Received: 1 November 2022 | Revised: 19 March 2023 | Accepted: 3 May 2023
DOI: 10.1002/mde.3906

RESEARCH ARTICLE

WILEY

Under (peer) pressure: Experimental evidence on team size and task performance

Adrian Chadi[1] | Konstantin Homolka[2]

[1]Department of Economics, University of Konstanz, Konstanz, Germany

[2]Trier University, Institute for Labour Law and Industrial Relations in the European Union (IAAEU), Trier, Germany

Correspondence
Adrian Chadi, Department of Economics, University of Konstanz, Universitätsstr. 10, D 78457 Konstanz, Germany.
Email: adrian.chadi@uni konstanz.de

Abstract

Causal evidence for the effect of team size on performance is lacking despite the high relevance of this question. From an economic perspective, one would expect performance incentives to decrease with increasing team size, but the psychological phenomenon of peer effects could mitigate the free-rider problem. To analyze the behavioral implications due to changes in team size, we exploit a rich dataset from a controlled experiment with teams of either two or three participants of a university research project performing a real-effort task. Our study provides three main findings. First, increasing team size does not change team performance on average, which is a robust result, be it across performance dimensions and even when introducing pay inequality between two work periods. Second, positive performance spillovers from peer to peer alleviate the free-rider problem when team size increases. Third, changes in peer pressure due to changes in the transparency of others' performance could explain the variation in peer effects across team size. In contrast to discussions in previous literature, our evidence points to a potentially negative role of peer pressure for team performance. While lower peer pressure in teams of three allows for more positive performance spillovers, a high-skilled peer in a team of two seems to pressure the other team member to produce more mistakes instead of more work output in high quality.

JEL CLASSIFICATION
J24, J33, C92, M52

1 | INTRODUCTION

Numerous practical guides and books have been published on the topic of how to improve team performance (Ben-Hafaïedh & Cooney, 2017; Chiocchio et al., 2015; Griffith & Dunham 2014; Salas et al., 2013). Although the number of team members is an important characteristic, the effect of team size on performance has not been extensively analyzed. One explanation for this research gap could be the difficulty to exogenously assign teams to obtain causal evidence. Another explanation could be that the effect of team size on productivity seems obvious because of the free-rider problem, which emerges when individuals have to bear the full costs of contributing to a group outcome that is shared equally between all members (Alchian & Demsetz, 1972; Holmstrom, 1982; Newhouse, 1973). Accordingly, if the number of team members rises, the incentive to free-ride increases as well. However, there is good reason to not expect free-riding as the dominant strategy of individuals working in a team. Kandel and Lazear (1992) point out that the incentive to shirk

This is an open access article under the terms of the Creative Commons Attribution NonCommercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non commercial and no modifications or adaptations are made.
© 2023 The Authors. Managerial and Decision Economics published by John Wiley & Sons Ltd.

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-2-flxpp4zvdwa68

may be countered by psychological mechanisms such as peer pressure. Experimental evidence from the field (Babcock et al., 2015) and laboratory (Corgnet et al., 2015) supports the idea that peer effects can mitigate the free-rider problem when comparing teamwork with individual work settings. Arguably, peer effects may be able to also compensate disincentives that result from increasing the size of the team. Yet, so far, research on this idea and the role of psychological mechanisms in this context is limited.

The aim of this study is threefold. First, we test the free-rider hypothesis, which implies that increasing the team size decreases individual performance. Second, we analyze whether peer effects can solve the free-rider problem that could occur when the number of team members increases. Third, we inspect what mechanisms are at play in this context. Given ongoing discussions in the literature, we are particularly interested in the role of peer pressure as a reason to contribute to teamwork. To address our aims, we examine the setting of a computer workstation at a German university with either two or three study participants performing a real-effort task. All participants and their actions are visible in this workplace setting, which could enable peer effects via exposure to the behavior and hence the performance of others. Thanks to a team piece rate, individual payments vary depending on the average individual output level across all team members, which implies stronger incentives to free-ride when working together with more beneficiaries of an individual's effort levels.[1]

To identify the effects caused by changes in team size, we experimentally manipulate whether two or three team members have to accomplish the task. We consider this comparison to be the most promising for our research goals for several reasons, apart from the practical aspect of being able to conduct a controlled experiment with a manageable sample size. From a theoretical perspective, adding one more person to a team of two implies the most severe consequences for both economic incentives and psychological drivers, compared to adding one more person to a larger team. Furthermore, from an empirical point of view, the question of whether a team of two should be expanded to include another person is probably the most relevant question. In fact, teams of two or three members are very common in many contexts, be it in academia (Hamermesh, 2013, 2017) or when it comes to entrepreneurial teams (Backes-Gellner et al., 2015; Shrivastava & Tamvada, 2011) or in workplace environments, which our study focuses on.

Although we acknowledge that comparing teams of two and three individuals does not inform us about the performance of larger teams with even more members, we are confident that the opportunity to study behavior in a controlled workplace setting allows us to establish credible evidence regarding our main research questions. In particular, we can exogenously assign teams to identify the causal effect on outcomes, such as worker productivity, which appears necessary for empirical identification given that endogenous team formation is a relevant phenomenon in workplace contexts (Bäker & Mertins, 2013; Büyükboyaci & Robbett, 2019; Chen & Gong, 2018; Fischer et al., 2020; Kölle, 2017; Kuhn & Villeval 2015). The experimental evidence informs us about possible free-riding after an increase in the team size and the occurrence of peer effects as a way to compensate for free-riding. Our setting also provides us with rich survey data on possible psychological mechanisms to find out why behavioral effects may or may not occur. Another feature of the workplace setting is the availability of precise information on performance in different dimensions, which allows examining the possibility that team members could be reducing the quality of work.

As our first main result, individual performance levels are on average almost the same across both team size conditions, which seems to contradict the free-rider hypothesis. This remains when we i) compare the results across two work periods, ii) analyze alternative performance indicators, and iii) differentiate between equal and unequal pay in consequence of random bonuses. In line with our main result, evidence from the survey confirms that on average there is no manipulation of any effort mechanism, such as willingness to free-ride.

Second, by analyzing effect heterogeneity across different levels of peer performance, we obtain evidence that psychological motives for putting in efforts can minimize the free-rider problem. A team member's performance is positively related to co-worker performance, suggesting motivational spillovers from peer to peer ('peer effects'), but this interacts with the team size. To empirically confirm the role of peer effects in our workplace context, we make use of survey data on computer skills, which is a strong predictor of performance in routine computer task. A split between high-skilled peers and low-skilled peers reveals that individual performance in teams of three is lower than that in two-person teams if peer skills are low. This shows that the free-rider problem exists but can be mitigated thanks to the positive peer effects induced by high-skilled and hence high-performing team members.

Third, we find that peer pressure could be a relevant mechanism in this context, given that it depends on the team's size. Interestingly, the evidence appears to be at odds with previous research describing a positive role of peer pressure. We find that individuals in teams of three perceive less pressure when peer skills are high, as compared to two-person teams. Furthermore, there is evidence that an increase in the number of team members goes along with less transparency of peer performance, as it becomes more difficult to correctly estimate the average team performance. In consequence, high peer pressure owing to more transparency of peer performance may prevent positive peer effects to occur in two-person teams, where we observe rather adverse effects in the performance data on the quality of work. By considering different performance dimensions, we find that the amount of incorrect output units increase in teams of two when peer skills are high, while the amount of correct work output declines. All this is inconsistent with the idea of peer pressure improving performance but rather suggests that negative peer effects could result from pressuring individuals to perform well and points to possible drawbacks of competitive work environments.

With our study, we contribute to several strands of research. In particular, we confront ideas from both empirical and theoretical work on team size with experimental evidence, which is an important contribution given that the personnel economics literature (for a review, see Lazear & Oyer, 2013) has been rather silent on the role of team size in performance outcomes so far. Furthermore, we contribute with

our findings to research on peer effects in workplace performance (for a review, see Herbst & Mas, 2015), which has not provided a consistent picture regarding the underlying mechanisms yet. Our finding that peer pressure could impair rather than improve outcomes may stimulate further research on the psychological patterns behind peer effects at the workplace and elsewhere.

In the following Section 2, we describe our experimental setup in more detail and discuss our hypotheses, while supplementary material and information are provided in Online Appendix A. In Section 3, we present the results of our empirical investigation, which we complement with additional evidence provided in Online Appendix B. Based on the summary of results, we discuss the limitations of our study and provide a broader discussion of our contributions to ongoing research on teams and peer effects in the concluding Section 4.

2 | EXPERIMENT

2.1 | Procedure

During the winter semester of 2016/2017, a large-scale research project took place at a research institute of a German university. Participation was possible throughout the entire semester from October 2016 to March 2017. The project was announced as a "study on work motivation at a computer workplace" to students via posters and flyers at the university campus as well as online using the university's email list. Each student could earn a flexible amount of money of approximately 12 Euro for a one-hour-long participation, which roughly corresponds to the hourly wage of student assistants employed at the university. To foster interest in participating, an additional bonus win of a 100 Euro lottery price was announced, the winner of which was determined and paid out after the study ended.

Recruitment of study participants followed clear rules and was organized by research assistants. Appendix A1 provides more information on the procedure, which aimed to minimize both chances of social ties between individuals participating in the same session and selectivity in individual characteristics across team-size conditions. After being recruited, each participant was invited to come to one out of three separate waiting rooms, which served the purpose of smoothing discrepancies in arrival times between participants. From a participant's point of view, it was unclear whether other individuals would participate in the same session, given that no such information was provided. Prior to the start of the session, participants were guided from the waiting rooms to the computer workplace by research assistants. All sessions took place at the same computer workstation (see Figure A1 for pictures).

We randomly assigned session slots as either two-person or three-person settings at least one week in advance. Three-person sessions had a higher chance to be drawn in the randomization process due to a simultaneously ongoing experiment (Chadi & Homolka, 2022), for which a control condition with a sizable number of observations from three-person sessions was needed. As part of the procedure (see Appendix A1 for more details), three-person sessions were not canceled in case one invitee did not show up. Instead, those sessions took place in the same way as regularly assigned two-person sessions. While this provides us with additional data, we pay particular attention to potential selection issues when analyzing our full sample, which includes the data from those 'transformed' two-person sessions. As the only difference between a regularly assigned and a transformed two-person session, we implemented a verbal treatment in the latter case (see Appendix A2 for the translated script of the instructions). Our idea here was to find out more about the possible psychological effects in the remaining participants when a third participant was absent, as individual behavior in transformed sessions might differ not only because of selectivity but also because of psychological effects, which we cannot fully rule out.[2]

The session host was a research assistant employed at the institute who guided the participants through the session by reading the instructions aloud. Accordingly, the participants were referred to as 'team players' and the session host took on the role of the 'team leader' during the session. Except for a few adjustments due to the different numbers of participants (two or three) across sessions, the instructions were always the same. At the beginning of each session, participants drew lots with either two or three numbers to determine their computer workplace. As can be seen in Figure A1, the computer screen of the third workplace was turned off and equipment was removed in case of two-person sessions.[3] Until the end of the session, oral communication was prohibited, but email communication was introduced as an alternative. While this could be seen as one option for participants to shirk by using emails to chat, another shirking option emerged out of the fact that all computers at the workstation were connected to the internet. This was necessary to allow participants to work on their tasks.

