Competitiveness in dynamic group contests: Evidence from combined field and lab data

Yann Girard Florian Hett*

April 2013

Abstract

We analyse data from a field setting in which students participate in a dynamic group contest with feedback. We combine this information with a laboratory measure of competitiveness. We find that competitive groups perform worse overall. In addition, we find that participants react to intermediate performance: A better rank in a given period increases the number of points in the subsequent period, even after controlling for group and time fixed effects. The effect is significantly stronger for competitive groups. We show that this difference in the sensitivity to dynamic incentives can explain the overall negative effect of competitiveness on performance.

Keywords: Dynamic contest, competitiveness, field experiments, lab experiments, rank feedback.

JEL Classification Codes: C91, C93, D03, D74, I21.

---

*Girard: Goethe University Frankfurt, GSEFM, Campus Westend, Grüneburgplatz 1 (House of Finance), 60323 Frankfurt, Germany, girard@uni-mainz.de. Hett: GSEFM and University of Mainz, Jakob-Welder-Weg 4, 55128 Mainz, Germany, florian.hett@uni-mainz.de. We are grateful for financial support from the "Inneruniversitäre Forschungsförderung" of the University of Mainz. We thank Florian Heiss, Katharina Klingler, Constantin Weiser and Amelie Wuppermann for sharing their data, as well as Michael Kosfeld, Markus Kröll, Mario Mechtel, and Isabel Schnabel for valuable comments and discussion. We also benefited from comments by participants of the economics workshop at the IAAEU (University of Trier), the Annual Meeting of the Verein für Socialpolitik 2012, the 2013 Annual Conference of the Royal Economic Society at Royal Holloway, University of London and the IMEBE 2013 in Madrid.

1 Introduction

Many aspects of life can be interpreted as contests, i. e. situations in which participants spend effort to compete over a given prize. Prominent examples include hirings, promotions and compensation in the workplace as well as admission decisions and performance evaluations in education. A large theoretical literature analyzes this huge variety of possible setups.[1] While ample empirical evidence largely supports the main theoretical predictions from this literature, *"there is significant dispersion in the behavior of individual subjects"* (Dechenaux et al., 2012) in these studies. Given the pervasiveness and importance of contest structures for many economic outcomes, it is crucial to investigate the sources of this heterogeneity of behavior to better assess the consequences of contest mechanisms in real settings.

In this paper, we shed light on this heterogeneity of the effect of contest incentives. We provide empirical evidence that incentive effects of relative performance schemes with feedback systematically depend on individual characteristics of contest participants. Using field data from a group contest among students, we find that groups whose members are more competitive perform worse overall, where competitiveness is a behavioral measure taken from the laboratory. Moreover, we show that this negative effect of competitiveness can actually be linked to the dynamic incentives induced by the contest: The reaction to intermediate performance, i. e. the sensitivity to dynamic incentives, is stronger for competitive groups than for non-competitive groups. This stronger path-dependence of performance for competitive participants represents a dynamic amplification mechanism, translating bad initial performance into even worse overall performance. We can even show that this amplification is the source of the overall worse performance of competitive groups.

Generally, our results confirm the theoretical notion that dynamic contests with feedback do create dynamic incentives. However, they also point to their ambivalent nature, being able to motivate but also to de-motivate participants. As a novel result, our analysis shows that individuals' responses to these dynamic incentives systematically vary. This implies that introducing contest schemes may not only have effects on efficiency and aggregated effort provision, but also distributional effects. This seems of particular importance in the already mentioned examples of education and the workplace. Taking our results at face value, they

[1] See Lazear and Rosen (1981) for an early contribution. Holmstrom (1999) and Holmstrom and Milgrom (1987) discuss dynamic settings and Holmstrom (1982) analyzes contests among teams. For the general theory of contests, see Konrad (2009) for a recent textbook treatment. Dechenaux et al. (2012)(Dechenaux et al., 2012) summarize the experimental evidence.

imply that the contest-like nature of many parts of life shifts individual outcomes not only with respect to obvious characteristics like ability, but also with respect to competitiveness.

Secondary to this main result, our setup allows to test and distinguish several theoretical predictions from the literature regarding the determinants of overall performance in group contests. In that matter, the theoretical literature provides arguments that group size can have either a positive or a negative effect on aggregated effort levels and thus performance. While the former follows the simple idea that larger groups can provide more aggregated effort for similar costs-per-member, the latter argues that free-riding problems increase in group size. In addition to group size, we also analyze whether unconditional and conditional cooperation (Fischbacher et al., 2001) affect group effort and success.[2] We find a statistically and economically significant positive effect of group size on performance. Supplemental survey evidence also suggests that individual effort is rather higher in larger groups than in smaller ones, working even more against the free-riding hypothesis. Considering the effect of conditional cooperation, we find a positive effect. However, this effect is not statistically significant, although the point estimate is quite large.

Our paper contributes to several other strands of the literature. First, we explore whether our behavioral measure of competitiveness, closely related to the experimental design of Niederle and Vesterlund (2007), is correlated with individuals' contest behavior. Despite the huge popularity of this measure, a field validation is still missing. Our finding that competitive groups react stronger to intermediate performance supports the usefulness of the measure, as stronger sensitivity to contest incentives seems to be a plausible operationalization of competitiveness from an economic perspective.

In addition, our paper adds to the debate on the reliability of behavioral measures as predictors for field behavior. Eliciting personal characteristics as determinants of decision making in a robust way is crucial for empirical research of decision behavior (Borghans et al., 2008; Heckman, 2011). However, there are only few studies analyzing the relationship of field behavior and laboratory measures (Camerer, 2011). Our results encourage the use of behavioral measures for the investigation of field behavior, as the data from the lab have significant predictive power. This is of particular importance as field setups are naturally limited with respect to the elicitation of precise behavioral measures. However, many questions depend on a field setting as specific contextual variables are often crucial for behavioral outcomes. In this sense, evidence on the reliability of the interchanged usage of these two types of data

[2]See Rustagi et al. (2010); Kosfeld and von Siemens (2011) for corresponding literature.

is valuable.

The paper is structured as follows: In section 2 we present the existing related literature in more detail, while section 3 presents details on the field contest we analyze and how the data have been obtained. Section 4 derives the hypotheses that are subsequently tested in section 5. In section 6 we conclude and discuss avenues for further research.

2 Related literature

This paper relates to three main strands of the literature, whose main contributions are presented in greater detail in this section. First, we discuss the empirical literature on contests and tournaments, beginning with static contests and later proceeding to dynamic contests. After that, we summarise the relevant literature with respect to competitiveness. Finally, we link our paper to other studies combining lab and field data.

A characteristic application of contests and tournaments is the provision of monetary incentives to agents in work relationships. For instance, Delfgaauw et al. (2012b) analyze the effect of a short-term sales competition in a Dutch retail store chain. Varying incentive pay and rank feedback based on store level performance, they do find a positive effect of competition on performance, but no effect associated with varying financial incentives. Blanes i Vidal and Nossol (2011) find that if workers are given private feedback on their individual rank of pay as well as productivity increases performance.[3] Barankay (2012a,b) conducts two field experiments where only rank feedback was provided without any additional direct financial incentives. In the first study, he finds that rank information about quality reduces performance in quantity, while quality remains unchanged. In the second study he finds, again, a negative effect associated with rank information and that this effect increases in the scope of effort substitution, i. e. the possibility to switch to other tasks, when performance is worse than expected.

While these studies focus on contests among individuals, some papers investigate contests among groups. Ryvkin (2011) shows theoretically how effort in such contests depends on the sorting of heterogeneous players into groups. Bandiera et al. (2012) provide field evidence on group contests, finding that the introduction of tournament incentives causes a change in group composition: Prior to the introduction of the contest group formation was

[3] The major drawback in this study is that they do not have a contemporaneous control group and therefore cannot clearly identify the feedback effect from a time effect.

predominantly driven by friendship. Once the contest started, workers of similar abilities were more likely to form groups. In line with that, Ivanova-Stenzel and Kübler (2005) find that gender differences in team performance depend on the composition of teammates.

