounts to

$$\bar{Y}(\text{hom}) = r \left( \frac{\bar{a}}{c} \right) \frac{4 - r}{4} P = r \left( \frac{a_S}{c_S} + \frac{a_W}{c_W} \right) \frac{4 - r}{8} P.$$

To prove that heterogeneity has a positive effect on total output, the relation $Y^{(SS\overline{WW})} > \bar{Y}(\text{hom})$ has to hold, i.e.

$$r \left[ \left( \frac{a_S}{c_S} \right) \cdot \frac{\phi^2 + (3 - r)\phi}{2[1 + \phi]^2} + \left( \frac{a_W}{c_W} \right) \cdot \frac{1 + (3 - r)\phi}{2[1 + \phi]^2} \right] P > r \left( \frac{a_S}{c_S} + \frac{a_W}{c_W} \right) \frac{4 - r}{8} P$$

$$\Leftrightarrow \left( \frac{a_S}{c_S} \right) (4\phi^2 + (12 - 4r)\phi) + \left( \frac{a_W}{c_W} \right) (4 + (12 - 4r)\phi) > \left( \frac{a_S}{c_S} + \frac{a_W}{c_W} \right) (4 - r)[1 + \phi]^2$$

To simplify the subsequent analysis, we make some normalizing assumptions: We assume that $a_w = c_w = c_S = 1$, which is without loss of generality as long as $a_S > 1$ does hold. Then, it follows from the definition of $\phi$ that $\phi = a_S^r$. This implies the following relation:

$$a_S(4a_S^{2r} + (12 - 4r)a_S^r) + 4 + (12 - 4r)a_S^r > (a_S + 1)(4 - r)[1 + a_S^2]$$

$$\Leftrightarrow (a_S^r - 1)(4a_S - r - ra_S + a_S^r(r + ra_S - 4)) > 0$$

$$\Leftrightarrow G(a_S, r) \equiv r(a_S^{r+1} - 1) - (4 - r)(a_S^r - a_s) > 0$$
Note that $G(a_S, r)$ is equal to zero if $a_S = 1$, while we are interested in the properties of $G(a_S, r)$ when $a_S > 1$ does hold. Therefore, we proceed in two steps: First, we will show that the slope of $G(a_S, r)$ is strictly positive when $a_S = 1$. Second, we will prove that the slope is strictly increasing, which implies that $G(a_S, r)$ is strictly increasing if $a_S > 1$. Since $G(a_S, r)$ is a continuous function, this will prove the claim that $G(a_S, r) > 0 \forall a_S > 1$.

(i) The first derivative of $G(a_S, r)$ with respect to $a_S$ reads

$$\frac{\partial G(a_S, r)}{\partial a_S} = r(r+1)a_S^r - (4-r)(ra_S^{r-1} - 1).$$

For $a_S = 1$, this derivative simplifies to the term $2(r-1)^2$, which is positive. Consequently, the slope of $G(a_S, r)$ at the point $a_S = 1$ is strictly positive.

(ii) The second derivative of $G(a_S, r)$ with respect to $a_S$ reads

$$\frac{\partial^2 G(a_S, r)}{\partial a_S^2} = r(r-4)(r-1)a_S^{r-2} + r^2(1+r)a_S^{r-1} = ra_S^{r-2}[r^2 - 5r + 4 + (r^2 + r)a_S]$$

Note that this second derivative is strictly greater than zero since $a_S > 1$. This proves that output is always strictly larger in the heterogeneous than in the homogeneous setting, since the measure of the difference between the two settings is always greater zero, i.e. $G(a_S, r) > 0 \forall a_S > 1$. Further, since $G(a_S, r)$ is also strictly in increasing in $a_S$, the strength of the positive effect of increases in the degree of heterogeneity increases.

Therefore, we find that total output is strictly increasing in the degree of heterogeneity between workers, which completes the proof.

Proof of Proposition 3: We start by proving the proposition for the static tournament model in part (a) of this proof, and subsequently consider the dynamic tournament model in part (b).

