The Efficacy of Tournaments for Non-Routine Team Tasks *

Florian Englmaier † Stefan Grimm ‡ Dominik Grothe §

David Schindler ¶ Simeon Schudy ‖

November 6, 2023

Abstract

Tournaments are often used to improve performance in innovation contexts. Tournaments provide monetary incentives but also render teams' identity and 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 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 the top performers. 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, IZA Bonn, SIOE 2021 and EEA-ESEM Virtual 2021 for providing helpful thoughts and comments. Comments by Peter Kuhn and three anonymous referees substantially improved an earlier version of this paper. David Schindler gratefully acknowledges funding by the Dutch Research Council (NWO) under the Talent Scheme (file number VI.Veni.211E.002). 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 enables a cost-effective design of incentives, the aim of the present study is twofold: 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

[1] Early examples of innovation competitions were the "longitude rewards", a system of inducement prizes offered by the Government of Great Britain 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.

conduct a large-scale field experiment to identify the causal effects of these components on team performance in real-life escape room challenges, in which teams have to solve a series of cognitively demanding tasks in order to succeed. These tasks are popular worldwide both among private teams seeking a complex team challenge and companies which use them for team building and recruiting purposes.[2] Escape challenges require cognitive skills, analytical and critical thinking, as well as social skills such as communication and collaboration. Thereby, they mirror defining features of many modern jobs (Deming and Kahn, 2018). Teams face a series of complex problems, need to collect and recombine information, and think outside the box. The tasks are interactive, as team members have to collaborate with each other, discuss possible actions, jointly develop ideas, and test their hypotheses. Hence, escape challenges encompass important elements of production in the "ideas sector" of the economy (see, e.g., Autor et al., 2003; Autor and Price, 2013) and require abilities, which modern employers consider of utmost importance (Deming, 2017; Casner-Lotto and Barrington, 2006; Jerald, 2009). Additionally, escape challenges allow for an objective measurement of team performance (the time spent until completion). At the same time, the team challenge provides space for team identity and image concerns to matter, as teams often proudly document their participation on a local "wall of fame" on site. Finally, the setting allows for exogenously manipulating important tournament characteristics such as the salience of identity, image, and instrumental concerns 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 (such that each additional step also comprised the treatment components of the previous step). To analyze the importance of salient team identity, we first compare a no intervention condition *Control*, in which teams have no team name, with a condition, in which we ensure salient team identity by asking teams to explic-

[2]Escape challenges are also used for education purposes of IT and Engineering students (see, e.g., Borrego et al., 2017) and 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]A more extensive discussion of the features of the setting and the task and the responsiveness of team performance to bonus incentives is provided in Englmaier et al. (2018).

itly discuss, and jointly choose, a team name they identify with. Since in most business contexts, teams already have some team (or brand) name they identify with, this condition also serves as a meaningful comparison group for the investigation of the additional effects of image concerns due to public rankings and instrumental concerns due to monetary prizes. Our second treatment condition focuses on image concerns and introduces a public ranking for all teams (using self-chosen team names) and our third treatment condition is a classical tournament, in which teams are publicly ranked (with self-chosen team names) and the best team receives a monetary prize.

We find that introducing salient team identity alone is not sufficient to improve team performance, but adding image concerns in the form of rankings appears to 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. Introducing a monetary prize in addition to the public ranking 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 tournament with a monetary prize and public ranking (using self-chosen team names) 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 field work on tournaments, incentives, and teamwork in non-routine analytical tasks and complement findings from laboratory experiments on "closed-form" creative tasks.[4] First and foremost, we provide novel field-experimental evidence on the causal effects of three major components innate to tournament incentives (salience of team identity, image concerns, and instrumental concerns) for performance in non-routine, analytical team tasks. In this way, we systematically advance earlier field work that studied rank versus monetary incentives in *routine* tasks. Findings in the context of routine tasks indicate that tournaments with and without prizes

[4] "Closed-form" creative tasks in the context of business innovation are for example characterized by specific goals such as enhancing a technological process, reducing costs, or refining an existing product. For a detailed discussion of open versus closed-form creativity see also Charness and Grieco (2019).

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 implement a treatment condition that exogenously assures salient team identity, and sheds light on how image concerns due to rankings, and instrumental concerns due to prizes affect performance in a non-routine task. Our results show that it is indeed the introduction of competition that fosters performance while assuring salient identity alone was ineffective.

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 led to stark selection into teams based on team members' performance potential (rather than friendship networks) and reduced performance, due to an increase in free-riding. Tournaments with prizes had similar effects in terms of selection, but yielded additional effort provision within teams, which offset 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 increase 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 in the laboratory. Their results indicate that monetary incentives effectively stimulate creativity 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. In another laboratory experiment that arose simultaneously to our work, Charness and Grieco (2021) analyze the relationship between performance pay, corporate culture, and "closed-form" creativity. Akin to the escape challenge, they use tasks with specific ex-ante goals, and compare (among others) a treatment without performance incentives (flat pay), a group-ranking treatment without performance incentives (flat pay + ranking) and a group treatment with group-ranking and performance pay proportional to the team's rank (performance pay + ranking). Similar to our results, they observe positive effects of rankings and additional monetary rewards.

Finally, we also link to field work on creative production. Gross (2020) documents that increased competition can foster creative production of individual logo designers, but heavy competition drives designers to stop producing logos altogether. In a similar vein, Casas-Arce and Martinez-Jerez (2009) show that the introduction of sales contests fosters effort, while incentives weaken with an increase in competition (i.e., more participants). Complementing the above findings, our results provide important field-experimental evidence on the efficacy of incentives for non-routine analytical team tasks. Focusing on teamwork that requires the forming and testing of hypotheses to come up with the solution to a complex closed-form problem, 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 line with field evidence that focuses on incentives for idea creation (Gibbs et al., 2017), and laboratory evidence on "closed-form" creative tasks, the findings from our natural field experiment suggest that incentives can foster performance in non-routine analytical team tasks with a specific goal. 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, Section 4 discusses other possible mechanisms through which the three non-control conditions could affect performance, and Section 5 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 and conducted our natural field experiment (Harrison and List, 2004) at the facilities of ETR in Munich, Germany.[5] The location offers three differently themed rooms and teams face a time limit of 60 minutes.[6] Teams can see their remaining time on a large screen in their room and if a team manages to succeed within the time limit, they win. If time runs out before the team completes all quests, they lose. Teams participate in these challenges with the aim of succeeding before the deadline, and are proud of finishing the task quickly, which is also reflected by the fact that many participants write their finishing times on the walls of the entrance area of our collaboration partner. Further, as teams do not know how many quests the challenge consists of, teams naturally aim for succeeding quickly.[7] If teams get stuck, they can request up to five 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, there are opportunity costs of asking for assistance. ETR staff provides hints upon request but never gives 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 offers teams the opportunity to purchase a voucher for future participation at a reduced rate.

ETR provides a rich setting containing the key characteristics of modern non-routine analytical teamwork. Teams have to carry out a series of cognitively demanding tasks in which they need to acquire and combine information and develop and exchange ideas with their team members. Akin to environments in innovation contexts, teams are proud to succeed but the observability of co-workers' cognitive effort provision is limited (ren-

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

[6] 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.

[7] Note that there is no entertainment value of simply waiting in the room without making any progress. In this setting, potential task utility merely stems from exploring the rooms and thereby making progress.

dering the task prone to free-riding). Thus, the setting leaves room for team identity and image concerns to matter and constitutes an excellent environment for a natural field experiment.