In reality, contests rarely follow the structure of a one-shot game, as usually information evolves over time and people have multiple occasions to exert effort. However, only few papers explicitly focus on the incentive effects of these dynamic structures. Delfgaauw et al. (2011) look at a two-stage elimination tournament with 208 Dutch retail stores and find that a more convex prize spread increases performance in the second round but at the expense of performance in the first round.[4] Casas-Arce and Martinez-Jerez (2009) analyze a multiperiod contest among retailers. They theoretically show that dynamic contests result in *"hump-shaped incentives"*: Participants with a strong initial lead reduce their effort, as well as those trailing far behind. Only those close to the cut-off provide higher effort. In fact, they also provide empirical evidence that effort decreases in 'leading' and 'trailing distance'. In another study by Delfgaauw et al. (2012a), they look at a four-week sales tournament on the store level with weekly performance feedback. They find that stores that fall far behind do not respond to the tournament anymore, but performance increases within the reach (probability) of winning the bonus. On average, tournament incentives do not lead to higher performance.

Second, our paper is related to the literature analyzing individual preferences for competition. Here, the primary goal of most articles has been studying gender differences in competition.[5] However, a common underlying of these papers is that they also implicitly investigate how competitiveness differs between subjects in general. The first experimental papers on performance differences under competition are by Gneezy et al. (2003) and Gneezy and Rustichini (2004). Later Niederle and Vesterlund (2007) introduce a design that is now commonly used. Here, subjects can choose between a competitive and non-competitive incentive scheme and thereby "reveal" their preference for competition. Recent papers on competitiveness discuss, for instance, when the competitiveness gap emerges (Sutter and Rützler, 2010), how competitiveness affects pro social behavior (Houser and Schunk, 2009) and how policy changes, like affirmative actions, change outcomes (Niederle et al., 2010; Balafoutas and Sutter, 2010; Dargnies, 2009).

Finally, our paper contributes to the debate on the generalizability of behavioral results

[4]This result is consistent with a) the theoretical results by Moldovanu and Sela (2001) regarding the prize structure on perfromance and b) with the discouragement effect by Konrad (2009).

[5]See Niederle and Vesterlund (2011) for a recent survey on the subject.

from the lab to field settings (Levitt and List, 2007a,b, 2008; Falk and Heckman, 2009; Camerer, 2011). Despite the current relevance of this discussion, we do not find but one paper with respect to contests or tournaments and competitive behavior. However, there is already a broad array of findings with respect to other characteristics and settings.

For example, Antonovics et al. (2009) study strategic elimination of opponents in the game show *The weakest link* and compare this with behavior in the lab. Other setups in which papers try to match lab and field settings as closely as possible are sports card trading (List, 2003, 2006), donation behavior (Benz and Meier, 2008), risk behavior in lotteries and game shows (Östling et al., 2011; Post et al., 2008), public good contribution (Stoop et al., 2010), bidding behavior (Isaac and Schnier, 2006), consumption behavior (Abeler and Marklein, 2010) and application of mixed strategies (Palacios-Huerta and Volij, 2008; Palacios-Huerta, 2003). Similar to our paper, there are several studies using lab measures to explicitly analyze how they are correlated with field behavior. Time- and risk measures, for example, are found to be correlated with contribution to pension funds (Barr and Packard, 2000) and present-biased preference are associated with larger credit card balances (Meier and Sprenger, 2010). Dictator game allocations are correlated with working time (Barr and Zeitlin, 2010) and predict whether subjects volunteer to help others (Carpenter and Myers, 2007). Behavior in trust games is correlated with worker income (Clark et al., 2010), the proportion of high reciprocators within a firm is strongly correlated with output per worker (Barr and Serneels, 2004) and less trustworthy players' default rates on microfinance loans are higher (Karlan, 2005). Behavior in laboratory public good games has predictive power in the field as well: Public goods contributions predict limited common resource pool extraction (Fehr and Leibbrandt, 2011; Lamba and Mace, 2011), group-level conditional cooperation is correlated with success in managing forest commons (Rustagi et al., 2010) and the willingness of actual leaders to enforce organizational norms in a third party punishment game is correlated with group outcomes (Kosfeld and Rustagi, 2011), as well.

3 Contest design, measurement and data collection

3.1 Design of the study group contest

Our analysis refers to the bachelor course "mathematics" ("math 101") in the winter term 2011/12 at the University of Mainz in Germany. In addition to regular tutorials there was the opportunity to voluntarily participate in a study group contest ("Übungsgruppenwettbewerb"). Students endogenously formed study groups with up to five members to submit weekly problem sets (quizzes). The problems covered material from the lectures. The answers were graded and each week a group level ranking was released. In the final week the best three study groups were awarded a certificate to honor their exceptional performance in mathematics. In the first lecture the study group contest was announced and rules and policies were presented to the students. All required information was available in detail on the chair's website, such that students not present were able to receive the necessary information as well. The instructor emphasised that there are no direct financial incentives, but that every member of the three "best" (performing) study groups would receive a certificate at the end of the course indicating the "exceptional performance in mathematics." Each member of a winning group received one certificate giving the contest prize the nature of a public good within a group. Although there were no direct financial benefits from the certificates, it was argued that they can be useful for applications for internships and the like.[6]

First, students had to decide whether to participate in the study group contest. Those who wanted to participate had to decide if they want to compete on their own or to form a group. There was a pre-announced deadline until which students had to subscribe to the contest in order to participate. Within these days, 78 study groups with 230 students in total had formed. Overall, there were 10 problem sets to solve. There was a common weekly deadline up to which each study group had to submit one and only one solution. Afterwards, the solutions were graded and points achieved were announced. Additionally, the current group ranking, based on aggregated points over the course of the contest, was published. All participants had access to all information about the history of the group ranking and achieved points of all groups.

The study group contest was organised in an online learning management system designed for education and professional training (ILIAS). In order to participate in the contest, every study group had to register once with its study group name. Problem sets were distributed via ILIAS, the solutions had to be entered by the group themselves, and the results, as well as the rankings, were also released in ILIAS. After the study group contest had been finished and the awards were handed out to the winners, participants and non-participants were asked

[6]One might argue about the characteristics of this type of reward being materialistic or not. On the one hand, the certificate might be seen as useful for application procedures and the like in which case it should induce incentives. On the other hand, there is also evidence that non-monetary awards (Kosfeld and Neckermann, 2011) can have substantial incentive effects.

to answer a questionnaire. In order to maintain a high response rate, three Amazon.de gift coupons were raffled off under all participants. The questionnaire was intended to elicit a self-assessment of their mathematical abilities, from non-participants the motivation not to participate and from participants their motivation, the group formation process, the working behavior within their group and the evaluation of the contest itself.

3.2 Behavioral measures

In addition to the data on the contest, we use laboratory measures about subjects' individual competitiveness and individual cooperativeness. These measures stem from a parallel study, in which we elicited behavioral measures for a sample of incoming students at two different points in time, October 2011 and July 2012. As not all participants of the study group contest took part in this study and of those who did, not all participated in both rounds, we end up with either zero, one or two observations per student. We proceed as follows: If there are two observations per student we take the newer one, i. e. the one from July 2012, otherwise we take the older observation from October 2011. If there is no observation, the respective student is excluded from those parts of the analysis where behavioral measures are included.

To elicit subjects' individual competitiveness we employ a real effort task.[7] The standard design by Niederle and Vesterlund (2007) implies that subjects choose between two different incentive schemes: A competitive and a non-competitive scheme. Under the latter subjects are paid according to a piece-rate for every question answered correctly, while under the former subjects compete against another randomly chosen subject over a given price.[8] We enhance this basic design in order to be able to further discriminate between competitive types: The piece-rate contained an ex-ante unknown fixed payment component F

$$U_i = 0.5e_i + F \qquad \text{with}\, F = \{3, \dots, 12\}.$$

Subjects had to state their reservation fixed payment with a list-price method that indicated at which fixed payment level subjects want to perform the real effort task under the piece-rate scheme instead of the tournament scheme

[7] As real-effort task we used adding five two-digit numbers as in Bartling et al. (2009).

