Experimental Investigations on Teamwork, Fundraising and Privacy Concerns

Inauguraldissertation zur Erlangung des Doktorgrades des Fachbereichs IV
– Wirtschafts- und Sozialwissenschaften, Mathematik und Informatikwissenschaften –
der Universität Trier

vorgelegt von
Konstantin Homolka

Trier, Mai 2021

Abstract

Teamwork is ubiquitous in the modern workplace. However, it is still unclear whether various behavioral economic factors de- or increase team performance. Therefore, Chapters 2 to 4 of this thesis aim to shed light on three research questions that address different determinants of team performance.

Chapter 2 investigates the idea of an honest workplace environment as a positive determinant of performance. In a work group, two out of three co-workers can obtain a bonus in a dice game. By misreporting a secret die roll, cheating without exposure is an option in the game. Contrary to claims on the importance of honesty at work, we do not observe a reduction in the third co-worker's performance, who is an uninvolved bystander when cheating takes place.

Chapter 3 analyzes the effect of team size on performance in a workplace environment in which either two or three individuals perform a real-effort task. Our main result shows that the difference in team size is not harmful to task performance on average. In our discussion of potential mechanisms, we provide evidence on ongoing peer effects. It appears that peers are able to alleviate the potential free-rider problem emerging out of working in a larger team.

In Chapter 4, the role of perceived co-worker attractiveness for performance is analyzed. The results show that task performance is lower, the higher the perceived attractiveness of co-workers, but only in opposite-sex constellations.

The following Chapter 5 analyzes the effect of offering an additional payment option in a fundraising context. Chapter 6 focuses on privacy concerns of research participants.

In Chapter 5, we conduct a field experiment in which, participants have the opportunity to donate for the continuation of an art exhibition by either cash or cash and an additional cashless payment option (CPO). The treatment manipulation is completed by framing the act of giving either as a donation or pay-what-you-want contribution. Our results show that donors shy away from using the CPO in all treatment conditions. Despite that, there is no negative effect of the CPO on the frequency of financial support and its magnitude.

In Chapter 6, I conduct an experiment to test whether increased transparency of data processing affects data disclosure and whether the results change if it is indicated that the implementation of the GDPR happened involuntarily. I find that increased transparency raises the number of participants who do not disclose personal data by 21 percent. However, this is not the case in the involuntary-signal treatment, where the share of non-disclosures is relatively high in both conditions.

Deutsche Kurzfassung
(German Abstract)

Teamarbeit ist in der modernen Arbeitswelt nicht mehr wegzudenken. Für verschiedene verhaltensökonomische Faktoren ist jedoch unklar, ob diese die Teamleistung an sich positiv oder negativ beeinflussen. Daher widmen sich die Kapitel 2 bis 4 dieser Dissertation drei Forschungsfragen, die sich mit unterschiedlichen Determinanten der Teamleistung befassen.

Kapitel 2 greift in einer Untersuchung die Frage auf, inwiefern ein ehrliches Arbeitsumfeld relevant für die Arbeitsleistung ist. In einer Arbeitsgruppe können zwei von drei Mitarbeitenden einen Bonus in einem Würfelspiel erspielen und während diesem durch eine Falschangabe unentdeckt betrügen. Entgegen der Annahme, dass Ehrlichkeit für die Arbeitsleistung ein wichtiger Faktor ist, stellen wir keine Leistungsminderung des dritten unbeteiligten Mitarbeitenden fest, wenn ein Betrugsverhalten stattfindet.

In Kapitel 3 wird der Effekt der Teamgröße auf die Arbeitsleistung untersucht. In einem Arbeitsplatzsetting bearbeiten dazu entweder zwei oder drei Personen eine realitätsnahe Arbeitsaufgabe. Unsere Hauptergebnisse zeigen, dass der Unterschied in der Teamgröße nicht die durchschnittliche Arbeitsleistung beeinträchtigt. In unserer anschließenden Diskussion über mögliche Mechanismen liefern wir Belege für fortwährende Peer-Effekte. Es zeigt sich, dass Peers, also Mitarbeitende, in der Lage sind, das potenzielle Trittbrettfahrerproblem abzuschwächen, das bei der Arbeit in einem größeren Team besteht.

In Kapitel 4 wird die Rolle der wahrgenommenen Attraktivität von Mitarbeiter/innen auf die Arbeitsleistung analysiert. Die Ergebnisse zeigen, dass die Arbeitsleistung umso geringer ist, je höher die wahrgenommene Attraktivität der anderen Mitarbeitenden ist, jedoch nur in Konstellationen unterschiedlichen Geschlechts.

Im folgenden Kapitel 5 dieser Arbeit wird die Auswirkung des Angebots einer zusätzlichen Zahlungsoption in einem Spenden-Kontext untersucht. Das anschließenden Kapitel 6 befasst sich mit Datenschutzbedenken von Probanden.

Im durchgeführten Feldexperiment in Kapitel 5 haben die Teilnehmenden die Möglichkeit, für die Fortsetzung einer Kunstausstellung entweder in bar oder in bar und mit einer zusätzlichen bargeldlosen Zahlungsoption zu spenden. Die experimentelle Manipulation wird dadurch vervollständigt, dass der Akt des Gebens entweder als Spende oder als „Zahle, so viel Du willst“ dargestellt wird. Unsere Ergebnisse zeigen, dass Spendende die Verwendung der bargeldlosen Zahlungsoption in allen Konditionen vermeiden. Dennoch gibt es keine negativen

Auswirkungen der bargeldlosen Zahlungsoption auf die Häufigkeit der finanziellen Unterstützung und deren Höhe.

In Kapitel 6 wird mithilfe eines Onlineexperiments getestet, ob eine erhöhte Transparenz des Datenverarbeitungsprozesses das Teilen von Daten mit dem Forschenden beeinflusst, sowie, ob sich die Ergebnisse ändern, wenn eine unfreiwillige Implementierung der DSGVO signalisiert wird. Als Ergebnis kann festgehalten werden, dass eine erhöhte Transparenz des Datenverarbeitungsprozesses die Anzahl der Teilnehmenden, die keine personenbezogenen Daten teilen, um 21 Prozent erhöht. Dies ist jedoch nicht der Fall, wenn die unfreiwillige Implementierung der DSGVO signalisiert wird. Hier ist der Anteil der Teilnehmenden, die nicht Daten teilen, in beiden Konditionen relativ hoch.

Danksagung

Für diese Dissertation waren eine große Portion Mut und Optimismus notwendig und es gibt eine Vielzahl von Personen, die mich in den letzten Jahren dabei unterstützt haben, Ideen zu Projekten werden zu lassen und aus diesen Projekten letztendlich diese Dissertation zu erstellen. An dieser Stelle möchte ich die Gelegenheit nutzen, diesem Personenkreis meinen allergrößten Dank auszusprechen.

An erster Stelle gilt mein Dank Professor Dr. Laszlo Goerke, der den Mut hatte, einem Masterabsolventen der Wirtschaftssoziologie eine Promotion in den Wirtschaftswissenschaften zuzutrauen und diese sowohl als Co-Autor als auch als Betreuender fortwährend durch präzise Kritik, ein hohes Engagement und auch dem Zugestehen des nötigen Freiraums zu unterstützen. Nicht weniger danke ich Professorin Dr. Katrin Muehlfeld für die Bereitschaft, diese Dissertation als Zweitprüferin zu betreuen. Zudem danke ich Professor Dr. Adrian Chadi, der als Haupt-Co-Autor unermüdlich den Fortschritt unserer gemeinsamen Projekte vorangetrieben hat und dabei nie seinen einmaligen Optimismus verloren hat. Außerdem danke ich Professorin Dr. Sabrina Jeworrek, meiner dritten Co-Autorin. Weiterhin wäre diese Arbeit nicht ohne die Zuarbeit von unzähligen Hilfskräften entstanden. Alle Namen finden sich jeweils am Anfang jedes Kapitels in der Fußzeile. Hier möchte ich kurz die wichtigsten nach chronologischem Erscheinen am Institut auflisten: Ruth, Knut, Tom, Lisa, René, Jessica und Dominic. Auch möchte ich den Kunstwissenschaftler/innen unter der Leitung von Professorin Dr. Ulrike Gehring danken, die uns im generator wahrscheinlich genauso staunend bei der Arbeit beobachtet haben, wie wir sie. Teils fachlich, teils freundschaftlich möchte ich all meinen lieben Kolleg/innen am Institut für Arbeitsrecht und Arbeitsbeziehungen in der Europäischen Union danken. Dies gilt insbesondere für Gabriel, Sven, Jonas; Natalia, Dominik, Katrin und natürlich Frau Mielke, aber auch alle weiteren, nicht genannten Kolleg/innen. Ihr habt ganz besonders dazu beigetragen, dass die Zeit am Institut eine ausgesprochen einmalige war.

Abseits davon möchte ich einen großen Dank an meine Eltern, meinen Bruder und meine liebste Laura aussprechen. Ohne Euch hätte ich nicht die Möglichkeit und die Kraft gehabt, zu erforschen, welche Regeln wir Menschen uns selbst auferlegen und welche uns auferlegt worden sind.

Contents

List of Tables

List of Tables in the Appendices

List of Figures

List of Figures in the Appendices

1 Introduction

In the modern workplace, almost every task requires cooperation. This also does not change in an increasingly digitalized working environment, where conference software applications and interactive work software are already an essential tool for everyday work life (Hofmann et al. 2020, Wolter et al. 2021). Additionally, the importance of working groups and teams is also reflected in the large embodiment of business guidebooks and public literature, which exist for this area.[1] Those strands of literature are loosely related to some classical but still puzzling personnel economic questions, such as ensuring that team performance remains continuously high or how to design efficient team wages. From an economic point of view, there are clear costs and benefits when forming teams at the workplace or implementing a team incentive scheme. On the one hand, free-riding becomes increasingly attractive for employees when only the entire team's performance is accountable for the wage determination and co-workers may still put enough effort into the combined work output to accomplish the job (Holmström 1982). This becomes an even more severe problem when shifting the focus on the principals' perspective, who cannot identify and penalize a free-rider due to the metering problem of individual performance within a team (Alchian and Demsetz 1972).[2] Thus, team performance may be disadvantageous compared to the performance of segregated workers. On the other hand, especially when it comes to the production of complex products, synergy effects due to complementarities, task specialization, or knowledge transfer between employees enhance team performance and make teamwork superior (Lazear 1995). Furthermore, even when the

[1] Lazear (1999), begins his classic „Globalisation and the Market for Team-Mates" with the statement "It is impossible to pick up a business publication these days without reading about the wonders of teamwork" (p. 15). Additionally, there are various recent publications focusing on different aspects of teamwork: Salas et al. (2013) illustrate teams within real-world scenarios from various angles, Chiocchio et al. (2015) highlight the psychological factors playing an important role for management and project teams, whereas Griffith and Dunham (2015) aim to transform "high-potential" groups into "high performing" teams and Ben-Hafaïedh and Cooney (2017) focus on entrepreneurial teams.

[2] Note that there are some approaches to solve the metering problem econometrically when co-workers have to switch between teams over time and team productivity is quantifiable (Müller and Upmann 2017, Bonhomme 2021).

complexity of work is relatively low as likewise the potential of synergy effects, implementing team incentives still increases performance as has been empirically shown in previous studies (Hamilton et al. 2003, Friebel et al. 2017). This may be due to behavioral factors, which come into play when forming teams and employees are able to cooperate, observe and react to each other's actions. For instance, Babcock et al. (2015) argue that guilt aversion, reciprocity, or altruism positively influence team performance. Though, for other factors, it is far from clear in the current state of the economic discussion, whether they are de- or increasing team performance. For instance, dishonest behavior within a team may spur the performance of the cheating co-worker but may lower the performance of a conscientious bystander in the team. In a larger team, the free-rider problem may become more severe compared to a smaller team, but at the same time, peer effects may counteract the free-riding effect due to mutual encouragement, which enhances performance in a larger team. Alternatively, the perceived attractiveness of a co-worker may distract or motivate an employee to put more effort into the work output. To investigate these three intriguing questions about team performance and its determinants in Chapters 2 to 4 of this thesis, Adrian Chadi and I use a realistic but well monitored workstation. As a result, we are able to measure team as well as individual performance and carefully analyze the causal relationship between team performance and the behavioral factors we manipulate. Thereby we rely on a real-effort task carried out by two or three co-workers individually, which is always compensated by a team piece rate. As has been stated by previous literature (Babcock et al. 2015), this facilitates capturing important behavioral team factors, which we have in mind to analyze, but prevents initiating uncontrollable team dynamics, which may result in idiosyncratic team behavior and an unobservable black box responsible for the overall team performance, for instance, due to growing social ties between co-workers (Bandiera et al. 2010). Supplementary carried out surveys during the experiment provide us with essential information about preferences and personal characteristics of participants as well as reveal the role of possible underlying mechanisms, which may explain the observed main result.

Further, these parts of this thesis are also an addition to research strands, besides classic team research addressing such fundamental issues as optimal team size (e.g. Backes-Gellner et al. 2015). First, the topic of cheating and dishonest behavior, which is one of the most growing topics in (experimental) economic literature (for reviews, see Rosenbaum et al. 2014, Jacobsen et al. 2018, Abeler et al. 2019, Garbarino et al. 2019) and has focused so far primarily on whether and how frequently individuals cheat, but also focusses increasingly on cheating

spillover effects (Pierce and Snyder 2008, Fosgaard et al. 2013, Innes and Mitra 2013). Second, the topic of peer effects and their impact on work productivity (for a review see Herbst and Mas 2015), including research dedicated to the positive (e.g. Kandel and Lazear 1992) and negative (e.g. Georganas et al. 2015) impact of peer pressure and lastly, the topic of perceived attractiveness (e.g. Baert and Decuypere 2014) and gender diversity at the workplace (e.g. Hoogendoorn et al. 2013). By combining those areas and embedding them in the greater frame of team research, this thesis offers interesting insights, which are presented in the following chapters.

Chapter 2, which is joint work with Adrian Chadi, initiates the empirical part of this thesis. As the first determinant of performance, we investigate cheating in work groups and assume that an honest workplace environment "leads to greater employee loyalty, higher productivity, and better organizational performance" (Losey et al. 2007). To do so, we analyze how dishonesty in a working group affects performance. Between two work periods, two out of three individuals have the opportunity to obtain a bonus in a dice game. By misreporting a secret die roll, cheating is allowed without the risk of exposure. Our focus lays particularly on the bystander's behavioral response as the potential witness to the dishonest action. To identify the implications of lies at work and not solely of income inequality, the bonus game rules were altered in the control treatment to randomly prevent cheating or not while holding the monetary consequences constant. In the analysis, we first estimate the mean lying rate and find that the opportunity-to-cheat is exploited in roughly 42% of cases. Contrary to claims on the importance of honesty at work, we do not observe a reduction in performance when cheating takes place, neither for the bystander nor for the whole team. Shifting the focus to possible underlying mechanisms and our survey data, bounded awareness comes into play, as we find substantial evidence for effect heterogeneity along the lines of information preferences. Bystanders with higher preferences for inconvenient information provide relatively low task performance compared to those with lower information preferences, who seem to turn a blind eye to their co-workers' dishonest action by putting increased effort into their work. In a follow-up experiment conducted after the end of the regular study, we find additional evidence for the relevance of our distinction between information preference types.

In **Chapter 3**, Adrian Chadi and I study behavior in a workplace environment where either two or three individuals perform the real-effort task that is compensated by a team piece rate. While this implies weaker incentives to provide effort in larger teams, the occurrence of peer effects could mitigate the free-rider problem and might interact with the number of team

members. To obtain a broad picture of the behavioral implications due to changes in team size, our setting informs us about individual performance levels as well as workplace perceptions in order to shed light on psychological mechanisms. The results show that an increase in team size is not harmful to team performance, which is a very robust finding, be it across performance dimensions and even when introducing pay inequality between two work periods. We find that positive performance spillovers from peer to peer, alleviate the free-rider problem when team size increases and that changes in peer pressure may help understand the heterogeneous nature of such peer effects. Contrary to discussions in previous research, however, our evidence on mechanisms underlying peer effects indicates that lower peer pressure in larger teams allows for more positive performance spillovers, whereas high-skilled peers in smaller teams seem to induce others to produce more mistakes instead of more work output in high quality.

Chapter 4 concludes the first part of this thesis, which is again joint work with Adrian Chadi. Here, previous research has highlighted the effect of self-attractiveness on labor market outcomes (e.g. Hamermesh and Biddle 1994), but has not highlighted the effect of perceived attractiveness of other co-workers on performance so far. Investigating this effect at the workstation with teams of two or three co-workers, we can show that the higher the perceived attractiveness of co-workers is in opposite-sex constellations, the lower the task performance is. Further, this drop in performance is especially noticeable in two-person scenarios. Here the individual performance level is reduced by roughly 15% if the opposite-sex co-worker is perceived as above-average attractive. Shifting the focus on potential mechanisms, we find no evidence that participants are distracted through social bonding activities with attractive co-workers. Instead, underperformance due to cognitively demanding self-regulation seems to be highly relevant to explain the performance drop found, which is reflected in increased emotional arousal.