Our setting reflects important characteristics of modern teamwork but also involves some caveats. First, teams solving the escape challenge choose to perform the task and likely derive task utility. While such selection is less common for traditional working environments, highly educated workers appear to deliberately self-select into occupations based on the interesting, non-routine nature of the tasks the occupation involves (Autor and Handel, 2013). Second, the effectiveness of tournament incentives may depend on a workers' motivation, which may not solely stem from the task itself, but also from salient greater goals that are missing in escape challenges. Importantly, Englmaier et al. (2018) show that monetary incentives are effective in the same escape setting, independent of differences in worker motivation and self-selection into the task. Finally, the escape challenge involves a complex problem with one final solution whereas complex problems in work environments may be multi-dimensional and in principle allow for several possible solutions. However, in innovation and business contexts, deadlines and budget-constraints often render one solution favorable, and the nature of the escape challenge mirrors this idea. It offers multiple ways to arrive at the (one) final solution and thereby allows us to study how tournament incentives motivate workers to produce the best possible solution within a given amount of time. As such, the escape challenge resembles the idea of closed-form creativity (Charness and Grieco, 2019), in which teams face a complex task with a well-delineated goal (as opposed to an open-form task that may not envision a specific final outcome). Thus, it can be reflective of modern work tasks in the context of business innovation, which may for example focus on the enhancement of a technology process or the development of new ideas that solve a well-defined problem subject to time constraints.

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.

To avoid contamination, we randomized treatment arms on a weekly level.[8] ETR staff implemented the different treatments after delivering the introduction. The choice of our experimental treatment variations was guided by the previous literature comparing tournaments with and without prizes (Barankay, 2012; Charness and Grieco, 2019), as well as by tournament designs in practice, which often involve rankings of team names that relate to teams' identity (e.g, the Netflix Prize). Hence, we focus on three components innate to tournaments: salient team identity (through team names), image concerns (due to being ranked), and instrumental concerns (due to prizes). Varying these three components independently would have resulted in a full factorial $2 \times 2 \times 2$ design with eight experimental conditions. However, our collaboration partner considered treatment variations in which we would i) publish team names without a ranking, ii) rank teams without a name teams can identify with, iii) or assign a prize to the best team without a public ranking with iv) or without team names (due to lack of transparency) incongruous. Thus, we opted for four experimental conditions which step-wise introduce team names, rankings, and prizes. These, we believe, cover also many applications relevant for practitioners, as prizes often involve public rankings, and public rankings usually require a unique and meaningful team identifier.

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 tourna-

[8] 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.

ments render team identity salient by explicitly identifying them by their name, brand, or company, our first treatment condition, *T1 (Identity)* (85 teams), was designed to increase the salience of team identity in a natural way, without adding any competitive aspects. Following the idea in Ai et al. (2022), in which the company DiDi (a leading transportation platform) explicitly used the creation of team names by team members to increase team identification, we asked teams to jointly deliberate on a team name to be used for communication during the task with ETR staff via the walkie-talkie.[9] Teams were free to choose any name all members identified with, and were actively engaged in choosing the team name.[10]

To study the effects of introducing image concerns through competition, we implement our second treatment condition *T2 (Identity, Rank)*, 94 teams. Based on the idea that people care about being ranked per se (Charness et al., 2014; Charness and Grieco, 2019), and thus also about the rank of their team, *T2 (Identity, Rank)* includes a weekly tournament for teams facing the exact same challenge (i.e., the same of three rooms) without a prize.[11] In the same manner as in *T1 (Identity)*, we also asked team members in *T2 (Identity, Rank)* to select a team name. In addition, we informed teams 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), where teams were ranked by room according to their finishing times with their team name. All teams that did not complete the task were assigned the same rank. Although the ranking did not reveal which team contained which members, team members were free to tell others about their team's performance, and some individuals indeed engaged with the weekly Facebook post using their real names (see also Figure A.1).

Lastly, treatment *T3 (Identity, Rank, Prize)*, with 87 teams, exhibited the same features as *T2 (Identity, 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

[9] In *Control*, ETR staff referred to the team member with the walkie-talkie as "you".

[10] Thus, our treatment rendered 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).

[11] As teams who booked the same room (usually several days in advance) do not encounter each other on site, and teams working in different rooms in overlapping time slots do not compete with each other, teams are unlikely to form informed priors about their potential competitors.

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.[12]

2.3 Outcome measures and sample characteristics

Our final sample consists of 373 teams (1,705 individuals, see Table 1).[13] 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.[14] 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 willingness to explore original solutions (which we measure inversely by the number of hints a team has taken) and a team's interest to perform a similar task again (which we measure using the probability of purchasing a voucher for future participation at ETR at a reduced rate immediately after performing the task).

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

[13]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.9.

[14]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)	*T1* *Identity* Mean (SD)	*T2* *Identity, Rank* Mean (SD)	*T3* *Identity, Rank, 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 {*T1 (Identity)*, *T2 (Identity, Rank)*, *T3 (Identity, Rank, Prize)*} at the ten percent level.

Table 1 provides an overview of team characteristics across treatments (team size, gender, age composition, team language: German or English, prior experience with escape rooms, and whether the team came as a private group or were part of a corporate team-building event). 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 will show regression results with and without team characteristics as controls.

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 for 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 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 team identity salient. Our approach reflects current business strategies pursued by companies relying on structures based on agile teams rather than strict hierarchical structures.[15] 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 as workers may care about their image, and change their behaviors based on how they are perceived by others. For instance, Tan and Bolle (2007) find that cooperation rates within teams (in laboratory public goods games) increase when outcomes are compared to other teams. Field-experimental evidence from individual routine tasks shows that non-instrumental rewards which encompass image value can substantially improve performance (Kosfeld and Neckermann, 2011). Further, Restivo and Van De Rijt (2012) show that informal rewards can raise contribution levels of high-performing individual contributors at Wikipedia. Studies on team performance in routine tasks suggest 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

[15] 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://web.archive.org/web/20210123010704/https://www.calabrio.com/wfo/workforce-management/boost-belonging-motivation-through-team-names/` and Ye et al., 2022).

(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 thus hypothesize that image concerns can boost performance (in addition to identity), also in non-routine tasks. Furthermore, prior research has documented that positive performance effects of symbolic rewards are particularly effective for top performers (Kosfeld and Neckermann, 2011). It thus seems reasonable to assume that image concerns may have different effects on teams depending on their relative likelihood of being ranked high. Teams that are expected to perform well based on observable characteristics (e.g., because they are particularly able, more experienced, or particularly motivated to perform well) may show a stronger reaction to the public ranking than teams that are expected to perform worse (e.g., because they are less able, less experienced or less motivated to perform well). As lower ranks in our weekly competitions were likely to pool several teams failing to complete the task, teams at the bottom end of the performance distribution are likely to expect lower marginal image returns to effort. We thus hypothesize that positive performance effects of rank incentives are observed particularly for the upper quantile 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 monetary prizes or other instrumental rewards 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 positive 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 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, we expect that the introduction of prizes further improves team performance (as compared to tournaments without prizes).