[8] The contest thereby simply compares the number of questions correctly answered and pays out a fixed prize to the one with the higher number.

$$U_i = \begin{cases} 20 & \text{if } e_i \geq e_{-i}, \\ 20 \text{ with p=0.5} & \text{if } e_i = e_{-i}, \\ 0 & \text{otherwise.} \end{cases}$$

A larger reservation fixed payment indicates a higher preference for competition, as the subject is willing to forgo higher monetary earnings associated with the piece-rate scheme.

In addition, we measure subjects' individual cooperativeness with a standard public good game (Ledyard, 1995). Four subjects, each endowed with EUR 10.00 decide how much to contribute to a common project. Subjects' payoffs are given by

$$U_i(e_i, e_{-i}) = 10 - e_i + 0.5 \sum_{j=1}^{4} e_j,$$

where e_i is the amount invested by individual i. This game represents the structure of a social dilemma: While the social optimum can be achieved if each player contributes his entire endowment to the project, individual payoff maximization implies to invest nothing. However, usually in public good games, we observe that people partially cooperate, i. e. invest at least some amount into the common project. One particular form of contribution behavior has caught the attention of the literature: Individuals who are willing to contribute more the more others contribute to the project, a behavior that has been coined "conditional cooperation." In order to grasp this concept we apply the design by Fischbacher et al. (2001): First, subjects decide how much to contribute to the project without further information. Thereafter they are asked to state their preferred contribution level for every average contribution level of the other group members. In addition to their unconditional contribution level, subjects can be classified according to their conditional contribution plan in different behavioral types. Although there are in principle more possible types, we focus on the distinction among only three of them, namely if an individual is an unconditional cooperator, a conditional cooperator or neither of the two.

4 Deriving hypotheses

4.1 Group size and contest performance

In our setup the contest unit varies in the number of their members and might impact group's performance in the contests. In that matter, the theoretical literature provides arguments that group size might have either a positive or a negative effect on aggregated effort levels and thus performance. The most obvious direct effect through which group size affects aggregated effort and performance is to assume homogenous members such that every individual faces the same optimization problem and thus chooses the same effort level. As aggregated effort is simply the sum of all individual effort levels, aggregated effort then strictly increases in group size.

Hypothesis 1a *(Positive group size effect) Larger groups exert higher aggregated effort and perform better in the study group contest.*

A conflicting argument was made by Olson (1965), arguing that smaller groups are more successful. Here the idea is that the free-rider problem is more severe within large groups compared to smaller groups. Therefore if this (indirect) negative free-riding effect dominates the positive direct group size effect then in equilibrium large groups exert *less* effort.

Hypothesis 1b *(Group size increases free riding incentives) Smaller groups exert higher aggregated effort and perform better in the study group contest.*

4.2 Dynamic incentives

We now explicitly take it into account and derive hypotheses about *dynamic* contest behavior. First, we present the framework introduced by Casas-Arce and Martinez-Jerez (2009, p. 1311) to derive the relationship between effort and past performance. Following that, we formalise the notion of competitiveness and derive how it affects the sensitivity to past performance.

Consider a tournament with N identical participants. A fraction μ of participants will win one of k identical prizes of value V. The prizes are awarded to the best performers on the basis of output produced over time. In the study contest output represents aggregated points. Each participant i produces

$$\tilde{y}_{it} = e_{it} + \tilde{u}_{it},$$

where $\tilde{y}_{it}$ is a measure of performance which is increasing in effort e_{it} and depends on a stochastic shock $\tilde{u}_{it}$, representing anything beyond the control of the participant. $\tilde{u}_{it}$ is assumed to be independent and identically distributed (i.i.d.), with a differentiable pdf ψ and cdf Ψ. Providing effort incurs cost of $c(e_{it})$, where it is assumed that $c(0) = 0$ and marginal costs increase in effort $c''(e_{it}) > 0$.

We assume that there are two time periods, $t = 1, 2$. At $t = 1$ all participants are identical and therefore they reach a symmetric equilibrium with $e_{i1} = e_1^*$ for all i. In $t = 2$ prizes will be allocated according to the distribution of realised performances

$$y_i = y_{i1} + y_{i2}.$$

Note that despite identical effort levels of all participants in the first round, realised performances will differ due to the realizations of idiosyncratic shocks. After the last period there exists a performance threshold $\bar{y}_\mu$ such that player i wins a prize iff $y_1 \geq \bar{y}_\mu$. If we assume that the number of participants is sufficiently large (infinity), the threshold is non stochastic and can be perfectly predicted by the law of large numbers.

Anticipating the required total performance $\bar{y}_\mu$ sufficient to win, a subject will maximise expected profits conditional on realised period 1 performance y_{i1}

$$U_i = p(y_{i1} + y_{i2} \geq \tilde{y}_\mu) \cdot V - c(e_{i2}).$$

To satisfy the first-order conditions marginal benefit has to equal marginal costs with $p(y_{i1} + y_{i2} \geq \tilde{y}_\mu) = 1 - \Psi(\tilde{y}_\mu - y_{i1} - e_{i2})$. Normalizing the prize to 1, the optimality condition is given by

$$\psi(\tilde{y}_\mu - y_{i1} - e_{i2}^*) = c'(e_{i2}^*).$$

This condition yields an optimal effort level e_{i2}^* in relation to past performance y_{i1}. Investigating how this optimal effort level is related to past performance leads to the following proposition:

Proposition 1. *The relationship between e_{i2}^* and y_{i1} is hump shaped.*

The corresponding proof can be found in the appendix and holds as long as ψ is unimodal. The intuition for the proposition is as follows: Participants who built up a strong lead after period 1 (i.e. had a large positive shock realised) will lose the contest only due to a large negative shock. For participants with a large gap (low y_{i1}) to the top need a relatively large positive shock to win the prize. A large positive as well as a large negative shock are very unlikely and effort cannot help to close the gap.

Therefore, the marginal impact of effort on the probability to win or lose is small which implies a small optimal effort level in period 2 as well. In contrast, if participants are close to the top, the marginal probability of winning is more sensitive to effort. Overall, this reasoning implies optimal effort to increase the narrower the gap to the winning spots of the contest gets. Proposition 1 therefore implies the following hypothesis for our empirical anaysis:

Hypothesis 2 *(Dynamic incentives) The better the rank of a group in a given period, the more points the group will achieve in the subsequent round.*

In principal, the model presented above not only predicts the last hypothesis, but also a negative effect on performance once a group has built up a strong lead. However, in our data this actually never happens and point differences at the top remain very narrow throughout the contest. As a consequence, using this data set we are simply not able to test this particular prediction of the model and thus not explicitly formulate it as a hypothesis.

4.3 Competitiveness and dynamic incentives

We are now interested in the effect heterogeneity of dynamic incentives with respect to competitiveness. However, given the lack of a formal definition of competitiveness in the literature, we will first provide an easy formalization that captures our understanding of what competitiveness represents. We adjust the already introduced utility function in the following way:

$$U_i = \theta_i \, p\left(y_{i1} + y_{i2} \geq \tilde{y}_\mu\right) \cdot V + (1 - \theta_i) \cdot \gamma \, e_{i2} - c(e_{i2}).$$

The pure utility from winning the price is now weighted by an individual specific factor θ_i. $(1 - \theta_i) \cdot \gamma \, e_{i2}$ represents the intrinsic utility from carrying out the task. In our case this could for example capture the idea that an individual's motivation to take part in the contest is a commitment to studying. Note that marginal utility gained from winning a contest prize and direct marginal utility from exerting effort are inversely related via θ_i. This could be the case if an individual's participation motive is rather explained by the fact that the contest might act as a study commitment device than by the desire to actually win the contest. Therefore, individuals differ with respect to their source of motivation to participate in the contest. Some subjects might be more motivated by the outlook to win the prize and others are more intrinsically motivated to carry out the task. The relative importance of these two sources of motivation is directly determined by the level of θ_i, while an increase in θ_i can be interpreted as an increase in subject's competitiveness, reflecting the fact that more competitive individuals care more about winning the contest than less competitive ones.