Chapter 5 and 6 of this thesis contribute to two different research strands: One analyzing individual behavior in a fundraising context. The other investigating privacy concerns when it comes to sharing personal data with the researcher. Whereas the analysis of Chapter 5 relies on data collected in a controlled field experiment, Chapter 6 analyzes data coming from an online experiment. Having said this, in both cases the chosen experimental setup relies very much on the subject of research.

In research on fundraising the most important aspect for the researcher is to gather data of realistic donor behavior in a non-artificial environment, which is the case in a field experiment (Harrison and List 2004). Although this comes with the advantage of a high external validity of

the findings, it is also stigmatized by a low internal validity. Hence, it is essential to compare the results with other (field-)studies focusing on the same issue and reaching, by replicating outcomes in the best case, a more generalizable finding. In Chapter 5, we partly aim to do so by investigating whether offering an additional payment option in a fundraising context has the potential to raise donations. As has been shown in previous research, offering an additional technology to attract supplementary donors may come with the risk that individuals stick with the well-established cash contributions or even shy away from donating at all (Soetevent 2011). However, whether this finding is unique and which behavioral aspects are responsible for this kind of avoidance behavior is still unclear and worthwhile to investigate.

In the case of analyzing privacy concerns, research so far focuses mainly on behavior in digitalized environments (for reviews see Goldfarb and Tucker 2019, Tucker 2016). This is in line with the intention of legislators, such as the GDPR, which aims to adopt data protection to technological developments (GDPR 2016, Article 6). However, when confronting the GDPR and the mechanisms, it requires businesses and individuals to apply to, with previous research, the findings hint toward difficulties those mechanisms may cause. For example, a higher perception of control over personal data can lead to even more careless data sharing (Brandimarte et al. 2013), or privacy concerns increase when data processing is highlighted regardless of whether the actual threat for data disclosure is high or low (Marreiros et al. 2017). Thus, to take a closer look at how the willingness to share personal data with a researcher is affected by the mechanisms the GDPR suggests, I conduct an online experiment. Furthermore, I aim to shed light on the decision process, which causes participants to disclose their personal data or not.

In **Chapter 5**, which is joint work with Adrian Chadi, Laszlo Goerke, and Sabrina Jeworrek, we follow the assumption that people who are asked face-to-face to contribute to charity are deterred from donating when presenting a cashless payment option (CPO). We investigate this phenomenon in more depth in a field experiment in the context of an art exhibition at a German university. As a first treatment, we randomly manipulate the opportunity to donate, either offering a CPO or not. Additionally, we expand our setting by implementing two pay-what-you-want (PWYW) treatments (with and without CPO) and, therefore, ask visitors at the end of the tour to compensate for the service of the guided tour instead of the art exhibition as such, which is the case in the donation treatments. Through this expansion, we are able to evaluate whether the findings are special to a donation context or also hold for a different fundraising approach. Supplementary, we ask visitors, after taking the guided tour, to fill out a feedback

survey. In line with Soetevent (2011), our results show that donors mainly shy away from using the CPO in the donation as well as the PWYW conditions. However, we do not observe a negative effect of the CPO, neither on the frequency of financial support nor on its magnitude. By analyzing subgroups of visitors, we find that the number of donations even increases when the CPO is offered for a substantial part, who are frequent debit card users. Nevertheless, donations still are given in cash and not by card.

In **Chapter 6**, I follow the assumption that scientific research is already committed to high data protection standards. Thus, I claim that complying with the General Data Protection Regulation (GDPR) and fostering the transparency of data processing should not impair the willingness of research participants to share personal data. In my ensuing investigation, I conduct an online experiment, where participants consent at the beginning to one out of four data protection pages. These pages vary in how explicit they display the data processing rule (increased transparency or not) and how voluntary the implantation of the GDPR was executed by the researcher (voluntary or involuntary). Afterwards participants fill out a questionnaire and decide whether they want to participate in an incentivized lottery and therefore have to disclose personal data. In the analysis, I find, in contrast to the expectation that increased transparency raises the number of participants who do not disclose personal data by 21 percent. However, this is not the case in the involuntary-signal treatment, where increased transparency of data processing does not further raise the share of non-disclosures. Furthermore, I show that participants base their disclosure decision in both treatments when data processing transparency is low, on the time they invest into the survey as well as the hypothetical value of time they state within the survey. Since previous research has struggled to identify a clear link between the stated hypothetical value of personal data and actual personal data sharing (for a review see Kokolakis 2017), the stated hypothetical value of time and actual time spent could be a valuable proxy to consider in future research settings to predict privacy decision-making.

Finally, **Chapter 7** provides some concluding remarks on the most important insights inferred from the preceding chapters and contains limitations and an outlook on future research needs.

2 Little lies and blind eyes – Experimental evidence on cheating and task performance in work groups*

We investigate cheating in work groups. Three individuals receive team-based performance pay for executing a real-effort task, but only two can obtain a bonus in a dice game, which allows cheating without exposure by misreporting a secret die roll. We find that the opportunity-to-cheat is exploited in roughly 42% of cases. We do not observe a reduction of bystander performance when cheating takes place. Bounded awareness could be an explanation, as we find substantial evidence for effect heterogeneity along the lines of information preferences. Bystanders with higher preferences for inconvenient information provide relatively low task performance, compared to those with lower information preferences, who seem to turn a blind eye to the dishonest action of their co-workers by putting increased effort into their work.

*This chapter is joint work with Adrian Chadi. A version of this chapter is currently submitted to the Journal of Economic Behavior and Organization. We are grateful for comments by Laszlo Goerke, Alberto Palermo as well as the participants of the TIBER symposium in Tilburg, the European ESA conference in Vienna, the 13th NCBEE conference in Odense, the RES conference 2019 in Warwick, the 31st EALE conference in Uppsala, the 2020 meeting of the German Economic Association, the UNI-GR meeting in Trier, the IAB/ZEW Workshop on Human Resource Management in Nuremberg, the IAAEU Workshop on Labour Economics, and seminar participants at universities in Hamburg, Konstanz and Trier. Martin Amann, Fabian Bührmann, Sam Butterick, Jonas Feld, Alexander Goldmann, Jessica Halle, Benjamin Hattemer, René Heinitz, Tom Hitzler, Simon Kleinert, Lisa Nagel, Ryan O'Leary, Ruth Regnauer, Dominic Reichert, and Knut Schumach provided excellent research assistance.

"[Honesty is] a key leadership strategy within an organization, resulting in greater trust and loyalty between managers and staff in a firm.
(S. Gaffney and C. Francis (2009) in 'Honesty Sells: How to Make More Money and Increase Business Profits')
"Most lies succeed because no one goes through the work to figure out how to catch them."
(P. Ekman (2009) in 'Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage')

2.1 Introduction

Lies are ubiquitous. Be it in politics, media, or sports, in numerous contexts individuals decide to disguise the truth. The workplace is no different, especially since even little lies can be tempting when there is a chance to improve economic outcomes by being dishonest. However, many employees abstain from such behavior, even when detection is unlikely. Apart from one's ethical considerations, an employee who stays honest complies with norms and rules, such as ethics codes, through which companies explicitly instruct their workforce to be honest.[3] This policy promises to prevent fraudulent actions, but it could also be serving as a positive factor in performance. In this regard, researchers of human resource management claim that an "honest workplace" is economically beneficial, as it "leads to greater employee loyalty, higher productivity, and better organizational performance" (Losey et al. 2007). While this may sound intuitive, empirical evidence for such benefits of promoting honesty at work is missing so far.

Consider the case of a working group with several employees who are encouraged by their company to put in effort through team-based piece rates. Research by personnel economists shows that such team incentives are able to increase performance (for reviews, see Lazear and Oyer 2013, Kocher et al. 2020). The researchers typically explain this by referring to psychological drivers of performance and argue, for instance, that individuals perform well in teams because of social pressure to put in effort for others (Babcock et al. 2015).[4] In consequence, the question arises whether employees still feel pressured to work hard for their team when the colleagues turn out to be dishonest individuals.

In this paper, we investigate dishonesty in work groups, for the first time regarding both its occurrence and its potential consequences for performance. To do so, we build upon a growing research field on cheating (for reviews, see Rosenbaum et al. 2014, Jacobsen et al. 2018, Abeler et al. 2019, Garbarino et al. 2019). Researchers in this field often analyze misreporting of

[3] For example, Barclays postulates having "zero tolerance" in their Compliance Code towards "dishonest actions perpetrated by bank employees." Goldman Sachs states in their Code of Business Conduct and Ethics: "Integrity and honesty are at the heart of our business." Daimler puts the goal of integrity explicitly into a team context ("integrity is trust into the team").
[4] Since the seminal contribution by Kandel and Lazear (1992) and even earlier contributions, economists discuss psychological motives of team performance, such as team spirit and loyalty, which according to Alchian and Demsetz (1972) could be enhanced by a "moral code of conduct" (p. 791).

outcomes in games to capture peoples' lies, an example of this is the die-roll paradigm by Fischbacher and Föllmi-Heusi (2013). Accordingly, individuals get the opportunity-to-cheat by misreporting a secret die roll, whereas the researcher then can prove the occurrence of dishonest behavior ex-post by aggregating reported numbers across individuals and checking for the mathematical plausibility. This idea is increasingly used in laboratory experiments to study cheating in groups (Conrads et al. 2013, Cadsby et al. 2016, Korbel 2017, Soraperra et al. 2017, Kocher et al. 2018) and its possible determinants, such as gender (Muehlheusser et al. 2015).

As a first contribution, we expand the research on cheating by setting up a team scenario of three individuals working together, in which only two benefit from cheating in a dice game. Our design ensures that the third individual suffers no economic disadvantage from cheating and hence, as a bystander, only plays the role of a possible witness to the dishonest act. Expecting that the two bonus players are tempted to increase their payoffs by exploiting the opportunity-to-cheat, our setting provides evidence on the performance effects of cheating when we focus on the behavioral response of the bystander. This second contribution is novel, given that research on cheating is rather silent on the possible implications of dishonest actions, i.e. when cheating is the determinant and not the outcome.[5] As our third contribution, we explore mechanisms underlying differences in performance and analyze heterogeneity in the effects of cheating. Thereby, we reveal the potentially important role of information preferences in behavioral responses to dishonest activities.

In fact, a growing amount of research suggests that individuals often ignore unpleasant information, which can be a deliberate decision or not. In the organizational behavior literature, researchers discuss 'bounded awareness' as a possible explanation for failures to recognize unethical actions (Bazerman and Sezer 2016). An example is the Bernie Madoff Ponzi scheme, which had financial returns that could hardly be explained by anything other than cheating, yet many turned a blind eye to it. In this case of likely (but not 100% verifiable) cheating, many individuals might prefer to believe what they consider a more desirable perception of reality. Related literature describes how people respond to inconvenient information, depending on whether they have high preferences for knowing the unpleasant or whether they prefer behaving like 'ostriches' (Golman et al. 2017). Assuming that, in the case of dishonesty among co-

[5] An exception is the research about possible spill-over effects of cheating on others' willingness to cheat (for a recent example and a literature review, see Benistant et al. 2021). Furthermore, Gawn and Innes (2018) show that lies by one person harm another person's trust, whereas Ederer and Fehr (2017) reveal that the effectiveness of performance incentives is undermined by principals who provide agents with dishonest performance appraisals. We are not aware of any study on the performance effects of cheating as practiced by co-workers.

workers, many employees prefer putting their heads in the sand, implications of cheating may differ across individuals with higher versus lower information preferences.

To gather causal evidence on the behavioral effects of dishonesty in work groups, a researcher faces several challenges in designing a research environment. First, experimentally manipulating the level of honesty in a workplace context without using deception via trained actors is possible, but only by offering individuals the opportunity-to-cheat and then relying on expectations of their decisions, which are not perfectly controlled. Second, disentangling the effect of dishonesty per se from other consequences, such as changes in relative income, requires a control scenario that provides the same payoff structure as with cheating. Third, endogenous team formation is an increasingly studied phenomenon (Kocher et al. 2006, Bandiera et al. 2013, Chen and Gong 2018, Büyükboyaci and Robbett 2019, Fischer et al. 2020), which however could modify the effects of dishonest behaviors by team members. To establish generalizable evidence, it is useful having control over the process determining team composition, so that selectivity, such as social ties between team members, can be prevented. While all these considerations suggest running a laboratory experiment, one could argue that a highly artificial setting may not be suitable for the goal of studying psychological mechanisms at work. Consequently, we present findings from an experiment in a controlled but authentic workplace environment, which we set up to allow behavioral responses to actions carried out by participants who are randomly given the opportunity-to-cheat (or not).

The setting of our experimental investigation is as follows. As part of a large-scale study project at a German research institute, several hundred university students are recruited to work in a computer workstation for approximately an hour. In groups of three, they carry out a routine task conducted by student assistants employed at the institute. Their incentive for performing the task is a team-performance-related pay that increases for all participants the higher the individual work output is, resulting in the same wage for all three. In addition to this wage component, two randomly chosen participants can win a monetary bonus via a dice game.[6] The bonus game takes place between the two work periods and we use it to implement our experimental manipulation. In the 'opportunity-to-cheat' treatment, the two bonus players receive a bonus if they report a perfect match after rolling their 20-sided die. In the randomly assigned 'control' treatment, the two bonus players also have the opportunity to win, allowing

[6] For experimental research on bonus-based compensation schemes with random components, see Gill et al. (2013) who point out that employees in companies often receive payments that are not determined by performance but rather by factors beyond the control of the individual.

us to observe the same pay inequalities, but, contrary to the other treatment, we prevent cheating by slightly altering the rules of the game. To address possible effect heterogeneity in our comprehensive analysis of behavioral responses to cheating, our setup allows for collecting survey data on traits, including preferences for inconvenient information. The survey also provides us with information on subjective perceptions to explore the mechanisms underlying individual responses to the dishonest actions of others at the workplace.

Our results are as follows. First, participants in the role of the decisive bonus player can be divided into two large groups, cheaters and non-cheaters, as we estimate an average lying rate in the opportunity-to-cheat treatment of 41.75%. Individual characteristics explain some of the variation in cheating, as, for example, men are twice as likely to become a 'partner in crime' compared to women. Second, we do not find evidence for detrimental impacts of cheating on performance. This is true for team performance when the opportunity-to-cheat is provided and for the bystander's performance when cheating takes place. Third, evidence on effect heterogeneity reveals that participants respond differently to cheating based on information preferences. Specifically, bystanders with lower preferences for inconvenient information react with comparatively high task performance and seem to turn a blind eye to the dishonest action of their co-workers, contrary to bystanders with higher information preferences. As an explanation of this finding, our survey evidence supports the idea that the ostriches benefit from putting in extra effort into their work, as it helps them deal with negative experiences. Further, they are less likely to believe that people lie for their own gain and thus appear as more naïve about dishonesty in society, compared to individuals with high information preferences. These findings support the narrative that some people have a desire to maintain (false) beliefs of others' willingness to lie, and respond by focusing more on their work, not less, when faced with co-workers engaged in suspicious activities at the workplace. In turn, this phenomenon could contribute to understanding why lies are ubiquitous at work, given that dishonest individuals can rely on their co-workers looking away while keeping performance high.

In Section 2.2, we describe the experiment and discuss our expectations in detail, whereas additional background information can be found in Appendix A.1 (including a description of the work task, full instructions, and supplementary figures). In Section 2.3, we present the empirical results and discuss our main findings, whereas additional statistics and analyses are provided in Appendix A.2, and a supplementary information experiment is illustrated in Appendix A.3. Section 2.4 concludes with practical implications.

2.2 The Experiment

2.2.1 Setting

2.2.1.1 Procedure

Students from a German university were invited to participate, via a university-wide announcement through emails and flyers placed around the campus, in a study on work motivation at a computer workstation (see Figure A.1.1 for a translated version of the original announcement poster). The project took place throughout the winter semester of 2016/2017, i.e., from October 2016 to March 2017, and was conducted at a labor market research institute connected to the university. On behalf of the institute, participation in this study was described as support for research and was announced to last for approximately one hour. Participants were informed that they could earn on average about 12 Euro (and separately had the chance of winning 100 Euro in a lottery).

To register for participation, each person had to fill out an online survey, which was set up for this project. Both the recruitment process and session organization were managed by research assistants, based on pre-defined rules. The assistants screened all registered persons to ensure i) proper language skills (German) and ii) the student status of those invited to the sessions. The available information from the webpage registration was also used to avoid inviting individuals possibly knowing each other to take part in the same session.

For our experiment, three students from the list of eligible persons were recruited and were randomly invited to different meeting rooms. After the arrival of all three invitees, research assistants brought them to the computer workstation where the sessions were taking place (see Figure A.1.2 for a map and a picture).[7] After the work session, each participant was paid separately from the others in the original meeting room.

2.2.1.2 Work session

During the session, participants assumed the role of 'team players' (numbered from one to three) working at a computer. The three computers were identically equipped, including the same software (internet browser, email program), and were randomly allocated at the beginning of the session. A research assistant assumed the role of the 'team leader' hosting the session by following instructions and using a big screen (see Figure A.1.2) for demonstration purposes.

[7] If invitees did not appear, sessions were carried out as two-person session for a companion study (or cancelled if only one person appeared). Note that some elements of the procedure and surveys are relevant for the companion study and thus serve here only as a distraction from the intentions behind our study on cheating.