Hypothesis 3 *Adding a monetary prize to the rank tournament improves team performance.*

The development of our hypotheses reflects the idea that salience of team identity, image concerns, and instrumental concerns are three major components innate to typical tournament incentives. We hypothesized that a public ranking introduces image concerns to a setting in which teams with salient team identity perform, and a monetary prize introduces instrumental concerns in settings in which teams otherwise compete for ranks. Alternatively, one could also hypothesize that image concerns through a public display of team names may interact with feelings of team identity and thus trigger an additional performance increase through stronger feelings of identity. Similarly, adding monetary prizes may additionally alter image concerns (or team identity). In other words, teams may perceive the value of appearing first in the public ranking differently, because monetary prizes may either crowd out parts of the image motivation or increase the image value of being first in the ranking. While we consider identity-strengthening aspects of additional image and instrumental concerns less likely in environments with otherwise salient team identity (like ours), our design does not exclude these potential interaction effects. We discuss these and other aspects related to differences across treatments further in Section 4.

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.[16] Our main analyses focuses on capturing the effects of introducing the three distinct components of a tournament, *Identity*, *Rank*, and *Prize*. That is, we focus on comparing each "subsequent" treatment group to the "prior" one. To do so, we code a dummy variable for each component based on whether this component existed in the treatment the observation stems from. For example, in treatment *T2 (Identity, Rank)*, the dummy "*Identity*" and "*Rank*" are equal to 1, whereas the dummy *Prize* is equal to 0. This coding allows us to cleanly identify the effect of introducing the respective component (as compared to the "prior" condition) on our outcome measures.[17] The results are shown in Table 2, and 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 assignment), and, because of the relatively low number of clusters, we provide p-values from score bootstrapping following Kline and Santos (2012).

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 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).

[16] Table A.1 shows summary statistics of the probability of completion, finishing time, number of hints and the probability of purchasing a voucher by treatment.

[17] 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
(making identity salient)	(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**
(adding a ranking)	(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*
(adding a prize)	(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 score bootstrapping following Kline and Santos (2012) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

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 than in *Control*. 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 significantly so along 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, see also Table A.2). 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 ranks and prizes influence teams across the entire performance spectrum using quantile regressions on residualized finishing times. We predict finish-

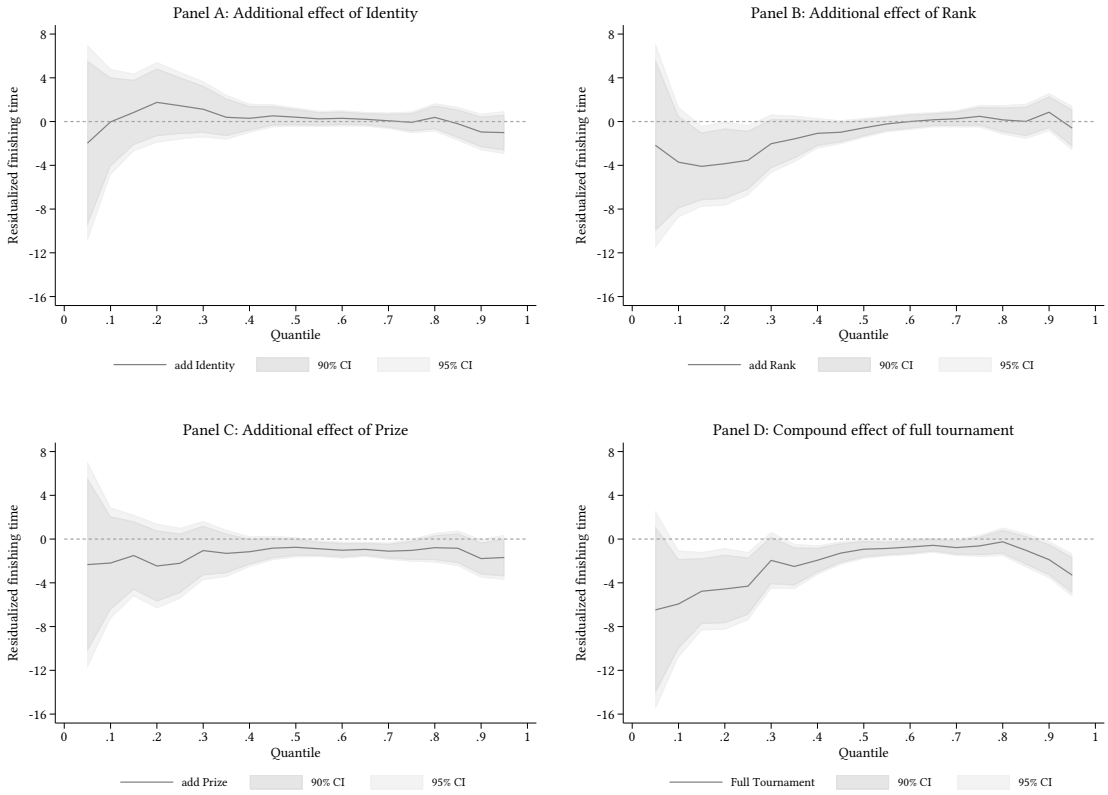
ing times and residuals for all teams using the same fully specified Tobit regression as in Table 2, Column (8), including team controls, room, staff and weekday fixed effects.[18]

Panel A of Figure 1 shows that asking teams to discuss and choose a team name jointly before working on the task does not affect performance along the whole performance distribution (confirming Result 1). Panel B of Figure 1 shows that adding a weekly competition with a public ranking to a setting in which teams jointly deliberate on team names reduced the finishing times of the top performers, i.e., the lowest quantiles. This 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 of teams being ranked and additionally eligible for a monetary prize with those of teams that are ranked but not eligible for a prize. Three interesting findings arise. First, adding a 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 adding a 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 *T3 (Identity, Rank, Prize)*, and thus the compound effect of implementing the full tournament including the ranking with team names and the monetary prize. This tournament improves performance along a large part of the distribution, so that teams facing salient team identity, image, and instrumental concerns perform 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 may additionally stimulate performance also below the very top.

In additional exploratory analyses, we also study possible heterogeneity in the observed treatment effects. To do so, we conducted additional regression analyses including

[18]The results in Table 2 did not show any performance improvement of *Identity* over *Control*. To increase statistical power, we therefore use observations from both *Identity* and *Control* for predicting finishing times. Using GLM (instead of Tobit) yields similar results (see Figure A.3).

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. Panel B shows the additional effect of a public ranking. Panel C shows the additional effect of a monetary prize. And Panel D shows the overall effect of a tournament with a monetary prize (compares *T3 (Identity, Rank, 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 adding component *Identity* (Panel A), *Rank* (Panel B), and *Prize* (Panel C), or due to adding all tournament features simultaneously (Panel D).

interaction terms between each treatment component (i.e., the dummy variables *Identity*, *Rank*, and *Prize*) 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 Willingness to explore original solutions 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 the willingness to explore original solutions on their own. Teams may thus substitute speed for such exploration, particularly when they face difficult problems (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 as well as on the timing of the hints. In general, the number of hints and finishing times are positively correlated (Spearman's $\rho = 0.5198$, $p < 0.001$), as worse teams are on average more likely to seek help by taking hints. 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 exposed to component *Prize* take on average slightly fewer hints (average number of hints taken in *Control*: 3.39, *T1 (Identity)*: 3.36, *T2 (Identity, Rank)*: 3.31, *T3 (Identity, Rank, Prize)*: 3.18). Additional analyses confirm that despite

the positive effect on performance, the addition of none of the components (*Identity*, *Rank*, or *Prize*) significantly increases the number of hints taken, nor their timing (see Appendix Section A.5). These results indicate that work environments in innovation contexts sharing the features of our team task are unlikely susceptible to a reduction in teams' inclination to explore own approaches due to tournaments.

