DISCUSSION PAPER SERIES

DP16360

The Efficacy of Tournaments for Non-Routine Team Tasks

Florian Englmaier, Stefan Grimm, Dominik Grothe, David Schindler and Simeon Schudy

ORGANIZATIONAL ECONOMICS

CEPR

ISSN 0265-8003

The Efficacy of Tournaments for Non-Routine Team Tasks

Florian Englmaier, Stefan Grimm, Dominik Grothe, David Schindler and Simeon Schudy

Discussion Paper DP16360
Published 14 July 2021
Submitted 09 July 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Organizational Economics

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Florian Englmaier, Stefan Grimm, Dominik Grothe, David Schindler and Simeon Schudy

The Efficacy of Tournaments for Non-Routine Team Tasks

Abstract

Tournaments are often used to improve performance in innovation contexts. Tournaments provide monetary incentives but also render teams' identity and social-image concerns salient. We study the effects of tournaments on team performance in a non-routine task and identify the importance of these behavioral aspects. In a natural field experiment (n > 1,700 participants), we vary the salience of team identity, social-image concerns, and whether teams face monetary incentives. Increased salience of team identity does not improve performance. Social-image motivates mainly the top-performing teams. Additional monetary incentives improve all teams' outcomes without crowding out teams' willingness to explore or perform similar tasks again.

JEL Classification: C93, D90, J24, J33, M52

Keywords: team work, tournaments, Rankings, incentives, identity, image concerns, Innovation, exploration, natural field experiment

Florian Englmaier - florian.englmaier@econ.lmu.de
LMU München and CEPR

Stefan Grimm - stefan.grimm@econ.lmu.de
LMU Munich

Dominik Grothe - dominik.grothe@econ.lmu.de
LMU München

David Schindler - d.schindler@tilburguniversity.edu
Tilburg University

Simeon Schudy - simeon.schudy@econ.lmu.de
LMU Munich

Acknowledgements
We thank Silvia Fernandez Castro, Tamina Straub, Oguzhan Bekar, and Benjamin Schumann for excellent research assistance. We are grateful to Josse Delfgaauw, Robert Dur, Daniel Gross, Nicola Lacetera, Susanne Neckermann, Marina Schröder, and seminar audiences in Munich, at the INSEAD RCT Days 2019 and IZA Bonn for providing helpful thoughts and comments. Dominik Grothe acknowledges funding by the Elite Network of Bavaria through IDK Evidence Based Economics. Further, financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119) is gratefully acknowledged. This study was approved by the Department of Economics' Institutional Review Board (IRB) at the University of Munich (Project 2015-11) and pre-registered as RCT AEARCTR-0002916 at https://www.socialscienceregistry.org/trials/2916/history/31141.

The Efficacy of Tournaments for Non-Routine Team Tasks *

Florian Englmaier † Stefan Grimm ‡ Dominik Grothe §
David Schindler ¶ Simeon Schudy ‖

July 9, 2021

Abstract

Tournaments are often used to improve performance in innovation contexts. Tournaments provide monetary incentives but also render teams' identity and social-image concerns salient. We study the effects of tournaments on team performance in a non-routine task and identify the importance of these behavioral aspects. In a natural field experiment (n>1,700 participants), we vary the salience of team identity, social-image concerns, and whether teams face monetary incentives. Increased salience of team identity does not improve performance. Social-image motivates mainly the top-performing teams. Additional monetary incentives improve all teams' outcomes without crowding out teams' willingness to explore or perform similar tasks again.

JEL codes: C93, D90, J24, J33, M52

Keywords: team work, tournaments, rankings, incentives, identity, image concerns, innovation, exploration, natural field experiment

*We thank Silvia Fernandez Castro, Tamina Straub, Oguzhan Bekar, and Benjamin Schumann for excellent research assistance. We are grateful to Josse Delfgaauw, Robert Dur, Daniel Gross, Nicola Lacetera, Susanne Neckermann, Marina Schröder, and seminar audiences in Munich, at the INSEAD RCT Days 2019 and IZA Bonn for providing helpful thoughts and comments. Dominik Grothe acknowledges funding by the Elite Network of Bavaria through IDK Evidence Based Economics. Further, financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119) is gratefully acknowledged. This study was approved by the Department of Economics' Institutional Review Board (IRB) at the University of Munich (Project 2015-11) and pre-registered as RCT AEARCTR-0002916 at `https://www.socialscienceregistry.org/trials/2916/history/31141`.

†florian.englmaier@econ.lmu.de, LMU Munich, Department of Economics & Organizations Research Group (ORG) & CESifo & CEPR & IZA.

‡stefan.grimm@econ.lmu.de, LMU Munich, Department of Economics.

§dominik.grothe@econ.lmu.de, LMU Munich, Department of Economics.

¶d.schindler@tilburguniversity.edu, Tilburg University, Department of Economics & CESifo.

‖simeon.schudy@econ.lmu.de, LMU Munich, Department of Economics & CESifo.

1 Introduction

Ever since the seminal contribution of Lazear and Rosen (1981), there has been great interest in tournaments to foster performance and innovation (cf. the overview in Lazear and Oyer, 2012).[1] Lazear and Rosen's original argument for the attractiveness of tournaments relied on the fact that tournaments can establish efficient outcomes at lower costs, since tournaments only require information on relative ranks instead of absolute performance. However, in innovation contexts, in which teams derive status from developing innovative solutions, tournaments include additional and important behavioral features, rendering them attractive for improving performance. First, tournaments naturally increase the salience of team identity because teams are explicitly identified (e.g., by a ranking of teams, departments, brand, or company names). Second, as the rankings are observable, tournaments may substantially intensify status-related image concerns. Prior research in psychology and economics has documented that both identity (see, e.g., Tajfel and Turner, 2001; Akerlof and Kranton, 2000; Chen and Chen, 2011) and image concerns (see, e.g., Kluger and DeNisi, 1996; Kosfeld and Neckermann, 2011; Fershtman et al., 2006; Ball et al., 2001; Moldovanu et al., 2007; Bursztyn and Jensen, 2017) can play a crucial role in human behavior. However, much less is known about their role in the efficacy of tournaments in complex, non-routine analytical tasks, which have become ubiquitous in modern economies and characterize many work environments in innovation contexts (see, e.g., Autor et al., 2003; Autor and Price, 2013). Since understanding the relative importance of these aspects allows a cost-effective design of incentives, the aim of the present study is both to investigate the efficacy of tournaments with prizes in non-routine team tasks and to determine the importance of behavioral aspects vis-a-vis monetary rewards.

This study exploits a unique field setting to understand the importance of salient identity, image concerns, and prizes in tournaments involving complex teamwork. We conducted a natural field experiment to identify the causal effects of these components on team performance in a real-life escape room setting, in which teams have to solve a se-

[1] Early examples of innovation competitions were the "longitude rewards", a system of inducement prizes offered by the Government of Great British for a practical and straightforward method to precisely determine a ship's longitude at sea. These rewards were granted by Parliament in 1714 and were administered by the newly created Board of Longitude. Brunt et al. (2012) and Khan (2015) provide more details on the role of inducement prizes in innovation.

ries of cognitively demanding tasks in order to succeed.[2] The setting is particularly suited to answer our research question, as it encompasses major characteristics of production in the "ideas sector" of the economy (see, e.g., Autor et al., 2003; Autor and Price, 2013), while simultaneously allowing an objective measurement of team performance (the time spent until completion) and for exogenously manipulating the salience of identity, image, and instrumental concerns. Teams face a series of complex problems, need to collect and recombine information, and think outside the box. The task is interactive, as team members have to collaborate with each other, discuss possible actions, jointly develop ideas, and test their hypotheses. Succeeding requires coordination and cooperation, and teams often proudly document their participation on a local "wall of fame" on site. Thus, this setting provides space for team identity and image concerns to matter, and, at the same time, allows for exogenous variation of the previously mentioned important tournament characteristics across a large number of teams.[3]

To identify the importance of team identity, image concerns, and monetary prizes, we randomly allocated participating teams to one of four conditions, which introduced these features in steps. Salience of team identity is an innate feature of tournaments (as in most tournaments, rankings identify competing teams by their team name). To analyze the relative importance of more salient team identity, we first compare a no intervention condition (*Control*), in which teams have no team name, with a condition in which we make team identity more salient by asking teams to explicitly discuss, and jointly choose, a team name they identify with (*Identity*). Since in most business contexts, teams already have some team (or brand) name they identify with, *Identity* also serves as a meaningful comparison group for the investigation of the additional effects of image concerns due to public rankings and or monetary prizes. Treatment *Rank* focuses on image concerns and introduces a public ranking for all teams (using self-chosen team names). Treatment *Prize* further adds a monetary prize. Thus, in treatment *Prize*, the teams partake in a classical

[2]Escape rooms are popular worldwide both among private teams seeking a complex team challenge and companies which us them for team building and recruiting purposes. They offer a unique opportunity to study problem-solving skills and the ability to work in a team in practice, and are nowadays also used in educating IT and Engineering students (Borrego et al., 2017). Prior research has used other unique opportunities to study competition in tournaments, e.g., data from sports (see, e.g., Brown, 2011; Brown and Minor, 2014).

[3]An extensive discussion of the task and the responsiveness of team performance to bonus incentives is provided in Englmaier et al. (2018).

tournament including monetary incentives and a public ranking (using self-chosen team names).

We find that strengthening the salience of team identity alone is not sufficient to improve team performance, but image concerns in terms of relative performance matter. When a treatment features a public ranking, teams tend, on average, to solve the task more quickly, which is mostly driven by the top performers. Those below the top quantiles are, however, similarly likely to complete the task compared to teams whose performance is not publicly ranked. An additional monetary prize substantially increases the likelihood of succeeding within the given time limit. Prizes boost performance at the top but also along the lower quantiles of the performance spectrum. Overall, the introduction of a tournament with a monetary prize increases completion rates by more than 20 percent (almost 12 percentage points) as compared to *Control* and reduced finishing times by more than 3 minutes (remaining times are almost doubled).

These findings contribute to the recent literature on tournaments, incentives, and teamwork in non-routine analytical tasks. First and foremost, we provide novel insights into the causal effects of three major components innate to tournament incentives: salience of team identity, image concerns, and prizes. In this way, we systematically advance earlier work that studied rank versus monetary incentives in *routine* tasks. Findings in the context of routine tasks indicate that tournaments with and without prizes can affect team performance, particularly when team identity is present. For instance, Delfgaauw et al. (2013) compare rank and monetary incentives in retail chains and document that sales competitions have a positive effect on sales growth, but only in stores where the store's manager and a sufficiently large fraction of the employees have the same gender (a proxy for stronger team identity). Our setting allows us to separate whether the salience of team identity can foster performance on its own, and sheds light on how salient identity, image concerns, and prizes affect performance in a non-routine task. Our results show that it is indeed the combination of salient identity and competition that fosters performance (not salient identity alone).