(a) Recall that total output in the heterogeneous setting, $Y(\text{SW})$ equals

$$Y(\text{SW}) = r \left( \frac{a_S}{c_S} + \frac{a_S}{c_g} \right) \frac{\phi}{[1+\phi]^2} P.$$ 

We know from Propositions 1 and 2 that increases in the ability level have a positive effect on total output, while a higher degree of heterogeneity has a negative effect. To prove the claim that the effect on total output of a change in the average ability level of workers dominates the

\[38\text{If } a_S \text{ were equal to one, one would obtain the relation } ra_S^{r-2}[2(r-1)^2 + 2] > 0. \text{ When } a_S > 1, \text{ the relation is even more positive.}\]

\[39\text{Note that higher values for } a_S \text{ immediately imply a higher degree of heterogeneity, since average ability is held constant by construction.}\]
corresponding effect of a change in the degree of heterogeneity if both ability and heterogeneity are changed simultaneously, we compare the heterogeneous setting with a homogeneous setting in which the average level of ability is lower. Proposition 3 is proven if we can show that total output is always higher in the heterogeneous setting than in the homogeneous setting with a lower average ability. Therefore, we show that total output in the heterogeneous setting $SW$ is always higher than in a situation with weak workers only ($WW$). Recall that total output in the latter case amounts to

$$Y(WW) = \frac{a_u}{c_u} \left( \frac{1}{2} \right)^P.$$ 

Consequently, we have to show that $Y(SW) > Y(WW)$ for all $\frac{a_u}{c_u} > \frac{a_y}{c_y}$. To simplify the subsequent analysis, we assume that $a_z = c_y = c_z = 1$, which is without loss of generality if the relation $0 < a_y < 1$ does hold. This gives

$$\Leftarrow r(1 + a_y) \frac{\phi}{(1 + \phi)^2} P > \frac{a_y}{2} P$$

$$\Leftarrow 2(1 + a_y) \frac{\phi}{(1 + \phi)^2} - a_y > 0$$

Now, recall that $\phi = \frac{1}{a_y}$ by definition. This gives $K = 2a_y - a_y^{2r+1} - a_y$. To complete the proof, we have to show that $K > 0$ for all $0 < a_y < 1$. We must take account of one additional contraint, since the derived equilibrium solution is only valid under Assumption 1. Therefore, we have to prove that

$$K = 2a_y - a_y^{2r+1} - a_y > 0$$

for all values of $a_y$ and $r$ that satisfy

$$(i) \quad 0 < a_y < 1$$

$$(ii) \quad r \leq 1 + a_y^r$$

We start by presenting a formal proof for the special case $r = 1$, which we use in our experiments and which is prominent in the literature. For $r = 1$, constraint (ii) is automatically satisfied. $K$ simplifies to

$$K = 2a_y - a_y^3 - a_y = a_y \left( 1 - a_y^2 \right) > 0.$$  

Since (ii) is satisfied by definition and $K > 0$ does hold, $Y(SW) > Y(WW)$ does hold for $r = 1$, independent of the degree of heterogeneity, which completes the proof.
An analytical proof for the general case with variable exponent $r$ and the nonlinear constraint (ii) is more involved. Therefore, we provide a graphical proof by ways of a range plot instead, which indicates for which values of $a_W$ and $r$ all conditions are simultaneously satisfied. Note that we can restrict attention to the parameter space where $0 < a_W < 1$ due to (i) and $0 < r < 2$ due to (ii). Figure 4 provides the plot. The relation $K > 0$ holds in both the light and the dark area, but not in the white one. Therefore, the plot shows that $K > 0$ is not satisfied for small $a_W$ and high values of $r$. Note, however, that condition (ii) excludes this area, since the pure strategy equilibrium which we consider throughout the paper does not exist for high values of $r$ when heterogeneity is high: The dark area indicates in the range in which condition (ii) holds. Consequently, the condition $K > 0$ is less restrictive than $r \leq 1 + a_W^r$, which proves the claim that $Y(SW) > Y(WW)$ does hold for all values of $a_W$ and $r$ for which the pure strategy equilibrium we consider exists (i.e., under Assumption 1).