Offering extrinsic incentives could also 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 facing the challenge) and thus a unique opportunity to test whether image and instrumental concerns innate to tournaments affect the intrinsic motivation to perform a similar task again. To evaluate whether the addition of any component 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.8). As such, our findings speak against a substantial crowding out of intrinsic motivation for future participation and underline the positive roles of image and instrumental concerns innate to tournaments.

4 Discussion

Our experimental treatments step-wise introduced three components innate to tournaments: salient team identity, image concerns, and instrumental concerns. However, we could not independently vary each of the three components as our collaboration partner

considered some treatments resulting from a full factorial design incongruous (see also Section 2.2). Being constrained to the three implemented treatment conditions comes with the caveat that we cannot explicitly study potential interactions between the three different components (e.g., we cannot directly measure potential identity-enhancing effects of the introduction of a public ranking) and requires a more detailed discussion of which other potentially relevant changes each treatment variation may bring about.

As compared to the *Control* condition, *T1 (Identity)* ensures salient team identity, but the limited treatment effect may have eventually resulted from team identity being also salient in *Control* (as about 80% of teams were composed of friends). This aspect leaves important room for future studies on the role of identity for team performance in non-routine tasks but renders potential additional identity enhancing effects of team competitions (by adding a ranking and a prize) in our setting less likely. Further, it is plausible that the introduction of public rankings may not only result in image concerns but also render time to completion a more relevant performance outcome. Similarly, adding a prize to the competition for ranks may not only introduce instrumental concerns but additionally render the role of finishing times salient, and such shifts in focus may improve team performance independently of image and instrumental concerns. Importantly, Englmaier et al. (2018, p.22) show that a focus on finishing times alone does not improve performance in escape challenges in the exact same setting, such that the observed performance improvements due to the introduction of the competition for ranks very likely result from additional image concerns, rather than from an interaction with salient team identity. Finally, introducing a prize may not only result in instrumental concerns but also alter image concerns. Teams may perceive the value of appearing first in the public ranking differently, because monetary prizes may either crowd out parts of the image motivation or increase the image value of being ranked high. As we observe that the introduction of ranks particularly boosts performance of teams at the top of the performance distribution while introducing a prize leads to improvements along the whole performance spectrum, we consider it less likely that the addition of a monetary reward substantially altered image concerns which then caused the observed performance improvements.

5 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 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 vary features innate to typical tournament incentives (salient team identity, team rankings, and prizes) treating a large number of teams performing a non-routine analytical task in a natural field experiment. We found that fully-fledged tournament incentives, in which teams compete for a monetary prize awarded to the best performing team listed in a public ranking of team names, substantially improved team performance. Public rankings of team names alone improved performances of teams expected to be at the top of the performance distribution but did not affect teams at the bottom. Lastly, rendering team identity salient by having teams jointly deliberate on their team name (see also Ai et al., 2022) was not enough to improve performance on its own.

Complementing this novel field-experimental evidence on the effects of tournaments for team performance in non-routine tasks, we further showed that performance improvements due to tournaments did not result in a reduction of teams' willingness to explore solutions on their own. Further, we found no indications of a reduction of teams' intrinsic motivation to perform similar tasks again in the future due to tournament incentives. As we elicited 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, this finding suggests that 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) likely result from actual, discouraging performance feedback for underperforming teams rather than from the anticipation of such feedback or competition per se. Avoiding such feedback, we thus found robust evidence for the important roles of image and instrumental concerns in the efficacy of tournaments in non-routine analytical team tasks.

Overall, our results make an important contribution to the literature on teamwork in non-routine analytical tasks with a clearly specified goal and deadline. We confirm and extend findings from laboratory experiments on closed-form creativity (Charness et al., 2014; Charness and Grieco, 2019) and show that tournaments can substantially improve performance in a novel and challenging field setting. Thereby, we provide basis for important future field work. One fruitful avenue for such research lies in studying whether image and instrumental concerns lead to adjustments in team organization. For example, Englmaier et al. (2018) find suggestive evidence that bonus incentives can alter team organization in the same setting and are accompanied by an increased demand for leadership. Following these results, it will be interesting to investigate whether tournament incentives and leadership are substitutes or complements. Further, it will be interesting to investigate the role competitions with and without prizes in field settings with open-form tasks.

References

Ai, W., Y. Chen, Q. Mei, J. Ye, and L. Zhang (2022). Putting teams into the gig economy: A field experiment at a ride-sharing platform. *Management Science (in press)*.

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.

Casas-Arce, P. and F. A. Martinez-Jerez (2009). Relative performance compensation, contests, and dynamic incentives. *Management Science 55*(8), 1306–1320.

Casner-Lotto, J. and L. Barrington (2006). *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century US workforce.* ERIC.

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

Charness, G. and D. Grieco (2021). Creativity and corporate culture. *Available at SSRN 3723865*.

Charness, G., D. Masclet, and M. C. Villeval (2014). The dark side of competition for status. *Management Science 60*(1), 38–55.

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).

Deming, D. and L. B. Kahn (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics 36*(S1), S337–S369.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics 132*(4), 1593–1640.

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.

Jerald, C. D. (2009). Defining a 21st century education. *Center for Public Education 16*.

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

Kline, P. and A. Santos (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods 1*(1), 23–41.

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.

Krupka, E. L. and R. A. Weber (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association 11*(3), 495–524.

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.

Restivo, M. and A. Van De Rijt (2012). Experimental study of informal rewards in peer production. *PloS one 7*(3), e34358.

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.

Tan, J. H. and F. Bolle (2007). Team competition and the public goods game. *Economics Letters 96*(1), 133–139.

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, J. Ye, and L. Zhang (2022). Virtual teams in a gig economy. *Proceedings of the National Academy of Sciences 119*(51), 1–12.

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

The Efficacy of Tournaments for Non-Routine Team Tasks

Web 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

Table A.1 shows summary statistics of the probability of completion, finishing time, number of hints and the probability of purchasing a voucher. Complementing our main analyses, which compares each subsequent component to treatments including the prior ones, Table A.2 compares each treatment directly to *Control*. By design, the results for treatment *T1 (Identity)* remain the same. Comparing *T2 (Identity, Rank)* to *Control*, we see that *T2 (Identity, Rank)* increases completion rates and lowers finishing times on average, but not significantly so, due to heterogeneity in reactions to the ranking (see Panel B in Figure 1). As compared to *Control*, Treatment *T3 (Identity, Rank, Prize)* significantly increases the likelihood of succeeding within 60 minutes and significantly lowers finishing times across all four specifications. 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.3, confirming the robustness of these findings. Since the salience of team identity is an innate feature of tournaments, Tables A.2 and A.3 further provide p-values from Wald tests for the differences between *T1 (Identity)* and *T2 (Identity, Rank)*, and *T1 (Identity)* and *T3 (Identity, Rank, 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 *T2 (Identity, Rank)* and *T3 (Identity, Rank, Prize)* when teams have an identity-related team name. These comparisons reveal that, on average, *T2 (Identity, Rank)* significantly improves teams' finishing times (see specifications (5) to (8)) as compared to *T1 (Identity)* ($0.028 < p < 0.078$), and *T3 (Identity, Rank, Prize)* improves both the likelihood of completion (specifications (1) to (4)) as well as finishing times (specifications (5) to (8), $0.004 < p < 0.039$).