Based on the modified utility function and after normalizing V to 1, the corresponding optimality condition is

$$\theta_i \psi(\tilde{y}_\mu - y_{i1} - e^*_{i2}) + (1 - \theta_i)\,\gamma = c'(e^*_{i2}).$$

In the appendix we show that proposition 3 still holds in this extended setup. In addition, we are now able to analyze how the relationship between current effort and past performance relates to the level competitiveness:

Proposition 2. *The sensitivity of e^*_{i2} to y_{i1} increases in θ_i.*

The proof can again be found in the appendix. The proposition states that the current effort is more sensitive to the ranking position (and thus past performance) for groups that are more competitive. The intuition behind this result directly follows from the way competitiveness is modelled: The basic dynamic incentive effect entirely builds on the difference in the marginal effect of effort on the probability of winning and how it changes with past performance. A higher level of competitiveness reflects a stronger weight put on actually winning the contest compared to the intrinsic valuation of carrying out the task. As a result, competitive groups will adjust their optimal effort level more strongly to their position in the ranking than non-competitive groups. The corresponding hypothesis for our empirical analysis therefore reads:

Hypothesis 3 *(Dynamic incentives and competitiveness) The effect of current rank on contemporaneous performance is stronger for competitive groups than for non-competitive groups.*

Table 1: Summary statistics

Variable	Obs	Mean	Std. Dev	Min	Max
Total points	78	203.21	133.16	0	429.25
Final rank	78	38.60	21.51	1	70
Period points	780	20.32	20.17	0	69
Period rank	780	30.64	19.47	1	70
Group size	78	2.95	1.47	1	5
D.Workedalone	78	0.23	0.42	0	1
D.Male	78	0.75	0.34	0	1
Cooperation	47	5.05	1.83	0	10
D.Conditional cooperation	47	0.51	0.33	0	1
Competitiveness	47	4.23	1.59	0	7
D.Competitiveness	47	0.49	0.50	0	1

5 Empirical analysis

5.1 Data overview

Table 1 shows group level descriptive statistics of the variables included in our analysis. Overall, 230 students organised in 78 groups took part in the contest. As Panel A in Figure 1 shows the distribution of group size was rather uniform giving us enough variation. Out of these 78 groups, 47 had at least one member that participated in our experimental session, yielding 47 groups with information on group member behavioral measures. Overall we have behavioral measures for 57.4% of the participants. The average coverage ratio (the share of members with information on competitiveness) among groups is 47.6%.[9] Restricting the sample to those groups on which we have at least one observation, i. e. to groups we can include into any regression with competitiveness as an explanatory variable, we have information on competitiveness for 78.3% of participants and the average coverage ratio among groups is 79.8%. Panel B depicts the actual distribution of coverage ratios. As all regressions in which competitiveness is included are carried out in the restricted sample, we are confident that the respective coverage ratio of almost 80% is sufficient.

D.Workedalone is a dummy variable taking on the value of 1 if the group consists just of one member, and 0 otherwise. Overall, there are 18 "groups" with students working on their own. *D.Competitiveness* is a dummy variable that takes on the value of 1 if the group

[9] The difference is explained due to the variation in group size.

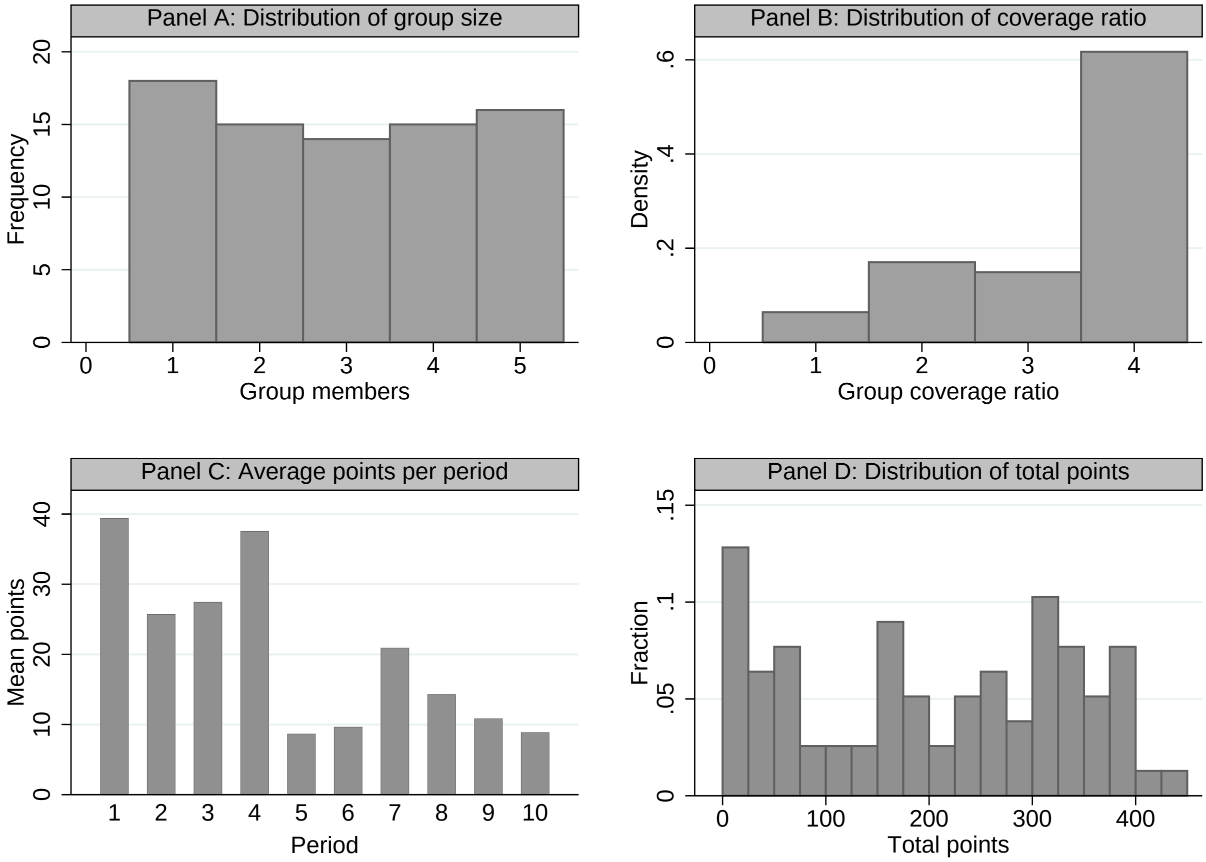

Figure 1: (Some) Distributions

average of *Competitiveness* is above the median and 0 otherwise. *Cooperation* is the group average amount out of an endowment of 10 units that is (unconditionally) contributed in a standard public good game. *Conditional cooperation* is the group average of a dummy variable on the individual level, which is 1 if the student is a conditional cooperator and 0 otherwise. The classification of conditional cooperation thereby follows the approach by Fischbacher et al. (2001).