(b) Recall that total output in the heterogeneous setting $Y(SSWW)$ equals

$$Y(SSWW) = r \left[ \left( \frac{a_S}{c_S} \right) \cdot \frac{\phi^2 + (3 - r)\phi}{2[1 + \phi]^2} + \left( \frac{a_W}{c_W} \right) \cdot \frac{1 + (3 - r)\phi}{2[1 + \phi]^2} \right] P.$$
We know from Propositions 1 and 2 that increases in the ability level as well as increases in
the degree of heterogeneity have a positive on total output. To prove the claim that the effect
on total output of a change in the average ability level of workers dominates the corresponding
effect of a change in the degree of heterogeneity if both ability and heterogeneity are changed
simultaneously, we compare the heterogeneous setting with a homogeneous setting in which the
average level of ability is higher. Proposition 3 is proven if we can show that total output is
always higher in the homogeneous setting with a higher average ability than in the heterogeneous
setting. Therefore, we show that total output in the homogeneous setting \( SSSS \) is always higher
than total output in a situation with both strong and weak workers \( SSWW \). Recall that total
output in the former situation amounts to

\[
Y(SSSS) = r \left( \frac{a_S}{c_S} \right) \frac{4 - r}{4} P.
\]

Consequently, we have to show that \( Y(SSSS) > Y(SSWW) \) for all \( \frac{a_S}{c_S} > \frac{a_W}{c_W} \). To simplify the subse-
quint analysis, we assume that \( a_S = c_S = c_W = 1 \), which is without loss of generality if the relation
\( 0 < a_W < 1 \) does hold. Recall that \( \phi = \frac{1}{a_W} \) by definition. Using these relations, we get

\[
\begin{align*}
Y(SSSS) &> Y(SSWW) \\
\iff \frac{4 - r}{4} P &> r \left[ \frac{1 + (3 - r)a_W^r}{2[1 + a_W^r]^2} + a_W^r + (3 - r)a_W^r \right] P \\
\iff (4 - r)[1 + a_W^r] &> 2 \left[ 1 + (3 - r)a_W^r + a_W^r(a_W^r + (3 - r)a_W^r) \right]
\end{align*}
\]

We must take account of one additional contraint, since the equilibrium solution derived is only
valid under Assumption 1. Consequently, we have to prove that

\[
M \equiv (4 - r)[1 + a_W^r]^2 - 2 \left[ 1 + (3 - r)a_W^r + a_W^r(a_W^r + (3 - r)a_W^r) \right] > 0
\]

for all values of \( a_W \) and \( r \) that satisfy

\[
(i) \quad 0 < a_W < 1 \\
(ii) \quad r \leq 1 + a_W^r
\]

We start again by presenting a formal proof for the special case \( r = 1 \), which we use in our
experiments and which is prominent in the literature. For \( r = 1 \), constraint (ii) is automatically
satisfied and $M$ simplifies to

$$M = 3(1 + a_W)^2 - 2[1 + 2a_W + a_W(a_W^2 + 2a_W)] = (2a_W + 1) (1 - a_W^2) > 0$$

Since (ii) is satisfied by definition and $M > 0$ does hold, this proves that $Y(SSSS) > Y(SSWW)$ does hold for $r = 1$, independent of the degree of heterogeneity.

An analytical proof for the general case with variable exponent $r$ and the nonlinear constraint (ii) is more involved. Therefore, we provide a graphical proof by ways of a range plot which indicates for which values of $a_W$ and $r$ all conditions are satisfied. Note that we can restrict attention to the parameter space where $0 < a_W < 1$ due to (i) and $0 < r < 2$ due to (ii). Figure 5 provides the plot. The relation $M > 0$ holds in both the light and the dark area. Therefore, the plot shows that $M > 0$ is satisfied over the whole range of parameter values which are allowed for the pure strategy equilibrium. The dark area indicates in the range in which condition (ii) holds. Condition $M > 0$ is thus less restrictive than $r \leq 1 + a_W^r$, which proves the claim that $Y(SSWW) > Y(WWWW)$ does hold for all values of $a_W$ and $r$ for which the pure strategy equilibrium we consider exists (i.e., under Assumption 1).
B Static Tournament with more than two Workers