Table A.1: Summary statistics

	Control - Mean (SD)	*T1* *Identity* Mean (SD)	*T2* *Identity, Rank* Mean (SD)	*T3* *Identity, Rank, Prize* Mean (SD)
Completion	0.53 (0.50)	0.42 (0.50)	0.56 (0.50)	0.65 (0.48)
Finishing time	56.47 (5.49)	56.93 (5.32)	55.03 (6.44)	53.80 (7.57)
Number of hints	3.39 (1.35)	3.36 (1.40)	3.31 (1.31)	3.18 (1.29)
Purchased a voucher	0.18 (0.38)	0.20 (0.43)	0.28 (0.52)	0.30 (0.66)
Observations	112	85	94	82

Notes: Completion denotes the share of teams that managed to complete the task within the given time limit of 60 minutes. Finishing time denotes the average time to complete the task (all teams that did not manage to complete the task within 60 minutes are assigned a finishing time of 60 minutes). Number of hints denotes the average number of hints teams' took. Purchased a voucher denotes the share of teams that purchased a voucher for future participation (at a reduced rate). Suggestive differences in completion probability between T1 and *Control* and voucher purchases in T3 and T2 vs. T1 and *Control* are not statistically significant in any of the regression results in which we correct for the influence of potential confounders in the form of fixed effects and control variables.

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

	Completed within 60 minutes				Finishing time			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T1	-0.086	-0.099	-0.048	-0.045	1.377	1.910	1.668	1.590
Identity	(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]
T2	0.019	-0.005	0.033	0.034	-1.411	-0.673	-0.906	-0.925
Identity, Rank	(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]
T3	0.110**	0.087*	0.112**	0.118**	-3.625**	-3.064**	-3.106**	-3.255**
Identity, Rank, Prize	(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]
T1 = T2	[0.126]	[0.182]	[0.230]	[0.188]	[0.055]	[0.051]	[0.034]	[0.034]
T2 = T3	[0.047]	[0.032]	[0.033]	[0.020]	[0.042]	[0.033]	[0.040]	[0.064]
T1 = T3	[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 score bootstrapping following Kline and Santos (2012) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

Table A.3: Team performance (completion and finishing times)

	Completed within 60 minutes				Finishing time			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T1	-0.091	-0.102	-0.050	-0.048	0.006	0.013	0.015	0.016
Identity	(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]
T2	0.014	-0.006	0.033	0.035	-0.019	-0.012	-0.011	-0.012
Identity, Rank	(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]
T3	0.105**	0.092**	0.118**	0.121***	-0.041**	-0.034*	-0.032*	-0.034*
Identity, Rank, Prize	(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]
T1 = T2	[0.126]	[0.177]	[0.291]	[0.259]	[0.078]	[0.046]	[0.032]	[0.028]
T2 = T3	[0.044]	[0.026]	[0.078]	[0.030]	[0.146]	[0.220]	[0.266]	[0.248]
T1 = T3	[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 score bootstrapping following Kline and Santos (2012) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

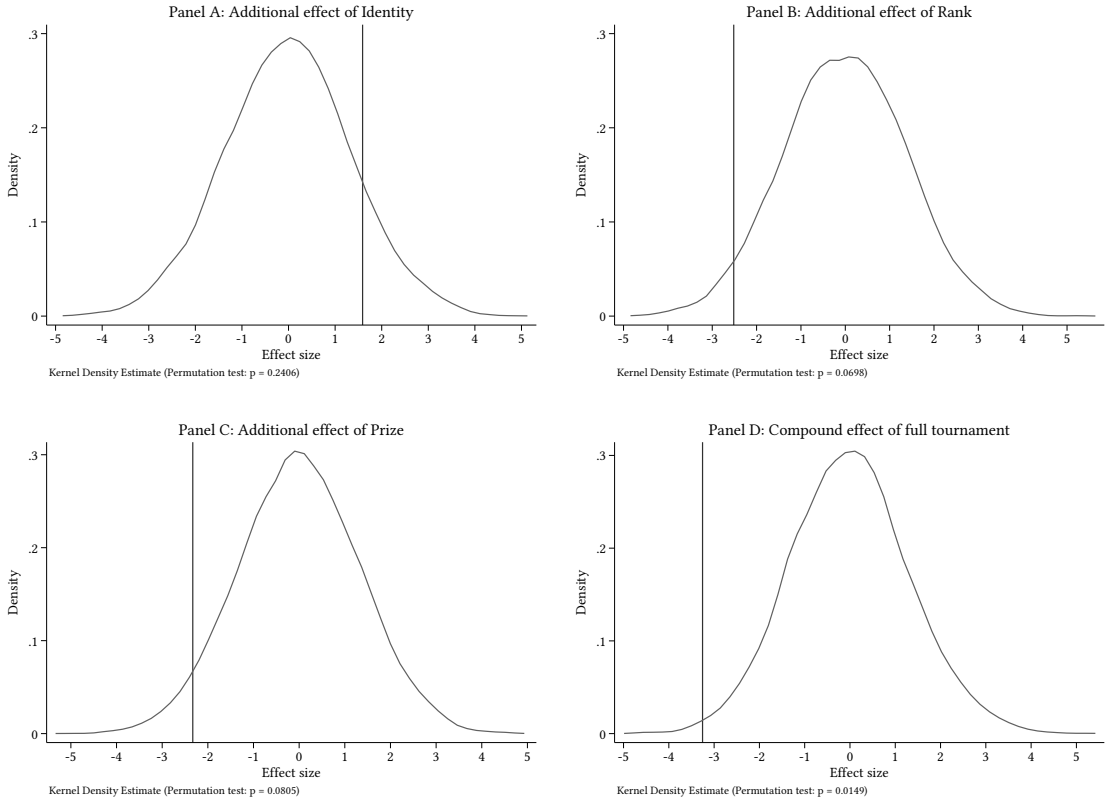
A.3 Randomization inference

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.

As in our main analyses in Section 3.1, we focus on comparing each "subsequent" treatment group to the "prior" one using dummy variables for each added component. 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 of *T3 (Identity, Rank, Prize)* relative to *Control* (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 that the true effect of *Identity* does not appear extreme, 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 additional 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

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), or for adding all tournament features simultaneously (Panel D).

have suggested ($p = 0.0805$). Lastly, Panel D shows the randomization distribution for the overall effect of *T3 (Identity, Rank, Prize)* compared to *Control*. The result supports our finding that a tournament with a monetary prize and a ranking of teams by their team name reduces finishing times substantially ($p = 0.0149$). To summarize, all four panels show that our previous results are robust to randomization inference.

A.4 Further heterogeneity analyses

Providing an alternative specification, Figure A.3 shows quantile regressions on residualized finishing times using a fully specified GLM regression (with log link). The results are similar (compared to Figure 1), confirming the robustness of these findings.