In terms of public rankings and prizes, we also complement work by Bandiera et al. (2013) which focused on the productivity of fruit-pickers. In their setting, team rankings lead to stark selection into teams based on team members' performance potential (rather than friendship networks) and reduces performance, due to an increase in free-riding. Tournaments with prizes have similar effects in terms of selection, but yield additional ef-

fort provision within teams, which offsets the negative effects of free-riding. Our study is novel and different to previous work in several important ways: First, we focus on a non-routine team task and vary incentives across the existing teams, excluding selection into teams based on incentives by design. Second, our setting allows us to vary the salience of team identity in a natural way without introducing competition. Third, while in previous work rankings are often informative about income differences (e.g., when teams are paid based on a piece-rate), our study isolates non-instrumental image concerns when introducing the public ranking. Excluding selection based on incentives and instrumental concerns, we find that introducing rank incentives has positive effects on performance. In contrast to studies on performance rankings in repeated settings (Blanes i Vidal and Nossol, 2011; Barankay, 2012; Ashraf et al., 2014; Bursztyn and Jensen, 2015; Delfgaauw et al., 2020; Ashraf, 2019; Blader et al., 2020), which sometimes document discouraging effects of relative performance rankings, we focus on the pure effect of the introduction of tournament incentives. Doing so, we show that the mere existence of tournament incentives (with and without prize) does not curb the preference for performing similar tasks again.

Studying non-routine tasks, we also complement recent laboratory studies focusing on the causal effects of incentives in creative tasks. Incentives have been discussed as potentially crowding out intrinsic motivation (e.g. Deci et al., 1999; Eckartz et al., 2012; Gerhart and Fang, 2015; Hennessey and Amabile, 2010). However, recent evidence suggests a more differentiated picture. In a laboratory experiment, Laske and Schroeder (2016) analyze incentives for the creativity of individuals, which they measure along three dimensions: quantity, quality, and originality of ideas. They compare piece-rate incentives for quantity alone, quantity combined with quality, and quantity in combination with originality, and a fixed wage condition. In their setting, incentives significantly affect the quantity and average quality of ideas, but not the average originality. Morgan et al. (2020) find that performance-based incentives affect team effort in Fermi problems (Ärlebäck and Albarracín, 2019) but do not result in better guesstimations. Bradler et al. (2019) use a large-scale laboratory experiment to analyze the impact of tournament incentives and wage gifts on creativity. While tournaments substantially increase creative output, with no evidence for crowding out of intrinsic motivation, wage gifts are ineffective. Charness and Grieco (2019) analyze incentives for "open-" and "closed-form" creative tasks. Their results indicate that monetary incentives effectively stimulate cre-

ativity only in tasks with specific ex-ante goals ("closed-form") but not in creative, yet less well-defined tasks ("open-form"), whereas a ranking is effective in both types of tasks. Gross (2020) documents that increased competition can foster creative production by individual logo designers but heavy competition drives designers to stop producing logos altogether. Complementing the above findings, our results provide important evidence on the efficacy of incentives for non-routine analytical tasks. Focusing on teamwork that requires the forming and testing of hypotheses, we show that tournaments can stimulate performance in these goal-oriented tasks, both due to concerns for social image and instrumental concerns.

We observe a robust performance-enhancing effect of rankings for the very top and of monetary prizes for all participating teams. At the same time, we do not observe negative side effects when offering these incentives. Teams neither request more external help to arrive at the solution nor do they request help earlier. In contrast with some of the results obtained for creative tasks, but in line with field evidence that focuses on incentives for idea creation (Gibbs et al., 2017), the findings from our natural field experiment suggest that incentives can foster performance in non-routine analytical team tasks. Lastly, we do not detect statistically significant effects on teams' revealed preferences for performing a similar task again: teams in conditions encompassing a ranking or a monetary prize are not less likely to purchase a voucher for future participation; if anything, our results point in the opposite direction.

The rest of this paper is structured as follows. Section 2 will describe the setting and our experimental design in more detail. Section 3 provides the results from the experiment and Section 4 concludes.

2 Experimental design

2.1 The field setting

For this study, we collaborated with *ExitTheRoom* (ETR), a provider of real-life escape room challenges.[4] Escape rooms have become increasingly popular over the last decade and can now be found in almost all major cities around the globe. Participating teams carry out a series of cognitively demanding, non-routine, and interactive tasks. They have

[4]For more information, see their website at `https://www.exittheroom.de/munich`.

to find clues, combine information, build or use objects in unusual ways, and develop and exchange innovative ideas with the other members of their team. If a team manages to succeed within a given amount of time, they win; if time runs out before the team completes all quests, they lose.

We conducted our natural field experiment (Harrison and List, 2004) at the facilities of ETR in Munich, Germany. The location offers three differently themed rooms.[5] Teams face a time limit of 60 minutes and can see their remaining time on a large screen in their room. If a team gets stuck, they can request hints via a walkie-talkie. Hint-taking involves no explicit costs (neither monetary nor in terms of the remaining time). However, as the number of allowed requests for a hint is limited to five, there are opportunity costs of asking for assistance. ETR staff provides hints upon request but never give the immediate solution to a (sub)task. Instead, they only include vague clues regarding the next required steps. At the very end, either after completing the task or reaching the time limit, ETR staff offer teams the opportunity to purchase a voucher for future participation at a reduced rate.

ETR provides a rich setting containing the key aspects of modern non-routine analytical team tasks, thus being reflective of many work environments in innovation contexts. Therefore, it constitutes an excellent environment for a natural field experiment to distinguish the roles of team identity, image concerns, and prizes, in team performance in tournaments. Most importantly, the setting requires cooperation among the team members, is potentially prone to free-riding concerns (as observability of co-workers' cognitive effort provision is limited), and leaves room for team identity and image concerns to matter. At the same time, participants remain unaware of being studied.

2.2 Procedures and treatments

Our field experiment was conducted with 1,728 customers in 378 teams at *ExitTheRoom's* Munich location between April and July 2018 during their regular opening hours from Monday to Friday. Teams booked and paid online in advance. Upon arrival on-site, ETR staff welcomed the teams and delivered a standard introduction, laying out the story behind the specific room and explaining the task's rules.

[5] In *Madness*, teams need to find the correct code to open a door to escape (ironically) before a mad researcher experiments on them. In *The Bomb*, a bomb and a code to defuse it have to be found. *Zombie Apocalypse* requires teams to find the correct mix of liquids, an anti-zombie potion, before time runs out.

To avoid contamination, we randomized treatment arms on a weekly level.[6] ETR staff implemented the different treatments after delivering the introduction. The general idea for our four experimental conditions is to step-wise introduce three elements innate to tournaments: salient team identity, relative image concerns, and prizes. In our *Control* condition (112 teams), teams were not subject to any intervention and started working on the task directly after receiving the standard introduction.

As tournaments render team identity salient by explicitly identifying competitors by the names of their team, brand, or company, our first treatment condition, *Identity* (85 teams), was designed to increase the salience of team identity in a natural way, without adding any competitive aspects. We achieved this by asking teams to come up with a team name they identified with, which was used for communication during the task with ETR staff (via the walkie-talkie). Teams were free to choose any name they identified with, and were actively engaged in jointly choosing the team name.[7] To study the effects of image concerns in addition to salient identity, our second treatment condition, *Rank* (94 teams), introduced a weekly tournament without a prize. Teams were again asked to select a team name (in the same manner as in *Identity*). In addition, we informed them that a ranking of the current week's teams would be publicly shown on ETR's Facebook account the following Monday (for an example, see Figure A.1).

Lastly, treatment *Prize* (87 teams) exhibited the same features as *Rank*, but in addition offered a prize of 150 Euro for the best team in a week (separately for each room). Winning teams were contacted by e-mail (simultaneously with the publication of the ranking) and invited to pick up the reward at the facilities of ETR at their earliest convenience. Incentives were large relative to the price paid for participation (which ranged between 99 and 129 Euro depending on the size of the team) and thus also salient.[8]

[6] ETR shared booking data from the first two weeks of our study period with us. This data reveals that more than 90% of the teams had already booked a slot in a given week before the first session in that week was conducted. Participating teams were not informed about the study and were thus unaware that we randomized at the weekly level as well as that there were different treatment arms. Learning about these aspects within the natural setting required repeated participation in at least two rooms in two different weeks, which disqualified the team's performance from our analyses. We identified six repeated (out of a total of 384) performances that are not included in our data.

[7] Thus, our treatment renders the sense of belonging to a group salient instead of exogenously assigning an arbitrary team identity (see also the discussion in Sen, 2007; Chowdhury et al., 2016).

[8] For the role of salience for incentives, see also Englmaier et al. (2017).

2.3 Outcome measures and sample characteristics

Our final sample consists of 373 teams (1,705 individuals, see Table 1).[9] We collected observable information related to team performance and background characteristics for all teams. These include time needed to complete the task, number and timing of requested hints, team size, gender and age composition of the team, team language (German or English), prior experience with escape rooms, and whether the customers came as a private group or were part of a corporate team-building event.[10] Further, we recorded the names of the teams in all treatments apart from *Control* (where teams did not choose a name).

Our primary outcome variable is team performance, which we measure by 1) whether teams completed the task within the time limit of 60 minutes, and 2) the time needed to complete the task. Exogenous variation in the salience of team identity, image concerns, and instrumental concerns allows us to estimate the causal effects on these outcomes. Furthermore, we analyze the impact on two secondary outcome variables: the originality of team performance (which we measure inversely by the number of hints a team has taken) and a team's willingness to perform a similar task again (which is measured by the probability of purchasing a voucher for future participation at ETR at a reduced rate immediately after performing the task).

Table 1 provides an overview of team characteristics across treatments. Accounting for multiple hypotheses testing following List et al. (2019), none of the observable characteristics differs significantly from Control. The only statistically significant difference (at the ten percent level) occurs for teams' median age (estimated by our RAs) when comparing *Identity* and *Rank*. We thus show regression results with and without team characteristics as controls.

[9] During data collection, ETR's operation became inhibited after suffering from water damage resulting from a burst pipe in the building. The water damaged the electronics in the room *The Bomb*, leading to its use between June 18 and June 20 being reduced. In total, five teams in treatment *Prize* were affected before full functionality could be restored. To avoid capturing any effects on performance this may have had, we exclude these observations from the main analyses. We provide robustness checks showing that our results do not hinge on this decision in Table A.8.

[10] To preserve the character of being a natural field experiment, we did not interfere with ETR's standard procedures. Therefore, we could not explicitly elicit the participants' ages. Instead, the age of each participant was estimated based on appearance to be either 1) below 18 years, 2) between 18 and 25 years, 3) between 26 and 35 years, 4) between 36 and 50 years, 5) 51 years or older. As we are interested in the behavior of adults (and in accordance with our IRB approval) we did not include teams with minors in our study.

Table 1: Sample size and characteristics

	Control Mean (SD)	*Identity* Mean (SD)	*Rank* Mean (SD)	*Prize* Mean (SD)
Group Size	4.52 (1.01)	4.41 (0.95)	4.69 (1.01)	4.67 (1.01)
Experience	0.62 (0.49)	0.78 (0.42)	0.71 (0.45)	0.68 (0.47)
Private	0.79 (0.41)	0.89 (0.31)	0.85 (0.36)	0.89 (0.31)
Men Share	0.47 (0.28)	0.41 (0.28)	0.49 (0.30)	0.44 (0.30)
Median Age	32.88 (9.81)	30.26 (7.64)^{b}	33.69 (8.47)^{a}	31.47 (9.37)
German	0.89 (0.31)	0.99 (0.11)	0.94 (0.25)	0.96 (0.19)
Observations	112	85	94	82

Notes: Rows report means on the group level. Group size denotes the number of team members. Experience is a dummy for teams with at least one member who experienced an escape room challenge before. Private is a dummy whether a team participates as a private event (1) or whether the team belongs to a team building event (0). Men Share refers to the share of male team members. Median Age is defined as the median of all participants' guessed age categories' midpoint in a team. German is a dummy for German-speaking (1) or English-speaking (0) teams. Standard deviations in parentheses. Stars indicate significant differences to Control (p-values adjusted for multiple hypothesis testing following List et al. (2019), with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$); $\{a,b,c\}$ indicate differences to {*Identity*, *Rank*, *Prize*} at the ten percent level.