The work task was to prepare scientific papers for research at the university by downloading and labeling journal articles from an online library, which was done in similar fashion by students employed as research assistants at the institute. During a session, participants had to work on a volume of a journal and prepare as many articles as possible within the working time. The session's journal and the different volumes for the computer workplaces were randomly drawn from a list prior to each session. To illustrate how to perform the task, the session host used a big screen (see Figure A1), while also reading the instructions out loud. The task itself consisted of two work periods, lasting ten minutes each, and work period one started only after all session participants had finished a test run of successfully preparing one article.

Figure 1 displays the timeline of a session. After the session, participants were paid separately in their waiting rooms. Pay was determined by the number of points received during the session, based on the following rule: 1 point equals 1 cent. As one out of several pay components, participants' performance in both work periods determined their earnings based on a team piece rate scheme. Accordingly, the number of downloaded and correctly labeled articles at the end of each work period was first aggregated across all participants and then divided by their number (two or three). Each participant was then rewarded with 10 points per average output unit so that the total

	Event	
1.	Welcome	Introduction to the rules regarding communication and pay
2.	Survey 1	Items include measures for reciprocity, trust levels and risk attitudes
3.	Instruction phase	Introduction to the task including test run
4.	Work period 1	Work task (10 minutes)
5.	Survey 2	Items include Big 5, skills, workplace perceptions and performance estimations
6.	Bonus game	Possible win for randomly selected bonus players
7.	Work period 2	Work task (10 minutes)
8.	Survey 3	Items include workplace perceptions, satisfaction and future work motivation
9.	Payout	Pay and final feedback in separate rooms

FIGURE 1 Timeline of the work session.

achievement was shared equally between all team members. Participants were informed about the team result, but not about individual performance, by the session host. In addition to the performance-related pay component, a bonus game took place between the two work periods (see Appendix A2 for more information on the procedure), which induces pay inequality through a random allocation of a bonus.[4] With a 50% chance, all but one team member received 300 points in the game. While we refer to this as the 'unequal pay' scenario, no one received a bonus in the 'equal pay' scenario, which also had a 50% chance of occurring. Finally, in addition to the flexible-income components, participants received fixed payments for showing-up (250 points) and for completing the three online surveys (200 points each).[5]

2.2 | Data

2.2.1 | Performance

Our workplace setup informs us about several performance indicators, which are based on the work output of digital files. First and foremost, we focus on what we refer to in the following as a 'task performance' indicator, which comes directly from the session records. This performance measure reflects both the quantity and quality of work output and was used as the basis for performance-related pay. The session host followed clear rules to determine each participant's task performance, including a spot check on possible mistakes in the work output: From all the articles downloaded by a participant in a work period, the session host checked a randomly chosen output unit for the correctness and, if incorrect, continued with up to two more checks until an article without inaccuracy was found. The total number of articles in the folder was then reduced by the number of output units identified as erroneous. In the case of three erroneous articles being detected, the work output of this team member was reduced to zero.[6]

In addition to the main task performance indicator, we inspect the quantitative dimension of performance by examining the total number of downloaded articles per person, independent of accurate labeling. We then exploit the opportunity to analyze more deeply the qualitative dimension of performance using computer software that conducts precise inspections of files produced in the sessions. By identifying deviations from fully correct article labels, it is possible to detect even small mistakes that a participant made when labeling articles, such as typing in incorrect author names. Using the software-generated data, we distinguish between different performance dimensions by splitting the 'total work output' into correct output units ('high-quality work output') and incorrect output units ('low-quality work output').

2.2.2 | Surveys

Given our aim to learn more about motives underlying the decision to put in effort in a team context, the participants in our experiment had to respond to a large set of questions via online surveys. To allow for a convenient survey length in the course of a session, the questions were separated into three blocks, i.e. before, between, and after the work periods (see Figure 1).

We included several question modules to find out about psychological drivers of effort based upon ideas and discussions in the literature on the work motivation of individuals working with others on a task. First, a question module on the potential channels of team performance asked participants after work period one about how they perceived teamwork, which included free-rider intentions, among other things (Table B1).[7] This module was repeated after work period two in order to allow for a before-and-after comparison regarding the effects of the bonus game on perceived unfairness. Second, participants were asked after the two work periods in a rather direct way about why they worked hard or not, which included for example whether feeling monitored was a reason for putting in the effort (Table B2). Third, we asked participants after work period one about perceptions regarding their peers and channels of peer effects, such as psychological pressure (Table B3). Hence, as shown in Figure 1, none of the three question modules on workplace perceptions were part of the first survey that took place before the task.

TABLE 1 Sample statistics.

	(1) Mean	(2) Mean	(3) Mean	(4) *p*-value	(5) Min	(6) Max
Age	25.02	24.93	25.14	0.915	19	53
Number of semesters	4.71	5.20	5.04	0.974	1	25
Bachelor's degree	0.33	0.33	0.36	0.883	0	1
Female	0.57	0.55	0.70	0.059	0	1
Skills: Math	0.36	0.13	0.20	0.561	−2	2
Skills: Probabilities	0.21	−0.09	−0.03	0.225	−2	2
Skills: Computer	0.55	0.16	0.41	0.083	−2	2
Big5: Extraversion	4.92	5.14	4.97	0.478	1.33	7
Big5: Agreeableness	5.62	5.43	5.41	0.555	2	7
Big5: Openness	5.29	4.77	5.03	0.078	2	7
Big5: Neuroticism	4.10	4.38	4.42	0.418	1	7
Big5: Conscientiousness	5.52	5.11	5.19	0.093	2.33	7
Reciprocity	6.21	6.09	6.06	0.957	2	7
Trust	3.48	3.46	3.52	0.518	1	5
Number of team members:	2	2	3			
Randomly assigned sessions:	Yes	No	Yes			
N	42	76	153			

Notes: Means are displayed in columns 1, 2, and 3. Column 4 shows *p*-values from Kruskal-Wallis equality-of-populations rank tests. Minimum and maximum values are reported in columns 4 and 5. The data used in column 1 are from 21 randomly assigned two-person sessions. The data used in column 2 are from 38 two-person sessions originally scheduled as three-person sessions and one person not showing up. The data used in column 3 are from 51 randomly assigned three-person sessions. Age, Number of semesters, and Bachelor's degree are variables reported by participants during webpage registration. Female is a dummy variable taken from the session records. Variables on skills are measured on a 5-point ordinal scale, ranging from "Very good" (score: 2) to "Very bad" (score: −2). The wording of the survey questions on math/probability/computer skills is "How good are you … at math?" / "… at judging probabilities?" / "… with computers?" Survey questions on Big5 personality dimensions (15 items), positive reciprocity (3 items), and trust (1 item) are taken from the German Socio-Economic Panel Study.

To collect further data on possible mechanisms, participants had to estimate both individual performance and team performance, and they were asked several questions about satisfaction and preferences for teamwork, including questions regarding the size of the team. Apart from subjective perceptions, the survey questions also aimed at collecting data on individual characteristics. Traits considered important for understanding behavior in a teamwork context included the Big Five personality dimensions (based on 15 items transformed into 5 factors) as well as reciprocity and trust attitudes. Finally, self-estimated skill levels in mathematics, probability estimation, and working with computers were obtained on a 5-point scale. To ease the interpretation of the effects in our empirical analysis, we redefine these variables to range from −2 (very lowly skilled) to 2 (very highly skilled), so that 0 reflects mid-level skills.

2.2.3 | Sample

We merged all the available data at the individual level, coming from the following sources: i) webpage registration, ii) session records, iii) online surveys conducted during sessions, and iv) software-generated performance data. Table 1 illustrates the sample by showing average statistics for all variables that we use as controls when we conduct regression analyses. Averages are shown for data from randomly assigned sessions of two persons (column 1), randomly assigned sessions of three persons (column 3), and transformed two-person sessions (column 2). In the following, we distinguish between the full sample with 271 observations and a smaller dataset with 195 observations, which come solely from sessions conducted as originally scheduled regarding the number of team members.

Columns 1 to 3 in Table 1 display mean values that are very similar across all variables shown. A closer inspection based on statistical testing nevertheless reveals a few weakly significant differences in observable characteristics when comparing individuals across all three subsamples. According to column 4, Kruskal–Wallis equality-of-populations rank tests do not reject the hypothesis of all subsamples coming from the same population, as we observe p-values indicating significant differences at the 10% level for the Big Five characteristics conscientiousness and openness as well as gender and computer skills. If we turn to pair-wise subsample comparisons and start with

the dataset excluding transformed sessions (N 195), as shown in columns 1 and 3, Kruskal–Wallis tests reveal one weakly significant result. This suggests that the data from randomly-assigned sessions come close to being perfectly randomized. If we focus on the full sample (N 271) and compare data from all two-person sessions (columns 1 and 2) versus three-person sessions (column 3), Kruskal–Wallis tests reveal significant gender differences, as the only statistically significant result.[8] Finally, by comparing data from transformed sessions (column 2) separately with the data from three-person sessions (column 3), we again find a statistically significant result in regard of gender and also a weakly significant difference in regard of computer skills.[9]

We conclude that there is reason to double-check our results regarding robustness across datasets. We hence proceed in the following analysis by first making use of the full sample with 271 observations and then reporting on whether results based on the dataset with 195 observations differ. To ease the flow of reading, we do so only if the results indeed turn out to be different in terms of statistical significance. Besides, we can check the robustness of our results after controlling for possible differences in characteristics via covariates in regression analyses.

2.3 | Hypotheses

Our setup allows us to test several hypotheses based on theories and findings from the literature. In the following, we postulate three main hypotheses that we focus on in our empirical analysis. We explain the idea behind each hypothesis by referencing research that is relevant to our investigation. In doing so, we also take into account the specific background of our investigation with a comparison of teams of two and three as an exogenous manipulation of team size.