The big screen was connected to the session host's computer and was used to display important information on, for example, how to carry out the task. Figure A.1.3 presents the timeline, which shows that at the beginning of the session, participants were instructed to use emails for communication (instead of verbal communication, which was prohibited).[8]

The main part of the session consisted of a work task that was split into two parts of ten minutes each. The task was to download scientific articles from various scientific journals with economic relevance and to label the files properly. These files are referred to as 'output units' in our paper, providing us with data on individual performance levels. At the time of the experiment, preparing scientific articles was a routine task conducted by student assistants contractually employed at the institute, in response to restrictions of access to publications. Hence, the task can be seen as authentic, as it is described as support for research by the session host during sessions. Appendix A.1 provides additional material and information about the workplace setting, including a detailed description of the task in Appendix A.1.1, the full instructions for the session host (translated from the original German instructions) in Appendix A.1.2, and supplementary figures about the experiment in Appendix A.1.3.

2.2.1.3 Payment details

After each work period, participants completed online surveys, while the session host conducted spot checks of the work output to determine the performance-related pay (see Appendix A.1.1). The total number of points earned by all participants combined was divided by three to determine the individual wage. For each output unit, the entire team received 10 points, which equaled 10 cent. For example, if the aggregate output was 300 points in the first work period, each participant earned 100 points, which equals 1 Euro, for this work period alone. The procedure was the same for the second work period.

In addition to performance pay, participants received fixed income components for appearing (250 points) and for completing all three surveys (200 points each), so that they could secure a minimum pay by simply following the procedure. Finally, two participants had the chance to win a bonus of 300 points in a dice game, conducted between the two work periods.

[8] Participants could send emails separately to each person or to all others via a mailing list. A test procedure was conducted to ensure that participants were able to use emails. Sanctions for violations of the verbal communication ban were announced at the beginning of the session, but it turned out that it was not necessary to penalize any person. Note that all participants were handed out an information document prior to the session start (Figure A.1.9), which clarified that data protection concerning email communication fully applied. Accordingly, neither third parties nor we as researchers were permitted to read the content of any email written by a participant, except for emails sent to the session host.

We used this bonus game to induce our experimental manipulation, as described in the following.

2.2.2 Experimental concept

Participants had a dice cup and a 20-sided die on their table to take part in the bonus game. The procedure of the game consisted of three steps: In step one, the session host randomly selected the first participant of the game ('bonus player A') through a secret die roll. In step two, bonus player A received an email with a request to report a number from one to twenty by rolling the 20-sided die. This number randomly determined who bonus player B would be, with a 50:50 chance for the two other participants.[9] In step three, bonus player B received an email request from the session host to report a number from one to twenty. For this purpose, the session host used the email response received by bonus player A and added the exact same inquiry text as the one used for bonus player A, before forwarding the email to bonus player B (see Figure A.1.4 for an exemplary screenshot).

The seminal rule of the game was that both bonus players obtained a bonus of 300 points in case of a match. The definition of a match was read aloud by the session host and was part of the information shown on the big screen, thus visible throughout the bonus game. After bonus player B reported a number, all participants, including the non-involved bystander, were informed about the outcome of the game via email (see Figure A.1.5 for two exemplary screenshots, with and without bonus win).

In the opportunity-to-cheat treatment (see Figure A.1.6), the bonus was paid out when B's number matched the number determined by bonus player A in step two of the bonus game, in the sense that both numbers were identical. In the control treatment (see Figure A.1.7), the bonus was paid out when B's number matched the number determined by the session host in step one of the bonus game, in the sense that both numbers had to be either even or odd. This rule change was the only difference between the two treatment conditions, and it ensured that in one treatment, an opportunity-to-cheat was prevented (control treatment), while no efforts were made in the other one to prevent cheating (opportunity-to-cheat treatment). Figure 2.1 shows the experimental design with its two conditions and case numbers, whereas information on sample statistics is provided in Appendix Table A.2.1.

[9] To fully ensure randomness in the selection of bonus player B, there was a rule through which one of the two other participants was selected based on the number reported by bonus player A. This rule was randomly determined by the session host in advance using a sheet (Figure A.1.11) unknown to the participants.

Figure 2.1 Experimental design

	Treatment conditions	
	Control	**Opportunity-to-cheat**
No bonus	50% random chance for a no-bonus session	Incidence of bonus sessions determined by participants
Bonus win	50% random chance for a bonus session	
	51 sessions with three participants → 153 individual observations	85 sessions with three participants →255 individual observations

The idea underlying our experimental setup is to randomly allow for cheating and to then study how individuals respond in their performance to dishonesty in a working group context. To avoid practicing dishonest actions as researchers and to allow for an authentic occurrence of actual misreporting, we manipulate the opportunity-to-cheat and let participants of the experiment decide on a possible treatment for others. Individuals in the role of bonus player B in our opportunity-to-cheat treatment make this decision. As the email request by the session host includes the information of bonus player A necessary to report a match, bonus player B can decide in favor of cheating independently of any communication with others. In our empirical analysis, we then focus on the non-involved bystander, for whom we assume that the occurrence of cheating is an exogenous incidence, which may or may not matter for perceptions in a work context and thereby potentially influence behavior.

A challenge in the analysis of how dishonesty affects the behavior of others is to disentangle the cheating effect from other effects that the action may trigger. If two persons cheat at the expense of a third person, which the literature refers to as a 'black' lie (Erat and Gneezy 2012), it is unclear whether behavioral responses are solely due to perceived dishonesty or due to undesirable economic implications, such as lower income. To shut this channel down, we focus on cheating that does not cause financial harm to the bystander. In addition to absolute income, relative income could be another relevant aspect for workplace outcomes (Clark and Oswald 1996, Clark et al. 2010, Card et al. 2012). Ockenfels et al. (2015) analyze the effects of bonuses in a multinational company and find that the performance of managers decreases when they fall behind their natural reference standard for a fair bonus payment. Hence, when A and B get a bonus through cheating in our setting, they receive more money in comparison to the bystander

whose behavioral response could be due to pay inequality rather than due to dishonest behaviors of others.[10]

The design in Figure 2.1 emerges out of these considerations, as it allows comparing situations with the same payoff structures from the perspective of the bystander. By conditioning on a bonus win for the other participants, we minimize differences across treatments, so that the bystander witnesses cheating (in our main treatment) or not (in the control treatment). The design helps us to also rule out other possible mechanisms, which could be income-related or not. For example, income-targets could play a role (see Houser et al. 2012), given that bystanders may consider it unfair if only others can play the bonus game in order to reach the desired income level. As another example, being excluded from a game might be a potentially negative experience per se. However, such considerations among bystanders should play the same role across experimental conditions.

Furthermore, Figure 2.1 reveals that we oversampled the opportunity-to-cheat treatment, compared to the control treatment. We did this to reduce the risk of gathering too few data from bonus win sessions, in case of only little cheating taking place. In the next section, we discuss our expectations regarding the occurrence of cheating in our setup comprehensively. Regardless of these considerations, note that no-bonus sessions are also useful, e.g. to analyze possible effects related to changes in the likelihood of a bonus win. Arguably, a lower chance of an honest win in the opportunity-to-cheat treatment could be seen as less fair. For bystanders who cannot win in this game, we expect that changes in the likelihood of a bonus win as such are irrelevant, but we can still investigate this idea using survey data on fairness perceptions.

2.2.3 Expectations

2.2.3.1 Cheating in work groups

A major finding in the empirical literature on cheating is that many individuals do not engage in dishonesty, even if they clearly benefited economically. Fischbacher and Föllmi-Heusi (2013) find in the laboratory that roughly 40% of subjects choose to be fully honest at their own expense. Abeler et al. (2014) even find that a large majority of people behave honestly when given the opportunity-to-cheat at the telephone, indicating stronger psychological costs of lying outside the laboratory. Accordingly, we would expect that participants in our workplace

[10] Note that pay inequality does not necessarily have adverse behavioral consequences in teams, according to experimental research in the laboratory (Bartling and von Siemens 2011).

scenario may abstain from cheating more often than in a laboratory context. Yet, there is an important factor inherent in our setup that could encourage dishonest behavior.

Research shows that cheating increases when others benefit from it, as potential cheaters can morally justify their behavior in case of 'altruistic' lies (Schweitzer and Hsee 2002, Gino and Pierce 2009, Wiltermuth 2011, Gino et al. 2013a). In an experiment with a bonus game similar to the one in this study (but without bystander), Weisel and Shalvi (2015) find that when both bonus players benefit, the likelihood of cheating increases compared to when only one player benefits. Having said that, the consequences of adding a third person observing others are less clear. Ewers and Zimmermann (2015) show experimentally that misreporting can be affected by an audience, which they link to image concerns. Van de Ven and Villeval (2015) find that scrutiny has little impact on dishonest actions when adding an observer to the two-player deception game setup by Gneezy (2005). Using a virtual reality environment, Mol et al. (2020) show that the presence of an observer reduces cheating if the decision-maker is stared at.

In conclusion, previous findings in the cheating literature suggest that for participants in our setup, there are various factors that could either discourage or encourage cheating. While the opposing roles of relevant factors make it difficult to make a clear prediction regarding the extent of cheating, we are nevertheless confident that we can expect two scenarios (honesty versus dishonesty) with potentially similar probabilities of occurrence. Due to its importance for the rest of our study on the behavioral implications of cheating, we start the presentation of results by providing evidence on its occurrence in Section 2.3.1. In addition, we examine cheating determinants in this part of our analysis. Due to the presence of altruistic motives to cheat for the benefit of others, one could expect that positive reciprocity increases lying in a team context, as one example. As pointed out by Weisel and Shalvi (2015), however, forcing others to become "partners in crime" could also deter some individuals from lying, which raises the question who appears as willing for collaborative cheating and who does not. Gender is a visible characteristic and a possible determinant of dishonesty that is intensely debated in the literature.[11]

[11] Some research on cheating suggests that men are more likely to cheat than women (Ward and Beck 1990, Dreber and Johannesson 2008, Bar-El and Tobol 2017, Grosch and Rau 2017), though many other studies do not find robust evidence of gender differences (Schwieren and Wechselbaumer 2010, Childs 2012, Gino et al. 2013b, Gylfason et al. 2013, Ezquerra et al. 2018). In case of altruistic lies that help others at one's own expense, women have been found to be more likely to cheat than men (Erat and Gneezy 2012). Yet, Muehlheusser et al. (2015) find that in groups, women are less likely to cheat under joint decision-making compared to men.

2.2.3.2 Performance implications of dishonesty at work

We now turn to the discussion of our expectations concerning the behavioral implications of dishonesty at the workplace. According to management researchers, organizational performance benefits from establishing an honest workplace environment (Losey et al. 2007). To test this empirically, we first analyze in Section 2.3.2 the effects of the opportunity-to-cheat for team performance. While this gives us an impression on the potential benefits of honesty at the workplace, in Section 2.3.3 we focus on the direct implications of dishonesty by examining the behavioral response of the bystander when others exploit the opportunity-to-cheat.

In our discussion why dishonesty at work could be harmful to performance, we refer to literature on social motives for individuals to contribute to team success. Field evidence on performance effects of team-related pay indicates that employees do not only consider their own payoffs as relevant but are also affected by psychological motives when working together with others (Hamilton et al. 2003, Delfgaauw et al. 2013, Friebel et al. 2017). As argued by Babcock et al. (2015), an important aspect could be social pressure, since individuals do not want to disappoint their team members.[12] However, any psychological motive to work hard for team achievement may be altered by perceptions of other team members. Arguably, perceiving others as dishonest could reduce the motivation to contribute to a team outcome, from which a negatively perceived team member would benefit. Evidence for the harmful effect of cheating on the perceptions of others comes from studies on people's willingness to punish liars, such as Brandts and Charness (2003) as well as Eisenkopf et al. (2017). Reuben and Stephenson (2013) conclude from their study that individuals "consider lying per se as normatively undesirable behavior that deserves to be punished."[13]

There are other psychological mechanisms of performance that could be relevant in the context of dishonest behaviors among co-workers, such as trust in the team members (Breuer et al. 2016, De Jong et al. 2016). This is particularly relevant given that dishonesty can reduce trust, which has been shown for black lies that go at the expense of another person (Gawn and Innes 2018). To provide a full picture of various reasons why cheating could affect performance, we make use of our rich survey data obtained from participants after the bonus game. This complements our comprehensive analysis of performance, for which we focus on the number of output units per individual, as provided in our workplace setting (Section 2.2.1).

[12] Note that we use the term 'team' in the context of individuals working together in groups, as in Babcock et al. (2015), who argue that the absence of typical team characteristics like complementarities helps to identify mechanisms of interest.

[13] For research on punishment of team members, see Carpenter et al. (2009) and Vranceanu et al. (2015).

In addition, we analyze the dimensions of performance to inspect possible cheating spillovers (Pierce and Snyder 2008, Fosgaard et al. 2013, Innes and Mitra 2013), given that individuals could cheat on their performance (Bäker and Mechtel 2019), which in our setting is possible by reducing the quality of work. Furthermore, we analyze both the dynamics of performance within work periods as well as future work motivation, as reported in our survey, to learn more about possible changes in treatment effects over time.

2.2.3.3 Bounded awareness and information preferences

The expectation of negative performance effects due to dishonesty at work crucially depends on whether individuals actually realize the occurrence of others' misbehavior. The literature on organizational behavior discusses cases in which many individuals turn a blind eye and do not recognize the cheating of others, a phenomenon referred to as 'bounded awareness' (Bazerman and Sezer 2016). If true, and some employees knowingly or unknowingly do not pay attention to undesirable forms of behavior in an organizational context, the question emerges whether performance implications could be moderated by preferences for inconvenient information.[14]

From an economic viewpoint, there are several explanations why some employees may wish to avoid an uncomfortable realization concerning others' misbehavior. A simple explanation is that the realization of a lie could require cognitive efforts that for some individuals are more costly than the benefits of knowing. There could even be psychological costs of knowing, as realizing something unpleasant may undermine one's well-being. This could be a particular issue in the context of teamwork and possible misbehavior of others, assuming that acknowledging dishonesty of team members may threaten the psychological benefit of putting in efforts for them. Furthermore, there are discussions in the literature on how information avoidance, be it active or passive, through ignorance or lack of attention, could actually be performance-enhancing (Bénabou and Tirole 2002, Kajackaite 2015, Hertwig and Engel 2016, Huck et al. 2018). In an experiment, Falk and Zimmermann (2017) show that individuals use the opportunity to distract themselves from inconvenient information by performing a task,

[14] While bounded awareness serves as an umbrella term for information avoidance independent of one's intentions, behavioral economists in particular discuss cases of more deliberate rather than unconscious actions to avoid information, using labels like 'willful ignorance' (Grossman and van der Weele 2017). The large body of research on cases in which people prefer avoiding inconvenient information includes Dana et al. (2007), Eil and Rao (2011), Oster et al. (2013), Ganguly and Tasoff (2017), d'Adda et al. (2018), Shalvi et al. (2019), and Freddi (2020), to name just a few examples. Notably, in addition to the case of active information avoidance, refusing to realize the uncomfortable by passively not paying attention to available information is also part of the growing research on information preferences (see Golman et al. 2017, Andries and Haddad 2020).

which suggests a motive for participants in our study to increase efforts when exposed to the potentially negative experience of dishonesty at work.

In summary, there are various reasons to expect heterogeneity in the way dishonesty at work affects co-worker performance along the lines of information preferences. To analyze whether individuals with high preferences for inconvenient information respond more negatively to cheating at work compared to those whom we refer to as ostriches in the following, we use a self-conceived survey item to identify and thereby distinguish information-preference types.[15]

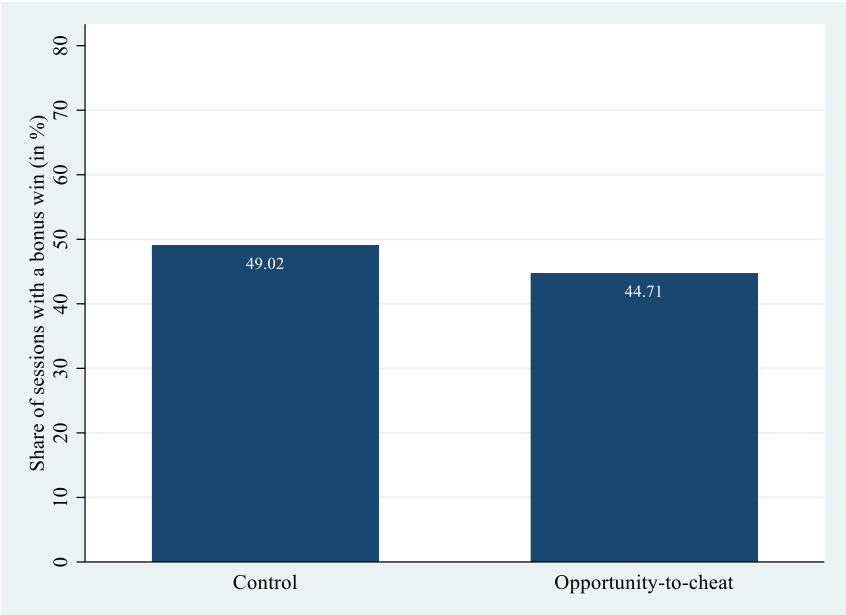
2.3 Empirical analysis

2.3.1 Occurrence of cheating

We start the presentation of results by providing visual evidence on the likelihood of a bonus win across treatments. In the control treatment, as shown by the left bar in Figure 2.2, the outcomes of the game are equally split due to randomization without any chance to influence the incidence of a match. We observe a bonus win in 49.02% of the cases (25 out of 51 control sessions). In line with expectations, the minor rule change in the bonus game across treatments goes along with a similar split of bonus-win sessions and those without. The bar labeled "opportunity-to-cheat" in Figure 2.2 reaches a level that comes close to the control treatment. In 44.71% of the cases (38 out of 85 opportunity-to-cheat sessions), the numbers reported by participants in the roles of bonus players A and B are a perfect match. Given a probability of 5% to win by chance, we acknowledge a high prevalence of cheating in these sessions. According to the 'Lying Calculator' (Garbarino et al. 2018), 41.75% of opportunity-to-cheat sessions can be considered as a session with cheating (95% confidence intervals range from 38.90% to 44.56%).