Assume that there are \( N \) participants in the tournament, where equal shares are strong and weak, i.e., there are \( \frac{N}{2} \) workers of each type, where \( N \geq 2 \). If \( r = 1 \), an arbitrary player \( s \) of the strong type faces the following optimization problem:

\[
\max_{x_s} \Pi_s = \frac{a_s x_s}{a_s x_s + a_s \sum_{i \neq s} x_i + a_w \sum_{j=1}^{N/2} x_j} - c_s x_s,
\]

i.e., player \( s \) chooses his effort \( x_s \) in such a way as to maximize his expected payoff \( \Pi_s \), taking as given the effort choices of his opponents of both player types. Similarly, an arbitrary weak player \( w \) chooses the optimal level of effort \( x_w \), and his maximization problem reads:

\[
\max_{x_w} \Pi_w = \frac{a_w x_w}{a_w x_w + a_s \sum_{i=1}^{N/2} x_i + a_w \sum_{j \neq w} x_j} - c_w x_w.
\]

Taking derivatives with respect to \( x_s \) and \( x_w \), respectively, gives the first order optimality conditions for an interior Nash equilibrium in which all agents participate:

\[
as_s \left( \frac{a_s \sum_{i \neq s} x_i + a_w \sum_{j=1}^{N/2} x_j}{(a_s \sum_{i=1}^{N/2} x_i + a_w \sum_{j=1}^{N/2} x_j)^2} \right) P - c_s = 0
\]

\[
a_w \left( \frac{a_s \sum_{i=1}^{N/2} x_i + a_w \sum_{j \neq w} x_j}{(a_s \sum_{i=1}^{N/2} x_i + a_w \sum_{j=1}^{N/2} x_j)^2} \right) P - c_w = 0.
\]

Next, symmetry among participants of the same type is imposed, i.e. \( x_i^* = x_s^* \forall i = 1, \ldots, \frac{N}{2} \), and \( x_j^* = x_w^* \forall j = 1, \ldots, \frac{N}{2} \). Equalizing the remaining two first order conditions and using the definition \( \phi = \frac{a_w}{a_s} \) gives the equilibrium ratio of efforts, which is characterized by the following relation:

\[
\Phi \equiv \frac{a_s x_s^*}{a_w x_w^*} = \frac{1}{\phi - \frac{N}{2} (\phi - 1)},
\]

Combining this relation with either of the first-order conditions above delivers equilibrium efforts

\[
x_s^* = \left( \frac{1}{c_s} \right) \left( \frac{N}{2} - 1 \right) + \Phi \frac{N}{2} P, \quad \text{and} \quad x_w^* = \left( \frac{1}{c_w} \right) \frac{N}{2} \frac{\Phi (\frac{N}{2} - 1)}{\phi - \frac{N}{2} (\phi - 1)}.
\]

As a consequence, total output in this heterogeneous tournament amounts to

\[
Y(\frac{N}{2} \cdot S, \frac{N}{2} \cdot W) = \left( \frac{a_s}{c_s} \right) \left( \frac{N}{2} - 1 \right) + \Phi \frac{N}{2} P + \left( \frac{a_w}{c_w} \right) \Phi \frac{N}{2} \left( \frac{N}{2} - 1 \right) + \Phi \frac{N}{2} P.
\]
From this, one can derive total output for a homogeneous specification with the same average ability. By imposing the condition \( \Phi = 1 \), we get

\[
Y(\text{hom}) = \left( \frac{a_s}{c_s} \right) \frac{N - 1}{2N} P + \left( \frac{a_w}{c_w} \right) \frac{N - 1}{2N} P.
\]

To determine the relative strength of the heterogeneity effect on total output in percent, one can compute the difference between output in the heterogeneous and the homogeneous specification, and normalize, i.e.

\[
\frac{Y\left( \frac{N}{2} \cdot S, \frac{N}{2} \cdot W \right) - Y(\text{hom})}{Y(\text{hom})}.
\]

Inserting the respective expression and simplifying gives

\[
Q = - \left[ \left( \frac{a_s}{c_s} \right) - \left( \frac{a_w}{c_w} \right) \right]^2 \left[ \left( \frac{a_s}{c_s} \right) + \left( \frac{a_w}{c_w} \right) \right].
\]

Assume, for example, that \( a_s = a_w = c_s = 1 \), while \( c_w = 1.5 \) as in the experimental treatment in section 3 of the paper. Then, the above expression shows, that total output is 4% lower in a heterogeneous setting where weak agents are 50% weaker than strong ones, compared to a setting with homogeneous participants who have the same average ability level. Obviously, \( Q \) is independent of \( N \), i.e., heterogeneity has the same relative strength when equal shares of both strong and weak types participate, independent of the overall number of participants. In particular, this implies that theory predicts the same effects for changes of the degree of heterogeneity in a tournament with two (as in section 2) or four participants (as in the experimental part of the paper).
C Experimental Instructions