2.4 Hypotheses

The sense of identity and belonging is a fundamental human need (see, e.g., Baumeister and Leary, 2017). Experimental evidence from the laboratory suggests that salient team identity can alter cooperation and coordination within groups as well as reciprocity among agents, all of which are crucial for successful performance in the task at hand. For instance, Chen and Li (2009) use a (near) minimal group design and find that participants are 19 percent more likely to reward an in-group match for good behavior but 13 percent less likely to punish an in-group match for misbehavior. Drouvelis and Nosenzo (2013) provide evidence that group identity is beneficial in contexts that allow leading by example, and Eckel and Grossman (2005) show that team identification may limit individual shirking and free-riding in environments with the character of a public good (in particular when paired with joint activities such as group problem-solving). Further, identity has been shown to affect group coordination and conflict (Chen and Chen, 2011; Chen et al., 2014; Chowdhury, 2021).

Our design focuses on increasing the salience of team identity. While pre-existing groups arrive on the premises of our collaboration partner, jointly elaborating on and choosing a team name renders identity more salient. Our approach reflects current business strategies pursued by companies relying on structures based on agile teams rather

than strict hierarchical structures.[11] In our context, we thus expect performance improvements when a team's identity is rendered more salient.

Hypothesis 1 *Rendering team identity more salient by asking team members to jointly deliberate on and choose a team name improves team performance.*

Competition between teams may reduce free-riding within each team due to image concerns. For instance, evidence from individual routine tasks shows that non-instrumental rewards which encompass image value can substantially improve performance, particularly for the upper quantile of the performance distribution (Kosfeld and Neckermann, 2011). Further, evidence on team performance from routine tasks suggests that rank incentives can substantially affect image concerns, and thereby team composition and performance. While changes driven by image concerns do not necessarily result in better performance (see, e.g., Bandiera et al., 2013; Kosfeld et al., 2017), positive effects have been observed in environments in which team identity was likely to be strong and salient (Delfgaauw et al., 2013). In line with these findings, we hypothesize that image concerns boost performance (in addition to identity), particularly for the upper end of the performance distribution.

Hypothesis 2 *Strengthening image concerns by implementing public rankings improves team performance, particularly for top-performing teams.*

Field experiments randomly assigning teams to tournaments with prizes have so far mainly focused on routine tasks. For example, Erev et al. (1993) showed that tournament incentives can help teams of orange pickers to overcome problems of free-riding innate to environments that require voluntary contributions. Blimpo (2014) extends this link to learning outcomes and finds substantial and positive effects of tournaments with monetary prizes when teams of students compete across schools. Similarly, positive effects are also observed when tournaments involve non-monetary prizes (grade improvements) that have instrumental value (Bigoni et al., 2015). In line with expected image and instrumental returns from effort, such tournaments increase the performance of good students

[11] Based on insights from social and applied psychology (see e.g. Van Knippenberg, 2000; Van Dick et al., 2006) suggesting a strong positive relation between organizational identification and organizational citizenship, many firms emphasize team identity as an important factor for success and explicitly encourage the choice of a team name (see for example *Calabrio*, `https://www.calabrio.com/wfo/workforce-management/boost-belonging-motivation-through-team-names/` and Ye et al., 2020).

while they often appear less effective for students at the lower end of the performance distribution (De Paola et al., 2012). In the context of production, Delfgaauw et al. (2013) provide evidence from sales team competitions with and without prizes in discount stores. They observe positive effects of competition, both for tournaments with ranks only and tournaments with prizes. However, they find no evidence that financial rewards led to additional performance improvements, potentially due to strong image concerns and related ceiling effects or due to perceived instrumental values of ranks for employees (e.g., better perceived career opportunities or lower likelihood of job loss). Given the evidence discussed above, vis-a-vis pure image concerns, we expect tournaments with prizes to further improve team performance, particularly for teams at the upper part of the performance distribution.

Hypothesis 3 *Adding a monetary prize to the tournament improves team performance, particularly for top-performing teams.*

3 Results

3.1 Team performance

We employ two outcome variables to measure team performance. First, to capture effects on the extensive margin, we consider whether a team manages to complete the task within the given time limit of 60 minutes. Second, we consider variation on the intensive margin by studying teams' finishing times, i.e., the time needed to complete the task. The effects of the three distinct components of a tournament, *Identity*, *Rank*, and *Prize* are shown in Table 2. Since these components are added to the different treatments in a step-wise fashion, we code a dummy variable for each component based on whether this component existed in the treatment the observation stems from. For example, treatment *Prize*, apart from a monetary prize (*Prize*), also contains the feature of a ranking (*Rank*), and the selection of a team name (*Identity*). This coding allows us to cleanly identify the *additional* effect of each component on our outcome measures.[12] All specifications include room fixed effects to take into account the differing levels of difficulty that each room bears. We cluster standard errors at the weekly level (the level of treatment assign-

[12] In Appendix Section A.2, we provide results from additional analyses in which we use treatment dummies instead. These are in line with the results presented in the main text.

Table 2: Team performance (completion and finishing time)

	Completed within 60 minutes				Finishing time			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
+ *Identity*	-0.086	-0.099	-0.048	-0.045	1.377	1.910	1.668	1.590
	(0.052)	(0.066)	(0.060)	(0.056)	(0.870)	(0.970)	(1.180)	(1.117)
	[0.198]	[0.241]	[0.434]	[0.447]	[0.219]	[0.145]	[0.206]	[0.218]
+ *Rank*	0.105	0.093	0.081	0.079	-2.788*	-2.583*	-2.575**	-2.515**
	(0.046)	(0.049)	(0.048)	(0.045)	(0.856)	(0.801)	(0.851)	(0.836)
	[0.126]	[0.182]	[0.230]	[0.188]	[0.055]	[0.051]	[0.034]	[0.034]
+ *Prize*	0.091**	0.092**	0.079**	0.084**	-2.214**	-2.391**	-2.200**	-2.330*
	(0.031)	(0.028)	(0.030)	(0.026)	(1.047)	(1.224)	(1.275)	(1.319)
	[0.047]	[0.032]	[0.033]	[0.020]	[0.042]	[0.033]	[0.040]	[0.064]
Mean in Control	0.527	0.527	0.527	0.527	56.470	56.470	56.470	56.470
Observations	373	373	373	373	373	373	373	373
Team Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Staff FE	No	No	Yes	Yes	No	No	Yes	Yes
Weekday FE	No	No	No	Yes	No	No	No	Yes

Notes: The table displays average marginal effects from Probit regressions of whether a team completed the task within 60 minutes (Columns (1) through (4)), and Tobit regressions of finishing time (Columns (5) through (8)). The main explanatory variables are indicators whether the observation stems from a treatment that included the component(s) *Identity*, *Rank*, or *Prize*. All columns include room fixed effects. Each column indicates whether team controls (group size, share of males, experience, median age, language, private), staff, and weekday fixed effects are included. Standard errors in parentheses are clustered at the week level. p-values from wild-cluster bootstrapping following Cameron et al. (2008) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

ment), and, because of the relatively low number of clusters, we provide p-values from wild-cluster bootstrapping following Cameron et al. (2008).

Columns (1) through (4) of Table 2 provide results from a series of Probit regressions, in which we estimate the marginal effects of each component on the probability of successfully completing the task. We control for team characteristics as per Table 1 starting in Column (2), and add fixed effects for the ETR staff member on duty from Column (3). Column (4) shows our preferred specification, which also includes a fixed effect for the day of the week. Columns (5) through (8) repeat the same step-wise inclusion of controls and fixed effects, but instead use the time a team needs to complete the task as the dependent variable in a series of Tobit regressions (with 60 minutes as the upper limit).

The top row shows the results from making the identity salient. Counter our expectations, teams in treatments encompassing the component *Identity* are not more likely to complete the task in 60 minutes, nor do they finish earlier. The coefficients are statistically insignificant and, if anything, teams in *Identity* were less successful than teams

in *Control*. Finally, the effect sizes of *Identity* are of relatively small magnitudes as compared to the effects of the other components when controlling for weekday fixed effects (Columns (4) and (8)). We conclude with Result 1:

Result 1 *Salient identity alone does not improve team performance.*

Adding a ranking (on top of making participants choose a team name) tends to make teams more likely to complete the task within 60 minutes (see Columns (1) through (4)) but the results are statistically insignificant due to the relatively large standard errors. However, adding a ranking significantly improves teams' finishing times by about 2.5 minutes (see Columns (5) through (8)). Hence, image concerns mainly enhance performance at the intensive margin (in line with the idea that mostly top-performing teams are affected). We summarize these findings in Result 2:

Result 2 *Adding a weekly competition for social image improves team performance along the intensive, but not the extensive margin.*

Adding a *Prize* to the weekly competition results in statistically significant performance improvements (see bottom row of Table 2). Teams are approximately 8 percentage points more likely to successfully complete the task within the time frame, and require 2.3 minutes less for completion. We conclude with Result 3:

Result 3 *Adding a prize to the weekly competition improves team performance along the extensive and intensive margins.*

As has become clear, we have found that tournaments can effectively improve team performance in non-routine tasks. Overall, the tournament with a prize increases the completion rate by more than 20 percent (almost 12 percentage points) and reduces finishing times by more than 3 minutes (remaining times are almost doubled). Additional robustness tests for our main results can be found in the Appendix. Section A.2 provides analyses based on treatment dummies instead of a component-based approach, with similar results. In Section A.3, we conduct a randomization inference exercise confirming our findings.