[15] Behavioral researchers recently proposed a scale to capture information preferences as an individual trait (Ho et al. 2021). When we conceived our experiment in 2016, we became aware of this and reached out to the researchers to whom we are very thankful for sharing information on a single-question item that we slightly adjusted and incorporated into our survey. Note that efforts to propose survey-based measures for information preferences go back decades. Miller (1987) distinguishes information seekers from information avoiders by asking whether subjects would like to obtain or avoid information in various hypothetical scenarios (e.g. in a plane flight). To allow for a discussion of the validity of our survey item, we incorporated a few questions on real-world scenarios in our survey. We discuss those in Section 2.3.4 together with other steps that we conducted in this regard, including a supplementary information experiment using invitations to a research seminar (Appendix A.3).

Figure 2.2 Share of bonus wins by treatment



Notes: Bars show percentages of sessions with a bonus win by treatment (left bar: control treatment, right bar: opportunity-to-cheat treatment).

In Table 2.1, we shed light on the possible determinants of cheating, relying on the idea that the vast majority of bonus wins in the opportunity-to-cheat treatment are the result of dishonest actions. For this analysis, we distinguish between individuals according to their different roles during the bonus game. We focus first on bonus player B, expecting that in our setting it is unlikely that (unobserved) characteristics of others could influence the decision to cheat. According to column 2, some characteristics of B are indeed relevant for the decision to cheat. Personality seems to matter, as agreeableness predicts the cheating incidence, whereas trust is negatively related to the willingness to cheat. The occurrence of altruistic lies to the benefit of another person, namely bonus player A, is not predicted by positive reciprocal attitudes of B. Interestingly, the gender of B does not matter, but the gender of the potential beneficiary bonus player A does. Column 1 shows an increase of 33.60 percentage points in the likelihood of a bonus win if the potential partner in crime is male. This implies a dramatic increase in the incidence of cheating, given that the average bonus probability is roughly 30% if A is female.

Table 2.1 Incidence of cheating

	(1) Bonus player A	(2) Bonus player B	(3) Bystander
Skills: math	-0.029	0.030	0.084
	(0.069)	(0.052)	(0.078)
Skills: probabilities	0.008	-0.051	-0.012
	(0.058)	(0.074)	(0.080)
Skills: computer	0.037	0.023	-0.018
	(0.062)	(0.058)	(0.074)
Big5: extraversion	-0.049	0.027	0.020
	(0.044)	(0.054)	(0.058)
Big5: agreeableness	0.188***	0.133***	-0.033
	(0.049)	(0.047)	(0.083)
Big5: openness	-0.066	-0.076	-0.003
	(0.048)	(0.053)	(0.054)
Big5: neuroticism	-0.010	-0.069	-0.009
	(0.054)	(0.061)	(0.042)
Big5: conscientiousness	0.004	-0.097	-0.028
	(0.055)	(0.065)	(0.059)
Reciprocity	0.069	0.005	-0.033
	(0.095)	(0.075)	(0.084)
Trust	0.019	-0.136**	-0.094
	(0.070)	(0.063)	(0.077)
Female	-0.336***	0.065	-0.156
	(0.116)	(0.138)	(0.140)
Age	0.013	-0.020	-0.003
	(0.015)	(0.021)	(0.027)
Number of semester	-0.036**	0.008	0.007
	(0.015)	(0.018)	(0.016)
Bachelor's degree	0.005	0.248*	0.170
	(0.118)	(0.142)	(0.155)
N	85	85	85
R^2	0.299	0.226	0.090

Notes: The dependent variable is a dummy with 1 for bonus win in this session, and 0 otherwise. See Table A.2.1 for more information on the independent variables used. Column 1 shows estimates for individuals in the role of bonus player A. Column 2 shows estimates for individuals in the role of bonus player B. Column 3 shows estimates for individuals as bystanders not taking part in the bonus game. Results from linear-probability estimations are shown. Robust standard errors are in parentheses. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We conclude that bonus player B's decision to cheat is hard to predict for others, given that the few significant factors are neither known nor visible. In this context, the effect of A's gender on the likelihood of cheating appears to be particularly intriguing, as this characteristic certainly is observable in our workplace context, contrary to the other two significant predictors regarding bonus player A (number of semesters and agreeableness, see column 1 in Table 2.1).

An explanation could be that the gender of the possible beneficiary of cheating signals relevant information to the decision-maker who might be less willing to make a woman become a partner in crime, which seems much more acceptable for a potential male beneficiary.[16] As another noteworthy observation, none of the factors related to the bystander play a significant role for the decision to cheat. An R^2 of less than 0.1 in column 3 in Table 1 empirically supports this finding, whereas the R^2 in the first two columns is above 0.2, implying more predictive power in the characteristics of bonus players A and B. Having found evidence of cheating in a large number of cases, we now turn to the potential implications for performance.

2.3.2 Opportunity-to-cheat and team performance

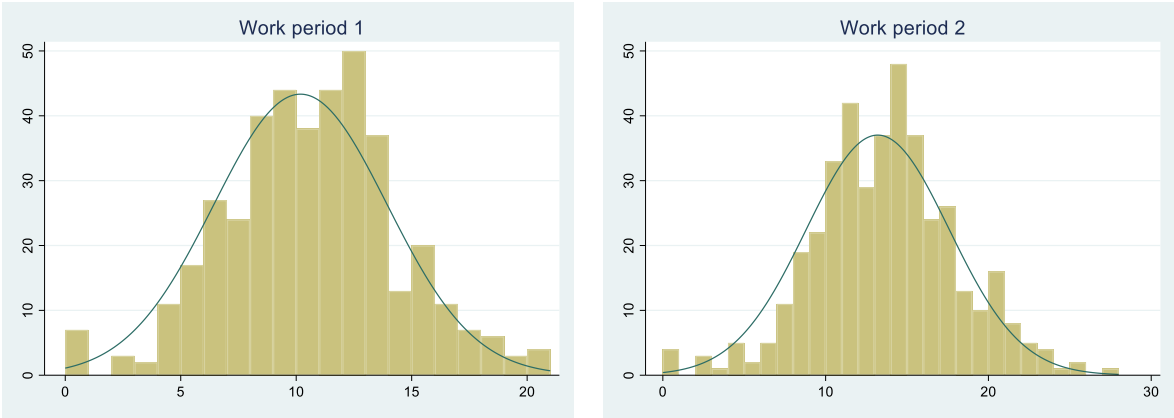
As a first step in our analysis of task performance during work sessions, we provide an illustration based on our main performance measure at the individual level. This measure comes from the session records and is the number of prepared articles used to determine the performance pay of the participants. Panel A of Figure 2.3 shows histograms to visualize the distribution of output units across work periods. For both work periods, we observe a variation in performance levels that broadly follows a normal distribution, with the median at 10 units of output in period 1 (mean: 10.19) and the median at 13 units in work period 2 (mean: 13.20). This implies an increase in performance from period 1 to period 2, as a possible learning effect, which amounts to 3 output units.

Panel B of Figure 2.3 visualizes the average team performance across work periods and treatment conditions. Bars are of similar size for both treatments, before and after the bonus game, suggesting that the opportunity-to-cheat as such is irrelevant for team performance. Since team averages may hide performance differences between participants according to their role in the bonus game, we split the performance data in the following and focus on within-person changes in individual performance from period 1 to period 2, thereby exploiting the longitudinal nature of the dataset.

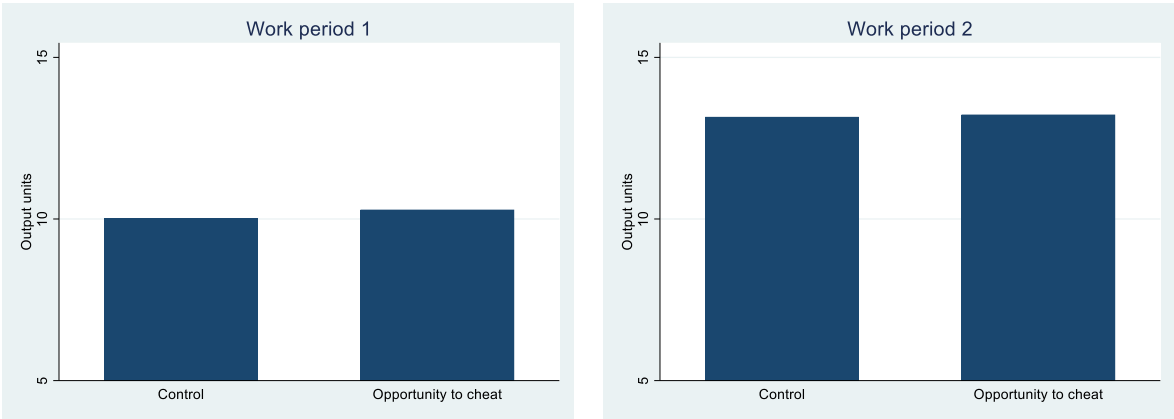
[16] To check this idea further, we also analyzed email communication between bonus players. While the content of the emails is unknown to us, we observe an increase in the frequency of emails sent from bonus player A to bonus player B in the opportunity-to-cheat treatment, compared to the control treatment. Furthermore, chances of a bonus win are higher in opportunity-to-cheat sessions when email communication from A to B took place. Remarkably, email communication does only increase chances of a bonus win if A is female. This suggests that the gender disparity in Table 2.1 is not due to effective email communication by males in the role of bonus player A, but rather that email communication actually reduced the gender gap.

Figure 2.3 Task performance across work periods

Panel A) Histograms of output units by work period



Panel B) Average team output by work period and by treatment



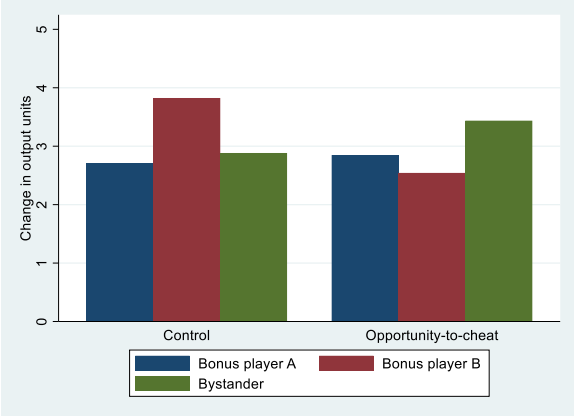
Notes: Panel A shows histograms of output units by work period (left picture: period 1, right picture: period 2). Bars in panel B show averages of output units by work period (left picture: period 1, right picture: period 2) and by treatment (left bar: control treatment, right bar: opportunity-to-cheat treatment).

By distinguishing task performance between all three participants according to their roles in the bonus game, the bars in Panel A of Figure 2.4 suggest that there is a negative opportunity-to-cheat effect for bonus player B. Indeed, there is a significant difference in how task performance changes from work period 1 to period 2 among individuals in the role of bonus player B, depending on the treatment condition (p=0.011).[17]

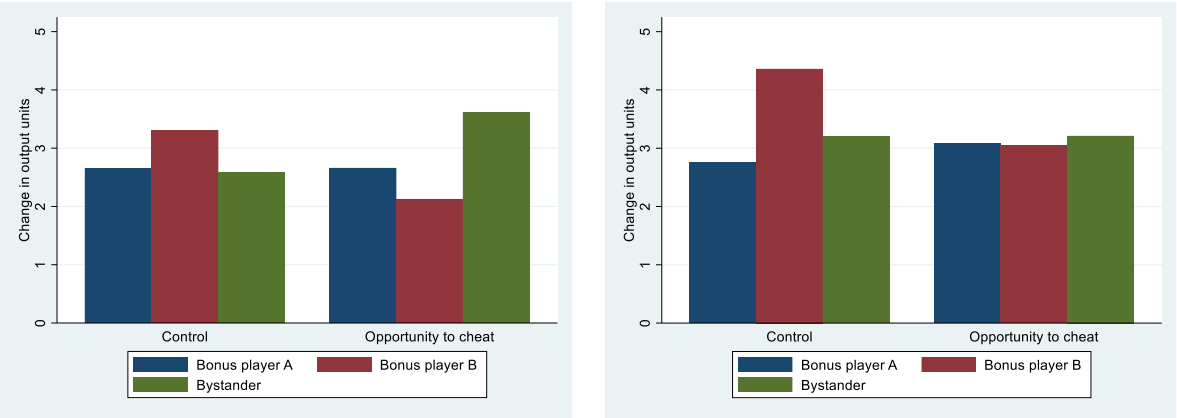
[17] p-values always come from two-sided t-tests that we conduct consistently throughout the paper. For non-performance outcomes, we additionally report p-values of non-parametric tests in the paper if the test result differs in a sense that a different significance threshold is reached.

Figure 2.4 Changes in task performance across work periods by treatment

Panel A) Full dataset (all sessions)



Panel B) Sample split: Sessions without bonus win (left) and sessions with bonus win (right)



Notes: Bars show averages of changes in output units from work period1 to work period 2 by treatment (left bars: control treatment, right bars: opportunity-to-cheat treatment) and by individual roles in the bonus game (bonus player A, bonus player B, and bystander). The illustration in Panel A shows bars for the full dataset. The left-side illustration in Panel B shows bars for data from sessions without a bonus win. The right-side illustration in Panel B shows bars for data from sessions with a bonus win.

On closer inspection, there is a comparatively strong increase in task performance among bonus players in the control treatment of 3.82 output units on average. This could be a 'lucky person' effect that did not occur among those bonus players who had to decide between cheating and honest reporting in the opportunity-to-cheat treatment (average increase in output units: 2.54). As seen in the illustration on the left side of Panel B, where we separate the data according to the outcome of the bonus game, this effect seems to be driven by 'successful' winners in the control treatment. Here, the increase in task performance reaches 4.36, compared to 3.05 output

units in the opportunity-to-cheat treatment (p=0.088). We also observe a performance gap between the two treatments for bonus players B without a bonus (right side of Panel B), which seems to reflect underperformance of B after deciding to report honestly on the outcome of the dice game (2.13 output units compared to 3.31 in the control, p=0.073).

In summary, there is some evidence for underperformance of bonus player B when the opportunity-to-cheat exists. This negative effect however is not strong enough to bring down average team performance, given that both bonus player A and the bystander keep their performance relatively high. With regard to the latter, the question emerges whether certain types of bystanders may respond with lower or higher performance to the occurrence of cheating at work. We investigate this deeper in the following section when we comprehensively analyze the implications of cheating for bystander performance and consider possible effect heterogeneity.

2.3.3 Cheating and bystander performance

2.3.3.1 Main results

To comprehensively investigate the effects of cheating on co-worker performance, we employ a 'Difference-in-Difference' (DiD)-analysis that we illustrate in the following. The idea is to compare differences in performance of individual i over time t, i.e. across work periods (t=1,2), and between treatment conditions. For the analysis, we restrict the data on bystanders and a successful bonus win, so that a bonus was paid out to the other two participants, but the bonus was paid out for a different reason, depending on the treatment condition. As we analyze changes in task performance across two work periods, before and after the bonus game, this longitudinal analysis eliminates any role of time-invariant factors, such as personality, which may predict differences in performance levels, but cannot explain differences in performance changes across periods.[18] Thus, we do not include any control variables in the following empirical model:

$$PERFORMANCE_{it} = \beta_1 TIME_{it} + \beta_2 OTC_i + \beta_3 IA1_{it} (OTC_i\_X\_TIME_{it})$$
$$+ \gamma_1 IP_i + \gamma_2 IA2_{it} (IP_i\_X\_TIME_{it}) + \gamma_3 IA3_i (IP_i\_X\_OTC_i) + \gamma_4 IA4_{it} (IP_i\_X\_IA1_{it})$$

[18] When we check randomization of the treatments in the full sample (N=408), by analyzing the variables shown in Appendix Table A.2.1, we find no significant difference with the exceptions of gender (p= 0.062) and bachelor degree (p=0.020). Focusing on the smaller dataset with only bystanders in bonus sessions (N=63), we find no significant treatment differences. Given that this could be because of sample size, we prefer conducting longitudinal analyses and exploiting the fact that we had arranged a panel setup with two work periods.