C.1 General Instructions

WELCOME TO THIS EXPERIMENT AND THANK YOU FOR YOUR PARTICIPATION

General Instructions:

You will participate in 3 different experiments today. Please stop talking to any other participant of this experiment from now on until the end of this session. In each of the three experiments, you will have to make certain decisions and may earn an appreciable amount of money. Your earnings will depend upon several factors: on your decisions, on the decisions of other participants, and on random components, i.e. chance. The following instructions explain how your earnings will be determined.

The experimental currency is denoted Taler. In addition to your Taler earnings in experiments 1 to 3, you receive 3 EURO show-up fee. You may increase your Taler earnings in experiments 1 to 3, where 2 Taler equal 1 Euro-Cent, i.e.

200 Taler correspond to 1 Euro.

At the end of this experimental session your Taler earnings will be converted into Euro and paid to you in cash.

Before the experimental session starts, you receive a card with your participant number. All your decisions in this experiment will be entered in a mask on the computer, the same holds for all other participants of the experiment. In addition, the computer will determine the random components which are needed in some of the experiments. All data collected in this experiment will be matched to your participant number, not to your name or student number. Your participant number will also be used for payment of your earnings at the end of the experimental session. Therefore, your decisions and the information provided in the experiments are completely anonymous; neither the experimenter nor anybody else can match these data to your identity.

We will start with experiment 1, followed by experiments 2 and 3. The instructions for experiments 2 and 3 will only be distributed right before the respective experiment starts, i.e. subsequent to experiments 1 and 2, respectively.

You will receive your earnings in cash at the end of the experimental session.
C.2 One-stage Treatment

Experiment 1

Overall, there are 30 decision rounds in Experiment 1. The course of events is the same in each decision round. You will be randomly and anonymously placed into a group of four participants in each round, and the identity of participants in your group changes with each decision round.

Course of events in an arbitrary decision round
All four participants of your group receive an endowment of 240 Taler at the beginning of a decision round. The endowment can be used to buy a certain amount of balls. The costs for the purchase of a ball are not the same for all participants:

There are equal shares of high (H) and low (L) cost types in each group of four participants, i.e. there are two participants of each type in a group of four. All experimental participants are informed about their type at the beginning of the experiment. Types do not change with decision rounds, such that you face either high or low costs in each of the 30 decision rounds.

Participants of type H have to pay 1.50 Taler for each ball they buy, i.e.

1 ball costs 1.50 Taler
2 balls cost 3.00 Taler
(and so on)

Participants of type L have to pay 1.00 Taler for each ball they buy, i.e.

1 ball costs 1.00 Taler
2 balls cost 2.00 Taler
(and so on)

Apart from differences in terms of costs per ball, there is no difference between participants of type H (high cost) or type L (low cost).

When deciding how many balls you want to buy, you do not know the decision of other participants. Also, your decision is not revealed to any other participant. All balls which were bought by four participants of a group are placed into a ballot box. One ball is randomly drawn from the ballot box, and each ball is drawn with the same probability. Assume, for example, that all balls which you bought are green colored. Then, the probability that one of your balls is drawn satisfies

\[
\text{probability(green ball is drawn)} = \frac{\#\text{green balls}}{\#\text{green balls} + \#\text{balls of other participants in your group}}
\]

where \# is short for number. The same probability rule does also hold for other participants in your group. Consequently, the probability that one of your balls in drawn is higher

- the more balls you purchased
- the less balls the other participants in your group purchased.

The computer simulates the random draw of a ball. If all participant of a group of four choose to buy zero balls, each participant wins with the same probability of 25%.