> **Hypothesis 1.** *Increasing the size of a team decreases the performance of team members*. The free-rider problem implies that as the team size increases, so do the incentives to rely on the performance of others (Alchian & Demsetz, 1972; Holmstrom, 1982; Newhouse, 1973). In our study, this could be reflected in lower individual task performance among the three team members, compared with the two-member teams. Having said that, there could be other ways through which free-rider incentives may affect behavior in a workplace setting, where team members may not know that the quality of their work is recorded. As a possibility, individuals may try to hide lower effort levels by reducing work quality while keeping performance high in quantitative terms to signal work willingness (Frick et al., 2013). In contrast to expectations of negative effects of increasing team size, the literature also discusses why free-rider incentives could be mitigated or even overcompensated, such as in the research on public goods where researchers study the role of contributions in groups (Isaac et al., 1994; Isaac & Walker, 1988). While contributing to a public good is a similar incentive problem as putting in the individual effort to raise the team outcome, motivations arguably can be different. For example, Zhang and Zhu (2011) show that a larger number of group members contribute more, which could be a consequence of greater social benefits. Whereas such motives could also play a role in a setting with different individuals working together, research in personnel economics rather suggests a role of performance spillovers between peers as a way to compensate for free-rider incentives.[10]

> **Hypothesis 2.** *Peer effects mitigate the free-rider problem when the size of a team increases*. Numerous studies have revealed positive performance effects when individuals work with peers on a task. Falk and Ichino (2006) experimentally show that productivity increases in two-person work settings compared to solo work. Further evidence on performance spillovers at work comes from field studies, which are often non-experimental, such as Mas and Moretti (2009) who show that visibility of peer behavior is an important factor. Building on this broader research on the role of peers at work, Sausgruber (2009) raises the question of whether peer effects also exist between teams, while others examine whether peer effects can solve the incentive problem in teams compared to settings with individual incentives. For example, Corgnet et al. (2015) show in a laboratory experiment, in which team production is compared with individual production, that a weak form of peer monitoring is an option to solve the free-rider problem by increasing peer pressure and to thereby raise team production to the level of individual production. In a field experiment at a university, Babcock et al. (2015) compare team incentives and individual incentives to show that free-riding can be prevented as team members pressure each other to keep motivation high. While this research shows that peer effects can minimize free-rider incentives in teams compared with individual incentives, experimental evidence comparing teams of two or three members as a modification of team size is lacking.

> **Hypothesis 3.** *Peer pressure is the mechanism underlying performance spillovers in teams*. The psychological phenomenon of peer pressure not only has received much attention in the above-mentioned research but also plays a seminal role in other discussions on how to tackle free-riding in teams, ranging from early contributions (Kandel & Lazear, 1992) to more recent ones, such as Backes-Gellner et al. (2015), who specifically focus on the role of team size. Hence, we consider the idea of peer pressure as the potential mechanism underlying differences in performance spillovers across team size,

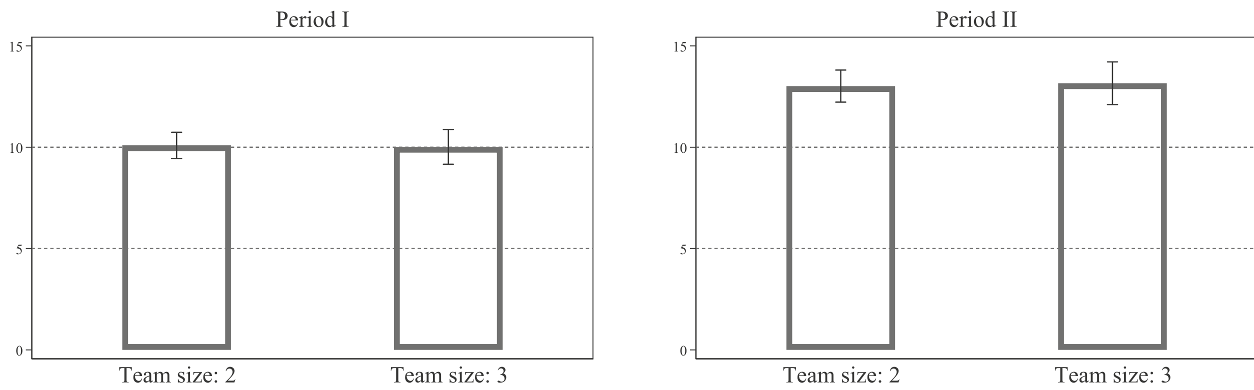


FIGURE 2 Team size and task performance. *Notes*: bars show average task performance by team size by work period. The left (right) illustration displays the first (second) work period. The task performance measure is the number of output units used for performance pay after a spot check on correctness was conducted during sessions. 95% confidence intervals are displayed. Total sample size is 271 observations.

although we acknowledge that other psychological factors could also be at play. In this context, the literature on peer effects and free-riding in teams discuss a variety of ideas and mechanisms that may be related to team size and the occurrence of peer pressure. For example, Vranceanu et al. (2015) discuss the role of punishment as part of a monitoring process to alleviate free-riding in teams. In another experiment on peer pressure, Bonein (2018) investigates how the display of co-worker effort matters for social comparisons with other workers. In related work, Carpenter et al. (2009) discuss the idea of establishing a norm of contributing to production as a means to prevent shirking in large teams, while Carpenter (2007) shows that monitoring could be of importance to tackle the free-rider problem when group size increases. For our workplace study, one could argue for or against a greater role of peer pressure owing to increased team size. On the one hand, individuals are observed by more peers and thus are more pressured to work. On the other hand, peers are less likely to identify a shirker with low work motivation when contributions to team outcomes are less transparent. In this context, Mohnen et al. (2008) show the importance of transparency as a prerequisite to enable peer pressure in teams, which could be an important factor in a workplace setting like ours and change the role of peer pressure across team size.

3 | EMPIRICAL ANALYSIS

3.1 | Team size, performance, and perceptions

3.1.1 | Average performance across team size

We start the presentation of our results by comparing average task performance across team sizes and work periods. To provide another perspective on the data, we present histograms of task performance in Figure B1. Note that in the following, we conduct t-tests for analyzing the statistical significance of any effect detected in performance-related measures. This allows us to take the actual magnitudes of differences in performance levels into account, instead of ranking only, while we focus on non-parametric testing below when we turn to subjective perceptions that are usually measured on ordinal scales. Also, note that we always conduct two-sided tests and routinely cluster the standard errors at the session level in regression analyses.

Figure 2 visualizes performance using bars and reveals that the average performance is about ten output units in the first work period (left illustration). The bars are almost identical for teams of two and teams of three (10.09 vs. 10.02 output units, p 0.866, t-test). Inspecting task performance in the second work period (right illustration), there is a visible increase in the overall performance of about three output units, compared to work period one. This difference may be due to a learning effect. Yet, both bars reflecting performance in period two across team size do not significantly differ (13.02 vs. 13.16 output units, p 0.794, t-test). Separate tests focusing on the change in performance over time do not reveal significant differences either (2.92 vs. 3.14 output units, p 0.554, t-test).[11]

Arguably, it may be interesting to discuss whether there is a small performance effect that could be detectable with even more statistical power. While we as applied researchers are interested in economically relevant effects, just like personnel management is, the fact of almost identical performance levels in Figure 2 does not leave much room for any meaningful effect. Thanks to the high precision of estimates, we have sufficient statistical power to reject even small performance effects.[12] Hence, the evidence appears to be at odds with the free-rider hypothesis.

3.1.2 | Performance dimensions and unequal pay

We continue to inspect the robustness of our first main result by analyzing team-size effects across different performance dimensions. For

this analysis, we differentiate between the total number of output units without any quality check and software-based quality-adjusted work output.[13] Figure B2 shows an increase in the total number of output units from work period one to two by roughly three (Panel A). This goes along with a similar-sized increase in the numbers of correct output units (Panel B), while the average number of erroneous output units is constant with about three in both work periods (Panel C). Regarding the impact of team size, all six illustrations show no differences in performance across dimensions and work periods.

Using another feature of our setup, we examine whether the lack of shirking in teams of three is robust even after inducing unequal pay. Between the two work periods, random bonus payments that are performance-independent are paid out to all but one team member. Given a 50:50 chance of a session with bonus pay, we examine period-two performance for both scenarios, with and without manipulation of the payment structure. To find out whether inducing pay inequality triggers psychological effects depending on team size, we first exploit the available survey data and find evidence of perceived unfairness in teams of three.[14] Despite the manipulation of fairness perceptions through unequal pay, there are no relevant differences in average changes in team performance across work periods when we separate by team size and by the outcome of the bonus game. If anything, Figure B3 shows that the increase in performance tends to be even stronger in teams of three when payments are unequal, but this is not statistically significant (2.95 vs. 3.44 output units, p 0.357, t-test).[15]

3.1.3 | Team size and perceptions

To better understand our first main result regarding the free-rider hypothesis and to determine whether there is a potential for peer effects that could explain the lack of free-riding, we analyze the survey data on workplace perceptions and start by inspecting channels of task performance. As a possibility, similar average performance levels across team size could be the result of different psychological motives for putting in the effort being at play but canceling each other out. When analyzing the data from the question battery on possible channels of team performance, shown in Table B1, we cannot find evidence for different channels responding to the manipulation of team size.[16] We come to the same conclusion when we analyze the data from the question battery on potential reasons for providing more or less effort, shown in Table B2.

While team size as such does neither affect average performance levels nor its drivers in general, contrary to our first hypothesis, participants may perceive working with peers differently depending on team size, as an indicator of a potential for heterogeneity regarding peer effects. Table B5 presents evidence from survey questions regarding satisfaction with teamwork and future work motivation by showing average scores, separated by team size, as well as p-values from rank-sum tests. While there are no significant differences in perceptions regarding the task or the team as such, three-member teams report significantly higher preferences for doing the task again compared to two-member teams. We obtain a similar result when we focus on the satisfaction with the team size.[17]

A separate survey question on perceptions of working with peers offers another option to study differences in perceptions of teamwork across team size. Participants had to assess on a scale whether their motivation is affected more negatively or more positively by doing the same work with others in a team. When we conduct a median-split of responses to establish a binary indicator reflecting more positive perceptions, we find a significant team-size effect in the likelihood of a positive assessment of working with peers (42.37% vs. 54.90%, p 0.041, rank-sum test).[18] In additional regression analyses, we confirm an increase of more than twelve percentage points in the likelihood of perceiving peers positively when team size increases (see Table B6). Arguably, more positive perceptions of peers could set the stage for more positive performance spillovers, in line with our second hypothesis, which we investigate next.