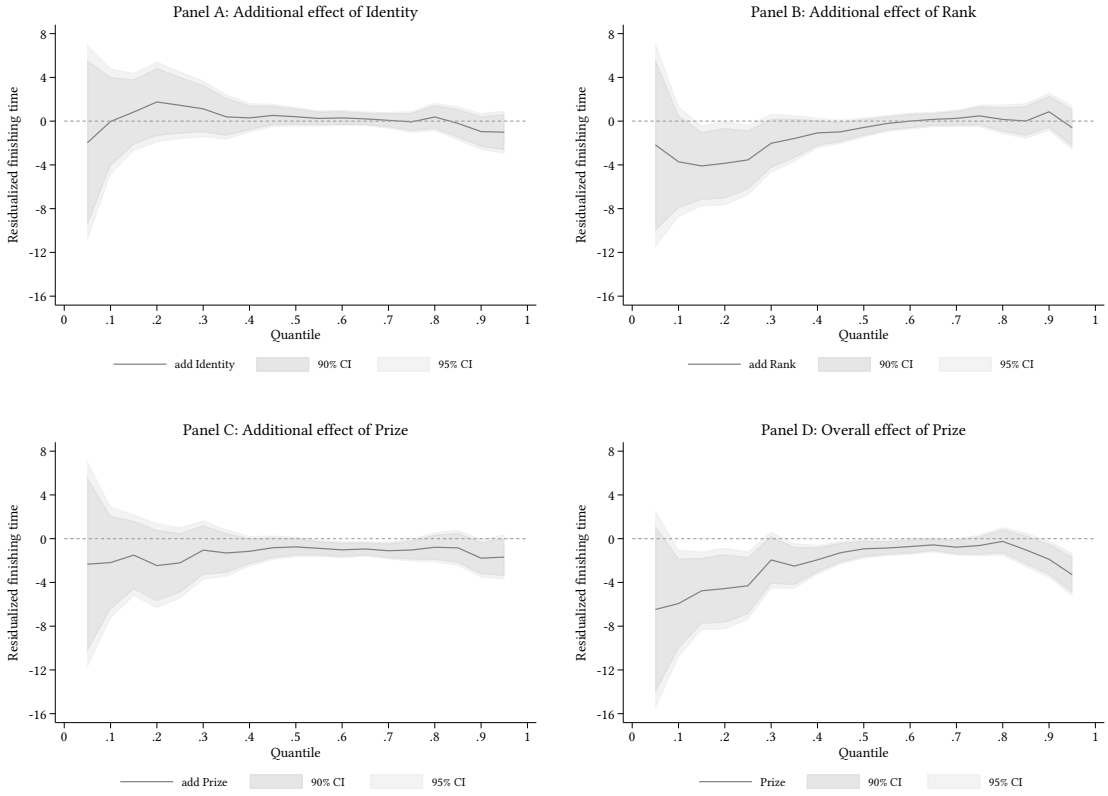
3.2 Team characteristics and the efficacy of tournaments

Competition for ranks and prizes may affect teams differently, due to their composition and potential for performance. To investigate such heterogeneity, we begin by illustrating in more detail how *Rank* and *Prize* influence teams across the entire performance spectrum using quantile regressions on residualized finishing times. We predict finishing times and residuals for all teams using the same fully specified Tobit regression as in Table 2, Column (8), including team controls, room, staff and weekday fixed effects.[13]

Panel A of Figure 1 shows that asking teams to discuss and choose a team name jointly before working on the task (*Identity*) does not affect performance along the whole performance distribution, confirming our earlier finding with respect to the comparison of *Control* and *Identity*. Panel B of Figure 1 shows that compared to *Identity*, the addition of a weekly competition with a public ranking (*Rank*) reduced the finishing times of the top performers, i.e., the lowest quantiles. The extra effect of rank incentives on the residualized finishing times declines along the performance distribution and becomes indistinguishable from zero around the 30% quantile. Panel C compares the residualized finishing times in *Rank* with those in *Prize*. Three interesting findings arise. First, *Prize* seems to further improve the finishing times of top performers substantially, but the effect lacks statistical significance due to the large confidence bands. Second, the positive impact of monetary prizes over rankings becomes significant around the 50% quantile and turns insignificant beyond the 75% quantile. Third, even though not always statistically significant, the estimated effects of *Prize* are all of similar magnitudes across the quantiles, suggesting a positive effect on the entire performance distribution. Panel D shows a comparison of residualized finishing times between *Control* and *Prize*, and thus the compound effect of implementing a tournament with a monetary prize. The results show that tournaments improve performance along a large part of the distribution, so that teams facing salient team identity, image, and instrumental concerns can rank better than similarly composed teams under the *Control* condition. In settings where top performance is particularly important, such as in many innovation contexts, public rankings, therefore, seem to be highly effective, whereas monetary prizes should be offered whenever performance below the very top is also important.

[13] The results in Table 2 did not show any performance improvement of *Identity* over *Control*. To increase the statistical power, we therefore use observations from both *Identity* and *Control* for predicting finishing times.

Figure 1: Quantile regressions on residualized finishing times



Notes: The figure shows quantile regressions on residualized finishing times. Panel A shows the additional effect of salient team identity (compares *Identity* to *Control*). Panel B shows the additional effect of a public ranking (compares *Rank* to *Identity*). Panel C shows the additional effect of a monetary prize (compares *Prize* to *Rank*). And Panel D shows the overall effect of a tournament with a monetary prize (compares *Prize* to *Control*). The line at zero marks residualized finishing times in the comparison group. Negative (positive) values indicate reductions (increases) in residualized finishing times due to *Identity* (Panel A), *Rank* (Panel B), and *Prize* (Panel C and Panel D).

To understand the role of team composition for possible heterogeneity in the observed treatment effects, we conducted additional regression analyses including interaction terms between our treatment indicators and observable team characteristics, presented in Appendix Section A.4. We do not find strong heterogeneity in the efficacy of our treatments, but suggestive evidence that rankings are particularly effective when teams are mostly composed of men (in line with the previous literature on competition and gender in routine tasks (Niederle and Vesterlund, 2011; Schram et al., 2019)).

3.3 Originality and potential crowding out

Prior research has suggested that incentives and competition may be ineffective (or even counterproductive) when production involves non-routine tasks that require thinking out of the box. Incentives may lead to focusing (Duncker, 1945), and thereby reduce thinking out of the box, and, in complex tasks, incentives may systematically discourage the exploration of new and original approaches (e.g. Amabile, 1996; Azoulay et al., 2011; Ederer and Manso, 2013; McCullers, 1978; McGraw, 1978). Furthermore, we incentivized performance in terms of teams' finishing times and not according to originality of solutions. Teams may thus substitute speed for originality (for an excellent discussion and evidence from the laboratory see also Laske and Schroeder, 2016).

Our setting offers the possibility of testing for such potential discouragement or substitution, as teams had the opportunity to seek external help using up to five hints, which did not negatively affect their rank in the tournament. We focus on the number of hints taken when adding different tournament components as well as on the timing of the hints. In general, the number of hints and finishing times are positively correlated (Spearman's $\rho = 0.5191$, $p < 0.001$), as worse teams are more likely to seek help (take more hints on average). However, the number of hints requested does not differ significantly across treatments (Kruskal–Wallis test, $p = 0.6041$). If anything, it appears as if teams in *Prize* take on average slightly fewer hints (*Control*: 3.39, *Identity*: 3.29, *Rank*: 3.25, *Prize*: 3.18).

Additional analyses of the number of hints and their timing (see Appendix Section A.5) confirm that despite the positive effect on performance, none of the treatments significantly increases the number of hints taken, nor their timing. Taken together, these results show that work environments in innovation contexts sharing the features of our non-routine analytical team task are not susceptible to a reduction in teams' inclination to pursue original approaches when presented with tournament incentives.

Offering extrinsic incentives could crowd out intrinsic motivation (e.g. Deci et al., 1999; Eckartz et al., 2012; Gerhart and Fang, 2015; Hennessey and Amabile, 2010) to perform the task at all. The challenging nature of non-routine analytical tasks renders them particularly exciting for intrinsically motivated workers (for a discussion see also Autor and Handel, 2013; Delfgaauw and Dur, 2010; Friebel and Giannetti, 2009), as in these settings workers can make new discoveries and experience progress jointly. Our setting provides us with teams that are highly motivated to perform the task (teams are even

willing to pay for it) and thus a unique opportunity to test whether the image and instrumental concerns innate to tournaments affect the intrinsic motivation to perform a similar task again. To evaluate whether our treatments indeed reduced a team's intrinsic motivation, we focus on a revealed preference measure. After completion of the task, all teams were offered the opportunity to buy a voucher at a reduced price allowing them to perform a new but comparable task again (at any branch of *ExitTheRoom*).

In contrast to the idea that tournaments may reduce a team's intrinsic motivation to work on a similar task again, we find small, positive, but statistically insignificant effects (see Appendix Table A.7). As such, our findings speak against a substantial crowding out of intrinsic motivation for future participation and underline the positive roles of the image and instrumental concerns innate to tournaments.

4 Conclusion

Tournaments are an important and often-used mechanism to foster innovation (Lindegaard, 2010; Terwiesch and Ulrich, 2009; Terwiesch and Xu, 2008; Scotchmer, 2004). They not only involve substantial instrumental incentives but also include important behavioral aspects that can foster team performance in non-routine tasks.

Our study exploited the unique opportunity to exogenously introduce step by step two behavioral features innate to tournament incentives (salient team identity and non-instrumental image concerns) as well as monetary prizes, treating a large number of teams in a natural field experiment. Thereby, we have been able to present causal evidence about the relative importance of these three components and the overall effect of tournaments with a monetary prize on team performance in non-routine tasks.

We have found that tournament incentives substantially improve team performance when tournaments involve team names, a public ranking, and a prize for the best performing team. Public rankings alone improve the top performances, but do not substantially affect teams at the lower end of the distribution of performance. Lastly, we have shown that simply increasing the salience of team identity is not enough to improve performance in our non-routine team tasks. Hence, our findings show that some behavioral aspects (such as identity) can be fragile predictors of performance, whereas others, such as image and instrumental concerns, appear more robust.

Apart from decomposing the effects of salient team identity, image concerns, and instrumental concerns, we have provided novel evidence about the efficacy of tournaments in non-routine team tasks. We observe no reduction when tournaments are introduced in the originality of the solutions or in the teams' intrinsic motivation to perform similar tasks again. Hence, we have confirmed that competition can be an important tool to stimulate thoughtful and original team performance. As we have observed a revealed preference measure of a team's willingness to work on a similar task, before the team receives actual feedback on its relative performance, our findings further suggest that the potentially negative effects of rank or tournament incentives observed in routine tasks (see e.g. Barankay, 2012; Ashraf et al., 2014; Ashraf, 2019; Blader et al., 2020) probably resulted from actual, discouraging performance feedback for under-performing teams rather than from the anticipation of such feedback or competition per se. Avoiding such feedback, we have obtained robust evidence for the important roles of image and instrumental concerns in the efficacy of tournaments in non-routine analytical team tasks.

References

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *Quarterly Journal of Economics 115*(3), 715–753.

Amabile, T. M. (1996). *Creativity in context: Update to the social psychology of creativity*. Boulder, Colorado: Westview Press.

Ärlebäck, J. B. and L. Albarracín (2019). The use and potential of fermi problems in the stem disciplines to support the development of twenty-first century competencies. *ZDM Mathematics Education 51*(6), 979–990.

Ashraf, A. (2019). Do performance ranks increase productivity? Evidence from a field experiment. *CRC TRR190, Discussion Paper No. 196*.

Ashraf, N., O. Bandiera, and S. S. Lee (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior & Organization 100*, 44–63.

Athey, S. and G. W. Imbens (2017). The econometrics of randomized experiments. In *Handbook of Economic Field Experiments*, Volume 1, pp. 73–140. Elsevier.

Autor, D. and B. Price (2013). The changing task composition of the US labor market: An update of autor, levy, and murnane (2003). *Working Paper*.

Autor, D. H. and M. J. Handel (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of Labor Economics 31*(S1), S59–S96.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics 118*(4), 1279–1333.

Azoulay, P., J. S. Graff Zivin, and G. Manso (2011). Incentives and creativity: Evidence from the academic life sciences. *RAND Journal of Economics 42*(3), 527–554.

Ball, S., C. Eckel, P. J. Grossman, and W. Zame (2001). Status in markets. *Quarterly Journal of Economics 116*(1), 161–188.

Bandiera, O., I. Barankay, and I. Rasul (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association 11*(5), 1079–1114.

Bandiera, O., G. Fischer, A. Prat, and E. Ytsma (2021). Do women respond less to performance pay? Building evidence from multiple experiments. *American Economic Review: Insights [forthcoming]*.

Barankay, I. (2012). Rank incentives: Evidence from a randomized workplace experiment.

Baumeister, R. F. and M. R. Leary (2017). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. In D. Canter, B. Laursen, and R. Zukauskiene (Eds.), *Interpersonal Development*. London: Routledge.