Table 2 shows correlation coefficients among the main variables. We see that conditional cooperation and cooperation have a significant correlation of approximately 0.4. This implies that in order to disentangle their individual effects, we should include them jointly in our regressions. Furthermore, we see that competitiveness is correlated with gender consistently with recent findings in the literature, namely that groups with more male students are also more competitive. Interestingly, we see that group size and cooperation are negatively

Table 2: Correlation matrix (main regressors)

	Cond. coop.	Coop.	Comp.	D.Male	Size	Points	Rank
Cond. coop	1						
Coop.	0.3995*	1					
Comp.	0.0686	-0.2376	1				
D.Male	0.0518	-0.1216	0.2846	1			
Size	-0.2016	-0.3014*	0.1911	-0.4085*	1		
Points	-0.0027	-0.1316	-0.3387*	-0.1894	0.2689*	1	
Rank	0.0077	0.1456	0.3333*	0.1705	-0.2525*	-0.9945*	1

Notes: $*$ denotes correlations significantly different from zero at the 5% level

correlated.

Panel C in Figure 1 depicts average points across periods, i. e. for different problem sets. We see a strong variation of average points per problem set, indicating the necessity to control for this by including time (i. e. period) fixed effects into the dynamic specifications. Finally, Panel D shows the distribution of total points, implying a substantial variation of overall performance among groups.

5.2 Static contest: Empirical strategy and results

Our main interest is to test if dynamic incentives have an effect on performance and how this effect varies with competitiveness. However, before we exploit the dynamic structure of our data, we first want to analyze determinants of aggregated performance, i. e. the relationship between group characteristics and their final rank (Hypothesis 1a and 1b).

To test this set of hypotheses we rely on a simple multivariate OLS regression including the respective variables and using heteroscedasticity-robust Huber-White standard errors. As dependent variable we use both, total points and final rank. With respect to the explanatory variables included, we estimate different specifications, varying the combination of regressors.

At this point, it is important to discuss the issue of causality. As described above, the group contest was voluntary and groups have *not* been composed by a random assignment procedure but self-selected. Of course, this raises several concerns regarding the interpretation of results stemming from the analysis just depicted. In addition it seems very likely, that important omitted variables like mathematical ability or motivation are correlated with included regressors and therefore might bias the results. Thus, we should not conclude from the analysis, that - for example - competitiveness has a causal influence on performance in

Table 3: Group size effect (static model)

	(1) Total points	(2) Final rank	(3) Total points	(4) Final rank	(5) Total points	(6) Final rank
D.Workedalone	-73.78**	11.57**			-25.76	4.722
	(36.50)	(5.769)			(53.97)	(8.684)
Group size			24.40**	-3.701**	18.96	-2.702
			(9.569)	(1.551)	(14.19)	(2.346)
Constant	220.2***	35.93***	131.2***	49.52***	153.3***	45.48***
	(16.50)	(2.699)	(34.24)	(5.458)	(56.31)	(9.223)
Observations	78	78	78	78	78	78
R^2	0.055	0.052	0.072	0.064	0.075	0.068

Notes: Robust standard errors in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

group contests. In fact, these results should rather be understood as descriptive, that need to be explored further, in order to understand their underlying mechanisms.

Regarding hypotheses 1a and 1b, i. e. the effect of group size, we either use a dummy for single-member groups or group size as such. We also look at the combination of both. This allows us to test if a potential effect of group size on performance is purely driven by "groups" with only one member, i. e. the difference between groups and individuals, or if it also holds for variation of group size above 2. Note that due to data availability regarding group characteristics, we can use a larger sample to test hypotheses 1a and 1b, as here no information from the lab is required.

We start with the effect of group size on overall performance. Regressions (1)-(4) in table 3 clearly reject hypothesis 1b in favor of hypothesis 1a, as $Group\,size$ as well as $D.Workedalone$ are both statistically significant and reflect a substantial positive effect of the number of group members on total points (a negative effect on final rank[10]). Following these results, working alone instead of within a group reduces total points on average by approximately 74 points, which implies a standardised coefficient of -0.24. Alternatively, it leads to a final rank almost 12 positions worse. Regressions (5) and (6) indicate that this positive effect of group size on performance cannot simply be assigned to the difference between individuals and groups: Although statistical significance vanishes in these two specifications, estimated coefficients of $D.Workedalone$ and $Group\,size$ both still point in the

[10]Note that a negative effect on final rank implies that larger groups perform better, as a lower rank reflects a better position.

Table 4: Competitiveness and cooperation effect (static model)

Panel A.	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Total points	Total points	Total points	Total points	Total points
Competitiveness	-25.51**	-23.61**		-31.36***	-29.19***
	(10.36)	(11.52)		(9.222)	(10.27)
D.Male		-29.99			-37.65
		(50.39)			(49.18)
Cooperation			-10.16	-18.75**	-19.39**
			(11.41)	(9.212)	(9.123)
D.Conditional cooperation			21.29	50.37	53.12
			(60.25)	(55.58)	(56.89)
Constant	327.4***	336.9***	259.8***	420.9***	435.6***
	(46.54)	(49.26)	(59.55)	(59.30)	(61.69)
Observations	47	47	47	47	47
R^2	0.115	0.122	0.020	0.178	0.190

Panel B.	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Final rank	Final rank	Final rank	Final rank	Final rank
Competitiveness	4.148**	3.904**		5.156***	4.859***
	(1.682)	(1.890)		(1.488)	(1.680)
D.Male		3.861			5.164
		(8.501)			(8.276)
Cooperation			1.833	3.246**	3.333**
			(1.832)	(1.485)	(1.482)
D.Conditional cooperation			-3.558	-8.340	-8.717
			(9.914)	(8.974)	(9.188)
Constant	18.72**	17.49**	28.85***	2.354	0.344
	(7.723)	(8.123)	(9.734)	(9.760)	(10.28)
Observations	47	47	47	47	47
R^2	0.111	0.115	0.024	0.181	0.188

Notes: Robust standard errors in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

same directions as in regressions (1)-(4).[11]

Next, we analyze the aggregated effects of group characteristics elicited from the laboratory, i. e. competitiveness and cooperative behavior. Panel A in Table 4 presents results of regressions where total points is the dependent variable, while in Panel B it is final rank.

Competitiveness has a statistically significant negative effect on performance in all specifications. An average increase of competitiveness by one unit leads to a decrease of total points by 24-31 points. This implies a standardised coefficient of at least 0.31. Cooperation

[11]The inclusion of both explanatory variables at the same time substantially reduces idiosyncratic variation available to estimate slope coefficients, as the correlation of the two is substantial (approximately 0.4). This drives up standard errors, thereby working against statistical significance without necessarily indicating a less pronounced effect.

also has a negative impact on performance, although it is not statistically significant across all specifications. The point estimate of conditional cooperation indicates a positive effect which is also economically meaningful, however the effect is not statistically significant.

5.3 Dynamic contest: Empirical strategy and results

Let us now consider the dynamic structure of the contest. To test the respective hypotheses, we can now fully exploit the dynamic structure of our data set. We estimate a panel regression with i groups and t contest rounds (periods), yielding 780 observations overall. The baseline model looks as follows.

$$Points_{i,t} = \alpha + \beta_1 Rank_{i,t-1} + \gamma_i + \delta_t + u_{i,t}, \tag{1}$$

i. e. we regress the points achieved by group i in period t on the rank of this group in the previous period $t-1$. As we include group and time fixed effects, any time invariant group specific influences as well as any time specific general shocks are controlled for. This is of particular importance due to several reasons. First, points achieved by a group in a given period should strongly depend on the abilities and characteristics of the groups' members. Thus, without group fixed effects, we cannot be sure to truly identify the dynamic incentive effect, i. e. the causal effect of the relative position within the contest on contemporary performance, by looking at the estimate of β_1. However, by including group fixed effects, we absorb the overall effect of all time invariant characteristics of a group, which reasonably should include the main drivers of overall performance in the contest.

Second, it might also be the case that the estimate of β_1 is driven by aggregate time specific effects. Namely, it seems possible that time available to invest into the contest might systematically vary as the semester proceeds or due to concurring events like exams or social occasions. Furthermore, we would like to control for differences in the difficulty of individual problem sets. To achieve all this, we also include time (period, problem set) specific dummies that capture all aggregated effects on overall points in a given period.