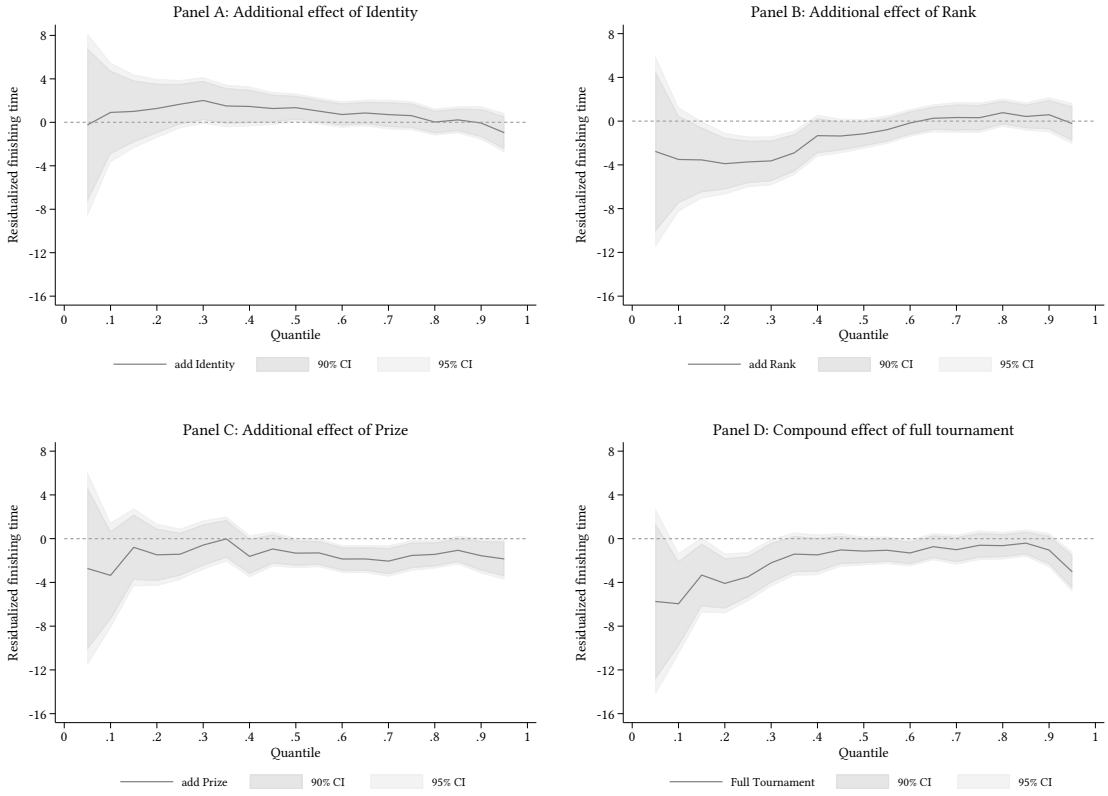
To understand the importance of the composition of a team for possible heterogeneity in the observed treatment effects, we estimate whether (and how fast) teams finish the task in linear probability (and Tobit) models by including interaction terms between each treatment component (i.e., the dummy variables *Identity*, *Rank*, and *Prize*) and observable team characteristics. Appendix Table A.4 shows that, for the probability of completing the task, adding a ranking (*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, introducing a *Prize* in addition to the ranking tends to increase team performance irrespective of the observed gender composition and other team characteristics. The latter provides 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.5 yield results in line with the above-mentioned interaction effect for *Rank*, although less precisely estimated. The more males there are in a team, the stronger 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 additional monetary incentive in *Prize* was particularly effective in stimulating the performance of private teams (regular booking).

One reason for the differential treatment effect of prizes for groups of colleagues could be driven by pessimistic expectations about the sharing norm among company

team members, who might expect not to be able to receive a fair share of the prize. To explore this argument, we conducted an additional survey in which we elicited social norms of prize sharing following the incentivized elicitation procedure of Krupka and Weber (2013). We recruited an online sample ($n = 209$) of subjects that had experience with real life escape challenges. We asked them about the appropriateness of different sharing norms across five scenarios. All scenarios were based on the situation in the actual escape challenge (in which winning teams of a given week were informed that they could send a team member to collect their prize money) and we varied how the prize was shared across the five scenarios. For each scenario and in randomized order, subjects had to evaluate the social appropriateness of how the prize is shared within a group of friends (taking part in their leisure time) or a group of colleagues (taking part in a team-building event) on a 4-point Likert-scale from "very socially inappropriate" to "very socially appropriate". One in one hundred participants was eligible for an additional payment and participants were informed that if they choose the same answer as the majority of all other survey participants in one randomly selected scenario, they would earn 50 Euro (if they were randomly selected for payment). The histograms in Figure A.4 show that the equal sharing norm is considered most appropriate, and, more importantly, that there are no systematic differences in sharing norms across types of teams. χ^2-tests comparing responses regarding groups of friends vs. groups of colleagues within each scenario cannot reject the equality of underlying distributions (p-values in brackets). *Scenario 1:* "The person who collects the prize receives all of the prize money (150 Euro)." (p-value: 0.986). *Scenario 2:* "The prize money (150 Euro) will be divided equally among all members of the group." (p-value: 0.681). *Scenario 3:* "The prize money (150 Euro) will be divided unequally among all members of the group." (p-value: 0.866). *Scenario 4:* "The person who collects the prize receives half of the prize money (75 Euro) and the rest will be divided equally among all members of the group." (p-value: 0.937). *Scenario 5:* "The person who collects the prize receives half of the prize money (75 Euro) and the rest will be divided unequally among all members of the group." (p-value: 0.783). Hence, differences in expected sharing norms are unlikely to explain differential treatment effects across private and company teams. Of course, there exist several other potential explanations for differ-

Figure A.3: 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. Panel B shows the additional effect of a public ranking. Panel C shows the additional effect of a monetary prize. And Panel D shows the overall effect of a tournament with a monetary prize (compares *T3 (Identity, Rank, 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), *Prize* (Panel C), or due to adding all tournament features simultaneously (Panel D).

ences in the observed coefficients for private and company teams. First, a primary reason for company teams to face the escape challenge may be bonding purposes as part of a team building event, which may render additional monetary incentives less effective. Second, there could be differences in income or wealth between teams of friends and teams of colleagues, that affect the perceived size of the incentive. Third, company teams may have formed less optimistic expectations about their subjective probability of winning the prize and therefore reacted less to incentives. As we observe only a relatively small number of team-building event groups in our sample, we see scope for future research on this exploratory finding and the potential additional channels discussed above.

Table A.4: 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
(making identity salient)	(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
(adding a ranking)	(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
(adding a prize)	(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					
(making identity salient)		(0.075)					
		[0.544]					
+ *Rank* x Group Size		0.077					
(adding a ranking)		(0.065)					
		[0.349]					
+ *Prize* x Group Size		0.019					
(adding a prize)		(0.045)					
		[0.753]					
+ *Identity* x Experience			0.117				
(making identity salient)			(0.193)				
			[0.633]				
+ *Rank* x Experience			-0.112				
(adding a ranking)			(0.163)				
			[0.615]				
+ *Prize* x Experience			0.006				
(adding a prize)			(0.134)				
			[0.971]				

... continued on next page

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

	Completed within 60 minutes						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
+ *Identity* x Private				0.185			
(making identity salient)				(0.184)			
				[0.552]			
+ *Rank* x Private				-0.161			
(adding a ranking)				(0.171)			
				[0.561]			
+ *Prize* x Private				0.158			
(adding a prize)				(0.135)			
				[0.287]			
+ *Identity* x Men Share					0.049		
(making identity salient)					(0.193)		
					[0.811]		
+ *Rank* x Men Share					0.319*		
(adding a ranking)					(0.124)		
					[0.059]		
+ *Prize* x Men Share					-0.256		
(adding a prize)					(0.200)		
					[0.377]		
+ *Identity* x Median Age						-0.002	
(making identity salient)						(0.004)	
						[0.851]	
+ *Rank* x Median Age						0.006	
(adding a ranking)						(0.006)	
						[0.482]	
+ *Prize* x Median Age						-0.003	
(adding a prize)						(0.011)	
						[0.823]	
+ *Identity* x German							0.690
(making identity salient)							(0.155)
							[0.236]
+ *Rank* x German							-0.318
(adding a ranking)							(0.167)
							[0.278]
+ *Prize* x German							-0.128
(adding a prize)							(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 score bootstrapping following Kline and Santos (2012) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