The model can be used in two variants: First, for a DiD analysis, we only use the first row of the model that includes three independent variables, which are a time dummy ($TIME_{it}$), a treatment dummy (OTC_i), and the interaction between the two. The latter reveals the average cheating effect, as our DiD result. Second, to study effect heterogeneity, further variables are added to the model. We expand the DiD-model to a 'Difference-in-Difference-in-Difference' (DiDiD) model by interacting all variables with a dummy for high information preferences (IP_i). To generate this dummy variable, we employ a median-sample split using the self-reported scores from the survey question on preferences for inconvenient information.[19]

The DiD-analysis in the first column in Table 2.2 shows an overall increase in performance from work period 1 to period 2 (row 1). This complies with our visual evidence for the main performance outcome, as presented in Figure 2.3. The treatment effect (row 2) is insignificant, which indicates that there are no differences in performance among bystanders across treatment conditions before the bonus game. Most importantly, the interaction of work period 2 and the treatment condition turns out to be positive and insignificantly different from zero (row 3).[20] Accordingly, dishonest behavior seems to have no detrimental impact on co-worker performance.[21]

[19] The original wording of the question is as follows: 'Do you want to be informed about something even when the information could be uncomfortable?' The response scale ranges from 1 ('No, under no circumstances') to 10 ('Yes, under all circumstances'). As roughly half of the participants reported scores of 9 or 10 on whether they prefer information, even if it could be uncomfortable, we consider those as 'high' information-preference types.

[20] From a theoretical perspective, one could ask whether there is a small effect that could be detectable with more statistical power. While lack of an effect certainly does not prove that the effect does not exist at all, we have to leave this question open in our study, in which we as applied researchers are interested in economically significant effects, just like personnel management is. To further address the issue of significance, we can determine meaningful effect sizes that can be ruled out. By testing the interaction in column 1 of Table 2.2 against minus two (instead of against zero), we are able to rule out a meaningful performance reduction by two output units as a consequence of co-worker cheating (p-value: 0.020).

[21] We use data from no-bonus sessions for additional analyses. First, we expand the DiD-setup by adding another interaction variable on the occurrence of a bonus win to the regression model and use a sample which includes all bystanders (N=136), independent of the bonus game outcome. Such a DiDiD-analysis relies on the strong assumption that differences in the payment structure between bonus and no-bonus sessions are not relevant for bystanders. This analysis shows again that bystanders witnessing cheating do not significantly underperform in period 2. Second, we use data from no-bonus sessions to inspect whether changes in the rules of the bonus game as such matter for the bystander. Yet, we do not observe differences in performance changes across treatment conditions when we repeat the DiD-analysis based on data from no-bonus sessions only. We also do not observe significant differences in fairness perceptions among bystanders in no-bonus sessions across treatment conditions when analyzing data from a survey battery that includes two items on perceived fairness (see Appendix Table A.2.5).

Table 2.2 Cheating and task performance

	(1)	(2)	(3)	(4)
Dependent variable:	Output units after check		All output units	
Work period 2	3.200***	1.444***	2.880***	0.667
	(0.673)	(0.369)	(0.681)	(0.461)
Opportunity-to-cheat	1.195	0.538	1.453	0.228
	(0.977)	(1.505)	(0.875)	(1.327)
DiD: Cheating treatment	0.011	2.029**	0.331	3.123***
(Opportunity-to-cheat X period 2)	(0.841)	(0.818)	(0.820)	(0.785)
Information preferences: High		0.660		-0.104
		(1.521)		(1.425)
DiD: Work period 2 X		2.743***		3.458***
Information preferences: High		(1.027)		(1.014)
DiD: Opportunity-to-cheat X		1.498		2.420
Information preferences: High		(2.032)		(1.816)
DiDiD: Cheating treatment X		-3.269**		-4.616***
Information preferences: High		(1.448)		(1.360)
N	126	126	126	126
R^2	0.130	0.179	0.172	0.231

Notes: The dependent variable is output units per individual observed in a work period. In columns 1 and 2, the output comes from the records of the team leader after a spot check of correctness. In columns 3 and 4, the output is the total number of articles prepared by each individual independent of quality. Opportunity-to-cheat is a dummy that is 1 if the individual was in the opportunity-to-cheat treatment, and 0 if the individual was in the control treatment. 'Information preferences: High' is a median-split dummy established based on reported scores on a 10-point scale ("No, under no circumstances" [1] - "Yes, under all circumstances" [10]), as provided in response to the question "Do you want to be informed about something even when the information could be uncomfortable?" Observations are from 63 individuals in the role of the bystander (individuals excluded from the bonus game), who are either observed in the cheating treatment or the control treatment. Estimates are from standard regressions. Robust standard errors are in parentheses. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Column 2 in Table 2.2 shows heterogeneity in the effect of cheating on co-worker performance along the lines of information preferences, as revealed by a significant DiDiD-effect (row 7). While bystanders with high preferences for inconvenient information provide relatively high performance in the control treatment (row 5), the picture changes in the cheating scenario, where ostriches improve their performance significantly (row 3).[22]

[22] One could ask whether ostriches are not just less inattentive towards dishonest actions by others, but are generally less attentive. The results of Table 2.2 do not conform to this notion, given that behavior differs substantially across conditions. Still, we use data that promises to be indicative of general differences in attention to find out more. For a first indicator of attention, we combine several rare incidences, namely falsely submitted emails (due to typos), missing test emails (which participants were asked to send but not forced to) and cases of lacking attention as noted by the session host (according to the session records). Second, we use information from the online registration on whether participants would like to be invited to further studies, as a proxy of interest in experimental research. Similarly, we use participants' responses to a survey question on their motivation to support research, as a third indicator. Fourth, another survey question asked during sessions and prior to the bonus game

The right-side columns in Table 2.2 show the results for a second outcome variable, namely the total number of output units without a check of correctness. The evidence for this alternative outcome variable appears to be similar. While our main performance indicator is the most relevant for the participants, as it determines their payout, the spot-check procedure carried out by the session host (Appendix A.1.1) certainly could be seen as imperfect and prone to human error. For a deeper investigation of performance, we use precise measures of correct and incorrect work output in the following.[23]

2.3.3.2 Additional analyses

Using our extensive dataset, we conduct additional analyses to learn more about our findings on the implications of cheating for performance. First, we inspect the idea of possible negative performance effects that could be hidden in the quality dimension, given that bystanders could respond negatively to cheating by reducing the quality of work as a form of cheating spillover. We then discuss the role of time to assess whether our findings could depend on the length of time allotted to each work period. For that purpose, we benefit from precise information on the rate at which the scientific articles were downloaded, which enables a dynamic analysis of performance in the course of working time. Furthermore, survey data allows shedding light on possible long-run effects of cheating for individual work motivation.

Performance dimensions

Thanks to computer software developed for this project, we can precisely assess the quality of work output. The software checks each article's file name, as prepared by the participant, by using information on the scientific article from a digital library. This provides us with two measures for each participant: the number of correct as well as the number of incorrect output units, the sum of which is the total number of output units, as analyzed in Table 2.2.

The evidence in Appendix Table A.2.2 confirms our main findings when we use the number of correctly prepared output units as the dependent variable. Accordingly, cheating has on

was about how much participants were focused on doing the work task. In all four cases, we find no significant link between the indicator of attention and our information preference measure. We also do not find a significant interaction between attention indicators and cheating, when we conduct DiDiD-analyses of individual performance.

[23] In further checks of the results shown in Table 2.2, we control for peer performance, which varies across work periods and could affect bystander performance. This does not change our findings. In another check, we interact the DiD-effect with gender composition of bonus players in a DiDiD-analysis, since the likelihood of cheating is higher when bonus player A is male (see Section 2.3.1). We find no significant interaction between the cheating treatment and any possible gender constellation of the two bonus players.

average no effect on performance, measured by high-quality work output, which means that we can reject the idea that cheating reduced performance in the quality dimension, as a potential form of cheating spillover. In line with this, there is no increase in erroneous output units when others cheat. As another finding, we observe effect heterogeneity for the quality-adjusted performance measure, which indicates that ostriches did not just increase performance quantitatively after cheating at the expense of reducing the quality of their work output.

Dynamic analysis

To find out more about how cheating could affect performance over time, we exploit timestamp information on the work output for a dynamic analysis. Each article has a second-precise timestamp reflecting the moment of the download, which allows us to distinguish between participants in bonus sessions who were more or less quick in work period 2. We do this for the first eight articles, thereby determining in each case the probability of a bystander being the fastest of the three participants.

First, we visualize how the likelihood that a bystander, in the cheating treatment, was the first to download an article develops over time, compared to bystanders in the control treatment. Figure A.2.1 shows a relatively high probability with roughly 40% throughout the work period, which is constantly higher than the probability of a bystander outperforming other participants in the control treatment. While this is further evidence against the idea of negative performance effects due to cheating, the stability of bystander performance over time also suggests that a longer time horizon would have revealed the same finding. Second, we analyze effect heterogeneity by using interactions in a regression analysis and thereby distinguish participants according to their information preferences. Appendix Table A.2.3 shows that ostriches respond rather positively to cheating, compared to bystanders in the control treatment, albeit statistical significance is reached only in the middle of the work period. As the picture hardly changes over time, the results are inconsistent with the idea of short-lived performance increases among bystanders, observable only at the beginning of the work period.

Future work motivation

To shed light on possible long-run implications of dishonest behavior at work, we examine how the experience of cheating affects survey responses on future work motivation. By employing regression analyses with interaction variables, we inspect effect heterogeneity along the lines

of information preferences. As the dependent variables, we exploit responses to three survey items and present the results in Appendix Table A.2.4, where we also show additional checks using control variables and provide further information on the variables in the table notes.

First, we consider an item on the willingness to perform the work task again in the future. For this indicator of bystanders' future work motivation, we observe a significantly positive effect of dishonesty on average, which is particularly strong among ostriches (Panel A, Table A.2.4). It appears that in line with our main results, some bystanders are very motivated and want to put in high efforts also in the future. However, they are not doing this in order to help their team as a form of altruistic behavior, as our next result shows. Here, we analyze an item on the willingness to work together with the same team in the future, which reveals no effect of dishonesty at all (Panel B, Table A.2.4). Finally, when analyzing an item on general preferences to work in teams later on in life, we once more observe a more positive effect of dishonesty on the motivation of ostriches compared to high-preference types (Panel C, Table A.2.4). Again, ostriches are and claim to be highly motivated to work, even when asked about the future, though not necessarily with the same group of people in the work session. We continue the discussion on reasons for putting in efforts in the following when we focus more on the mechanisms of task performance.

2.3.3.3 Mechanisms

To inspect potential mechanisms how cheating may or may not affect co-worker performance, we analyze subjective perceptions of the teamwork context using a survey battery with various items, including the willingness to free-ride ('lean back and let others do the job'), social pressure ('a desire to not disappoint others'), and trust ('trusting others at the workplace'). Comparing average scores and median-splits reveals no evidence for manipulation of team perceptions due to cheating, as can be seen in Appendix Table A.2.5. We observe a weakly significant result, according to which the likelihood that the working atmosphere is assessed as good is higher in the cheating scenario compared to the control ($p=0.090$), although this effect is not robust to the test procedure (two-sided Fisher's exact test $p=0.120$). When we run regressions using an interaction between treatment and information preferences to inspect effect heterogeneity, we cannot find any significant result that would support the idea that team spirit or similar team-related motives spurred ostriches to improve their performance after cheating.

Another explanation for higher task performance among ostriches could be their desire to distract themselves from a possible negative experience by putting in more effort into their work (see Section 2.2.3.3). To test this idea, we included a survey item (wording: "I have put in efforts because it helps me better deal with negative experiences.") into a battery of reasons for putting in effort in a team context, and we asked participants whether they agree or disagree. Appendix Table A.2.6 provides information on all items and shows that there are no average treatment effects in survey responses. To inspect heterogeneity along the lines of information preferences, we conduct a regression analysis with a median-split reflecting more or less agreement with the statement as the dependent variable.

In Table 2.3, we first confirm that there is no average effect in the whole population of bystanders when asked about whether they put in efforts to better deal with negative experiences, independent of whether we use control variables or not (columns 1 and 2).

Table 2.3 Reason for putting in efforts

	(1)	(2)	(3)	(4)
Dependent variable:	Dealing with negative experiences			
Cheating treatment	-0.019	0.109	0.363**	0.368**
	(0.130)	(0.122)	(0.160)	(0.183)
Information preferences: High			0.514***	0.385*
			(0.165)	(0.226)
Cheating treatment X Information preferences: High			-0.619**	-0.456*
			(0.233)	(0.268)
Control variables	No	Yes	No	Yes
R^2	0.000	0.287	0.106	0.328

Notes: The dependent variable in each panel is a median-split dummy established based on reported scores on 7-point scales ("I have put in effort, because it helps me better deal with negative experiences."). See Table 2.2 for more information on the variables used. Observations are from 63 individuals in the role of the bystander, who are either observed in the cheating treatment or the control treatment, in each case after a bonus win. Results from linear-probability estimations are shown. Control variables are those shown in Table 2.1. Robust standard errors are in parentheses. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

When analyzing effect heterogeneity, however, we find that, in the cheating scenario, ostriches are more likely to agree with this idea, compared to individuals with high preferences for inconvenient information (columns 3 and 4). This is a remarkable result, given that in additional analyses, we do not find such evidence for effect heterogeneity regarding any of the other eight reasons why individuals put effort into their work. The list of reasons includes for

example the idea of doing “something for others.” We conclude that ostriches increase efforts not to help their co-workers, but rather to avoid acknowledging their dishonest activities.

2.3.4 Discussion of information preferences

Information preferences play a crucial role in our findings, including the first result of cheating taking place in a large number of cases. Exploiting the opportunity-to-cheat arguably becomes less difficult for bonus players when they expect others to turn a blind eye to dishonest behavior. This leads to a few remaining questions regarding information preferences and their role in the context of dishonesty that we aim to address in the following based on additional evidence. We start with the issue of whether we actually measure preferences for inconvenient information accurately with our survey item, before we conclude our investigation by inspecting the link between information preferences and beliefs about dishonest behavior.

In one of several attempts to empirically check the validity of our information preference measure, we conceived survey questions where we asked participants whether they would like to know more about possible inconvenient activities of others in several hypothetical scenarios: i) a student cheating in an exam, ii) a sports athlete practicing doping, and iii) a former partner being unfaithful. In Appendix Table A.2.7, we compare correlations in different samples and find that reported scores of participants are always correlated positively with our information preferences measure, though insignificantly in scenario i), weakly in scenario ii), and strongly in scenario iii). To further exploit the survey context, we asked participants whether they want to know how they were rated by other participants. Ratings were part of the survey prior to the work task and included various aspects such as attractiveness. As can be seen in Appendix Table A.2.7, the participants’ desire to obtain sensitive information relates significantly to our information preference item.

In another effort to verify the information-preference measure, we conceived a follow-up experiment using our university setting. After all sessions were finished, we invited participants to take part in a brown bag seminar where we presented our findings from the experimental research project. We modified the invitation emails to randomly announce an option to get information about one’s performance in the work task at the location of the seminar. Inspired by research on information avoidance behavior (Karlsson et al. 2009, Sicherman et al. 2016), the idea was to investigate whether participants change their decision to take part in the seminar when facing potentially (un)pleasant information about (low) performance, depending on their

information preferences. The results in Appendix Table A.3.1 support this idea, as announcing information on the performance rank renders preferences for inconvenient information to become a positive predictor of seminar participation if we condition on low performers. We conclude that our information preference measure is relevant and able to predict real-world behavior (see Appendix A.3 for experimental details and further discussions of the findings, including checks).

To learn more about the characteristics of individuals with different preferences for inconvenient information, we inspect how types differ from each other, according to our data, and whether this aligns with findings in the literature. For that purpose, we compare ostriches and individuals with high information preferences in Table 2.4, where we also check the robustness of our findings across samples (left side: full sample, right side: bystanders in bonus sessions).

Table 2.4 shows that most of the characteristics, such as gender, do not play any role in describing differences between information-preference-types. From several potential ability proxies, only one (skills for probability estimations) seems to differ, but this finding is not robust across samples. This fits well with previous research considering information preferences as a stable characteristic trait with no clear link to intelligence or any standard socio-demographic factor (Golman et al. 2017, Ho et al. 2021). We do find some differences regarding personality, as it seems that ostriches are less prosocial than the high-preference-types. The latter also score higher on the Big5 ‘openness’ item, which is in line with Ho et al. (2021), who describe individuals with high information preferences as particular curious and open to new experiences. We conclude from this comparison that our survey item allows for a reasonable distinction of individual types that fits to previous findings in the research on information preferences.