Consequentially, the variation that is left in order to estimate the dynamic incentive effect does not include group specific characteristics or specific time dynamics. The estimation of β_1 only exploits the within-group variation of the previous rank for a given period. Thus, we can rule out that omitted time-invariant factors like group characteristics, or a specific time structure of aggregated shocks is what drives our results. Therefore we can confidently

Table 5: Competitiveness effect (dynamic model)

	(1) Points	(2) Points	(3) Points	(4) log(Points)	(5) log(Points)	(6) log(Points)
$Rank_{t-1}$	-0.0785	-0.0330	0.299*	-0.0266**	-0.0258	0.0113
	(0.136)	(0.165)	(0.149)	(0.0122)	(0.0160)	(0.0112)
$Rank_{t-1} * D.Comp.$			-0.986***			-0.110***
			(0.251)			(0.0195)
Constant	20.39***	32.56***	32.95***	3.143***	3.901***	3.944***
	(3.283)	(2.678)	(2.154)	(0.267)	(0.259)	(0.176)
$Rank_{t-1} + Rank_{t-1} * D.Comp.$			-0.687***			-0.099***
			(0.21)			(0.018)
Group Fixed effects	yes	yes	yes	yes	yes	yes
Period Fixed effects	yes	yes	yes	yes	yes	yes
Observations	450	296	296	450	296	296
Number of groups	71	45	45	71	45	45
R^2	0.655	0.710	0.732	0.316	0.373	0.445

Notes: Robust standard errors in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

interpret results as the causal effect of past performance on current performance.

To address hypothesis 3 - effect heterogeneity with respect to competitiveness - we rely on the variable *D.Competitiveness*, which is a dummy variable that takes on the value of 1 if the group average of *Competitiveness* is above the median and 0 otherwise. We interact our main regressor, i. e. rank in the previous period, with this dummy variable. This allows checking if the causal effect of position in the contest on contemporaneous performance differs systematically between competitive and non-competitive groups. Note that although competitiveness is a time-invariant variable, its interaction with rank in the previous period (a time-varying variable) is not absorbed by the group fixed effect and can therefore be jointly estimated.

One remaining concern is, that a potentially negative correlation between previous rank and current performance is simply driven by the relatively large number of groups that drop out during the contest, i. e. do not hand in any solution at all anymore after a certain round. Once a group drops out of the contest, its rank will subsequently increase (become worse) in each round left. As by not participating they receive just zero points, one might worry that this is what is driving our result. We therefore exclude all observations where a group gained zero points in the previous and the current round, inferring that they have simply dropped out of the contest.

Table 5 presents results of the dynamic specification using disaggregated data on the period level.[12] In the standard linear specification that does not distinguish between competitive and non-competitive groups (regression (1)), we see that the influence of the position in the contest on contemporaneous performance is rather small (-0.0785) and statistically not significant. In regression (2) we restrict the sample to those groups on which we have information about competitiveness. This allows us to rule out that any differences in the results following the inclusion of competitiveness is driven by a change of the sample rather than the explanatory power of the variable competitiveness itself. We see that the effect of competitiveness remains slightly negative (-0.033) and statistically insignificant. Next we run the same regressions, the only difference being that the dependent variable is now the logarithm of points. In regression (4) this leads to a negative and statistically significant effect of contest position on performance. However, the size of the effect is still rather small: An increase in the ranking by one position leads on average to an increase of points by approximately 2.7%. Once we restrict the sample again to those groups on which we have information about competitiveness (regression (5)), the effect turns statistically insignificant. This is mainly driven by the increase of the standard error following the reduction of the sample size by one third instead of a decreased size of the slope coefficient. Overall, these regressions, that do not distinguish groups with respect to their degree of competitiveness, indicate only a small impact of contest position on performance if any. Thus, we cannot confidently confirm hypothesis 2.

Regression (3) and (6) now investigate the heterogeneity of the effect of contest position on performance by allowing for different sensitivities to feedback for competitive and non-competitive groups as hypothesis 3 stated. In both regressions, the coefficient for non-competitive groups ($Rank_{t-1}$) even turns positive, however without being statistically significant beyond the 10% level. In turn, the effect for competitive groups becomes substantially stronger and statistically highly significant: Regressions (3) predicts an increase of approximately 0.7 points per rank. Regression (6) claims an increase of points of almost 10% per rank. As the coefficients of $Rank_{t-1} * D.Comp.$ show, this difference in sensitivity to contest position between competitive and non-competitive groups is statistically significant at the 1% level. This implies that the small negative overall effect we saw in regressions (1), (2), (4), and (5) has been entirely driven by those groups in the sample classified as competitive.

[12] Note that in the following a *negative* effect means, that there is a negative relationship between rank and points. However, as a higher rank (i. e. a greater number) reflects a worse position in the ranking, a negative relationship implies, that groups react *positively* to better previous performance.

Summarizing these results we can state that there is a systematic difference in the behavior of competitive and non-competitive groups. Competitive groups react systematically stronger to feedback on intermediate performance. This different sensitivity is thereby a consistent reaction to changes in the probability of winning the contest as predicted by our model: A better (worse) position in the ranking implies a higher (lower) chance of reaching one of the top three spots. This higher (lower) probability of winning increases (decreases) the marginal productivity of effort, which in turn implies a higher (lower) optimal level of effort. This effect is the stronger, the more weight is put on winning the contest, which is how competitiveness is captured in our model. Complementary evidence from our post-contest questionnaire provides further support to this explanation: In a simple OLS regression, competitiveness has a positive effect on the probability of participants stating to take part in the contest because of the outlook for the certificate, which can be interpreted as a higher weight on winning the contest. The effect is significant at the 5% level.

5.4 Dynamic incentives and the negative aggregated effect of competitiveness

Given the results presented so far, one might wonder if there is a connection between the negative aggregated effect of being competitive and the difference in the sensitivity to dynamic effects between competitive and non-competitive groups.

At first glance, it seems surprising that competitiveness actually leads to worse performance in a competitive environment. However, while the dynamic result confirms the notion that competitive groups are motivated by the competition and thus increase performance if the probability of winning increases, it also gives rise to an opposing effect: If a competitive group is in a bad position and thus its chances of winning are reduced, the reduction in effort is even stronger than for non-competitive groups. In general, a stronger sensitivity to dynamic incentives therefore leads to a stronger path-dependence of performance and represents a dynamic amplification mechanism. Given any achieved performance level in early rounds, competitive groups' subsequent effort depends more strongly on it than non-competitive groups' effort. On the one hand, this implies that a good initial performance leads to an even better overall performance. On the other hand, a bad initial performance is translated into an even worse overall performance.

With respect to our analysis this raises the question whether the mechanism just de-

Table 6: Fixed effects and competitiveness

	(1)	(2)	(3)	(4)	(5)	(6)
	Est. FE	Est. FE	Est. FE	Est. FE	Est. FE	Est. FE
D.Competitiveness	-6.002	-5.982	13.20***	-0.494	-0.492	1.649***
	(3.641)	(3.565)	(3.750)	(0.320)	(0.312)	(0.350)
Constant	-1.540	-1.730	-8.979***	-0.134	-0.176	-0.985***
	(2.464)	(2.408)	(2.715)	(0.230)	(0.223)	(0.233)
Observations	45	45	45	45	45	45
R^2	0.060	0.062	0.221	0.052	0.054	0.342

Notes: Robust standard errors in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

scribed is responsible for the overall worse performance of competitive groups. Testing this requires two steps: First, we check if competitive groups actually do perform worse in the first period, giving rise to the possibility of negative dynamic incentive effects. Second, we check if once one controls for dynamic effects, the negative effect of being competitive still remains, i. e. if there is anything else with competitive groups making them performing worse. Regarding the first step, we find that competitive compared to non-competitive groups indeed have, on average, 5.61 points less in the first round, where the average number of points across all groups was 39.38 and the respective standard deviation 18.59. So we see that competitive groups are indeed worse after the first period, possibly triggering a decrease in relative effort compared to non-competitive groups, given the dynamic incentive effect.