3.2 | Peer effects and team size

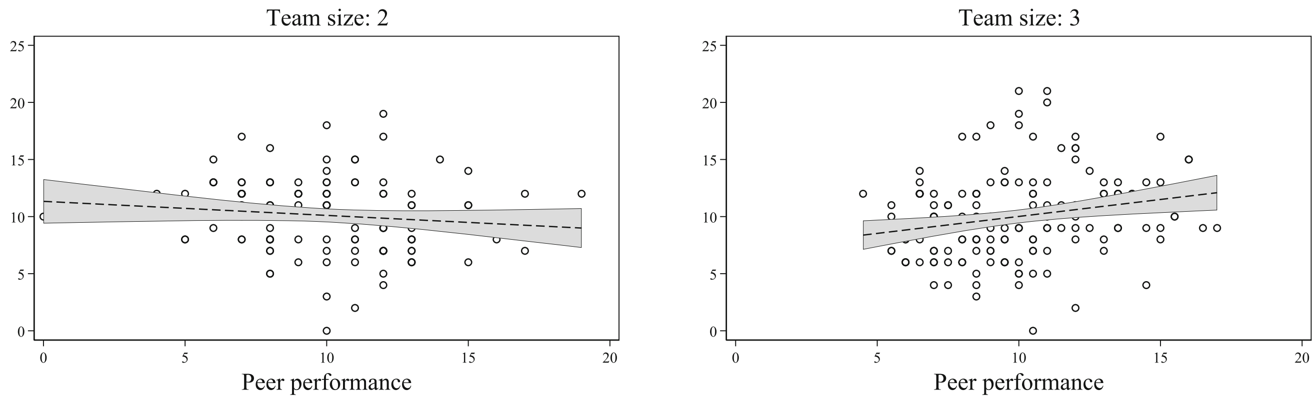
3.2.1 | Performance spillovers across team size

In the first step to investigate the occurrence of peer effects in our workplace setting, we show the raw performance data by means of scatterplots. A positive link between peer performance and own performance could suggest possible performance spillovers. This provides us with first evidence on whether the behavior of peers is relevant for one's own performance depending on the team's size, thereby alleviating a potential free-rider problem, in line with hypothesis two.

Figure 3 visualizes the relationship between individual task performance (on the y-axis) and average peer performance (on the x-axis) across team size and across work period. Based on linear regressions without control variables, we draw lines through all scatterplots, reflecting the slope parameter, for which we then check whether it differs significantly from zero. The upper-right illustration shows for the first work period a significantly positive relationship between average peer performance and own performance in teams of three (coefficient: 0.297, p 0.005). As the link is statistically insignificant and negative (coefficient: −0.123, p 0.184) for two-member teams, as shown in the upper-left illustration, the comparison of both illustrations conforms to the idea of an interaction between peer effects and team size. As can be seen in the bottom-left illustration, the relationship between peer performance and own performance in period two remains insignificant (coefficient: −0.002, p 0.982).[19] For teams of three, the bottom-right illustration on work period two is similar to work period one, with respect to both coefficient (0.277) and statistical significance (p 0.009).

While the evidence in Figure 3 is consistent with our second hypothesis and the idea of performance spillovers between peers reducing free-riding in teams of three, we perform further analysis to address methodological challenges in identifying peer effects in groups. Our setup addresses common problems of identification like self-selection; yet, the reflection problem that occurs when regressing

(a) Work period I



(b) Work period II

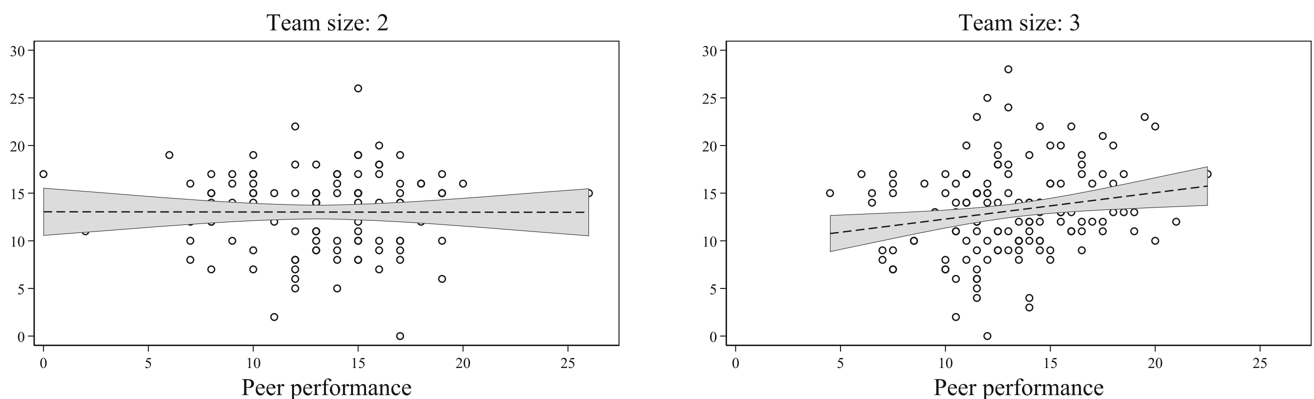


FIGURE 3 Team size and peer effects in task performance. *Notes*: panels a and b shows scatterplots of individual performance and peer performance in the first (second) work period by team size. 95% confidence intervals for the linear fits are shown based on a linear regression without control variables. In the left (right) illustrations, the team size is 2 (3) and peer performance is the (average) performance of the other (two) team members. The task performance measure is the number of output units used for performance pay after a spot check on correctness was conducted during sessions. Total sample size in each panel is 271 observations.

outcomes on outcomes (Manski, 1993) deserves attention. We address this issue by using information from the survey on computer skills, as a separate indicator for the performance potential of peers. This relies on the idea that from the perspective of each participant, the skills of a different team member are an exogenous driver of this person's performance. By assuming that i) own computer skills do not affect the computer skills of others, and ii) computer skills are a strong determinant of performance in this task of downloading and renaming scientific articles from the internet, we can consider the computer skills of peers ('peer skills') as an independent trigger of peer effects in the following.[20]

Table 2 shows the results from regression analyses when using peer skills as a determinant of individual task performance interacted with team size to test for heterogeneous peer effects. Across four specifications, we show results for both work periods as well as for two outcome variables: our task performance indicator and the total number of output units. In all columns, the results are consistent with the visual evidence in Figure 3 regarding heterogeneity in peer effects across team size. While better peer skills do not go along with higher task performance in teams of two, there is evidence for effect heterogeneity, as the role of peer skills is more positive in teams of three. This is demonstrated by the interaction effect between team size and peer skills. In a separate row, Table 2 also shows the total peer effect in teams of three (i.e. peer skills effect plus interaction term effect). While statistical significance varies slightly across work periods and outcome variables, the effect sizes are in each case very similar. We come to the same conclusion when using control variables and aggregated performance data from both work periods as the dependent variable (see Table B7).[21]

3.2.2 | Heterogeneous peer skills and performance spillovers

Since we observe peer effects in both work periods, our results also inform us about the nature of performance spillovers. Given that

TABLE 2 Team size and heterogeneous peer effects.

Dependent variable:	(1) Task performance	(2)	(3) Total work output	(4)
Work period:	I	II	I	II
Team size 3	−0.492	−0.317	−0.481	−0.339
	(0.454)	(0.576)	(0.426)	(0.566)
Peer skills	−0.193	−0.080	−0.038	−0.084
	(0.276)	(0.348)	(0.255)	(0.349)
Interaction: Peer skills X team size 3	1.070**	1.131*	1.043**	1.195*
	(0.529)	(0.666)	(0.486)	(0.635)
Total effect of peer skills in teams of 3	0.878*	1.051*	1.005**	1.111**
	(0.451)	(0.568)	(0.413)	(0.531)
R^2	0.015	0.013	0.020	0.016

Levels of statistical significance:
*p < 0.1.**p < 0.05.***p < 0.01.*Notes*: Linear regressions are used. The dependent variable in columns 1 and 2 is the number of output units used for performance pay after a spot check on correctness was conducted during sessions. The dependent variable in columns 3 and 4 is the total number of output units without any checks of correctness. Team size 3 is a variable that is 1 if the team size is three and 0 if the team size is two. Peer skills are a variable ranging from −2 ("very poor skills") to 2 ("very good skills"). All columns show first the results for separate interaction effects (between peer skills and team size) and second, in the lower row, the total interaction effects for peer skills in teams of larger size (as a combination of the peer-skills effects and the separate interaction effect). The full sample (N = 271) is used. Clustered standard errors (at the session level) are in parentheses.

participants neither had information about the level of peer skills nor did they obtain information about performance levels prior to the end of the first work period, it is very likely that one's performance was indirectly affected by peer skills via observable peer behavior. Imagine observing a person with high computer skills who seems confident to perform well in the task versus a person struggling with the computer equipment and hence performing not so well. While it may seem plausible to assume that the low performer rather than the high performer is the driving force behind peer effects, our setting allows us to examine such possible heterogeneity.

To learn more about the performance spillovers between peers in teams of three, we modify the analysis of Figure 3 by distinguishing between the performance levels of the two remaining team members. Examining the roles of better-performing peers and worse-performing peers separately promises to shed light on whether high performers spur others to provide high performance or whether low performers make it easier for others to shirk. This analysis reveals for work period one a coefficient that becomes slightly larger (0.320, p 0.002) when instead of average peer performance we use minimum peer performance, i.e. the performance of the worse-performing peer of the two other team members. The relationship between own and peer performance becomes weaker (0.179, p 0.037) when defining the latter as the performance of the better-performing peer. This gap in the performance correlations becomes even more pronounced in work period two, as the correlation between own performance and the performance of the better-performing peer becomes insignificant (coefficient: 0.114, p 0.206), while it remains strong for the worse-performing peer (coefficient: 0.353, p 0.000).

These findings based on the performance data are in line with the notion of low performers encouraging others to underperform. We confirm this interpretation when again using the information on peer skills, in line with the analysis in Table 2. By distinguishing between the skills of the more skilled and the less skilled peer, we find that the minimum peer skills clearly drive the occurrence of peer effects in teams of three (see Table B9).