Bigoni, M., M. Fort, M. Nardotto, and T. G. Reggiani (2015). Cooperation or competition? a field experiment on non-monetary learning incentives. *The BE Journal of Economic Analysis & Policy 15*(4), 1753–1792.

Blader, S., C. Gartenberg, and A. Prat (2020). The contingent effect of management practices. *Review of Economic Studies 87*(2), 721–749.

Blanes i Vidal, J. and M. Nossol (2011). Tournaments without prizes: Evidence from personnel records. *Management Science 57*(10), 1721–1736.

Blimpo, M. P. (2014). Team incentives for education in developing countries: A randomized field experiment in Benin. *American Economic Journal: Applied Economics 6*(4), 90–109.

Borrego, C., C. Fernández, I. Blanes, and S. Robles (2017). Room escape at class: Escape games activities to facilitate the motivation and learning in computer science. *Journal of Technology and Science Education 7*(2), 162–171.

Bradler, C., S. Neckermann, and A. J. Warnke (2019). Incentivizing creativity: A large-scale experiment with performance bonuses and gifts. *Journal of Labor Economics 37*(3), 793–851.

Brown, J. (2011). Quitters never win: The (adverse) incentive effects of competing with superstars. *Journal of Political Economy 119*(5), 982–1013.

Brown, J. and D. B. Minor (2014). Selecting the best? Spillover and shadows in elimination tournaments. *Management Science 60*(12), 3087–3102.

Brunt, L., J. Lerner, and T. Nicholas (2012). Inducement prizes and innovation. *Journal of Industrial Economics 60*(4), 657–696.

Bursztyn, L. and R. Jensen (2015). How does peer pressure affect educational investments? *Quarterly Journal of Economics 130*(3), 1329–1367.

Bursztyn, L. and R. Jensen (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics 9*, 131–153.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics 90*(3), 414–427.

Charness, G. and D. Grieco (2019). Creativity and incentives. *Journal of the European Economic Association 17*(2), 454–496.

Chen, R. and Y. Chen (2011). The potential of social identity for equilibrium selection. *American Economic Review 101*(6), 2562–89.

Chen, Y. and S. X. Li (2009). Group identity and social preferences. *American Economic Review 99*(1), 431–57.

Chen, Y., S. X. Li, T. X. Liu, and M. Shih (2014). Which hat to wear? Impact of natural identities on coordination and cooperation. *Games and Economic Behavior 84*, 58–86.

Chowdhury, S. M. (2021). The economics of identity and conflict. In *Oxford Research Encyclopedia of Economics and Finance.*

Chowdhury, S. M., J. Y. Jeon, and A. Ramalingam (2016). Identity and group conflict. *European Economic Review 90*, 107–121.

De Paola, M., V. Scoppa, and R. Nisticò (2012). Monetary incentives and student achievement in a depressed labor market: Results from a randomized experiment. *Journal of Human Capital 6*(1), 56–85.

Deci, E. L., R. Koestner, and R. M. Ryan (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin 125*(6), 627–668.

Delfgaauw, J. and R. Dur (2010). Managerial talent, motivation, and self-selection into public management. *Journal of Public Economics 94*(9), 654–660.

Delfgaauw, J., R. Dur, J. Sol, and W. Verbeke (2013). Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics 31*(2), 305–326.

Delfgaauw, J., R. Dur, and M. Souverijn (2020). Team incentives, task assignment, and performance: A field experiment. *Leadership Quarterly 31*(3).

Drouvelis, M. and D. Nosenzo (2013). Group identity and leading-by-example. *Journal of Economic Psychology 39*, 414–425.

Duncker, K. (1945). On problem-solving. *Psychological Monographs 58*(5), i–113.

Eckartz, K., O. Kirchkamp, and D. Schunk (2012). How do incentives affect creativity? *CESifo Working Paper Series No. 4049*.

Eckel, C. C. and P. J. Grossman (2005). Managing diversity by creating team identity. *Journal of Economic Behavior & Organization 58*(3), 371–392.

Ederer, F. and G. Manso (2013). Is pay for performance detrimental to innovation? *Management Science 59*(7), 1496–1513.

Englmaier, F., S. Grimm, D. Schindler, and S. Schudy (2018). The effect of incentives in non-routine analytical team tasks-evidence from a field experiment. *CEPR Discussion Paper 13226*.

Englmaier, F., A. Roider, and U. Sunde (2017). The role of communication of performance schemes: Evidence from a field experiment. *Management Science 63*(12), 4061–4080.

Erev, I., G. Bornstein, and R. Galili (1993). Constructive intergroup competition as a solution to the free rider problem: A field experiment. *Journal of Experimental Social Psychology 29*(6), 463–478.

Fershtman, C., H. K. Hvide, and Y. Weiss (2006). Cultural diversity, status concerns and the organization of work. *Research in Labor Economics 24*, 361–396.

Friebel, G. and M. Giannetti (2009). Fighting for talent: Risk-taking, corporate volatility and organisation change. *Economic Journal 119*(540), 1344–1373.

Gerhart, B. and M. Fang (2015). Pay, intrinsic motivation, extrinsic motivation, performance, and creativity in the workplace: Revisiting long-held beliefs. *Annual Review of Organizational Psychology and Organizational Behavior 2*(1), 489–521.

Gibbs, M., S. Neckermann, and C. Siemroth (2017). A field experiment in motivating employee ideas. *Review of Economics and Statistics 99*(4), 577–590.

Gross, D. P. (2020). Creativity under fire: The effects of competition on creative production. *Review of Economics and Statistics 102*(3), 583–599.

Harrison, G. W. and J. A. List (2004). Field experiments. *Journal of Economic Literature 42*(4), 1009–1055.

Hennessey, B. A. and T. M. Amabile (2010). Creativity. *Annual Review of Psychology 61*(1), 569–598.

Khan, B. Z. (2015). Inventing prizes: a historical perspective on innovation awards and technology policy. *Business History Review 89*(4), 631–660.

Kluger, A. N. and A. DeNisi (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin 119*(2), 254.

Kosfeld, M. and S. Neckermann (2011). Getting more work for nothing? symbolic awards and worker performance. *American Economic Journal: Microeconomics 3*(3), 86–99.

Kosfeld, M., S. Neckermann, and X. Yang (2017). The effects of financial and recognition incentives across work contexts: The role of meaning. *Economic Inquiry 55*(1), 237–247.

Laske, K. and M. Schroeder (2016). Quantity, quality, and originality: The effects of incentives on creativity. *Cologne Graduate School Working Paper Series 07-01*.

Lazear, E. P. and P. Oyer (2012). Chapter 12: Personnel economics. In *The Handbook of Organizational Economics*, pp. 479–519. Princeton University Press.

Lazear, E. P. and S. Rosen (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy 89*(5), 841–864.

Lindegaard, S. (2010). *The open innovation revolution: essentials, roadblocks, and leadership skills*. Wiley.

List, J. A., A. M. Shaikh, and Y. Xu (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics 22*(4), 773–793.

McCullers, J. C. (1978). Issues in learning and motivation. In M. R. Lepper and D. Greene (Eds.), *The Hidden Costs of Reward: New perspectives on the psychology of human motivation*, pp. 5–18. New York: Psychology Press.

McGraw, K. O. (1978). The detrimental effects of reward on performance: A literature review and a prediction model. In M. R. Lepper and D. Green (Eds.), *The Hidden Costs of Reward: New perspectives on the psychology of human motivation*, pp. 33–60. New York: Psychology Press.

Moldovanu, B., A. Sela, and X. Shi (2007). Contests for status. *Journal of Political Economy 115*(2), 338–363.

Morgan, J., S. Neckermann, and D. Sisak (2020). Peer evaluation and team performance: An experiment on complex problem solving. mimeo.

Niederle, M. and L. Vesterlund (2011). Gender and competition. *Annual Review of Economics 3*(1), 601–630.

Rigdon, J. and M. G. Hudgens (2015). Randomization inference for treatment effects on a binary outcome. *Statistics in Medicine 34*(6), 924–935.

Schram, A., J. Brandts, and K. Gërxhani (2019). Social-status ranking: A hidden channel to gender inequality under competition. *Experimental Economics 22*(2), 396–418.

Scotchmer, S. (2004). *Innovation and Incentives*. MIT press.

Sen, A. (2007). *Identity and violence: The illusion of destiny*. Penguin Books India.

Tajfel, H. and J. Turner (2001). An integrative theory of intergroup conflict. In M. A. Hogg and D. Abrams (Eds.), *Key Readings in Social Psychology*, pp. 94–109. New York: Psychology Press.

Terwiesch, C. and K. T. Ulrich (2009). *Innovation tournaments: Creating and selecting exceptional opportunities*. Harvard Business Press.

Terwiesch, C. and Y. Xu (2008). Innovation contests, open innovation, and multiagent problem solving. *Management Science 54*(9), 1529–1543.

Van Dick, R., M. W. Grojean, O. Christ, and J. Wieseke (2006). Identity and the extra mile: Relationships between organizational identification and organizational citizenship behaviour. *British Journal of Management 17*(4), 283–301.

Van Knippenberg, D. (2000). Work motivation and performance: A social identity perspective. *Applied Psychology 49*(3), 357–371.

Ye, T., W. Ai, Y. Chen, Q. Mei, and J. Ye (2020). Virtual teams for the modern workforce. mimeo.

A Appendix

A.1 Screenshot of an actual ranking on Facebook

Figure A.1: Screenshot of an actual ranking on Facebook (in German)

Notes: The figure shows a screenshot of an actual ranking on Facebook (in German). Teams are ranked according to their finishing times and all teams that did not complete the task are assigned to the same rank.

A.2 Direct treatment comparisons

Complementing our main analyses, which show the additional effects of each tournament component, Table A.1 compares each treatment directly to *Control*. By design, the results for treatment *Identity* remain the same. Introducing rank incentives (*Rank*) does not lead to statistically significant different completion rates or lower finishing times on average compared to *Control*, which is driven by the heterogenous effect of a public ranking. In contrast, treatment *Prize* is significantly associated with a higher likelihood of succeeding within 60 minutes as well as lower finishing times across all four specifications, indicating that introducing a combination of rank incentives and a prize for the best team of the week improves team performances along the whole distribution of performances. The completion rate increases by more than 20 percent (almost 12 percentage points) and the remaining time is almost doubled (more than 3 minutes lower finishing times). Further, we provide alternative specifications using linear regressions and GLM models with log link in Table A.2, confirming the robustness of these findings. Since the salience of team identity is an innate feature of tournaments, Tables A.1 and A.2 further provide test statistics for the differences between *Identity* and *Rank*, and *Identity* and *Prize* (see rows 4 to 6). Akin to business contexts in which team identity is already salient (e.g., due to existing names for the team or brand), this comparison reveals the effects of *Rank* and *Prize*, while holding the salience of *Identity* in our setting constant. These comparisons reveal that, on average, *Rank* significantly improves teams' finishing times as compared to *Identity* ($0.028 < p < 0.078$), and *Prize* improves both the likelihood of completion as well as finishing times ($0.004 < p < 0.039$).