Table A.5: 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
(making identity salient)	(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
(adding a ranking)	(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
(adding a prize)	(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					
(making identity salient)		(1.417)					
		[0.122]					
+ *Rank* x Group Size		-3.399					
(adding a ranking)		(1.820)					
		[0.124]					
+ *Prize* x Group Size		1.649					
(adding a prize)		(1.313)					
		[0.238]					
+ *Identity* x Experience			-1.456				
(making identity salient)			(3.558)				
			[0.684]				
+ *Rank* x Experience			0.660				
(adding a ranking)			(3.416)				
			[0.898]				
+ *Prize* x Experience			2.513				
(adding a prize)			(3.411)				
			[0.422]				

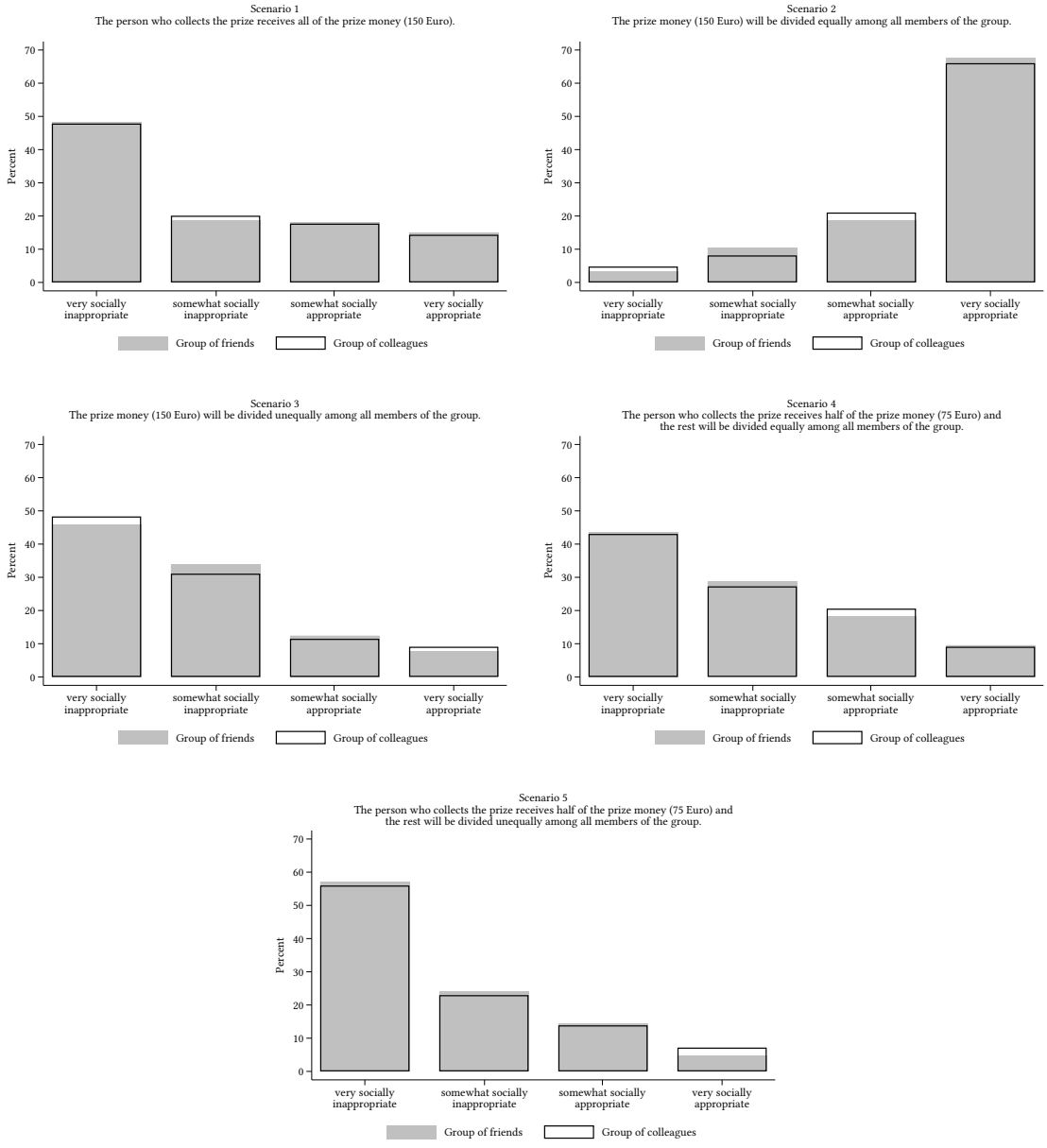
... continued on next page

Table A.5: Team performance (finishing times, interactions) - continued

	Finishing time						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
+ *Identity* x Private				-2.759			
(making identity salient)				(2.673)			
				[0.472]			
+ *Rank* x Private				6.968*			
(adding a ranking)				(2.022)			
				[0.085]			
+ *Prize* x Private				-9.695***			
(adding a prize)				(2.260)			
				[0.005]			
+ *Identity* x Men Share					3.300		
(making identity salient)					(4.566)		
					[0.503]		
+ *Rank* x Men Share					-7.144		
(adding a ranking)					(3.515)		
					[0.120]		
+ *Prize* x Men Share					2.490		
(adding a prize)					(3.340)		
					[0.605]		
+ *Identity* x Median Age						-0.021	
(making identity salient)						(0.092)	
						[0.814]	
+ *Rank* x Median Age						-0.187	
(adding a ranking)						(0.101)	
						[0.218]	
+ *Prize* x Median Age						0.186	
(adding a prize)						(0.177)	
						[0.378]	
+ *Identity* x German							-34.494
(making identity salient)							(4.018)
							[0.389]
+ *Rank* x German							31.953
(adding a ranking)							(3.615)
							[0.323]
+ *Prize* x German							-4.954
(adding a prize)							(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 score bootstrapping following Kline and Santos (2012) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

Figure A.4: Social norms of splitting a prize between friends and colleagues



Notes: The figure shows histograms of survey answers on the social appropriateness of splitting a monetary prize within a group of friends (taking part in their leisure time) or a group of colleagues (taking part in a team-building event). For each of the five scenarios, subjects had to evaluate the social appropriateness on a 4-point Likert-scale from "very socially inappropriate" to "very socially appropriate".

A.5 Willingness to explore original solutions and potential crowding out

Figure A.5 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.6) confirm the non-parametric finding that neither component, *Identity*, *Rank* nor *Prize* affect the willingness to explore original 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.