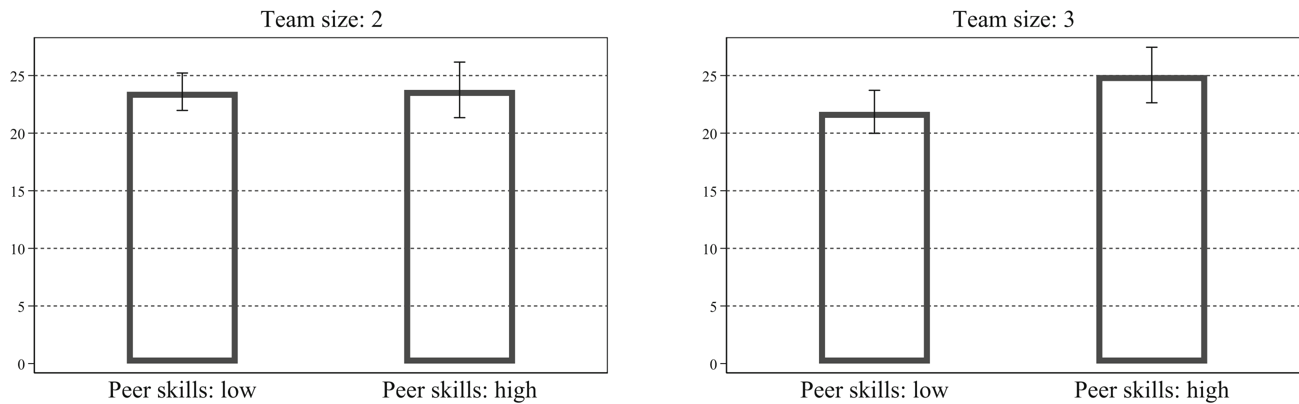
3.2.3 | Free riding, peer skills, and team size

Our analysis of peer effects allows us to examine another important question. If it is true that positive performance spillovers compensate the economic disincentive of working in a team of three, we would expect to detect evidence of free-riding when we focus on those constellations where a negative performance spillover is more likely. Based on the results in Table 2 with its interaction-term analysis, we can conclude if and when free-riding in teams could occur. According to the estimates for teams of three, each score point on the peer-skills scale (from −2 to 2) increases the work output by roughly one output unit in each work period. This means that in the case of low peer skills, a negative performance effect of increasing team size becomes likely. In fact, for mid-level peer skills (i.e. a value of zero, see Table 1), all results show a negative coefficient for the team size variable. One of our robustness checks even reveals a weakly significant negative team-size effect (see column 1 in Table B8). This suggests that our finding of similar average performance levels across team size (see Section 3.1.1) could be due to a relatively large share of participants with sufficient skills to perform well in a simple computer task while free-riding only occurs in teams of three if peer skills are low.

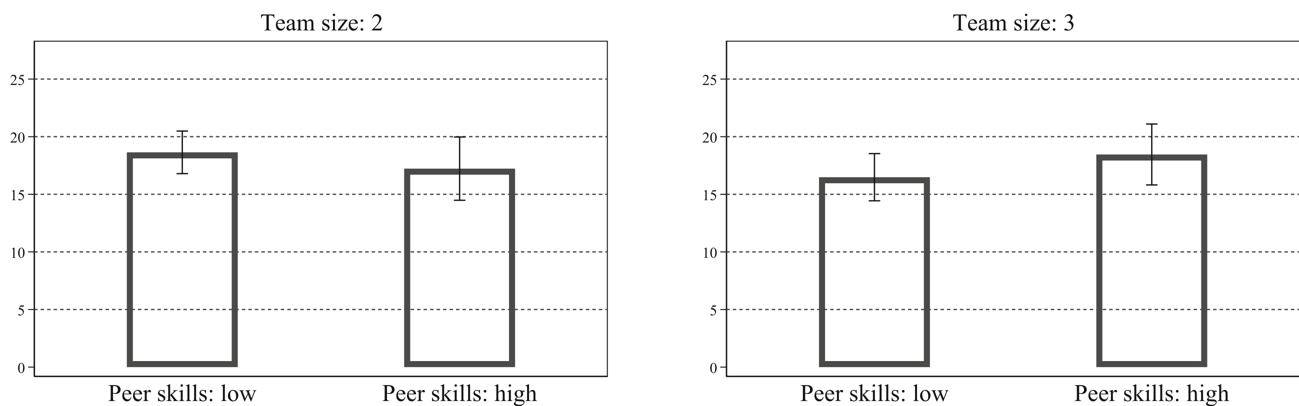
To find out more, we focus on the small number of individuals (N 49) working in sessions where peer skills were below the mid-level score (i.e. negative values). Despite the severe loss of observations in this separate analysis, we find that the team-size effect in task performance for work period one becomes strongly negative (−1.73 work output units) and statistically significant (p 0.031, t-test). This result holds in a slightly increased sample (N 69), where the minimum peer skills have a negative value (−1.80 work output units, p 0.019, t-test), indicating that both peer effects and free-riding are driven by peers with relatively low skills.

Further evidence of the idea of free-riding being prevented by high peer skills comes from our survey data on potential mechanisms. For an analysis based on similarly sized subsamples, we distinguish between settings with "low" peer skills and settings with "high" peer skills, which we define as positive values in the following. As can be seen in Table B1 (columns 3 and 4 of Panel A), the indicator of free-rider intentions shows higher scores for teams of three (1.45) than for teams of two (1.28), when we condition on individuals working together with low-skilled peers only, which is weakly significant (p 0.096, rank-sum test).[22] While this effect is not robust across work periods, we consider this as indicative of the idea that free-

(a) Performance measure: Total work output



(b) Performance measure: High-quality work output



(c) Performance measure: Low-quality work output

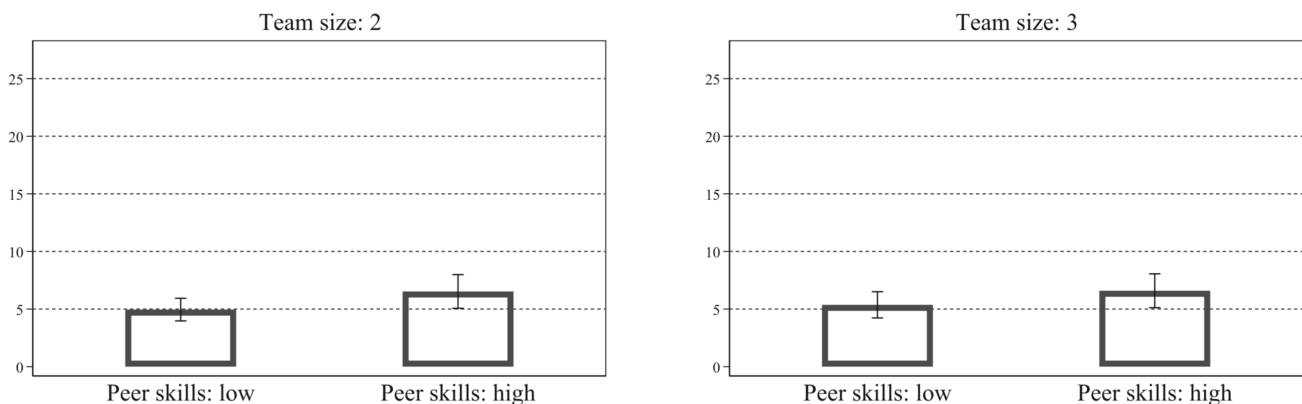


FIGURE 4 Peer effects and performance dimensions. *Notes*: bars show average performance for both work periods combined. The left (right) illustration displays average performance for teams of 2 (3). Peer skills is a variable ranging from −2 ("very poor skills") to 2 ("very good skills"), which is defined as high if values are above 0, and low otherwise. Panel a displays the average number of output units, independent of correctness. Panels b and c displays the average number of correct (incorrect) output units, according to software-based analyses. 95% confidence intervals are displayed. Total sample size in each panel is 267 observations.

riding may be more likely to occur when team size increases in a scenario without high-skilled peers.

3.2.4 | Performance dimensions and peer skills

To learn more about the nature of performance spillovers, we analyze whether peer effects vary across performance dimensions. For this analysis, we differentiate between team size and settings with "high" versus "low"' peer skills, as in the previous subsection, and we consider three different outcome variables. Accordingly, we distinguish between numbers of high-quality work output units and low-quality work output units, while the sum of both is the total number of work output units, independent of correctness. In all three cases, we examine the combined work output which is aggregated across the two work periods.

Figure 4 shows that differences in peer effects depending on team size become even more striking when using a binary distinction into high and low peer skills instead of a linear peer skills variable, as in the analysis of Table 2. For the total work output (Panel A), we find a significantly positive peer effect in teams of three when comparing high-skilled and low-skilled peers (21.9 vs. 25.0, p 0.010, t-test). No such effect is observed in teams of two (23.6 vs. 23.8, p 0.895, t-test). The software-based performance indicators reveal how changes in total work output translate into changes in both high-quality and low-quality work output (Panels B and C). For teams of three, it appears that both go up due to peer effects, as, the 3.2 increase in total output units can be separated into 2.0 more correct output units (p 0.143, t-test) and 1.2 more incorrect output units (p 0.103, t-test). In contrast, we observe clear differences in the direction of peer effects in teams of two. While there is an increase of 1.6 low-quality output units when peers are highly skilled compared to when they are not (p 0.035, t-test), the number of high-quality output units decreases by 1.4 output units (p 0.310, t-test), which explains the absence of any peer effects on performance, as measured in total output units. Strikingly, while members of two-person teams do not increase their work output quantitatively when confronted with a high-skilled peer, our analysis of performance dimensions indicates that they make more mistakes, thereby providing the same quantitative amount of output but in lower quality.

3.3 | Peer pressure and team size

3.3.1 | Subjective perceptions

Our finding of heterogeneous peer effects confirms the second hypothesis and provides an explanation for the lack of evidence in support of the first hypothesis, as performance spillovers in teams of three limit the free-rider problem. This conclusion also fits well with the above evidence presented, according to which working with others is perceived more positively when team size increases (see Section 3.1.3). However, the evidence in the previous subsection on increased mistakes when peer skills are high in teams of two does not seem to conform to the idea of a performance-enhancing role of peer pressure.

To address our third hypothesis and to learn more about peer pressure as the potential mechanism underlying performance spillovers in teams, we analyze responses from a question battery on workplace perceptions related to the occurrence of peer effects. In Table B3, we compare average scores, as reported for various items, across team size and peer skills. For most of the usual suspects discussed in the literature on peer effects, including peer monitoring, the scores have rather similar average values across all four subgroups, suggesting that those mechanisms do not play a role in the occurrence of peer effects in our context, with the exception of two items. First, there is an increase in perceived distraction (from 1.77 to 2.27, p 0.059, rank-sum test) when team size increases and peer skills are high (columns 5 and 6). Second, the same subgroup comparison indicates that the setting with high-skilled peers is perceived as less competitive when team size increases (from 3.64 to 3.10, p 0.071, rank-sum test). Both findings speak against the idea of well-performing peers in three-member teams exerting pressure on others to also perform well.