Table A.1: Team performance (completion and finishing times)

	Completed within 60 minutes				Finishing time			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Identity	-0.086	-0.099	-0.048	-0.045	1.377	1.910	1.668	1.590
	(0.052)	(0.066)	(0.060)	(0.056)	(0.870)	(0.970)	(1.180)	(1.117)
	[0.198]	[0.241]	[0.434]	[0.447]	[0.219]	[0.145]	[0.206]	[0.218]
Rank	0.019	-0.005	0.033	0.034	-1.411	-0.673	-0.906	-0.925
	(0.045)	(0.050)	(0.050)	(0.041)	(0.883)	(0.892)	(1.082)	(1.008)
	[0.690]	[0.918]	[0.524]	[0.481]	[0.214]	[0.484]	[0.437]	[0.396]
Prize	0.110**	0.087*	0.112**	0.118**	-3.625**	-3.064**	-3.106**	-3.255**
	(0.038)	(0.045)	(0.044)	(0.043)	(1.026)	(1.320)	(1.333)	(1.349)
	[0.019]	[0.080]	[0.032]	[0.026]	[0.033]	[0.039]	[0.036]	[0.032]
Identity = Rank	[0.126]	[0.182]	[0.230]	[0.188]	[0.055]	[0.051]	[0.034]	[0.034]
Rank = Prize	[0.047]	[0.032]	[0.033]	[0.020]	[0.042]	[0.033]	[0.040]	[0.064]
Identity = Prize	[0.026]	[0.039]	[0.025]	[0.030]	[0.012]	[0.011]	[0.008]	[0.013]
Mean in Control	0.527	0.527	0.527	0.527	56.470	56.470	56.470	56.470
Observations	373	373	373	373	373	373	373	373
Team Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Staff FE	No	No	Yes	Yes	No	No	Yes	Yes
Weekday FE	No	No	No	Yes	No	No	No	Yes

Notes: The table displays average marginal effects from Probit regressions of whether a team completed the task within 60 minutes (Columns (1) through (4)), and Tobit regressions of finishing time (Columns (5) through (8)). All columns include room fixed effects. Each column indicates whether team controls (group size, share of males, experience, median age, language, private), staff, and weekday fixed effects are included. Standard errors in parentheses are clustered at the week level. *p*-values from wild-cluster bootstrapping following Cameron et al. (2008) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

Table A.2: Team performance (completion and finishing times)

	Completed within 60 minutes				Finishing time			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Identity	-0.091	-0.102	-0.050	-0.048	0.006	0.013	0.015	0.016
	(0.052)	(0.066)	(0.061)	(0.057)	(0.009)	(0.010)	(0.012)	(0.012)
	[0.210]	[0.261]	[0.522]	[0.504]	[0.538]	[0.231]	[0.254]	[0.272]
Rank	0.014	-0.006	0.033	0.035	-0.019	-0.012	-0.011	-0.012
	(0.043)	(0.049)	(0.049)	(0.039)	(0.011)	(0.010)	(0.012)	(0.013)
	[0.819]	[0.908]	[0.586]	[0.452]	[0.165]	[0.278]	[0.407]	[0.389]
Prize	0.105**	0.092**	0.118**	0.121***	-0.041**	-0.034*	-0.032*	-0.034*
	(0.037)	(0.043)	(0.043)	(0.042)	(0.014)	(0.017)	(0.016)	(0.017)
	[0.039]	[0.043]	[0.013]	[0.010]	[0.033]	[0.089]	[0.058]	[0.087]
Identity = Rank	[0.126]	[0.177]	[0.291]	[0.259]	[0.078]	[0.046]	[0.032]	[0.028]
Rank = Prize	[0.044]	[0.026]	[0.078]	[0.030]	[0.146]	[0.220]	[0.266]	[0.248]
Identity = Prize	[0.036]	[0.037]	[0.037]	[0.027]	[0.004]	[0.012]	[0.017]	[0.031]
Mean in Control	0.527	0.527	0.527	0.527	56.470	56.470	56.470	56.470
Observations	373	373	373	373	373	373	373	373
Team Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Staff FE	No	No	Yes	Yes	No	No	Yes	Yes
Weekday FE	No	No	No	Yes	No	No	No	Yes

Notes: The table displays average marginal effects from OLS regressions of whether a team completed the task within 60 minutes (Columns (1) through (4)), and GLM regressions (with log link) of finishing time (Columns (5) through (8)). All columns include room fixed effects. Each column indicates whether team controls (group size, share of males, experience, median age, language, private), staff, and weekday fixed effects are included. Standard errors in parentheses are clustered at the week level. p-values from wild-cluster bootstrapping following Cameron et al. (2008) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

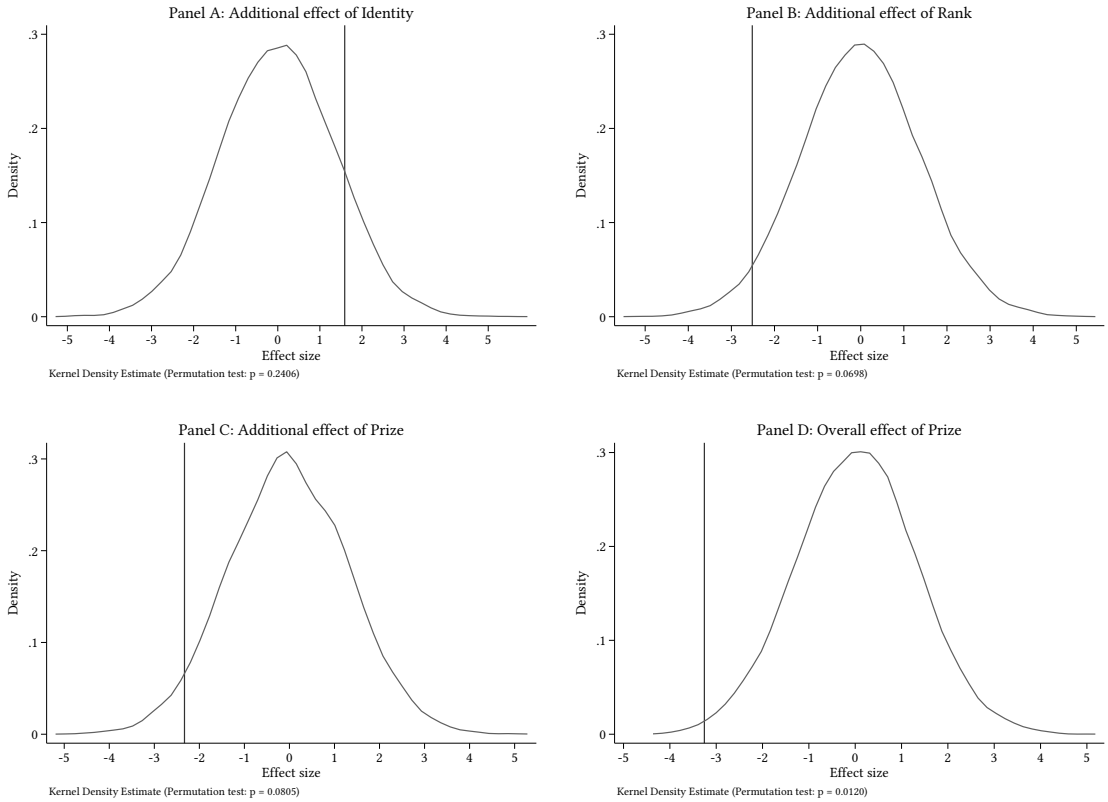
A.3 Randomization inference

In addition, we have also carried out a randomization inference exercise (Athey and Imbens, 2017). Because a treatment effect may also arise due to the randomness of who gets assigned to which condition, we want to establish the probability that our findings indeed result from the treatment. Intuitively, randomization inference asks what would have occurred not only under the actual random assignment, but whether the result would also hold under all possible random assignments of treatments to data. We randomly assigned treatment status (preserving the original ratio between treatments) to observations and estimated our regression equation of interest. By repeating this procedure 10,000 times, we obtain a distribution of counterfactual estimates to which we can compare our actual estimates. The resulting randomization inference p-value is equivalent to the proportion of times the placebo treatment effect was more extreme than the estimated actual treatment effect.

Figure A.2 plots the randomization distributions of the effect sizes of adding *Identity* (Panel A), adding *Rank* (Panel B), adding *Prize* (Panel C) and the overall effect size of *Prize* (Panel D) on finishing time. We abstain from a randomization inference exercise on the probability of finishing the task, because the necessary additivity assumption for constructing a confidence interval is unlikely to be fulfilled for binary outcome variables (Rigdon and Hudgens, 2015).

In each panel, the vertical, solid lines indicate the actually observed effect. Panel A shows the randomization distribution of finishing times when team *Identity* plays a role. The true effect does not appear excessive, and with $p = 0.2406$, we cannot reject the null hypothesis of no individual effect. This is different in Panel B, where we plot the distributions for teams that are subjected to a ranking (*Rank*) in addition. With $p = 0.0698$, the true effect of a reduced finishing time seems unlikely to be a statistical artefact. Panel C shows the randomization distribution for teams with the opportunity to win a monetary prize (on top of being ranked). These teams are much quicker than a random distribution of treatments across observations would have suggested ($p = 0.0805$). Lastly, Panel D shows the randomization distribution for the overall effect of *Prize* (compared to *Control*). The result supports our finding, that a tournament with a monetary prize reduces finishing times substantially ($p = 0.0120$). To summarize, all four panels show that our previous results are robust to randomization inference.

Figure A.2: Randomization distributions of effect sizes



Notes: The figure plots the randomization distributions (10,000 resampling replications) of finishing times. The vertical line in each graph shows the observed effect size for adding *Identity* (Panel A), adding *Rank* (Panel B), adding *Prize* (Panel C) and the overall effect of *Prize* (Panel D).

A.4 Further heterogeneity analyses

To understand the importance of the composition of a team for possible heterogeneity in the observed treatment effects, we estimate whether (and how fast) teams finish the task in linear probability (and Tobit) models by including interaction terms between our treatment indicators and observable team characteristics. Appendix Table A.3 shows that, for the probability of completing the task, *Rank* interacts positively with the share of males in a team. This is not only in line with the recent literature on gender differences in the willingness to compete (Niederle and Vesterlund, 2011), but also with recent evidence from laboratory experiments studying the role of gender in individual competition without prizes in a routine task (Schram et al., 2019). In contrast, condition *Prize* tends to increase team performance irrespective of the observed gender composition and other team characteristics, providing suggestive evidence for agency theory (irrespective of gender) from individual (and mostly routine) tasks (see Bandiera et al., 2021).

Tobit regressions on finishing times as reported in Table A.4 yield results in line with the above-mentioned interaction effect for *Rank*, although less precisely estimated. The more males are in a team, the stronger is the reduction in finishing times due to the competition introduced in *Rank*. Further, they reveal a more nuanced picture in terms of image and instrumental concerns. It turns out that the image concerns prevalent in *Rank* are particularly effective in reducing the finishing times of teams that performed the task with their colleagues (company booking), whereas the monetary incentive in *Prize* was particularly effective in stimulating the performance of private teams (regular booking).