Only the participant whose ball is drawn from the ballot box receives a prize of 240 Taler in a given decision round. The other participants do not receive any prize.
Your Payoff
Assume that you bought X balls in some decision round. There are two possibilities for your payoff:

1) one of your balls was drawn from the ballot box

\[ \text{Your Payoff} = \text{endowment} - X \times \text{your cost/ball} + \text{prize} \]
\[ = 240 \text{ Taler} - X \times \text{your cost/ball} + 240 \text{ Taler} \]

2) none of your balls was drawn from the ballot box

\[ \text{Your Payoff} = \text{endowment} - X \times \text{your cost/ball} \]
\[ = 240 \text{ Taler} - X \times \text{your cost/ball} \]

Note that your cost/ball are 1.00 Taler (if you are of type L) or 1.50 Taler (if you are of type H), respectively.
Therefore, your payoff is determined by the following components: by the number of balls you buy (X); by your cost type (high or low); by a random draw (one of your balls is (not) drawn). The same holds for any other participants of the experiment. Note, however, that costs per ball differ between participants.

Information: You will learn your type before the first decision round starts. The information will be provided on the computer screen. Your type (cost per ball) be same in all 30 decision rounds. At the end of each decision round, you will learn whether or not one of your balls was randomly drawn and how many balls the other participants in your group bought in total. In addition, you will be informed about your payoff.

Decision: In each of the 30 decision rounds you have to decide how many balls you want to buy. You have to enter this number into the respective field on the computer screen. When making this decision, you do know your own type (high or low costs) and the type of the other participants in your group. An example of the decision screen is shown below.
Your Total Payoff: Four out of 30 decision rounds are paid. These rounds are randomly determined, i.e. the probability that some decision round is paid is identical ex-ante for all 30 decision rounds. You will receive the sum of payoffs for the respective decision rounds at the end of the experiment.

Remember:
You receive an endowment of 240 Taler at the beginning of each decision round and have to decide how many balls you want to buy. Overall, there are three additional participants in each group who face the same problem. The identity of these participants is randomly determined in each decision round. However, it always holds that equal shares of participants in a given group are of type L (1.00 Taler per ball) and type H (1.50 Taler per ball), respectively.

If you have any questions, please raise your hand now!
C.3 Two-stage Treatment

Experiment 1

Overall, there are 30 decision rounds with two stages each in Experiment 1. The course of events is the same in each decision round. You will be randomly and anonymously placed into a group of four participants in each round, and the identity of participants in your group changes with each decision round.

Course of events in an arbitrary decision round

All four participants of each group receive an endowment of 240 Taler at the beginning of a decision round. The endowment can be used to buy a certain amount of balls in two subsequent stages of a decision round. It is important to note that you receive one endowment only which must suffice to buy balls in both stages. The costs for the purchase of a ball are not the same for all participants:

There are equal shares of high (H) and low (L) cost types in each group of four participants, i.e. there are two participants of each type in a group of four. All experimental participants are informed about their type at the beginning of the experiment. Types do not change with decision rounds, such that you are either a high or a low cost type in each of the 30 decision rounds. The same holds for all other participants of the experiment.

Participants of type H have to pay 1.50 Taler for each ball they buy in stage 1 and stage 2, i.e.

- 1 ball costs 1.50 Taler
- 2 balls cost 3.00 Taler
  (and so on)

Participants of type L have to pay 1.00 Taler for each ball they buy in stage 1 and stage 2, i.e.

- 1 ball costs 1.00 Taler
- 2 balls cost 2.00 Taler
  (and so on)

Apart from differences in terms of costs per ball, there is no difference between participants of type H (high cost) or type L (low costs).

When deciding how many balls you want to buy, you do not know the decision of other participants. Also, your decision is not revealed to any other participant.

All interactions in the experiment are pairwise. Assume that you are in one group with participant A, participant B, and participant C. Then, you interact with participant A in stage 1, while participants B and C simultaneously meet each other in the second stage 1 interaction. If you reach stage 2, you will interact either with participant B or C, depending on the outcome in the second stage 1 interaction. In stage 1, there are two ballot boxes:

- all balls bought by you or participant A are placed in ballot box 1
- all balls bought by participants B and C are placed in ballot box 2

One ball is randomly drawn from each ballot box, and each ball drawn with the same probability. The two participants whose balls are drawn from ballot box 1 and 2, respectively, reach stage 2; the decision round is over for the other two participants (whose balls were not drawn), i.e. they drop out from this decision round. Any participant has to pay the balls he or she bought in stage 1, whether or not he/she reached stage 2. The respective amount is deducted from the endowment.
The two participants who reached stage 2 do again buy a certain number of balls, using whatever remains from the endowment they received after costs for balls in stage 1 were deducted. The balls are then placed into ballot box 3. One ball is randomly drawn from ballot box 3. The participant whose ball is drawn receives a prize of 240 Taler. The other participants do not receive any prize in this decision round. Independent of whether or not a participant receives the prize, he/she does always have to pay for the balls bought in stage 2.