Regarding the second step, we proceed as follows: Note, that the dynamic regressions in Table 5 decompose the variance of *Period points* into three parts: First, the variation that can be explained by within-group variation of past performance, second, the time-invariant between-group variation which is captured in the fixed effects, and third, the idiosyncratic unexplainable time-varying error term $u_{i.t}$. In contrast, in the cross-sectional regressions, the average difference of the time-varying effect between competitive and non-competitive groups was also captured in the regressor competitiveness. The question now is, if this difference in dynamic incentive effects among competitive and non-competitive groups was driving the overall difference in their performance. To test this, we investigate the estimated fixed effects of the dynamic regressions. They represent the between-group differences in overall performance that remains once one controls for dynamic incentive effects. If we find that after controlling for dynamic incentive effects, the difference in overall performance between competitive and non-competitive groups vanishes, we can confidently conclude, that they have been the reason for this difference in the first place.

Table 6 presents the corresponding results. Each column represents a regression of the estimated fixed effects from the respective regression in Table 5. We see that in regressions (1), (2), (4), and (5) competitiveness has no statistically significant effect anymore on overall between-group performance differences, i. e. estimated fixed effects. These are the regressions that allowed for dynamic incentives, however assumed a homogeneous effect across groups. If we turn to regression (3) and (6), which are based on those regressions allowing for heterogeneity in the effect of dynamic incentives, the effect of competitiveness even turns positive and in addition also becomes statistically significant.

As a result, we find that once controlling for dynamic incentive effects, the negative effect of competitiveness on overall performance vanishes. Once one additionally allows for heterogeneity of the dynamic incentive effect, competitive groups even perform significantly better. These results are consistent with the explanation that the main reason for the overall worse performance of competitive groups was indeed the combination of their bad performance in the first period and the subsequent path-dependence induced by dynamic incentives. Once one controls for this, competitive groups are no longer worse than non-competitive ones but perform even better.

6 Conclusion

In this paper we present results from the analysis of combined data from a dynamic group contest among students and from the laboratory. We argue that incentive effects implied by dynamic contests with feedback are heterogeneous with respect to differences in individual characteristics of contest participants. We show that the reaction to intermediate performance, i. e. the strength of dynamic incentive effects, is stronger for groups whose members have a high behavioral measure of competitiveness. Furthermore, we show that this increased sensitivity to dynamic incentives can turn out to harm participants, as the dynamic incentive effect can induce a stronger de-motivation from initial bad performances. In our case, this even creates an overall worse performance of competitive groups, which can be attributed to dynamic incentive effects. Overall, a stronger sensitivity to dynamic incentives generally amplifies initial performances and thus can have both, good and bad overall effects. Secondary to this main result, we also test several theoretical predictions from the literature regarding the determinants of overall performance in group contests. We find that large groups perform better, and do not suffer from increased free-riding problems.

In future research, we like to investigate more deeply the nature of competitiveness and how it affects behavior in contest and tournament situations. Despite the robustness of the results presented here, it seems important to further validate them and check if they also hold in different environments to get an idea about their context sensitivity. A natural starting point would therefore be to investigate behavior in dynamic contests of competitive and non-competitive groups and individuals in the lab. This furthermore allows to get an even deeper understanding of the mechanisms driving our results by varying specific parameters of the contest. For example, we would like to check whether competitive participants differ in their reaction to changes of the price structure of a contest.

Overall, this paper underscores the usefulness of combined field-lab approaches and improves confidence in the reliability of laboratory measures for the prediction of field behavior. Furthermore, our results imply that dynamic contest type structures with feedback might be a valuable tool in order to incentivise individuals and groups to provide effort. However, the heterogeneity in the effect demands for caution and context awareness when designing and implementing such incentive mechanisms in practice. Our results provide a possible explanation for the strong behavioral dispersion of behavior in previous results and stresses the potential downside of dynamic incentive schemes.

References

Abeler, J. and F. Marklein (2010): “Fungibility, Labels, and Consumption,” IZA Discussion Papers 3500, Institute for the Study of Labor (IZA).

Antonovics, K., P. Arcidiacono, and R. Walsh (2009): “The Effects of Gender Interactions in the Lab and in the Field,” *The Review of Economics and Statistics*, 91, 152–162.

Balafoutas, L. and M. Sutter (2010): “Gender, competition and the efficiency of policy interventions,” Working Papers 2010-12, Faculty of Economics and Statistics, University of Innsbruck.

Bandiera, O., I. Barankay, and I. Rasul (2012): “Team Incentives: Evidence from a Firm Level Experiment,” IZA Discussion Papers 6279, Institute for the Study of Labor (IZA).

Barankay, I. (2012a): “Rank Incentives Evidence from a Randomized Workplace Experiment,” Working paper.

——— (2012b): “Rankings and Social Tournaments: Evidence from a Crowd-Sourcing Experiment,” Working paper.

Barr, A. and T. Packard (2000): “Revealed and Concealed Preferences in the Chilean Pension System: An Experimental Investigation,” Economics Series Working Papers 053, University of Oxford, Department of Economics.

Barr, A. and P. Serneels (2004): “Wages and Reciprocity in the Workplace,” CSAE Working Paper Series 2004-18, Centre for the Study of African Economies, University of Oxford.

Barr, A. and A. Zeitlin (2010): “Dictator games in the lab and in nature: External validity tested and investigated in Ugandan primary schools,” CSAE Working Paper Series 2010-11, Centre for the Study of African Economies, University of Oxford.

Bartling, B., E. Fehr, M. Marechal, and D. Schunk (2009): “Egalitarianism and Competitiveness,” *American Economic Review*, 99, 93–98.

Benz, M. and S. Meier (2008): "Do people behave in experiments as in the field? - Evidence from donations," *Experimental Economics*, 11, 268–281.

Blanes i Vidal, J. and M. Nossol (2011): "Tournaments Without Prizes: Evidence from Personnel Records," *Management Science*, 57, 1721–1736.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43, 972–1059.

Camerer, C. (2011): "The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List," *SSRN eLibrary*.

Carpenter, J. and C. K. Myers (2007): "Why Volunteer? Evidence on the Role of Altruism, Reputation, and Incentives," IZA Discussion Papers 3021, Institute for the Study of Labor (IZA).

Casas-Arce, P. and A. Martinez-Jerez (2009): "Relative Performance Compensation, Contests, and Dynamic Incentives." *Management Science*, 55, 1306–1320.

Clark, A., D. Masclet, and M.-C. Villeval (2010): "Effort and Comparison Income: Experimental and Survey Evidence," *Industrial and Labor Relations Review*, 63, 407–426.

Dargnies, M.-P. (2009): "Does team competition eliminate the gender gap in entry in competitive environments?" Documents de travail du Centre d'Economie de la Sorbonne 09006, Universite Pantheon-Sorbonne (Paris 1), Centre d'Economie de la Sorbonne.

Dechenaux, E., D. Kovenock, and R. Sheremeta (2012): "A Survey of Experimental Research on Contests, All-Pay Auctions and Tournaments," Working Papers 12-22, Chapman University, Economic Science Institute.

Delfgaauw, J., R. Dur, A. Non, and W. Verbeke (2011): "The Effects of Prize Spread and Noise in Elimination Tournaments: A Natural Field Experiment," Tinbergen Institute Discussion Papers 11-120/1, Tinbergen Institute.

——— (2012a): "Dynamic Incentive Effects of Relative Performance Pay: A Field Experiment," Tinbergen Institute Discussion Papers 10-124/1, Tinbergen Institute.

Delfgaauw, J., R. Dur, J. Sol, and W. Verbeke (2012b): "Tournament Incentives in the Field: Gender Differences in the Workplace," IZA Discussion Papers 4395, Institute for the Study of Labor (IZA).