Table 2.4 Comparison of information preference types

	(1) Mean	(2) Mean	(3) *p*-value	(4) Mean	(5) Mean	(6) *p*-value
Skills: math	3.13	3.26	0.196	3.25	3.17	0.777
Skills: probabilities	2.87	3.08	0.038	3.00	3.03	0.918
Skills: computer	3.43	3.48	0.626	3.25	3.51	0.281
Big5: extraversion	4.73	4.98	0.044	4.62	4.96	0.285
Big5: agreeableness	5.28	5.46	0.079	5.08	5.32	0.386
Big5: openness	4.84	5.11	0.015	4.18	5.34	0.000
Big5: neuroticism	4.51	4.38	0.279	4.70	4.16	0.098
Big5: conscientiousness	5.00	5.29	0.006	4.92	5.16	0.362
Reciprocity	5.88	6.22	0.000	5.58	6.31	0.000
Trust	3.49	3.44	0.520	3.57	3.40	0.403
Female	0.64	0.64	0.998	0.57	0.51	0.657
Age	24.67	24.83	0.683	24.54	25.17	0.398
Number of semesters	5.13	5.06	0.878	4.46	5.94	0.204
Bachelor's degree	0.29	0.29	0.899	0.32	0.26	0.582
Performance work period 1	10.10	10.26	0.665	9.14	10.54	0.163
Information preferences: High	No	Yes		No	Yes	
Bonus win	No	No		Yes	Yes	
Bystander only	No	No		Yes	Yes	
N	218	190		35	28	

Notes: See Table A.2.1 for more information on the variables shown. 'Information preferences: High' is a median-split dummy established based on reported scores on a 10-point scale ("No, under no circumstances" [1] - "Yes, under all circumstances" [10]), as provided in response to the question "Do you want to be informed about something even when the information could be uncomfortable?" The full sample is used in the left half of the table. The sample in the right half of the table is restricted to individuals in the role of the bystander observed in sessions with a bonus win. Means are displayed in columns 1 and 2 as well as columns 4 and5. *p*-values of two-sided t-tests are shown in columns 3 and 6, in each case testing for significant differences between information-preference types.

Finally, we inspect how information preferences may or may not be related to personal attitudes towards the dishonest actions of other people. So far, we have assumed that realizing the misbehavior of others is inconvenient and information preferences matter in this context, but research indicates that people's preferences could be context-dependent (Ho et al. 2021). To find out more, we incorporated several items on the topic of dishonest behavior into a survey module on the topic of trust. The module included different statements, which respondents had to assess by reporting whether they agree or disagree. Appendix Table A.2.8 shows results for correlations of items about honesty and information preferences, revealing a robust link for this

statement: “Most people would lie to gain an advantage.” If we compare ostriches and individuals with high information preferences using the full dataset, we find that the probability of agreeing with this statement increases from 57.89% to 64.68% while it increases from 5.26% to 16.06% for strong agreement (i.e. highest category). This evidence indicates that higher preferences for inconvenient information go along with more pessimistic beliefs regarding others’ willingness to cheat in life.

We conclude that information preferences in the way measured and analyzed in our paper are behaviorally valid and are related to views on dishonest actions, as a potentially inconvenient experience. Our results are also consistent with the idea that those individuals who are more naïve in regard to other people’s honesty may take action to maintain their beliefs both in our experiment and outside in their real lives.

2.4 Conclusion

This experimental study analyzes cheating among members of a work group executing a real-effort task in a workplace context. By studying co-worker behavior when cheating takes place, compared to when it does not, we analyze the possible implications for performance as well as heterogeneity in individual responses to the dishonest act. Despite the idea that dishonesty at the workplace could be detrimental for performance, in more than 40 % of the cases, participants decide to cheat. The likelihood of dishonest behavior increases if the beneficiary is male, suggesting that decision-makers consider females as less willing to become a partner in crime. In regard to performance, we cannot find evidence for negative implications when cheating takes place, which holds for a variety of performance measures as well as for survey responses on future work motivation. If anything, we observe a tendency among co-workers to show higher work motivation after the occurrence of cheating, which we relate to the role of information preferences. Individuals with lower preferences for inconvenient information respond to cheating with higher effort, compared to the behavioral response by high-preference-types. Our survey data provides an explanation for this, as ostriches are more likely to report that distraction through working harder helps them dealing with potentially negative experiences. Furthermore, the survey data shows that information preferences are negatively related to naïveté regarding the dishonest activities of others. We conclude that some individuals turn a blind eye to other people’s lies by focusing more on their task and thereby maintain their (false) beliefs about honesty in human behavior.

Apart from contributing to ongoing research, our study also offers practical implications that could be of interest to companies and personnel management. Although individuals differ substantially in whether they exploit opportunities-to-cheat, most of the variation in choices cannot be explained by our data. Hence, instead of focusing on specific characteristics of individuals to limit the occurrence of cheating at work, companies are probably better advised to enforce a climate of honesty for that purpose. This might be a particularly good idea given our finding that many individuals prefer to look away rather than to focus on co-workers who behave dishonestly. Such a policy of establishing an honest workplace environment, however, does not appear as a promising measure to also enhance performance. For this purpose, personnel departments may rather focus on those non-monetary factors that have been shown in previous studies to be potentially effective, such as awards, gifts and motivational talk (Kube et al. 2012, Ashraf et al. 2014, Kvaløy et al. 2015, Bradler et al. 2016, Gerhards and Siemer 2016) or task meaning and mission (Fehrler and Kosfeld 2014, Gerhards 2015, Carpenter and Gong 2016, Chadi et al. 2017, Kosfeld et al. 2017). In comparison, having more or less honesty at the workplace does not seem to matter as much as one may believe, which could also help explain why little lies are ubiquitous and most likely will continue to be so.

Having said that, the behavioral implications of cheating at work could be very different if dishonesty goes at the expense of others. While misbehaviors towards co-workers could trigger severe consequences in organizations and hence may be less common, any investigation into the effects of black lies at the workplace certainly could be interesting. Another issue that we leave to future research concerns the way that the cheating is carried out. In our setup, individuals cheat secretly without exposure, implying some uncertainty regarding the occurrence of a lie, similar to other recent work in the context of cheating and information (Benistant et al. 2021). While we argue that this aligns well with the situation in organizations where telling lies secretly is easier for a potential cheater and hence more common, it is certainly conceivable to arrange a setup in which the observer knows for sure that cheating took place. Finally, cheating in our case takes place through lying about a randomly determined outcome, whereas cheating could also be carried out as part of performing a real-effort task (Kajackaite 2018). While we prefer the die-roll game to cleanly identify the effect of dishonesty on performance by ruling out psychological mechanisms that could be related to others' performance, we do acknowledge interesting avenues for further research.

2.5 Appendix A.1: Additional background information

Appendix A.1.1: Task

Here we provide details on the work task of downloading and labelling scientific articles from scientific journals, which was at the center of the experimental research project. Note that we obtained the approval of a major database of academic research materials to download scientific articles online from their digital library as part of our experimental study. As a consequence of this agreement, we could only invite students enrolled at the university due to subscription-based access to the database.

To carry out the task of preparing scientific articles, the participants first had to use keyboard and mouse to go to a journal webpage and download one article after another from a pre-selected volume. A link to this volume was always available via a bookmark, which, prior to the session start, was configured by research assistants. In a second step, participants had to label the file correctly regarding page number, author names, and title, and then move it to an existing folder. Each participant in a session had to work on a different volume from the same journal. Before the start of the project, several journals, which the university at the time of the study had access to, and three volumes from each journal were chosen. Both the selection of the journal for a work session and the allocation of the volume, which each participant had to work on, were determined randomly prior to the session.

To ensure that all participants were fully aware of how to do the task, the session host used the big screen, as can be seen in Figure A.1.2, for demonstration purposes. At the beginning of each work period, a summary of how to do the task was shown. The original text of the procedure (translated from the original German instructions) is shown in Figure A.1.8. Prior to the first work period, the session host discussed the procedure in detail and also carried out a test run together with all participants. Accordingly, the session host downloaded and labeled one exemplary article; the participants did so as well. When all three succeeded by correctly preparing one article, the first work period started. Participants were instructed to continue with their volume by going down the list of articles, downloading one after another. After the completion of one issue, they had to continue with the next issue of the same volume. In work period 2, participants had to continue with the issue they had worked on at the end of work period 1. While both work periods had the same length of ten minutes each, the second one was not announced until its start.

After each work period, the session host conducted spot checks on the work output to determine the performance-related pay. The spot-check procedure on correctness, as defined by the rules on how to do the task (Figure A.1.8), was conducted as follows. For each participant's output, the session host checked a randomly chosen output unit for correctness. Only if incorrect, the procedure continued with up to two more checks until an article without inaccuracy was found. The total number of output units was reduced by the number of incorrect units. In cases when three erroneous articles were found, the participant's output was considered to be zero. Note that checks were possible because each computer in the room was connected via an internal network, allowing an inspection of the work output during the session.

Appendix A.1.2: Instructions

[All the text in quotation marks is to be read aloud by the session host. Supplementary instructions for the session host are in italics, which include demonstrations of procedures using the available big screen. The instructions for the session host also include references to material, which can be found in this Appendix (marked by '➔').]

Welcome participants and ask them to read the general information document [➔ Figure A.1.9], *switch mobile phones to flight mode as well as put them aside together with their bags. When all participants have been arrived at the computer workstation assign them to their places by drawing lots.*

"Welcome to this scientific study, which is moderated by me, the team leader. You will be the work team, consisting of ..."

Introduce every person with the correct team player number according to the lot drawn before.

"First of all, the most important rule that you must follow is: Apart from what I say, verbal communication is prohibited until the end of the session. Misconduct leads cuts in payouts. We will communicate via email. In the following, we will test the email program installed on each computer. You can see the email addresses on the tables for everyone to read. We assigned an email address to each one of you."

Refer to each sign by hand.

"My address, as the team leader, is chef@iaaeu.de. I will send a test email to everyone now using the mailing list team@iaaeu.de."

Show the preparation of the first test email on the big screen and send it [➔ Email template 1 in Figure A.1.10].

"I will separately send out a second test email to myself and all of you by putting team@iaaeu.de in blind copy (BCC)."

Show this on the big screen using the mouse cursor.

"By putting you in BCC, you will receive an email without being able to see the other recipients of the email."

Send the second test email [➔ Email template 1 in Figure A.1.10].

"To send an email, click on 'New email' in the upper left corner of your email program. Now please write me a test email to chef@iaaeu.de and set the team mailing list, which is team@iaaeu.de, in BCC."

Wait until all emails are received.

"You can use the email function from now on. If you have any questions, you can always send me an email to chef@iaaeu.de.

In this study, the aim is to collect as many points as possible. For every point you score, you will be paid one Cent at the end. Thus, for 100 points, you will receive 1 Euro. You will begin with 250 points for your attendance, so you are guaranteed to earn 2 Euro and 50 Cent."

Send separate emails with a link to Survey 1 [➔ Email template 2 in Figure A.1.10].

"By filling out an online questionnaire completely and conscientiously, you will earn an additional 200 points. I have sent you a link via email. You have to click on the link to get to the questionnaire. Once everyone has completed the questionnaire, I will receive a confirmation and then we can continue. Please complete the questionnaire now."

Wait until all confirmation emails for survey 1 are received.

"Everyone has completed the online questionnaire. You can close your browser window. I will now explain your work task. The following is about supporting research at our institute by procuring literature. As you may know, the university has limited access to scientific journals, which are very important for research. The limited availability impedes the research at the institute. In order to make studies of certain journals directly available for research purposes, your task is now to download as many scientific articles as possible and label them correctly. On every computer desktop you find an open Explorer window that shows a folder."

Show folder using the Explorer window on the big screen.

"Behind me, you can see my folder. In this folder, the first two articles of a volume from a specific journal have been downloaded and correctly labeled. Please note that the journal's volume year is different on each computer. If you accidentally close your folder, you can find a shortcut on your desktop. Alternatively, you can use a bookmark via the 'Favorites' function in the upper left corner of the Explorer window."

Show the 'Favorites' function using the mouse cursor.

"I will now show you the website of the journal."

Show the journal website in the browser window and show the relevant information using the mouse cursor in the following.

"Your task is to continue downloading articles and labeling them correctly. As you can see on the big screen behind me, the articles are ordered from top to bottom according to the page numbers. It is important to download the articles in order, which means you should download the articles one by one, moving from top to bottom. When you have finished downloading all of the articles from a volume's issue, click on 'next issue' located at the top right of the page, and continue with downloading the articles from top to bottom. When you are finished with all of the issues from your volume, you can download the following volume's articles.
Note that you do not have to download the documents named 'Front Matter' and 'Back Matter'. The standard procedure is summarized on the big screen behind me."

Show information slide 'task instructions' on the big screen [➔ Figure A.1.8].

“Before informing you about the remuneration for the work task, I would like you all to practice the task I have just described. We will go through the process of downloading the respective third article together. Please keep in mind that we all work on different volumes and that the third article in my case is, therefore, different from yours.”

Show all steps on the big screen in the following.

“Now we will all go to the journal’s website. You will find the link as a bookmark in the upper left corner of your internet browser.

1. Click on the third article
2. Click on ‘Download PDF’
3. Confirm the terms and conditions of use, if necessary
4. Click on ‘Save document‘ on the top right side
5. Select ‘Save File’
6. The folder to store the file should already be selected. If not, you can find the folder by using a bookmark via the ‘Favorites’ function in the upper left corner of your internet browser.
7. Assign the correct file name: Page number of the article’s first page, underscore, surnames of the authors, each separated by a comma, underscore, the title of the study. You can use ‘copy & paste’ for this purpose.
8. Click on ‘Save’ to finish.”

Show the information slide ‘task instructions’ on the big screen again [➔ Figure A.1.8].

“Note that special characters such as colons, question marks, or quotation marks can cause error messages. Therefore, you have to omit these when assigning the file names.”

Check all folders to see if the files were successfully downloaded and help with problems. If a problem is not solvable via email communication, you can resort to verbal communication.

“Your remuneration is determined as follows. For each additional downloaded article, the team receives 10 points, as long as I cannot find mistakes in the labeling. The total amount of points at the end of the task is divided by the number of team players, so everybody receives the same earnings from this task. If all of you prepare three articles each, this means 30 points on average and thus 30 cent per team player. Note that the number of points you score affects your earnings but has otherwise no further consequences for you. Please keep in mind that talking is not allowed and is punished with point deductions. Of course, you can communicate at any time using the email program. If you want to ask me a question, the best way to do so is to reply to the email which I am sending you now. In this email, I am asking you to estimate the number of articles you will be able to download and label correctly during the following 10 minutes of working time as well as your estimation regarding the average performance of your team.”

Send the email about performance estimations [➔ Email template 3 in Figure A.1.10].

“As soon as I have all your estimates, I will give the starting signal.”

Ensure that the information slide 'task instructions' is on the big screen [➔ Figure A 1.8] *and wait until all confirmation emails are received. (If this is not the case after 30 seconds, say:* "I have not yet received an answer from all team players." *If this is still not the case after another 15 seconds, contact the team player directly:* "I am still waiting for your email."*)*

Work period 1

"We will now start with the work period. I will now start the timer."

Start stopwatch. Prepare the next emails during the work period and begin checking articles of the participants while they work. Announce the end of the work period after 10 minutes.

"Time is up. Please finish the last process."

Wait a moment. Send separate emails with a link to Survey 2 [➔ Email template 2 in Figure A.1.10].

"While I count and check the articles to determine the remuneration, you can receive another 200 points by completing a second questionnaire conscientiously. Please click on the link again that you received via email and start filling out the questionnaire immediately."

Determine the results by using the calculation sheet and prepare the next email. Wait until the first confirmation email for survey 2 is received.

"As soon as everyone has completed the questionnaire, I can send you the results of the work period via email."

Wait until the second confirmation email for survey 2 is received.

"In the email, I will also ask a question that you should answer via email."

Send the email about the remuneration [➔ Email template 4 in Figure A.1.10] *and wait briefly for the responses.*

Bonus game

"We will now start with a bonus game. In front of you is a 20-sided die and a dice cup, which allows us to roll the die secretly so that no one else can see your number."

Show information slide 'Bonus game instructions' on the big screen [➔ Figure A.1.6 respectively Figure A.1.7, depending on the treatment condition].

"The game consists of three steps. In the first step, I, as the team leader, roll the die to determine which one of you participates in the game as bonus player A. In the second step, player A rolls the die and replies to my email with a number from 1 to 20. Based on a randomly chosen rule, the number player A has rolled will determine who bonus player B is. In the third step, player B rolls the die and replies to my email with a number of 1 to 20. This roll determines whether or not a bonus is paid out. You can see the rule on the big screen behind me."

Show the rule by using the mouse cursor and read it aloud. Wait briefly.

"Note that a different distribution of points is not possible."

Practice round

"We will go through the bonus game one time to practice. The first step, I will now determine who is player A."

Roll secretly and simultaneously both 6-sided and 20-sided dice to determine the correct row and the correct column in the player selection sheet for the 'Selection of A' [➔ Figure A.1.11]. *Send the email to A* [➔ Email template 5 in Figure A.1.10].

"According to the number I rolled, I have determined who bonus player A is in this practice round. It is team player ...
Now step two: I have sent player A an email. As soon as my email is received, player A rolls the die and replies with a number from 1 to 20."

Wait for A's reply and continue after receiving it.

"I have received a reply. Based on the number sent to me by player A, I can now determine bonus player B. There is a 50% probability for either team player ... or ...."

Choose the same column as before in the player selection sheet and consider the reported number of A to determine the correct row in the player selection sheet for the 'Selection of B' [➔ Figure A.1.11].