Let’s take a closer look at the random draw of balls from ballot boxes. Assume, for example, that all balls which you bought are green colored, and that you interact with participant A in stage 1. Then, the probability that one of your balls is drawn (such that you make it to stage 2) satisfies

\[
\text{probability(green ball is drawn)} = \frac{\# \text{ green balls}}{\# \text{ green balls} + \# \text{ balls by participant A}}
\]

where \# is short for number. The same probability rule does also hold for other participants in your group. Consequently, the probability that one of your balls in drawn is higher

- the more balls you purchased
- the less balls the other participant with whom you interact purchased.

The computer simulates the random draw of a ball. If all participant of a group of four choose to buy zero balls, each participant wins with the same probability of 25%.
Your Payoff

Assume that you bought "X1" balls in stage 1, and that you buy "X2" balls whenever you reach stage 2. Then, there are three possibilities for your payoff:

1) None of your balls is drawn in stage 1

Your Payoff = endowment − X1 * your cost/ball
= 240 Taler − X1 * your cost/ball

2) one of your balls is drawn from the ballot box in stage 1; in stage 2, none of your balls is drawn

Your Payoff = endowment − X1 * your cost/ball − X2 * your cost/ball
= 240 Taler − X1 * your cost/ball − X2 * your cost/ball

3) one of your balls is drawn from the ballot box in stage 1; also, one of your balls is drawn in stage 2

Your Payoff = endowment − X1 * your cost/ball − X2 * your cost/ball + prize
= 240 Taler − X1 * your cost/ball − X2 * your cost/ball + 240 Taler

Note that your cost/ball are 1.00 Taler (if you are of type L) or 1.50 Taler (if you are of type H), respectively.

Therefore, your payoff is determined by the following components: by the number of balls you buy in stage 1 ("X1"); by the number of balls you buy in stage 2 ("X2") if you reach it; by your cost type (high or low); by up to two random draws (one of your balls is drawn/not drawn in stage 1 and potentially stage 2). The same holds for any other participants of the experiment. Note, however, that costs per ball differ between participants.

Information: You will learn your type before the first decision round starts. The information will be provided on the computer screen. Your type (cost per ball) will be same in all 30 decision rounds.

- Before making the first decision in stage 1, you will learn the type of participant A whom you meet in stage 1, i.e. you learn whether participant A has to pay 1.00 Taler (type L) or 1.50 Taler (type H) for each ball he/she buys.
- After you made your decision in stage 1, you are informed whether or not you can participate in stage 2, i.e. whether or not one of your balls was drawn from ballot box 1.
- If you did not reach stage 2, you are informed about how many balls participant A bought in stage 1.
- If you reach stage 2, you receive information about the remaining endowment (after costs for the purchase in stage 1 are deducted), and about the type of the other participant whom you meet in stage 2.
- After you made your decision in stage 2, you learn whether or not one of your balls was drawn from ballot box 3 and how many balls the participants who you met in stages 1 and 2, respectively, bought. Further, you learn your payoff for the respective decision round.

Decision: In each of the 30 decision rounds you have to decide how many balls you want to buy in stage 1. If you reach stage 2, you face a similar decision in stage 2. In both cases, you have to enter a number into a field on the computer screen. An example of the decision screen in stage 1 is shown below.
**Your Total Payoff:** Four out of 30 decision rounds are paid. These rounds are randomly determined, i.e., the probability that some decision round is paid is identical ex-ante for all 30 decision rounds. You will receive the sum of payoffs for the respective decision rounds.

**Remember:**
You receive an endowment of 240 Taler at the beginning of each decision round and have to decide how many balls you want to buy in stage 1; if you reach stage 2, you have to decide again. Overall, there are three additional participants in each group who face the same problem. The identity of these participants is randomly determined in each decision round. However, it always holds that equal shares of participants in a given group are of type L (1.00 Taler per ball) and type H (1.50 Taler per ball), respectively.

If you have any questions, please raise your hand now!