Next, we conduct a regression analysis using the data from the same survey battery (Table B3) and employing interaction terms between team size and peer skills. The results in Table 3 again show that peer skills increase the perception of a competitive environment in teams of two but not in teams of three, as can be seen in a separate row for the total effect. Furthermore, peer skills significantly interact with team size when we turn to an item measuring the perceived pressure to succeed. This suggests that peer pressure could play a negative role in the emergence of peer effects since less pressure goes along with more positive performance spillovers from peer to peer (as shown for illustration purposes in the first specification of Table 3). To further inspect this notion of peer pressure as a possibly negative factor in task performance, we correlate the two variables and find no evidence of a positive relationship.[23]

The idea of a performance-reducing role of peer pressure receives further empirical support when we use the survey data on reasons for putting in effort (Table B2) to inspect differences in perceptions when peers are either highly skilled or not. We focus on teams of three (columns 4 and 6) to find out more about the occurrence of positive peer effects, which reveals a highly significant negative effect of high peer skills on 'competition for best performance' as a reason to provide effort (decline from 4.50 to 3.51, p 0.001, rank-sum test). In teams of two (columns 3 and 5), we do not observe a decline in average scores, as reported for the same item, which again supports the idea of heterogeneity in peer pressure across team size and suggests that reduced peer pressure could help prevent competitive situations from occurring.

3.3.2 | Transparency of peer performance

Finally, we provide an explanation for the heterogeneity in peer pressure across team size using data from assessments of others'

TABLE 3 Channels of peer effects and team size.

Dependent variable:	(1) Work output	(2) Monitored	(3) Observing	(4) Distracted
Team size 3	−0.481	0.222	−0.115	0.261*
	(0.426)	(0.226)	(0.211)	(0.154)
Peer skills	−0.038	0.066	0.065	−0.133
	(0.255)	(0.175)	(0.141)	(0.097)
Interaction: Peer skills X team size 3	1.043**	−0.034	−0.044	0.023
	(0.486)	(0.266)	(0.229)	(0.182)
Total effect of peer skills in teams of 3	1.005**	0.032	0.021	−0.110
	(0.413)	(0.201)	(0.180)	(0.155)
R^2	0.020	0.004	0.002	0.018

Dependent variable:	(5) Competition	(6) Pressure	(7) Strangers	(8) Shirking
Team size 3	−0.108	0.311	0.162	0.113
	(0.212)	(0.239)	(0.222)	(0.218)
Peer skills	0.324**	0.087	0.020	0.182
	(0.146)	(0.142)	(0.144)	(0.156)
Interaction: Peer skills X team size 3	−0.365	−0.518**	−0.091	−0.232
	(0.282)	(0.258)	(0.253)	(0.232)
Total effect of peer skills in teams of 3	−0.040	−0.431**	−0.071	−0.050
	(0.241)	(0.215)	(0.209)	(0.172)
R^2	0.019	0.015	0.002	0.006

Levels of statistical significance:
*p < 0.1.**p < 0.05.***p < 0.01.*Notes*: Linear regressions without control variables are used. The dependent variable in the specification (1) is the total number of output units in work period one. The dependent variables in specifications (2) to (8) are based on responses to survey questions regarding teamwork. The question battery starts with: "What is your view regarding these statements about the previous teamwork?" The statements are:

- "I feel monitored when working on the task." (2)
- "I pay attention to the others while working on the task." (3)
- "I feel distracted from working on my task." (4)
- "I experience a competitive situation." (5)
- "I feel the pressure to succeed." (6.)
- "I feel uncomfortable when I have to work with strangers." (7)
- "I think it's problematic that not everyone is making an effort." (8).

The response scale ranges from "Totally disagree" (score: 1) to "Totally agree" (score: 7). Team size 3 is a variable that is 1 if the team size is three and 0 if the team size is two. Peer skills are a variable ranging from −2 ("very poor skills") to 2 ("very good skills"). Columns 1 to 8 show first the results for separate interaction effects (between peer skills and team size) and second, in the lower row, the total interaction effects for peer skills in teams of larger size (as a combination of the peer-skills effects and the separate interaction effect). The full sample (N = 271) is used. Clustered standard errors (at the session level) are in parentheses.

performance. This informs us about a possible role of transparency based on the idea that estimating peer performance could be more difficult if there are more team members, which then might imply differences in peer pressure and thus in the way that peer effects occur across team size. In the following, we exploit the fact that participants were informed about the team performance after work period one, whereas they were asked in a survey beforehand to estimate both their individual performance level and the average team performance level. This allows us to examine the magnitudes of estimation errors by determining (absolute) differences between estimates and actual performance levels.

Figure 5 illustrates via bars how performance estimation errors differ across team size. In the left illustration, the focus is on performance estimation errors at the individual level, which reveals no significant differences in errors when estimating one's own performance across team size (2.04 vs. 1.73, p 0.147, t-test). In comparison, the right illustration shows that it is significantly more difficult to correctly estimate the average team performance in teams of three compared to teams of two, as the estimation error goes up from 6.56 to 10.22 (p 0.000, t-test).[24] Accordingly, peer performance levels appear less clear when team size increases, which aligns with the idea that less transparency regarding the performance of others reduces peer pressure.

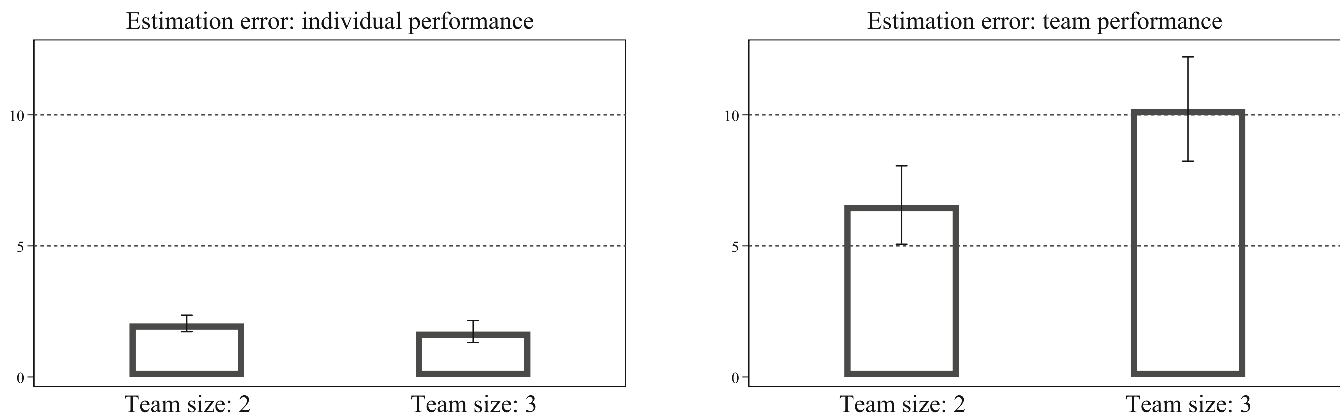


FIGURE 5 Team size and performance estimations. *Notes*: bars show average estimation errors based on absolute deviations between performance estimates and actual performance in the first work period by team size. Illustrations on the left show estimation errors for the individual performance level. Illustrations on the right show estimation errors for the average team performance level. The task performance measure is the number of output units used for performance pay after a spot check on correctness was conducted during sessions. 95% confidence intervals are displayed. Total sample size is 271 observations.

4 | CONCLUSION

4.1 | Summary and limitations of the study

Our analysis informs both scholars and practitioners, as we conclude that increasing team size does not necessarily decrease the performance of team members and that peer effects could play a role in this finding. Our evidence indicates that peer skills affect performance and thus can compensate for disincentives linked to the free-rider problem. Vice versa, this means that if peer skills are low, free-riding is a possibility, leading to worse outcomes when team size increases. Therefore, it makes sense for personnel management to ensure that positive peer effects are established in work teams in order to prevent free-riding, which is particularly important when teams are expanded to include low-skilled workers. As another lesson learned, our study points to possible adverse effects when high-skilled peers pressure co-workers too much, leading to negative peer effects among team members under (peer) pressure.

Our experimental study has several limitations, including the short-run nature of our investigation, which per se might limit the occurrence of free-riding. However, since we do observe free-riding in the low-skilled-peers scenario, we believe that our main finding of free-riding being limited via peer effects could be generalizable to long-run settings. The issue of generalizability rather seems to be whether peer effects are also effective in the long run, which could be the case according to research in field settings, such as Mas and Moretti (2009).

One major limitation of our experimental study based on a controlled work environment is that we are limited to teams of either two or three members. From various angles, this could be seen as a very relevant comparison, as we argue in the introduction of our paper (see Section 1). Nevertheless, an important question arises about whether and how our findings are informative about workplace setups with larger numbers of team members. From an economic viewpoint, one would expect that the phenomenon of free-riding becomes increasingly relevant in increasingly larger teams. While this is consistent with our finding of a negative team-size effect observed when peer skills are low, our results also indicate that the phenomenon of peer effects could become more relevant when team size increases. Considering, for example, the role of social ties, it might be plausible to assume that the incentives to free-ride increase in larger teams if team members know each other well; at the same time, the potential for peer effects as a way to mitigate free-riding could also be increasing. Hence, the question is which one of the two phenomena—free-riding or peer effects—dominates. Arguably, this depends on a variety of factors, such as the level of peer skills or the nature of the work task; thus, the question becomes increasingly context-specific and subject to the characteristics of the workplace setup, implying a need for further research on the topic of large teams.

4.2 | Contributions and outlook

With our study, we contribute to several strands of ongoing research. First and foremost, we add new findings to the research on teams, which in economics often focuses on team incentives (Corgnet et al., 2015; Delfgaauw et al., 2020; Englmaier et al., 2018; Friebel et al., 2017; Hamilton et al., 2003) but pays less attention to the role of team size for workplace performance. We contribute to this research by providing causal evidence from a controlled experimental comparison of workplace behavior across teams of either two or three members.

Apart from this novel contribution, we believe that our findings fit well with recent non-experimental results in the research on team size and peer effects, building upon the seminal contribution by Kandel and Lazear (1992). In a comprehensive study of teams in a large transportation company, Steinbach and Tatsi (2018) find that peer effects interact with team size, which aligns with our finding from a

controlled setting with random assignment of teams. In another related study, Backes-Gellner et al. (2015) analyze data on start-up teams to empirically test hypotheses derived from a theoretical model. By focusing on the role of team size in explaining differences in effort, they extend the work of Kandel and Lazear (1992) and argue that peer effects may offset disincentives in larger teams. However, they also point out that peer pressure may not counterbalance free-riding in increasingly large teams, while social ties could be an important factor in this entrepreneurial context.