"According to the number sent by player A, I have determined who takes the role of player B in this practice round. It is team player ...
In the same way I would now ask player B to reply with an email including a number from 1 to 20. This number then decides whether a bonus is paid out or not, according to the rule."

Read the rule aloud again [➔ Figure A6 respectively Figure A.1.7].

Main round

"Now we will play for money. In the first step, I will determine who is player A."

Roll secretly and simultaneously both 6-sided and 20-sided dice to determine the correct row and the correct column in the player selection sheet for the 'Selection of A' [➔ Figure A.1.11]. *Send the email to A* [➔ Email template 5 in Figure A.1.10].

"According to the number I rolled, I determined who will be player A in this main round. As soon as my email is received, player A rolls the die and replies with a number from 1 to 20."

Wait for A's reply and continue after receiving it.

"I have received a reply. Based on the number sent to me by player A, I will determine now who will be player B in this main round of the bonus game."

Choose the same column as before in the player selection sheet and consider the reported number of A to determine the correct row in the player selection sheet for the 'Selection of B' [➔ Figure A.1.11].

Send an email to player B by forwarding the email reply from A and adding the sentence: 'Please reply to this email with a number from 1 to 20'.

"As soon as my email is received, player B rolls the die and replies to my email with a number from 1 to 20."

Wait for B's reply and continue after receiving it.

"I have received an answer from player B. I will now announce the result of the bonus game via email."

Send the email with the result [➔ Email template 6 in Figure A.1.10].

Work period 2

"In the email, I will also ask you a question about the second work period that is about to start. Once you have replied to the question via email, you may continue with the work task. You will start from where you finished in the first period. This means that the next article according to the number of pages must be downloaded and labeled correctly. I will start the timer now"

Start stopwatch. Show the information slide 'task instructions' on the big screen [➔ Figure A.1.8]. *Prepare the next emails during the work period and begin checking articles of the participants while they work. Announce the end of the work period after 10 minutes.*

"Time is up. Please finish the last process."

Wait a moment. Send separate emails with a link to Survey 3 [➔ Email template 2 in Figure A.1.10].

"While I count and check the articles to determine the remuneration, you can receive another 200 points by completing the third questionnaire conscientiously. Please click on the link again that you received via email and start filling out the questionnaire immediately. It is important to fill out this third questionnaire completely. The study will only be finished after you have completed the questionnaire. This is necessary for receiving the payout."

Determine the results by using the calculation sheet and prepare the receipts for the payout. Wait until the first confirmation email for survey 3 is received.

"To be on the safe side, we ask you to delete all emails in your email program, in your inbox, in the sent folder, and the Recycle Bin."

Wait until all confirmation emails for survey 3 are received.

"The third and last online questionnaire has been completed by everyone. I have created the receipts with the amounts of money. With these you will receive your payout at the meeting room where you first arrived today. Thank you very much for participating in this study."

Appendix A.1.3: Supplementary figures

Figure A.1.1 Original announcement (translated version)

Earn money easily and support research at the same time!

IAAEU
Institute for Labour Law and Industrial Relations in the European Union

The IAAEU on Campus II is looking for participants for a study on work motivation at a computer workstation.

Requirements? Good German language skills

When? On each working day until 6 p.m.

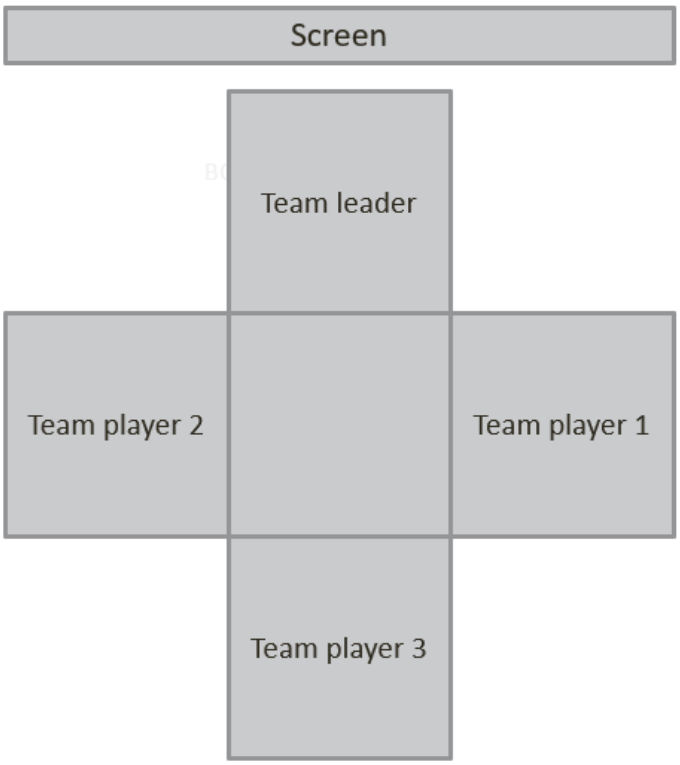
- **Duration:** approx. 1 hour
- **Payment**
  - Participation is remunerated on a variable basis, on average 12€
  - With luck and skill higher payouts are possible
  - Additionally: A lottery of 100€ among all participants!

Interested?
Register online:

www.iaaeu.de/studie

Figure A.1.2 The computer workstation

Panel A) Map

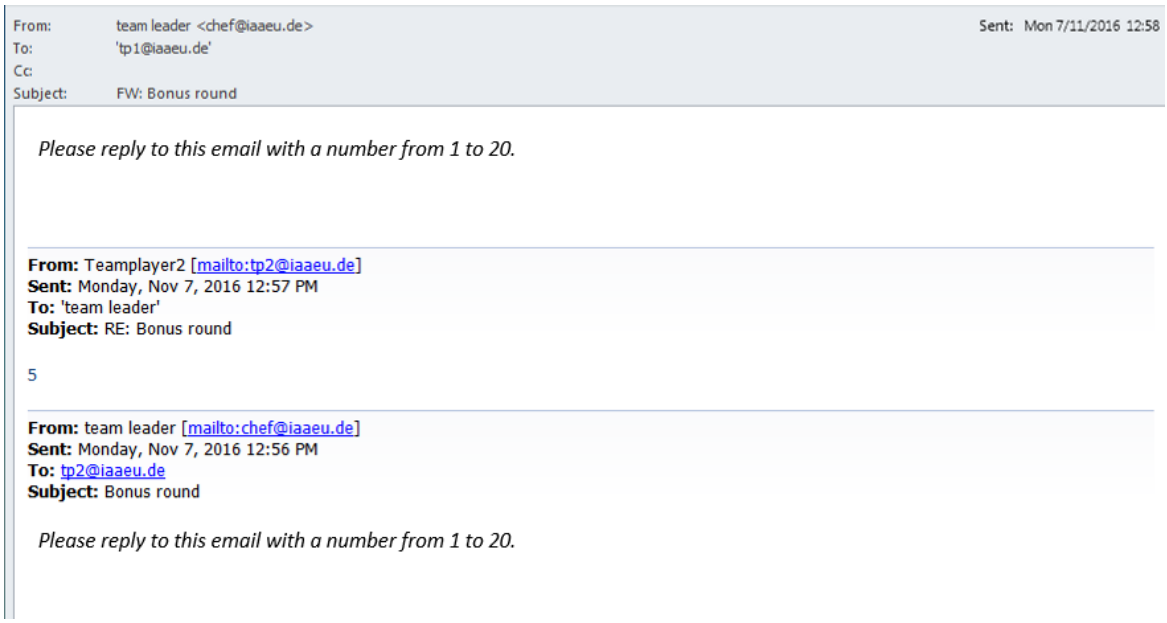


Panel B) Photo

Figure A.1.3 Timeline of the work session

	Event	
1.	Welcome	Introduction to the rules regarding communication and pay
2.	Survey 1	Items include questions on information preferences and attitudes
3.	Work instruction	Introduction to the task
4.	Work period 1	Work task (10 minutes)
5.	Survey 2	Items include questions on skills and personality
6.	Bonus game	Possible win for randomly selected bonus players A and B (including *experimental manipulation* of decision rule)
7.	Work period 2	Work task (10 minutes)
8.	Survey 3	Items include questions on mechanisms and future work motivation
9.	Payout	Pay and final feedback at the meeting room

Figure A.1.4 Email request example (translated version)



From: team leader <chef@iaaeu.de> Sent: Mon 7/11/2016 12:58
To: 'tp1@iaaeu.de'
Cc:
Subject: FW: Bonus round

Please reply to this email with a number from 1 to 20.

From: Teamplayer2 [mailto:tp2@iaaeu.de]
Sent: Monday, Nov 7, 2016 12:57 PM
To: 'team leader'
Subject: RE: Bonus round

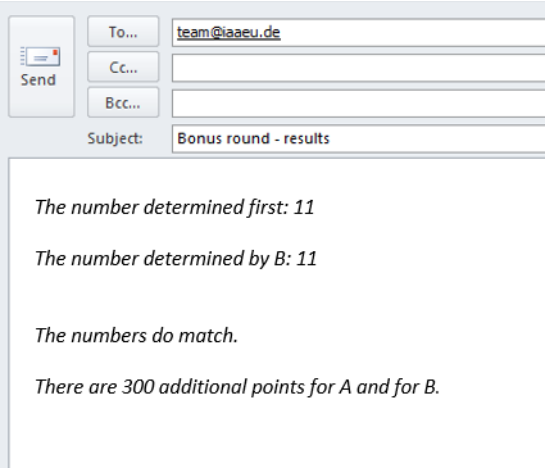
5

From: team leader [mailto:chef@iaaeu.de]
Sent: Monday, Nov 7, 2016 12:56 PM
To: tp2@iaaeu.de
Subject: Bonus round

Please reply to this email with a number from 1 to 20.

Figure A.1.5 Email example for the bonus game outcome announcement (translated version)

Bonus win



No bonus win

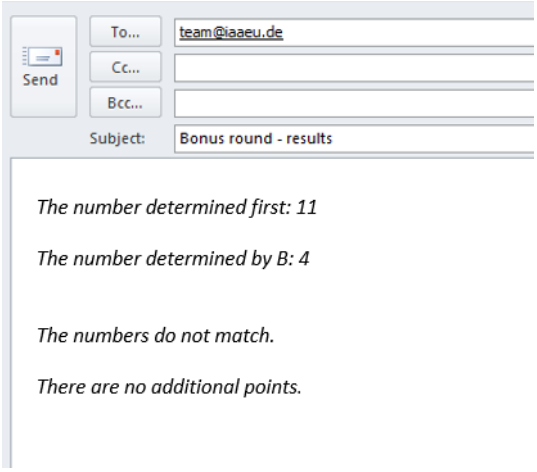


Figure A.1.6 Bonus game instructions (opportunity-to-cheat treatment)

Step 1: **Who is A?**
The team leader rolls the die.
This number decides who A is, based on a pre-determined rule.

Step 2: **Who is B?**
A rolls the die and replies to the team leader‘s email with a number from 1-20.
This number decides who B is, based on a pre-determined rule.

Step 3: **Will there be a bonus?**
B rolls the die and replies to the team leader‘s email with a number from 1-20.

Rule:
Only if the numbers of B and A match
(i.e. both numbers are identical),
a bonus is paid out (otherwise not).

Bonus:
300 points for A, 300 points for B

Figure A.1.7 Bonus game instructions (control treatment)

Step 1: **Who is A?**
The team leader rolls the die.
This number decides who A is, based on a pre-determined rule.

Step 2: **Who is B?**
A rolls the die and replies to the team leader‘s email with a number from 1-20.
This number decides who B is, based on a pre-determined rule.

Step 3: **Will there be a bonus?**
B rolls the die and replies to the team leader‘s email with a number from 1-20.

Rule:
Only if the numbers of B and the team leader match
(i.e. both numbers are even or both numbers are uneven),
a bonus is paid out (otherwise not).

Bonus:
300 points for A, 300 points for B

Figure A.1.8 Task instructions

Step 1: Download PDF

Step 2: Enter PDF title -> **Page_author1, author2, author3, etc._title**

(Example: 1471_Jolls,Sunstein,Thaler_A behavioral approach to law and economics)

Rules:

1. NO ABBREVIATIONS when listing the authors

➔ "et al." is a taboo!

2. UPPERCASE can be retained, in general, there are no specifications of upper and lower case

➔ copy & paste the title of the study retrieved from the journal's website is generally allowed

3. SPECIAL CHARACTERS partly have to be left out when entering titles

➔ This refers particularly to colons (": "), question marks (" ? ") and quotation marks (" " ")

4. SUBTITLES can be left out

➔ convert "Norm-Based Trade Union Membership: Evidence for Germany" into

"Norm-Based Trade Union Membership"

Figure A.1.9 Information document

Important information about this experiment

First of all, we would like to thank all who are participating in this study and thus supporting the research at the Institute for Labour Law and Industrial Relations in the European Union (IAAEU). We would like to point out two important aspects.

Data protection

With the registration for this study, consent was given that individual data can be generated and subsequently analyzed anonymously. All data collected will be treated in strict confidence and the dataset generated in the course of this project will be stored anonymously. All observations will differ only in terms of identification numbers. It is therefore impossible to draw conclusions about specific individuals. In addition, the rules and laws of communication customary in Germany will also apply to this experiment at the computer workstation. If, for example, communication between study participants takes place via email or in chat programs without the explicit involvement of the session host, neither we as researchers nor third parties are permitted to read any such communication. This does not include emails sent to the session host. We will delete all other emails sent via the installed email program after the experiment. For the sake of data protection, study participants can of course also delete the emails themselves.

Ethical principles

This is a study conducted by economists. In contrast to experimental research in other fields such as psychology, the ethical principle in experimental economics is that the researchers do not deceive participants. In concrete terms, this means that all statements made by the session host are truthful. In accordance with this ethical principle, it also applies that no participant in the study has received instructions in advance, for example to act as an actor. Finally, we would like to point out that the concept for this experimental study was presented to the ethics committee of the University of Trier and was approved.

Figure A.1.10 Email templates used during sessions

1

Subject:	*Test*
Content:	*Test*

2

Subject:	*Survey*
Content:	*Please click on this link* _____

3

Subject:	*Estimation*
Content:	*Reply to this email with an estimate of how many articles your team will download and correctly label in 10 minutes. Additionally estimate, in brackets after the first estimate, how many articles you will download and label correctly in the work period that is about to start.*

4

Subject:	*Remuneration*
Content:	*Your team has prepared the following number of articles: XXX* *You have thus earned the following points: XXX* *How satisfied are you with this remuneration?* *[Please answer with a value from 1 (for "completely dissatisfied") to 7 (for "completely satisfied").]*

5

Subject:	*Bonus game*
Content:	*Please reply to this email with a number from 1 to 20.*

6

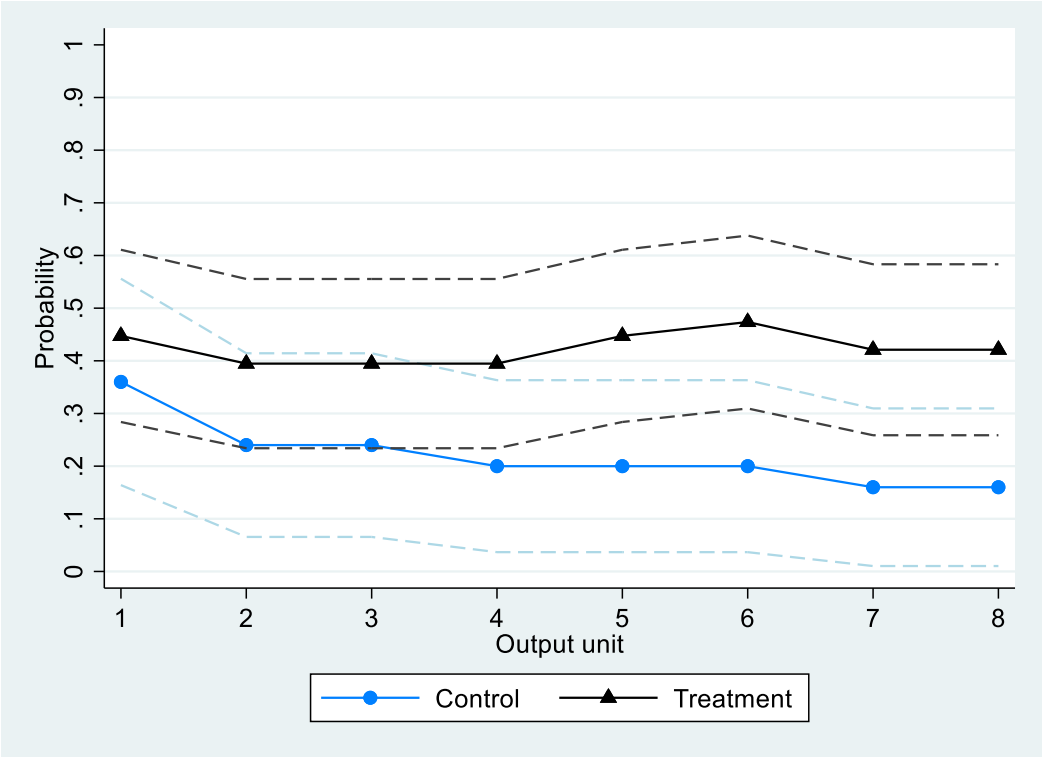
Subject:	*Bonus game - Result*
Content:	*The number determined first: XXX* *The number determined by B: XXX* *The numbers do [not] match.* *There are 300 additional points for A and for B. / There are no additional points.* *Question: How many articles will you download and label correctly in the upcoming work period?*

Figure A.1.11 Player selection sheet

	Column selection (20-sided die of session host)			
	1 to 5	6 to 10	11 to 15	16 to 20
Selection of A (6-sided die of session host)	1,2→team player 1 3,4→team player 3 5,6→team player 2		1,2→team player 2 3,4→team player 3 5,6→team player 1	
Selection of B (20-sided die of A)	even number → team player on the left odd number → team player on the right	even number →team player on the right odd number →team player on the left	even number →team player on the left odd number →team player on the right	even number → team player on the right odd number → team player on the left

2.6 Appendix A.2: Supplementary statistics and analyses

Figure A.2.1 Dynamic analysis



Notes: The illustration shows likelihoods for being the first in a session to download an article (i.e. output unit). Probabilities are determined separately for the first output unit, the second unit, and so on, in work period 2. Estimates are from standard regressions. The sample is restricted to sessions with a bonus win. Observations are from 63 individuals in the role of the bystander, who are either observed in the cheating treatment or the control treatment, in each case after a bonus win. 95% confidence intervals are displayed using dashed lines.