Table A.3: Team performance (completion, interactions)

	Completed within 60 minutes						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
+ *Identity*	-0.048	0.167	-0.133	-0.211	-0.071	0.008	-0.715
	(0.057)	(0.344)	(0.143)	(0.182)	(0.130)	(0.105)	(0.151)
	[0.504]	[0.651]	[0.444]	[0.588]	[0.676]	[0.946]	[0.294]
+ *Rank*	0.083	-0.264	0.169	0.224	-0.066	-0.109	0.409
	(0.047)	(0.294)	(0.159)	(0.174)	(0.090)	(0.183)	(0.152)
	[0.259]	[0.458]	[0.475]	[0.511]	[0.563]	[0.583]	[0.240]
+ *Prize*	0.086**	-0.005	0.080	-0.052	0.207	0.184	0.204
	(0.029)	(0.226)	(0.115)	(0.103)	(0.113)	(0.362)	(0.112)
	[0.030]	[0.985]	[0.566]	[0.629]	[0.160]	[0.687]	[0.202]
Group Size	0.049**	0.041	0.049**	0.049**	0.045*	0.048**	0.052**
	(0.020)	(0.033)	(0.019)	(0.019)	(0.020)	(0.020)	(0.019)
	[0.034]	[0.301]	[0.033]	[0.025]	[0.057]	[0.041]	[0.022]
Experience	0.136	0.131	0.110	0.142*	0.132*	0.137	0.139
	(0.071)	(0.069)	(0.156)	(0.070)	(0.067)	(0.072)	(0.075)
	[0.106]	[0.107]	[0.652]	[0.081]	[0.077]	[0.107]	[0.116]
Private	0.101	0.099	0.099	0.026	0.108*	0.103	0.111*
	(0.059)	(0.062)	(0.064)	(0.096)	(0.055)	(0.061)	(0.063)
	[0.115]	[0.139]	[0.142]	[0.859]	[0.069]	[0.116]	[0.086]
Men Share	0.039	0.032	0.042	0.044	-0.092	0.040	0.037
	(0.088)	(0.089)	(0.088)	(0.083)	(0.169)	(0.089)	(0.087)
	[0.677]	[0.746]	[0.656]	[0.606]	[0.655]	[0.673]	[0.679]
Median Age	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)	(0.003)
	[0.811]	[0.732]	[0.774]	[0.698]	[0.812]	[0.107]	[0.744]
German	-0.128	-0.132	-0.126	-0.119	-0.132	-0.130	-0.302**
	(0.080)	(0.081)	(0.082)	(0.087)	(0.085)	(0.081)	(0.073)
	[0.203]	[0.194]	[0.211]	[0.229]	[0.208]	[0.209]	[0.017]
+ *Identity* x Group Size		-0.049					
		(0.075)					
		[0.544]					
+ *Rank* x Group Size		0.077					
		(0.065)					
		[0.349]					
+ *Prize* x Group Size		0.019					
		(0.045)					
		[0.753]					
+ *Identity* x Experience			0.117				
			(0.193)				
			[0.633]				
+ *Rank* x Experience			-0.112				
			(0.163)				
			[0.615]				
+ *Prize* x Experience			0.006				
			(0.134)				
			[0.971]				

... continued on next page

Table A.3: Team performance (completion, interactions) - continued

	Completed within 60 minutes						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
+ *Identity* x Private				0.185			
				(0.184)			
				[0.552]			
+ *Rank* x Private				-0.161			
				(0.171)			
				[0.561]			
+ *Prize* x Private				0.158			
				(0.135)			
				[0.287]			
+ *Identity* x Men Share					0.049		
					(0.193)		
					[0.811]		
+ *Rank* x Men Share					0.319*		
					(0.124)		
					[0.059]		
+ *Prize* x Men Share					-0.256		
					(0.200)		
					[0.377]		
+ *Identity* x Median Age						-0.002	
						(0.004)	
						[0.851]	
+ *Rank* x Median Age						0.006	
						(0.006)	
						[0.482]	
+ *Prize* x Median Age						-0.003	
						(0.011)	
						[0.823]	
+ *Identity* x German							0.690
							(0.155)
							[0.236]
+ *Rank* x German							-0.318
							(0.167)
							[0.278]
+ *Prize* x German							-0.128
							(0.133)
							[0.392]
Mean in Control	0.527	0.527	0.527	0.527	0.527	0.527	0.527
Observations	373	373	373	373	373	373	373
Team Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Staff FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weekday FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table displays coefficients from OLS regressions of whether a team completed the task within 60 minutes. All columns include room fixed effects. Each column indicates whether team controls (group size, share of males, experience, median age, language, private), staff, and weekday fixed effects are included. Standard errors in parentheses are clustered at the week level. *p*-values from wild-cluster bootstrapping following Cameron et al. (2008) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

Table A.4: Team performance (finishing times, interactions)

	Finishing time						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
+ *Identity*	1.590	-9.216	2.750	4.175	0.208	2.361	35.840
	(1.117)	(6.773)	(2.683)	(2.505)	(2.713)	(3.358)	(3.740)
	[0.218]	[0.210]	[0.329]	[0.312]	[0.945]	[0.491]	[0.387]
+ *Rank*	-2.515**	12.942	-3.046	-8.517*	0.746	3.708	-34.370
	(0.836)	(8.091)	(3.011)	(2.315)	(2.129)	(3.039)	(3.554)
	[0.034]	[0.178]	[0.348]	[0.058]	[0.730]	[0.355]	[0.296]
+ *Prize*	-2.330*	-10.133	-4.148	6.096**	-3.601	-8.542	2.409
	(1.319)	(6.405)	(3.395)	(1.560)	(2.351)	(6.488)	(2.264)
	[0.064]	[0.102]	[0.228]	[0.018]	[0.206]	[0.255]	[0.384]
Group Size	-1.408**	-1.811	-1.416**	-1.475**	-1.363**	-1.382**	-1.447**
	(0.508)	(0.771)	(0.495)	(0.491)	(0.519)	(0.520)	(0.501)
	[0.032]	[0.230]	[0.026]	[0.022]	[0.039]	[0.037]	[0.029]
Experience	-4.334**	-4.229**	-4.351	-4.590***	-4.256**	-4.282**	-4.437**
	(1.384)	(1.324)	(2.157)	(1.362)	(1.347)	(1.398)	(1.426)
	[0.011]	[0.012]	[0.186]	[0.010]	[0.011]	[0.011]	[0.011]
Private	-2.175*	-1.975	-1.992	-1.852	-2.203*	-2.316**	-2.003*
	(1.090)	(1.187)	(1.159)	(1.668)	(1.111)	(1.110)	(1.098)
	[0.067]	[0.115]	[0.117]	[0.334]	[0.070]	[0.046]	[0.077]
Men Share	-1.462	-1.256	-1.623	-1.474	-0.793	-1.503	-1.474
	(1.530)	(1.539)	(1.563)	(1.333)	(3.223)	(1.467)	(1.494)
	[0.325]	[0.403]	[0.286]	[0.270]	[0.821]	[0.302]	[0.316]
Median Age	0.060	0.073	0.063	0.076	0.058	0.120*	0.067
	(0.058)	(0.060)	(0.060)	(0.055)	(0.056)	(0.068)	(0.058)
	[0.250]	[0.151]	[0.243]	[0.105]	[0.263]	[0.077]	[0.194]
German	0.711	0.520	0.690	1.141	0.526	0.854	2.830
	(1.692)	(1.682)	(1.661)	(1.395)	(1.729)	(1.618)	(2.313)
	[0.702]	[0.767]	[0.708]	[0.513]	[0.773]	[0.645]	[0.356]
+ *Identity* x Group Size		2.451					
		(1.417)					
		[0.122]					
+ *Rank* x Group Size		-3.399					
		(1.820)					
		[0.124]					
+ *Prize* x Group Size		1.649					
		(1.313)					
		[0.238]					
+ *Identity* x Experience			-1.456				
			(3.558)				
			[0.684]				
+ *Rank* x Experience			0.660				
			(3.416)				
			[0.898]				
+ *Prize* x Experience			2.513				
			(3.411)				
			[0.422]				

... continued on next page

Table A.4: Team performance (finishing times, interactions) - continued

	Finishing time						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
+ *Identity* x Private				-2.759			
				(2.673)			
				[0.472]			
+ *Rank* x Private				6.968*			
				(2.022)			
				[0.085]			
+ *Prize* x Private				-9.695***			
				(2.260)			
				[0.005]			
+ *Identity* x Men Share					3.300		
					(4.566)		
					[0.503]		
+ *Rank* x Men Share					-7.144		
					(3.515)		
					[0.120]		
+ *Prize* x Men Share					2.490		
					(3.340)		
					[0.605]		
+ *Identity* x Median Age						-0.021	
						(0.092)	
						[0.814]	
+ *Rank* x Median Age						-0.187	
						(0.101)	
						[0.218]	
+ *Prize* x Median Age						0.186	
						(0.177)	
						[0.378]	
+ *Identity* x German							-34.494
							(4.018)
							[0.389]
+ *Rank* x German							31.953
							(3.615)
							[0.323]
+ *Prize* x German							-4.954
							(2.731)
							[0.205]
Mean in Control	56.470	56.470	56.470	56.470	56.470	56.470	56.470
Observations	373	373	373	373	373	373	373
Team Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Staff FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weekday FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table displays coefficients from Tobit regressions of finishing times. All columns include room fixed effects. Each column indicates whether team controls (group size, share of males, experience, median age, language, private), staff, and weekday fixed effects are included. Standard errors in parentheses are clustered at the week level. *p*-values from wild-cluster bootstrapping following Cameron et al. (2008) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

A.5 Originality and potential crowding out

Figure A.3 illustrates the hint taking behavior over time and across treatments. In all treatments, teams request a similar number of hints. If anything, teams in *Prize* tend to take slightly fewer hints. OLS regressions on the number of hints (Table A.5) confirm the non-parametric finding that neither component, *Identity*, *Rank* nor *Prize* affect the originality of the solutions, also when controlling for team characteristics, adding staff, or weekday fixed effects. In fact, all coefficients are small in magnitude, sometimes switch to the opposite sign, and are far from statistically significant.

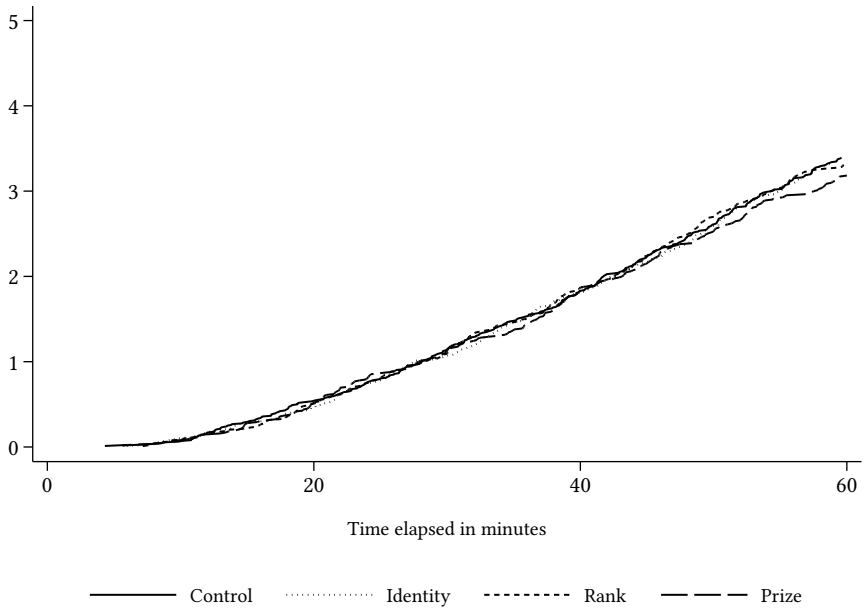
Even though the willingness to explore new and original solutions does not seem to be crowded out if measured by the total number of hints requested, it would still be conceivable that teams request their hints earlier. This would effectively also allow them to rely on external help early on and thus arrive at the solution quicker. Table A.6 shows the coefficients of Tobit regressions on the timing of hints[14] using treatment components as explanatory variables. The results are again small in magnitude and indistinguishable from zero. The step-wise introduction of additional controls and fixed effects does not affect this result.