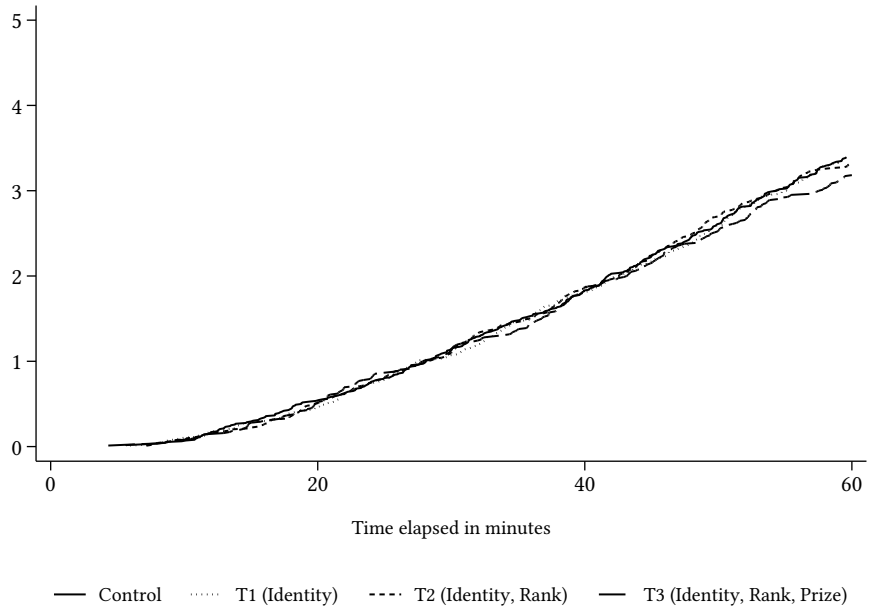
Even though the willingness to explore new and original solutions does not seem to be crowded out if measured by the total number of hints requested, it would still be conceivable that teams request their hints earlier. This would effectively also allow them to rely on external help early on and thus arrive at the solution quicker. Table A.7 shows the coefficients of Tobit regressions on the timing of hints using treatment components as explanatory variables.[19] 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 their willingness to explore original solutions, we also present results from linear regressions within quantiles (based on residualized finishing times) in Figure A.6. Panel A shows the difference in the number of hints taken in *Identity* as compared to *Control*. Panel B compares adding *Rank* to only having component *Identity*. Panel C compares the addition of *Prize* on top of *Rank*. Panel D provides the comparison between *Control* and *T3 (Identity, Rank, Prize)*. 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.8 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

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

Figure A.5: Hint taking over time



Notes: The figure shows the cumulative distribution of hints by minute in *Control*, *T1 (Identity)*, *T2 (Identity, Rank)*, and *T3 (Identity, Rank, Prize)*.

controls and fixed effects in each column. The results speak clearly against any crowding out of intrinsic motivation for future participation.

Table A.6: Willingness to explore original solutions (number of hints)

	Number of hints			
	(1)	(2)	(3)	(4)
+ Identity	-0.058	-0.038	0.048	0.075
(making identity salient)	(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
(adding a ranking)	(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
(adding a prize)	(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

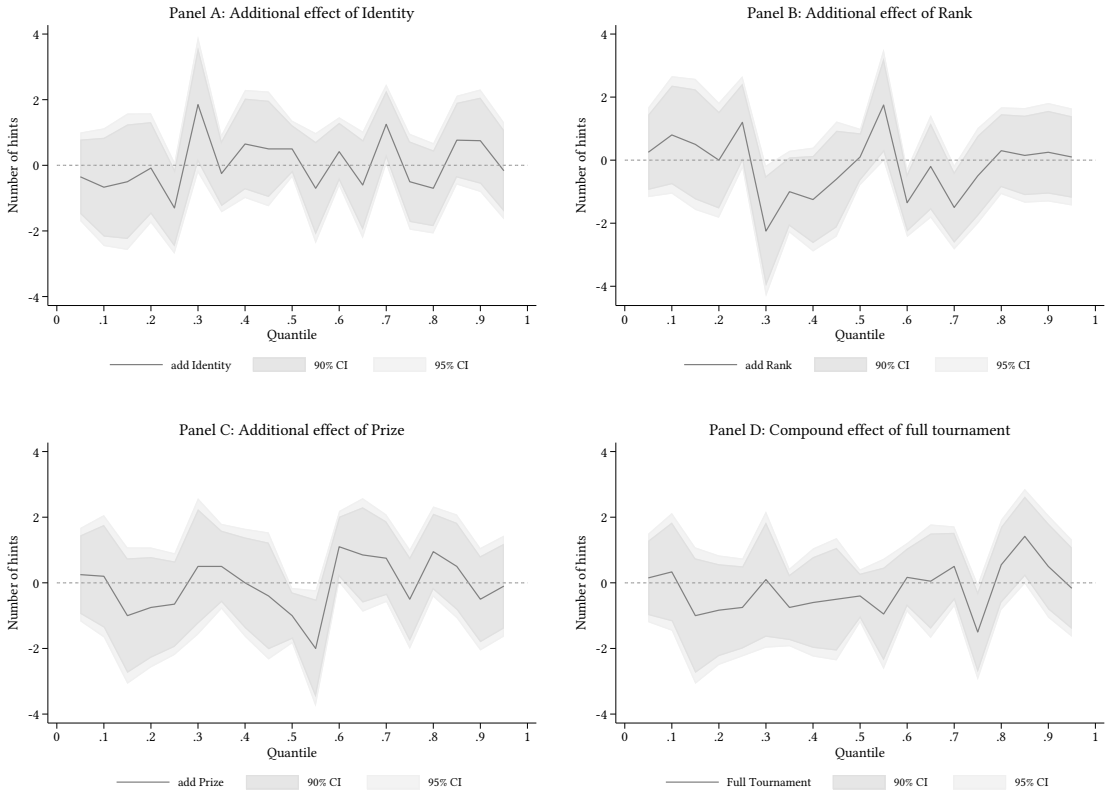
Notes: The table displays coefficients from OLS regressions of number of hints. The main explanatory variables are indicators whether the observation stems from a treatment that included the component(s) *Identity*, *Rank*, or *Prize*. All columns include room fixed effects. Each column indicates whether 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 score bootstrapping following Kline and Santos (2012) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

Table A.7: Willingness to explore original solutions (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
(making identity salient)	(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
(adding a ranking)	(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
(adding a prize)	(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 score bootstrapping following Kline and Santos (2012) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

Figure A.6: 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. Panel B shows the additional effect of a public ranking. Panel C shows the additional effect of a monetary prize. And Panel D shows the overall effect of a tournament with a monetary prize (compares *T3 (Identity, Rank, 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), *Prize* (Panel C), or due to adding all tournament features simultaneously (Panel D).

Table A.8: Purchased a voucher

	Purchased a voucher			
	(1)	(2)	(3)	(4)
+ *Identity*	0.012	-0.004	0.005	0.009
(making identity salient)	(0.025)	(0.028)	(0.024)	(0.027)
	[0.575]	[0.907]	[0.839]	[0.745]
+ *Rank*	0.053	0.041	0.019	0.014
(adding a ranking)	(0.038)	(0.033)	(0.028)	(0.026)
	[0.270]	[0.314]	[0.546]	[0.607]
+ *Prize*	-0.009	0.004	0.001	0.010
(adding a prize)	(0.057)	(0.048)	(0.042)	(0.042)
	[0.896]	[0.925]	[0.981]	[0.782]
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 score bootstrapping following Kline and Santos (2012) 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.9 repeats the specifications from Table 2 but includes the five omitted data points. The results are very similar.

Table A.9: 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
(making identity salient)	(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**
(adding a ranking)	(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*
(adding a prize)	(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 score bootstrapping following Kline and Santos (2012) are listed in square brackets, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.