Finally, we contribute to the growing research on peer effects at work (Batheja, 2022; Buechel et al., 2018; Cornelissen et al., 2017; Gerhards & Gravert, 2020; Van Veldhuizen et al., 2018), where researchers are still looking for explanations regarding why such psychological phenomena occur in the first place. According to Beugnot et al. (2019), competitive rivalry could play an important role in the occurrence of peer effects. In this respect, our evidence suggests a potential for negative peer effects due to perceived competition and corresponding pressure to succeed. Notably, there is evidence on negative peer effects in specific settings, be it in sports (Emerson & Hill, 2018) or in an educational context (Brady et al., 2017), but when and why this occurs is still rather unclear. Some researchers discuss the idea of a ‘choking-under-pressure’ phenomenon (Baumeister, 1984; Dohmen, 2008), which is also discussed in the context of peer pressure at work (Georganas et al., 2015). Bellemare et al. (2010) argue that reduced self-motivation explains negative effects owing to very high levels of peer pressure. We contribute to this research by providing further evidence of adverse effects of high peer pressure in teams and by showing the importance of transparency regarding the behavior of others as a driver of peer pressure (Mohnen et al., 2008). Thereby, we also inform the debate on performance feedback as a source of peer effects (for a recent review, see Villeval, 2020), where researchers discuss possible ‘negative quality peer effects’ (Eriksson et al., 2009). While it is certainly a worthwhile objective for future research to continue the discussion on negative peer effects and the underlying mechanisms, the same is true for studying the role of free-riding and peer effects in other workplace settings with a larger number of team members.

ACKNOWLEDGMENTS

The authors are grateful for comments by Laszlo Goerke, two anonymous referees, and participants of the AKempor conference in Salzburg 2018, the team dynamics and peer effects workshop in Trier 2018, the (virtual) conference of the European Association of Labour Economists 2021 as well as seminars in Konstanz and Trier. Martin Amann, Fabian Bührmann, Sam Butterick, Jonas Feld, Alexander Goldmann, Jessica Halle, Benjamin Hattemer, René Heinitz, Tom Hitzler, Johanna Kluge, Simon Kleinert, Lisa Nagel, Ryan O'Leary, Ruth Regnauer, Dominic Reichert, and Knut Schumach provided excellent research assistance. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

Data are available upon request from the corresponding author.

ORCID

Adrian Chadi https://orcid.org/0000-0003-2008-0653

ENDNOTES

[1] Note that we follow other research on the determinants of team performance (Babcock et al., 2015) by minimizing the role of team factors, such as task complementarities, as described in the textbooks on personnel economics. Arguably, doing so solves the measurement problem to identify individual performance levels in teams and also helps uncovering more general effects as well as the underlying mechanisms, as any such team factor could not only influence the optimal team size but may also interact with other variables of interest. The same holds for social ties (Bandiera et al., 2010), as another potentially relevant factor that we attempt to rule out by design.

[2] As part of the instructions, the session host randomly (with 50% probability) informed the two participants at the beginning of a transformed session that a third person did not show up. Although our analyses of this manipulation based on survey data and performance data do not support the idea of psychological effects, this could be attributable to our decision to implement a rather weak manipulation with just one statement.

[3] To check whether the procedure to organize the sessions was in any way revealing regarding our experimental manipulation, we asked the research assistants to carefully and comprehensively take notes on any striking features regarding the participants' behaviors. In particular, when being paid out, participants were asked for feedback concerning the session before. No indication of awareness regarding the manipulation was reported. Furthermore, during the subsequent summer semester of 2017, we invited all participants to a brown bag seminar where we presented the first results of the study. In this seminar, attended by more than 50 participants, not one person answered the question in the affirmative regarding awareness of the manipulation of team size.

[4] In a separate treatment condition, which is not part of the dataset analyzed here, a modified version of the game was used for a different purpose, as part of the above-mentioned parallel experiment. Accordingly, there are also a few other features of the procedure, such as some of the survey questions on subjective perceptions, which are not of prime interest for our investigation into peer effects and the role of team size here, and, hence, rather serve the purpose of distracting the participants from these very objectives.

[5] It turned out that there was no case in which a participant's pay had to be reduced due to incomplete surveys or due to sanctions, which were announced as a possible response to violations of the ban on verbal communication.

[6] Cases of zero performance were rare (see Figure B1). Note that spot checks were possible thanks to an internal network connecting every computer in the room and allowing a review of work output during the session.

[7] As an example, the module included one item to test whether team members are more or less likely to identify themselves with their team. This refers to the discussion on the role of team identity in preventing shirking (Eckel & Grossman, 2005), which may interact with team size.

[8] Females were on average more likely to take part as study participants, and we decided to not intervene in this natural selection process, for example by conducting stratification regarding gender. We hence did not ask participants about their gender during webpage registration. Instead, a research assistant noted the gender in the session records based on observable information during the session. Checking for measurement errors in this variable, we cannot find evidence in support of this in deeper inspections of the data.

[9] To learn more about possible selection effects, we analyze the available data from the webpage registration, which yields no evidence of

selectivity regarding individuals not showing up. Using the same data, we conduct another analysis by comparing individuals as originally assigned to either two-person sessions or three-person sessions, which reveals no evidence indicating imperfect randomization.

[10] Note that the free-rider hypothesis may be more plausible in settings where peer effects through presence of other team members are ruled out by design. For example, in research on teams outside of economics, Mao et al. (2016) study virtual teams and find evidence for free-riding in effort when team size increases.

[11] For robustness checks, we conduct regression analyses with and without control variables where we also compare results across the full sample ($N = 271$) and the dataset without transformed sessions ($N = 195$). In line with Figure 2, Table B4 shows that for both work periods one (Panel A) and two (Panel B), the average team-size effect is close to zero, which is also true when we aggregate the performance across both work periods (Panel C).

[12] For example, based on the team-size effect for the aggregate performance levels (23.11 vs. 23.18 output units), we are able to rule out an average performance reduction by two output units per period with a p-value of 0.000. To provide another perspective, we can conduct an ex-post power analysis to determine the hypothetical sample size that would yield statistical significance at the 5% level. Using again the aggregate performance levels and assuming a similar standard deviation as in our performance data, the analysis reveals that more than 50,000 observations are required to yield a statistically significant performance effect of the team-size increase.

[13] Note that we lose four observations when we analyze the software-based performance indicators, given that the program needs the original work output units (i.e. digital files), which were not stored properly in these few cases.

[14] According to Table B1, the perception of an unfair situation increases when comparing the responses after the bonus game (Panel B) to those before (Panel A). Regarding differences across team size (columns 1 and 2), we find that two-member teams report less unfairness after the bonus game, compared to three-member teams (1.75 vs. 2.24, $p = 0.017$, rank-sum test). This gap is somewhat larger in sessions when bonuses were paid out (columns 5 and 6). Further evidence comes from a question on the perception of being treated unfairly. Responses show the highest average score in teams of three when bonuses are paid out (column 6), which is driven by those who did not get the bonus in the unequal pay scenario. These participants report stronger perceptions of being treated unfairly compared to participants without bonus in two-member teams (1.69 vs. 2.80, $p = 0.015$ rank-sum test).

[15] Arguably, the random bonus can be seen as a technical feature in our experimental setup, which is not capable of inducing a behaviorally relevant psychological manipulation, as observed in other research on the consequences of pay inequality (e.g. Cohn et al., 2014). As an alternative idea for future research, a randomized treatment where the ratio between performance and pay changes for some team members could be more promising in this regard.

[16] Apart from the above-discussed changes in fairness perceptions after work period two, we find no significant differences across team size in Table B1, which is also true for perceptions reported after work period one. Average scores are very similar for each channel, including free-rider motives, where several thousand observations would be required for a statistically significant effect of the team size increase according to our ex-post power analysis.

[17] Furthermore, three-member teams seem to report being less satisfied with communication, which is interesting given that there was a ban on verbal communication for all teams. This finding could point to a stronger desire to chat in what might be a more convenient atmosphere if more people are present, but we are cautious with interpretations given that our checks do not confirm this result as robust.

[18] In the dataset without transformed sessions ($N = 195$), the average probability of assessing the work with peers positively increases from 33.33% in teams of two to 54.90% in teams of three ($p = 0.098$, rank-sum test).

[19] In the dataset without transformed sessions, we even find a statistically significant negative relationship between peer performance and own performance in two-person teams in work period one (coefficient: -0.446, $p = .003$), while the correlation remains negative but becomes insignificant in work period two (coefficient: -0.054, $p = 0.736$).

[20] Examining task performance across both work periods, we observe a work output of individuals with high computer skills (defined as above-zero scores) that amounts to 25.91 output units in the full sample, which is 5.34 more than we observe for the remaining individuals. This implies a performance improvement of roughly 25% linked to better skills. To check whether computer skills are reported in an unbiased fashion, we analyze the link between self-reported skills and skills as reported by the teammates. Neither do we find any evidence of a significant link between these two variables, nor do we detect any significant interaction with team size. To fully rule out concerns regarding the validity of the skills measure, we also make use of data from the test run to establish a separate performance indicator. Here, peers could be faster or slower when preparing the test article, which we identify by using second-exact time stamps from the original digital files. We cannot find that this alternative indicator of others' performance is related to one's own reported computer skills, and we again do not find any significant interaction with team size.

[21] We confirm our main finding in several additional robustness checks, which include regression analyses based on the dataset without transformed sessions (see Table B8). For another check, we exclude individuals above the age of 40. Furthermore, we exclude a few individuals who reported in the survey on knowing a member of their team. We asked participants about this to check for a potential role of social ties within teams, despite efforts in the recruiting procedure to avoid that (see Appendix A1).

[22] The effect of team size becomes more pronounced ($p = 0.028$, rank-sum test) when we use the dataset without transformed sessions, as the average score of self-reported free-rider intentions is 1.11 in teams of two.

[23] When inspecting the link between pressure and performance, we find that individuals who report feeling high pressure to succeed (scores from 5 to 7 on the 7-point scale) produce 1.7 work output units less than the remaining individuals ($p = 0.062$, t-test).

[24] The median error goes up from 4 to 6, when team size increases, indicating that the finding is not caused by outliers. We can also employ log-transformations of the estimation-error variables without changing the results.

REFERENCES

Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *The American Economic Review*, *62*(5), 777 795.

Babcock, P., Bedard, K., Charness, G., Hartman, J., & Royer, H. (2015). Letting down the team? Social effects of team incentives. *Journal of the European Economic Association*, *13*(5), 841 870. https://doi.org/10.1111/jeea.12131

Backes-Gellner, U., Werner, A., & Mohnen, A. (2015). Effort provision in entrepreneurial teams: Effects of team size, free-riding and peer pressure. *Journal of Business Economics*, *85*(3), 205 230. https://doi.org/10.1007/s11573-014-0749-x

Bäker, A., & Mertins, V. (2013). Risk-sorting and preference for team piece rates. *Journal of Economic Psychology*, *34*, 285 300. https://doi.org/10.1016/j.joep.2012.10.010

Bandiera, O., Barankay, I., & Rasul, I. (2010). Social incentives in the workplace. *Review of Economic Studies*, *77*(2), 417 458. https://doi.org/10.1111/j.1467-937X.2009.00574.x

Batheja, D. (2022). *Gender peer effects in the workplace: A field experiment in Indian call centers*. Mimeo.

Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, *46*(3), 610 620. https://doi.org/10.1037/0022-3514.46.3.610

Bellemare, C., Lepage, P., & Shearer, B. (2010). Peer pressure, incentives, and gender: An experimental analysis of motivation in the workplace. *Labour Economics*, *17*(1), 276 283. https://doi.org/10.1016/j.labeco.2009.07.004

Ben-Hafaïedh, C., & Cooney, T. M. (Eds.). (2017). *Research handbook on entrepreneurial teams: Theory and practice*. Edward Elgar Publishing. https://doi.org/10.4337/9781784713263

Beugnot, J., Fortin, B., Lacroix, G., & Villeval, M. C. (2019). Gender and peer effects on performance in social networks. *European Economic Review*, *113*, 207 224. https://doi.org/10.1016/j.euroecorev.2019.01.002

Bonein, A. (2018). Peer pressure and social comparisons with heterogeneous ability. *Managerial and Decision Economics*, *39*(2), 142 157. https://doi.org/10.1002/mde.2876

Brady, R. R., Insler, M. A., & Rahman, A. S. (2017). Bad company: Understanding negative peer effects in college achievement. *European Economic Review*, *98*, 144 168. https://doi.org/10.1016/j.euroecorev.2017.06.013

Buechel, B., Mechtenberg, L., & Petersen, J. (2018). If I can do it, so can you! Peer effects on perseverance. *Journal of Economic Behavior & Organization*, *155*, 301 314. https://doi.org/10.1016/j.jebo.2018.09.004

Büyükboyaci, M., & Robbett, A. (2019). Team formation with complementary skills. *Journal of Economics and Management Strategy*, *28*(4), 713 733.

Carpenter, J., Bowles, S., Gintis, H., & Hwang, S. H. (2009). Strong reciprocity and team production: Theory and evidence. *Journal of Economic Behavior & Organization*, *71*(2), 221 232. https://doi.org/10.1016/j.jebo.2009.03.011

Carpenter, J. P. (2007). Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, *60*(1), 31 51. https://doi.org/10.1016/j.geb.2006.08.011

Chadi, A., & Homolka, K. (2022). Little lies and blind eyes experimental evidence on cheating and task performance in work groups. *Journal of Economic Behavior & Organization*, *199*, 122 159.

Chen, R., & Gong, J. (2018). Can self selection create high-performing teams? *Journal of Economic Behavior & Organization*, *148*, 20 33. https://doi.org/10.1016/j.jebo.2018.02.004

Chiocchio, F., Kelloway, E. K., & Hobbs, B. (Eds.). (2015). *The psychology and management of project teams*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199861378.001.0001

Cohn, A., Fehr, E., Herrmann, B., & Schneider, F. (2014). Social comparison and effort provision: Evidence from a field experiment. *Journal of the European Economic Association*, *12*(4), 877 898. https://doi.org/10.1111/jeea.12079

Corgnet, B., Hernán-González, R., & Rassenti, S. (2015). Peer pressure and moral hazard in teams: Experimental evidence. *Review of Behavioral Economics*, *2*(4), 379 403. https://doi.org/10.1561/105.00000040

Corgnet, B., Hernán-González, R., & Schniter, E. (2015). Why real leisure really matters: Incentive effects on real effort in the laboratory. *Experimental Economics*, *18*(2), 284 301. https://doi.org/10.1007/s10683-014-9401-4

Cornelissen, T., Dustmann, C., & Schönberg, U. (2017). Peer effects in the workplace. *American Economic Review*, *107*(2), 425 456. https://doi.org/10.1257/aer.20141300

Delfgaauw, J., Dur, R., & Souverijn, M. (2020). Team incentives, task assignment, and performance: A field experiment. *The Leadership Quarterly*, *31*(3), 101241. https://doi.org/10.1016/j.leaqua.2018.03.003

Dohmen, T. J. (2008). Do professionals choke under pressure? *Journal of Economic Behavior & Organization*, *65*(3 4), 636 653. https://doi.org/10.1016/j.jebo.2005.12.004

Eckel, C. C., & Grossman, P. J. (2005). Managing diversity by creating team identity. *Journal of Economic Behavior & Organization*, *58*(3), 371 392. https://doi.org/10.1016/j.jebo.2004.01.003

Emerson, J., & Hill, B. (2018). Peer effects in marathon racing: The role of pace setters. *Labour Economics*, *52*, 74 82. https://doi.org/10.1016/j.labeco.2018.03.008

Englmaier, F., Grimm, S., Schindler, D., & Schudy, S. (2018). *The effect of incentives in non-routine analytical teams tasks-evidence from a field experiment*. CESifo Working Paper No. 6903. CESifo Munich. https://doi.org/10.2139/ssrn.3164800

Eriksson, T., Poulsen, A., & Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, *16*(6), 679 688. https://doi.org/10.1016/j.labeco.2009.08.006

Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, *24*(1), 39 57. https://doi.org/10.1086/497818

Fischer, M., Rilke, R. M., & Yurtoglu, B. B. (2020). *Two field experiments on self-selection, collaboration intensity, and team performance*. IZA Discussion Paper No. 13201. IZA Bonn. https://doi.org/10.2139/ssrn.3590900

Frick, B. J., Goetzen, U., & Simmons, R. (2013). The hidden costs of high-performance work practices: Evidence from a large German steel company. *ILR Review*, *66*(1), 198 224. https://doi.org/10.1177/001979391306600108

Friebel, G., Heinz, M., Krueger, M., & Zubanov, N. (2017). Team incentives and performance: Evidence from a retail chain. *American Economic Review*, *107*(8), 2168 2203. https://doi.org/10.1257/aer.20160788

Georganas, S., Tonin, M., & Vlassopoulos, M. (2015). Peer pressure and productivity: The role of observing and being observed. *Journal of Economic Behavior & Organization*, *117*, 223 232. https://doi.org/10.1016/j.jebo.2015.06.014

Gerhards, L., & Gravert, C. (2020). Because of you I did not give up Peer effects in perseverance. *Journal of Economic Psychology*, *81*, 102316. https://doi.org/10.1016/j.joep.2020.102316

Griffith, B. A., & Dunham, E. B. (2014). *Working in teams: Moving from high potential to high performance*. Sage Publications.

Hamermesh, D. S. (2013). Six decades of top economics publishing: Who and how? *Journal of Economic Literature*, *51*(1), 162 172. https://doi.org/10.1257/jel.51.1.162

Hamermesh, D. S. (2017). Age, cohort and co-authorship: The statistics of collaboration. In *Collaborative research in economics* (pp. 65 93). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-52800-7 5

Hamilton, B. H., Nickerson, J. A., & Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, *111*(3), 465 497. https://doi.org/10.1086/374182

Herbst, D., & Mas, A. (2015). Peer effects on worker output in the laboratory generalize to the field. *Science*, *350*(6260), 545 549. https://doi.org/10.1126/science.aac9555

Holmstrom, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, *13*, 324 340. https://doi.org/10.2307/3003457

Isaac, R. M., & Walker, J. M. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *Quarterly Journal of Economics*, *103*(1), 179 199. https://doi.org/10.2307/1882648

Isaac, R. M., Walker, J. M., & Williams, A. W. (1994). Group size and the voluntary provision of public goods: Experimental evidence utilizing large groups. *Journal of Public Economics*, *54*(1), 1 36. https://doi.org/10.1016/0047-2727(94)90068-X

Kandel, E., & Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy*, *100*(4), 801 817. https://doi.org/10.1086/261840

Kölle, F. (2017). Affirmative action, cooperation, and the willingness to work in teams. *Journal of Economic Psychology*, *62*, 50 62. https://doi.org/10.1016/j.joep.2017.06.004

Kuhn, P., & Villeval, M. C. (2015). Are women more attracted to co-operation than men? *Economic Journal*, *125*(582), 115 140. https://doi.org/10.1111/ecoj.12122

Lazear, P., & Oyer, P. (2013). Personnel Economics. In R. Gibbons & J. Roberts (Eds.), *Handbook of organizational economics* (pp. 479 519). Princeton: Princeton University Press.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, *60*(3), 531 542. https://doi.org/10.2307/2298123

Mao, A., Mason, W., Suri, S., & Watts, D. J. (2016). An experimental study of team size and performance on a complex task. *PLoS ONE*, *11*(4), e0153048. https://doi.org/10.1371/journal.pone.0153048

Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review*, *99*(1), 112 145. https://doi.org/10.1257/aer.99.1.112

Mohnen, A., Pokorny, K., & Sliwka, D. (2008). Transparency, inequity aversion, and the dynamics of peer pressure in teams: Theory and evidence. *Journal of Labor Economics*, *26*(4), 693 720. https://doi.org/10.1086/591116

Newhouse, J. P. (1973). The economics of group practice. *Journal of Human Resources*, *8*, 37 56. https://doi.org/10.2307/144634

Salas, E., Tannenbaum, S., Cohen, D., & Latham, G. (Eds.). (2013). *Developing and enhancing teamwork in organizations: Evidence-based best practices and guidelines* (Vol. 33). Wiley & Sons.

Sausgruber, R. (2009). A note on peer effects between teams. *Experimental Economics*, *12*(2), 193 201. https://doi.org/10.1007/s10683-008-9209-1

Shrivastava, M., & Tamvada, J. P. (2011). Entrepreneurial teams, optimal team size, and founder exits. Mimeo.

Steinbach, D., & Tatsi, E. (2018). Peer effects, free-riding and team diversity. Mimeo.

Van Veldhuizen, R., Oosterbeek, H., & Sonnemans, J. (2018). Peers at work: Evidence from the lab. *PLoS ONE*, *13*(2), e0192038. https://doi.org/10.1371/journal.pone.0192038

Villeval, M. C. (2020). Performance feedback and peer effects. In *Handbook of labor, human resources and population* (pp. 1 38). Springer International Publishing. https://doi.org/10.1007/978-3-319-57365-6 126-1

Vranceanu, R., El Ouardighi, F., & Dubart, D. (2015). Team production with punishment option: Insights from a real-effort experiment. *Managerial and Decision Economics*, *36*(6), 408 420. https://doi.org/10.1002/mde.2678

Zhang, X. M., & Zhu, F. (2011). Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review*, *101*(4), 1601 1615. https://doi.org/10.1257/aer.101.4.1601

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Chadi, A., & Homolka, K. (2023). Under (peer) pressure: Experimental evidence on team size and task performance. *Managerial and Decision Economics*, *44*(7), 3769–3786. https://doi.org/10.1002/mde.3906