Falk, A. and J. Heckman (2009): "Lab Experiments Are a Major Source of Knowledge in the Social Sciences," *Science*, 326, 535–538.

Fehr, E. and A. Leibbrandt (2011): "A field study on cooperativeness and impatience in the Tragedy of the Commons," *Journal of Public Economics*, 95, 1144 – 1155.

Fischbacher, U., S. Gächter, and E. Fehr (2001): "Are people conditionally cooperative? Evidence from a public goods experiment," *Economics Letters*, 71, 397–404.

Gneezy, U., M. Niederle, and A. Rustichini (2003): "Performance In Competitive Environments: Gender Differences," *The Quarterly Journal of Economics*, 118, 1049–1074.

Gneezy, U. and A. Rustichini (2004): "Gender and Competition at a Young Age," *American Economic Review*, 94, 377–381.

Heckman, J. (2011): "Integrating Personality Psychology into Economics," NBER Working Papers 17378, National Bureau of Economic Research, Inc.

Holmstrom, B. (1982): "Moral Hazard in Teams," *The Bell Journal of Economics*, 13, 324–340.

——— (1999): "Managerial Incentive Problems: A Dynamic Perspective," *The Review of Economic Studies*, 66, pp. 169–182.

Holmstrom, B. and P. Milgrom (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55, pp. 303–328.

Houser, D. and D. Schunk (2009): "Social environments with competitive pressure: Gender effects in the decisions of German schoolchildren," *Journal of Economic Psychology*, 30, 634–641.

Isaac, M. and K. Schnier (2006): "Sealed Bid Variations on the Silent Auction," in *Experiments Investigating Fundraising and Charitable Contributors (Research in Experimental Economics, Volume 11)*, ed. by M. Isaac and D. Davis, Emerald Group Publishing Limited, 31–46.

Ivanova-Stenzel, R. and D. Kübler (2005): "Courtesy and Idleness: Gender Differences in Team Work and Team Competition," Discussion paper series of sfb/tr 15 governance and the efficiency of economic systems, Free University of Berlin, Humboldt University of Berlin, University of Bonn, University of Mannheim, University of Munich.

Karlan, D. (2005): "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions," *American Economic Review*, 95, 1688–1699.

Konrad, K. (2009): *Strategy and Dynamics in Contests*, Oxford University Press, USA.

Kosfeld, M. and S. Neckermann (2011): "Getting More Work for Nothing? Symbolic Awards and Worker Performance," *American Economic Journal: Microeconomics*, 3, 86–99.

Kosfeld, M. and D. Rustagi (2011): "Leadership and Cooperation," Working paper, Goethe University Frankfurt.

Kosfeld, M. and F. von Siemens (2011): "Competition, cooperation, and corporate culture," *RAND Journal of Economics*, 42, 23–43.

Lamba, S. and R. Mace (2011): "Demography and ecology drive variation in cooperation across human populations," *Proceedings of the National Academy of Sciences*, 108, 14426–14430.

Lazear, E. and S. Rosen (1981): "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89, 841–64.

Ledyard, J. (1995): *Public Goods: A Survey of Experimental Research*, Princeton University Press, 111–194.

Levitt, S. and J. List (2007a): "Viewpoint: On the generalizability of lab behaviour to the field," *Canadian Journal of Economics*, 40, 347–370.

——— (2007b): "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives*, 21, 153–174.

——— (2008): "Homo economicus Evolves," *Science*, 319, 909–910.

List, J. (2003): "Does Market Experience Eliminate Market Anomalies?" *The Quarterly Journal of Economics*, 118, 41–71.

——— (2006): "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions," *Journal of Political Economy*, 114, 1–37.

Meier, S. and C. Sprenger (2010): "Stability of Time Preferences," IZA Discussion Papers 4756, Institute for the Study of Labor (IZA).

Moldovanu, B. and A. Sela (2001): "The Optimal Allocation of Prizes in Contests," *American Economic Review*, 91, 542–558.

Niederle, M., C. Segal, and L. Vesterlund (2010): "How Costly is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness," NBER Working Papers 13923, National Bureau of Economic Research, Inc.

Niederle, M. and L. Vesterlund (2007): "Do Women Shy Away from Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics*, 122, 1067–1101.

——— (2011): "Gender and Competition," *Annual Review of Economics*, 3, 601–630.

Olson, M. (1965): *The Logic of Collective Action: Public Goods and the Theory of Groups*, Harvard University Press.

Östling, R., J. Wang, E. Chou, and C. Camerer (2011): "Testing Game Theory in the Field: Swedish LUPI Lottery Games," *American Economic Journal: Microeconomics*, 3, 1–33.

Palacios-Huerta, I. (2003): "Professionals Play Minimax," *Review of Economic Studies*, 70, 395–415.

Palacios-Huerta, I. and O. Volij (2008): "Experientia Docet: Professionals Play Minimax in Laboratory Experiments," *Econometrica*, 76, 71–115.

Post, T., M. van den Assem, G. Baltussen, and R. Thaler (2008): "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show," *American Economic Review*, 98, 38–71.

Rustagi, D., S. Engel, and M. Kosfeld (2010): "Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management," *Science*, 330, 961–965.

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

Stoop, J., C. Noussair, and D. van Soest (2010): "From the lab to the field: Cooperation among fishermen," MPRA Paper 28924, University Library of Munich, Germany.

Sutter, M. and D. Rützler (2010): "Gender differences in competition emerge early in life," Working Papers 2010-14, Faculty of Economics and Statistics, University of Innsbruck.

APPENDIX

A Proof of proposition 1

Proposition *If ψ is unimodal, then effort $e^*_{i2}(y_{i1})$ has a hump shape.*

Proof. Following Casas-Arce and Martinez-Jerez (2009, p. 1312) and assuming a concave objective function[13], the maximization problem has a unique solution characterised by the above first-order condition, and by the implicit function, theorem $e^*_{i2}(y_{i1})$ is differentiable and satisfies

$$\frac{\partial e^*_{i2}(y_{i1})}{\partial y_{i1}} = -\frac{\psi'(\tilde{y}_\mu - y_{i1} - e^*_{i2})}{\psi'(\tilde{y}_\mu - y_{i1} - e^*_{i2}) - c''(e^*_{i2})} > -1.$$

For a concave objective function the denominator is positive, and therefore, the sign of $\frac{\partial e^*_{i2}(y_{i1})}{\partial y_{i1}}$ is the opposite of the sign of ψ'. Furthermore, $y_{i1}+e^*_{i2}$ is increasing in y_{i1}. Hence there exists a constant $\hat{y}_{i1}$ such that effort is increasing in past output if $y_{i1} \leq \hat{y}_{i1}$ and decreasing otherwise. □

B Proofs of proposition 2

Proposition *The sensitivity of e^*_{i2} to y_{i1} increases in θ_i.*

Proof. Following the same steps as above, we get the following result from the first-order condition

$$\frac{\partial e^*_{i2}(y_{i1})}{\partial y_{i1}} = -\frac{\theta_i\psi'(\tilde{y}_\mu - y_{i1} - e^*_{i2})}{\theta_i\psi'(\tilde{y}_\mu - y_{i1} - e^*_{i2}) - c''(e^*_{i2})} > -1.$$

Analogously to the reasoning above, there exists again a constant $\hat{y}_{i1}$ with the same characteristics and implications as before and proposition 1 still holds for this adjusted setup. Taking the derivative of $\frac{\partial e^*_{i2}(y_{i1})}{\partial y_{i1}}$ with respect to θ_i we get the following expression

$$\frac{\partial^2 e^*_{i2}(y_{i1})}{\partial y_{i1}\partial\theta_i} = \frac{c''}{(\theta_i\,\psi' - c'')^2},$$

[13] A stochastic (Tullock) CSF satisfies these condition.

which is strictly positive, as the denominator is always positive and c'' is positive by assumption. $\square$