To shed light on whether particularly (un)successful teams differ in the originality of their solutions, we also present results from linear regressions within quantiles (based on residualized finishing times) in Figure A.4. Panel A shows the difference in the number of hints taken in *Identity* as compared to *Control*. Panel B compares *Rank* with *Identity*. Panel C (D) compares *Prize* with *Rank* (*Control*). No clear and consistent picture emerges: none of the components seem to affect the number of hints taken across the entire performance spectrum.

To analyze whether our treatments reduced a team's intrinsic motivation, Table A.7 presents results from Probit regressions on the marginal effects of the *Identity*, *Rank*, and *Prize* components on purchasing a voucher. As in previous analyses, we add additional controls and fixed effects in each column. The results speak clearly against any crowding out of intrinsic motivation for future participation.

[14] We assigned a time of 60 minutes for all unused hints.

Figure A.3: Hint taking over time



Notes: The figure shows the cumulative distribution of hints by minute in *Control*, *Identity*, *Rank*, and *Prize*.

Table A.5: Originality (number of hints)

	Number of hints			
	(1)	(2)	(3)	(4)
+ *Identity*	-0.058	-0.038	0.048	0.075
	(0.308)	(0.297)	(0.299)	(0.300)
	[0.868]	[0.916]	[0.901]	[0.845]
+ *Rank*	0.028	0.040	0.102	0.100
	(0.342)	(0.305)	(0.296)	(0.288)
	[0.924]	[0.895]	[0.821]	[0.835]
+ *Prize*	-0.142	-0.171	-0.214	-0.213
	(0.279)	(0.248)	(0.211)	(0.187)
	[0.642]	[0.530]	[0.415]	[0.398]
Mean in Control	3.393	3.393	3.393	3.393
Observations	373	373	373	373
Team Controls	No	Yes	Yes	Yes
Staff FE	No	No	Yes	Yes
Weekday FE	No	No	No	Yes

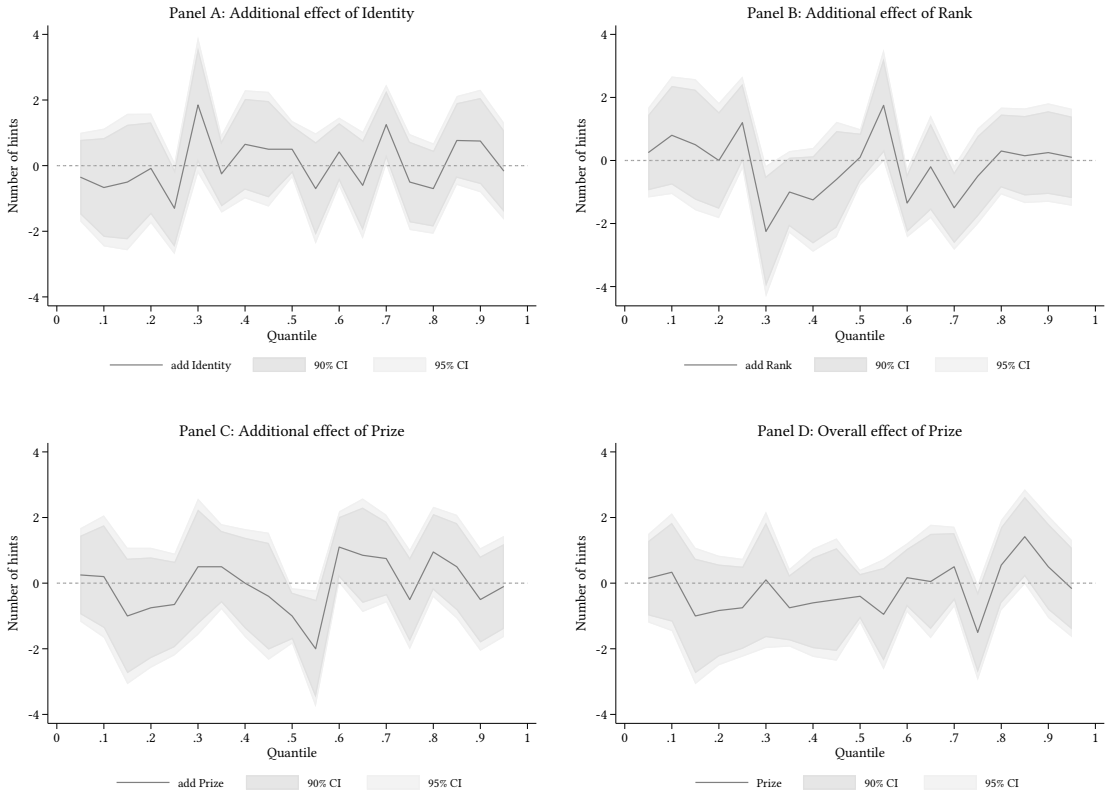
Notes: The table displays coefficients from OLS regressions of number of hints. The main explanatory variables are indicators whether the observation stems from a treatment that included the component(s) *Identity*, *Rank*, or *Prize*. All columns include room fixed effects. Each column indicates whether team controls (group size, share of males, experience, median age, language, private), staff, and weekday fixed effects are included. Standard errors in parentheses are clustered at the week level. *p*-values from wild-cluster bootstrapping following Cameron et al. (2008) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

Table A.6: Originality (timing of hints)

	Timing of hints				
	1st hint	2nd hint	3rd hint	4th hint	5th hint
	(1)	(2)	(3)	(4)	(5)
+ *Identity*	1.279	-0.358	0.105	-0.034	-1.636
	(1.259)	(1.899)	(2.199)	(2.238)	(2.241)
	[0.368]	[0.866]	[0.968]	[0.985]	[0.527]
+ *Rank*	-2.306	-1.977	-1.524	-2.573	0.421
	(1.674)	(2.317)	(2.436)	(2.814)	(2.605)
	[0.203]	[0.447]	[0.567]	[0.376]	[0.881]
+ *Prize*	-0.155	1.808	2.960	3.504	2.619
	(1.829)	(2.108)	(2.184)	(2.402)	(1.999)
	[0.965]	[0.454]	[0.225]	[0.250]	[0.301]
Mean in Control	22.990	37.243	47.715	55.072	58.448
Observations	373	373	373	373	373
Team Controls	Yes	Yes	Yes	Yes	Yes
Staff FE	Yes	Yes	Yes	Yes	Yes
Weekday FE	Yes	Yes	Yes	Yes	Yes

Notes: The table displays coefficients from Tobit regressions of timing of hints. The main explanatory variables are indicators whether the observation stems from a treatment that included the component(s) *Identity*, *Rank*, or *Prize*. All columns include room fixed effects. Each column indicates whether team controls (group size, share of males, experience, median age, language, private), staff, and weekday fixed effects are included. Standard errors in parentheses are clustered at the week level. *p*-values from wild-cluster bootstrapping following Cameron et al. (2008) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

Figure A.4: OLS regressions on number of hints (within quantiles)



Notes: The figure shows OLS regressions (within quantiles sorted by residualized finishing time) on number of hints. Panel A shows the additional effect of salient team identity (compares *Identity* to *Control*). Panel B shows the additional effect of a public ranking (compares *Rank* to *Identity*). Panel C shows the additional effect of a monetary prize (compares *Prize* to *Rank*). And Panel D shows the overall effect of a tournament with a monetary prize (compares *Prize* to *Control*). The line at zero marks the number of hints in the comparison group. Negative (positive) values indicate reductions (increases) in the number of hints due to *Identity* (Panel A), *Rank* (Panel B), and *Prize* (Panel C and Panel D).

Table A.7: Purchased a voucher

	Purchased a voucher			
	(1)	(2)	(3)	(4)
+ *Identity*	0.012	-0.004	0.005	0.009
	(0.025)	(0.028)	(0.024)	(0.027)
	[0.575]	[0.906]	[0.831]	[0.742]
+ *Rank*	0.053	0.041	0.019	0.014
	(0.038)	(0.033)	(0.028)	(0.026)
	[0.276]	[0.332]	[0.551]	[0.621]
+ *Prize*	-0.009	0.004	0.001	0.010
	(0.057)	(0.048)	(0.042)	(0.042)
	[0.890]	[0.928]	[0.984]	[0.794]
Mean in Control	0.179	0.179	0.179	0.179
Observations	373	373	373	373
Team Controls	No	Yes	Yes	Yes
Staff FE	No	No	Yes	Yes
Weekday FE	No	No	No	Yes

Notes: The table displays average marginal effects from Probit regressions of whether a team purchased a voucher. The main explanatory variables are indicators whether the observation stems from a treatment that included the component(s) *Identity*, *Rank*, or *Prize*. All columns include room fixed effects. Each column indicates whether team controls (group size, share of males, experience, median age, language, private), staff, and weekday fixed effects are included. Standard errors in parentheses are clustered at the week level. *p*-values from wild-cluster bootstrapping following Cameron et al. (2008) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

A.6 Water damage

For our main data analysis, we removed five observations because of water damage to ETR's equipment resulting from a burst pipe. Table A.8 repeats the specifications from Table 2 but includes the five omitted data points. The results are very similar.

Table A.8: Team performance (including observations affected by water damage)

	Completed within 60 minutes				Finishing time			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
+ *Identity*	0.085	-0.097	-0.048	-0.044	1.378	1.910	1.668	1.590
	(0.051)	(0.065)	(0.059)	(0.055)	(0.870)	(0.970)	(1.180)	(1.117)
	[0.198]	[0.240]	[0.434]	[0.446]	[0.220]	[0.146]	[0.205]	[0.217]
+ *Rank*	0.103	0.092	0.080	0.078	-2.789*	-2.583*	-2.575**	-2.515**
	(0.045)	(0.048)	(0.048)	(0.044)	(0.856)	(0.801)	(0.851)	(0.836)
	[0.127]	[0.181]	[0.229]	[0.184]	[0.057]	[0.051]	[0.034]	[0.034]
+ *Prize*	0.090**	0.091**	0.078**	0.083**	-2.214**	-2.391**	-2.200**	-2.330*
	(0.031)	(0.028)	(0.030)	(0.026)	(1.047)	(1.224)	(1.275)	(1.319)
	[0.047]	[0.032]	[0.032]	[0.020]	[0.042]	[0.033]	[0.040]	[0.064]
Mean in Control	0.527	0.527	0.527	0.527	56.470	56.470	56.470	56.470
Observations	378	378	378	378	378	378	378	378
Team Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Staff FE	No	No	Yes	Yes	No	No	Yes	Yes
Weekday FE	No	No	No	Yes	No	No	No	Yes

Notes: The table displays average marginal effects from Probit regressions of whether a team completed the task within 60 minutes (Columns (1) through (4)), and Tobit regressions of finishing time (Columns (5) through (8)). The main explanatory variables are indicators whether the observation stems from a treatment that included the component(s) *Identity*, *Rank*, or *Prize*. All columns include room fixed effects. Each column indicates whether team controls (group size, share of males, experience, median age, language, private), staff, and weekday fixed effects are included. Standard errors in parentheses are clustered at the week level. *p*-values from wild-cluster bootstrapping following Cameron et al. (2008) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.