Table A.2.1 Sample statistics

	(1) Mean	(2) Mean	(3) Mean	(4) Mean	(5) Min	(6) Max
Skills: math	3.20	3.20	3.24	3.18	1	5
Skills: probabilities	2.97	2.99	2.92	3.08	1	5
Skills: computer	3.41	3.48	3.28	3.47	1	5
Big5: extraversion	4.97	4.80	4.73	4.86	1.33	7
Big5: agreeableness	5.41	5.35	5.20	5.23	1.67	7
Big5: openness	5.03	4.96	4.88	4.79	1.33	7
Big5: neuroticism	4.42	4.45	4.36	4.43	1	7
Big5: conscientiousness	5.19	5.13	5.11	5.02	1.33	7
Reciprocity	3.52	3.42	3.40	3.53	1	5
Trust	6.06	6.06	6.09	5.92	2.33	7
Female	0.70	0.61	0.60	0.50	0	1
Age	25.14	24.53	24.56	25.11	18	53
Number of semester	5.04	5.13	4.32	5.92	1	25
Bachelor's degree	0.36	0.25	0.24	0.32	0	1
Opportunity-to-cheat:	No	Yes	No	Yes		
Bonus win:	No	No	Yes	Yes		
Bystander only:	No	No	Yes	Yes		
N	153	255	25	38		

Notes: Means are displayed in columns 1, 2, 3, and 4. Minimum and maximum values are reported in columns 5 and 6. The data used in columns 1 and 3 are from control sessions. The data used in columns 2 and 4 are from opportunity-to-cheat sessions. The data used in columns 3 and 4 are restricted to sessions with a bonus win and to individuals in the role of the bystander (individuals excluded from the bonus game). Variables on skills are measured on a 5-point ordinal scale, ranging from "Very bad" (score: 1) to "Very good" (score: 5). 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 the same as in the German Socio-Economic Panel Study. Female is a dummy variable taken from the session records. Age, Number of semesters, Bachelor's degree are variables reported by participants during webpage registration.

Table A.2.2 Cheating and quality of work

	(1)	(2)	(3)	(4)
Dependent variable:	Correct output units		Incorrect output units	
Work period 2	2.217***	0.000	0.957***	1.000
	(0.796)	(0.727)	(0.352)	(0.696)
Opportunity-to-cheat	0.515	-1.767	0.668	1.090
	(0.945)	(1.153)	(0.421)	(0.671)
DiD: Cheating treatment	0.177	2.842**	-0.141	-0.053
(Opportunity-to-cheat X period 2)	(0.987)	(1.151)	(0.538)	(0.913)
Information preferences: High		-1.464		0.455
		(1.263)		(0.690)
DiD: Work period 2 X		3.187**		-0.062
Information preferences: High		(1.255)		(0.811)
DiD: Opportunity-to-cheat X		3.991**		-0.666
Information preferences: High		(1.796)		(0.888)
DiDiD: Cheating treatment X		-4.082**		-0.201
Information preferences: High		(1.721)		(1.158)
N	122	122	122	122
R^2	0.083	0.135	0.081	0.092

Notes: The dependent variable in the left (right) two columns is the number of correctly (incorrectly) prepared articles per individual observed in a work period, according to software-based calculations. See Table 2.2 for more information on the variables used. The sample is restricted to sessions with a bonus win. Observations are from 61 individuals in the role of the bystander (individuals excluded from the bonus game), who are either observed in the cheating treatment or the control treatment and for whom exact performance data is available. Compared to Table 2.2, the sample size is smaller because in two cases, the original files were not stored properly. Estimates are from standard regressions. Robust standard errors are in parentheses. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2.3 Dynamic analysis of effect heterogeneity in effort

	(1)	(2)	(3)	(4)
Dependent variable:	1st place (article 1)	1st place (article 2)	1st place (article 3)	1st place (article 4)
Cheating treatment	0.140	0.257	0.257	0.257
	(0.201)	(0.157)	(0.157)	(0.157)
Information preferences: High	0.042	0.201	0.201	0.139
	(0.205)	(0.161)	(0.161)	(0.156)
Cheating treatment X Information preferences: High	-0.094	-0.149	-0.149	-0.086
	(0.264)	(0.230)	(0.230)	(0.226)
R^2	0.010	0.044	0.044	0.052

	(5)	(6)	(7)	(8)
Dependent variable:	1st place (article 5)	1st place (article 6)	1st place (article 7)	1st place (article 8)
Cheating treatment	0.363**	0.363**	0.257	0.257
	(0.160)	(0.160)	(0.157)	(0.157)
Information preferences: High	0.139	0.139	0.076	0.076
	(0.156)	(0.156)	(0.148)	(0.148)
Cheating treatment X Information preferences: High	-0.192	-0.139	0.029	0.029
	(0.228)	(0.229)	(0.221)	(0.221)
R^2	0.074	0.085	0.085	0.085

Notes: The dependent variable is a dummy that is 1 if the bystander was the first in a session to download a specific article (i.e. output unit), and 0 otherwise. First places are determined separately for the first output unit (column 1), the second unit (column 2), and so on, in work period 2. See Table 2.2 for more information on the variables used. The sample is restricted to sessions with a bonus win. Observations are from 63 individuals in the role of the bystander, who are either observed in the cheating treatment or the control treatment, in each case after a bonus win. Results from linear-probability estimations are shown. Robust standard errors are in parentheses. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2.4 Cheating and future work motivation

Panel A) Dependent variable:	Motivation to perform the task again			
Cheating treatment	0.244*	0.290**	0.409**	0.512*
	(0.127)	(0.134)	(0.183)	(0.260)
Information preferences: High			0.340*	0.351
			(0.192)	(0.261)
Cheating treatment X Information preferences: High			-0.235	-0.388
			(0.247)	(0.336)
Control variables	No	Yes	No	Yes
R^2	0.059	0.253	0.109	0.285

Panel B) Dependent variable:	Motivation to work in the same team			
Cheating treatment	0.059	0.070	-0.135	-0.076
	(0.130)	(0.133)	(0.207)	(0.233)
Information preferences: High			-0.056	-0.077
			(0.214)	(0.278)
Cheating treatment X Information preferences: High			0.371	0.282
			(0.266)	(0.322)
Control variables	No	Yes	No	Yes
R^2	0.003	0.176	0.065	0.196

Panel C) Dependent variable:	Preferences for teamwork in the future			
Cheating treatment	0.083	0.063	0.234	0.351**
	(0.118)	(0.121)	(0.197)	(0.174)
Information preferences: High			0.194	0.370
			(0.204)	(0.234)
Cheating treatment X Information preferences: High			-0.247	-0.518**
			(0.249)	(0.256)
Control variables	No	Yes	No	Yes
R^2	0.008	0.308	0.028	0.369

Notes: The dependent variable in each panel is a median-split dummy established based on reported scores on 5-point scales, as provided in response to these questions: Panel A) "Would you want to do this task again?" ("No, under no circumstances" [1] - "Yes, definitely" [5]). Panel B) "Would you want to work with this team constellation again?" ("No, under no circumstances" [1] - "Yes, definitely" [5]). Panel C) "Would you prefer to work in a team or alone in your later professional life?" ("Definitely alone" [1] - "Definitely in a team" [5]). See Table 2.2 for more information on the variables used. Observations are from 63 individuals in the role of the bystander, who are either observed in the cheating treatment or the control treatment, in each case after a bonus win. Control variables are those shown in Table 2.1. Results from linear-probability estimations are shown. Robust standard errors are in parentheses. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2.5 Channels of team performance (full survey battery)

Variable definition:	Full score			Median-split		
	(1) Mean	(2) Mean	(3) *p*-value	(4) Mean	(5) Mean	(6) *p*-value
I) Trust in the team	4.72	4.90	0.639	0.56	0.68	0.324
II) Free-rider intention	1.32	1.37	0.775	0.20	0.32	0.319
III) Unfair situation	2.68	2.76	0.861	0.44	0.50	0.647
IV) Treated unfairly	2.80	2.66	0.764	0.44	0.50	0.647
V) Assess my performance	3.68	3.76	0.819	0.48	0.58	0.449
VI) Team identification	3.64	3.71	0.852	0.52	0.58	0.651
VII) Not let down the team	4.92	4.97	0.907	0.40	0.45	0.716
VIII) Work atmosphere	4.64	5.03	0.252	0.44	0.66	0.090
Opportunity-to-cheat:	No	Yes		No	Yes	
N	25	38		25	38	

Notes: Observations are from 63 individuals in the role of the bystander, who are either observed in the cheating treatment or the control treatment, in each case after a bonus win. Means of survey responses are displayed in columns 1 and 2 (full score) as well as columns 4 and 5 (median-split dummy). *p*-values of two-sided t-tests are shown in columns 3 and 6, in each case testing for significant differences between control and cheating treatment. The question battery starts with: "What is your view regarding this statement about the previous teamwork?" The statements are:

I) "I have great trust in the other persons in this room."
II) "I lean back and let others do the work for me."
III) "I think what is happening here is unfair."
IV) "Personally, I feel treated unfairly."
V) "I can assess my performance within the team well."
VI) "I identify myself with my team."
VII) "I do not want to let down my team."
VIII) "The work atmosphere is good."

The response scale ranges from "Totally disagree" (score: 1) to "Totally agree" (score: 7).

Table A.2.6 Reasons for work effort (full survey battery)

Variable definition:	Full score			Median-split		
	(1) Mean	(2) Mean	(3) *p*-value	(4) Mean	(5) Mean	(6) *p*-value
I) Important contribution	5.00	5.16	0.667	0.40	0.55	0.243
II) Felt monitored	2.80	2.87	0.871	0.44	0.47	0.797
III) Image concerns	3.16	3.26	0.808	0.44	0.47	0.797
IV) Competition	3.88	4.42	0.221	0.48	0.61	0.336
V) Role model	3.88	3.90	0.972	0.36	0.37	0.947
VI) Inspired	4.52	4.45	0.863	0.64	0.55	0.499
VII) Exclusion fear	3.28	3.21	0.899	0.48	0.55	0.579
VIII) Negative experience	2.72	2.47	0.555	0.44	0.42	0.884
IX) Help others	4.80	4.97	0.670	0.32	0.42	0.427
Opportunity-to-cheat:	No	Yes		No	Yes	
N	25	38		25	38	

Notes: Observations are from 63 individuals in the role of the bystander, who are either observed in the cheating treatment or the control treatment, in each case after a bonus win. Means of survey responses are displayed in columns 1 and 2 (full score) as well as columns 4 and 5 (median-split dummy). *p*-values of two-sided t-tests are shown in columns 3 and 6, in each case testing for significant differences between control and cheating treatment. The question battery starts with "Now we want to find out about why you have put in more or less effort. I have put in effort, because ..." The statements are:

I) "...my performance was of great importance to the team results."
II) "...I felt like I was observed."
III) "... it was important to me what others might think about me."
IV) "... the competition for the best performance motivated me."
V) "...I want to be the team's role model through my performance."
VI) "...the team's performance level motivated me."
VII) "... I did not want to be excluded from the team."
VIII) "...it helps me better deal with negative experiences."
IX) "... I like to do something for others."

The response scale ranges from "Totally disagree" (score: 1) to "Totally agree" (score: 7).

Table A.2.7 Information preferences and information avoidance (correlation matrix)

Panel A) N=408	**(I)**	**(II)**	**(III)**	**(IV)**	**(V)**
I) Case: exam cheating	1				
II) Case: doping in sports	0.526***	1			
	[0.000]				
III) Case: partner betrayal	0.134***	0.087*	1		
	[0.007]	[0.080]			
IV) Desire to know ratings	0.215***	0.171***	0.118**	1	
	[0.000]	[0.001]	[0.017]		
V) Information preferences	0.011	0.053	0.193***	0.152***	1
	[0.828]	[0.289]	[0.000]	[0.002]	

Panel B) N=63	**(I)**	**(II)**	**(III)**	**(IV)**	**(V)**
I) Case: exam cheating	1				
II) Case: doping in sports	0.573***	1			
	[0.000]				
III) Case: partner betrayal	0.271**	0.289**	1		
	[0.032]	[0.021]			
IV) Desire to know ratings	0.197	0.345***	0.480***	1	
	[0.123]	[0.006]	[0.000]		
V) Information preferences	0.133	0.217*	0.420***	0.334***	1
	[0.299]	[0.087]	[0.001]	[0.007]	

Notes: Full sample is used in Panel A). The sample in Panel B) is restricted to bystanders observed in sessions with a bonus win. Correlations are shown. *p*-values are in brackets. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The wording of the questions is as follows:

I) "Imagine you have written an exam and could observe how two rows in front of you, a fellow student used a suspicious piece of paper. Do you want to know if the student cheated?"

II) "Imagine that an Olympic champion is suspected of doping, but is declared innocent by the authorities. Do you want to know if the athlete cheated?"

III) "Imagine you are recently single because a long-lasting relationship has come to an end. Do you want to know if you were cheated on in that relationship?"

IV) "Finally, we would like to remind you that the data is collected anonymously. If this were not the case, would you want to know which assessments have just been given about you?"

V) "Do you want to be informed about something even when the information could be uncomfortable?"

Responses are reported on a 10-point scale from 1 ('No, under no circumstances') to 10 ('Yes, under all circumstances'), except for a 5-point scale in case of question IV).

Table A.2.8 Information preferences and attitudes towards honesty (correlation matrix)

Panel A) N=408	**(I)**	**(II)**	**(III)**	**(IV)**	**(V)**	
I) Most people lie for gain	1					
II) Lying to help me is good	0.053	1				
	[0.283]					
III) Others should be honest	-0.071	-0.267***	1			
	[0.150]	[0.000]				
IV) I feel bad about lying	-0.051	-0.301***	0.476***	1		
	[0.301]	[0.000]	[0.000]			
V) People are honest to me	-0.269***	-0.077	0.137***	0.117**	1	
	[0.000]	[0.120]	[0.006]	[0.018]		
VI) Information preferences	0.127**	-0.112**	0.182***	0.163***	0.020	1
	[0.010]	[0.023]	[0.000]	[0.001]	[0.683]	

Panel B) N=63	**(I)**	**(II)**	**(III)**	**(IV)**	**(V)**	
I) Most people lie for gain	1					
II) Lying to help me is good	-0.057	1				
	[0.656]					
III) Others should be honest	-0.107	-0.286**	1			
	[0.403]	[0.023]				
IV) I feel bad about lying	-0.179	-0.438***	0.413***	1		
	[0.160]	[0.000]	[0.001]			
V) People are honest to me	-0.420***	-0.140	0.280**	0.152	1	
	[0.001]	[0.273]	[0.026]	[0.234]		
VI) Information preferences	0.283**	0.202	-0.037	-0.144	0.000	1
	[0.025]	[0.112]	[0.775]	[0.259]	[1.000]	

Notes: Full sample is used in Panel A). The sample in Panel B) is restricted to bystanders observed in sessions with a bonus win. Correlations are shown. *p*-values are in brackets. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Variables are from a survey module that starts as follows: "Trust is an important topic in economic research. How much do you agree with the following statements?" The first statement (left out in the correlation analyses of this table) is "In general, you can trust people." The wording of the remaining statements of this survey module is as follows:

I) "Most people would lie to gain an advantage."
II) "When someone lies to help me, I appreciate it."
III) "It is important to me that others are honest."
IV) "If I had to lie, I would feel bad."
V) "Most people are honest with me."

The response scale ranges from "Totally disagree" (score: 1) to "Totally agree" (score: 5). For information on item VI) Information preferences, see Table A.2.7.

2.7 Appendix A.3: Information experiment

In May 2017, we organized a brown bag seminar for which we invited all participants of the experimental research project. The idea was to publicly determine the winner of the 100-Euro lottery, which served as an additional incentive for students to participate in the experiment, as well as to present our first results to those who were interested. As we were aware that some participants wanted to know the results of the study and to receive performance feedback, we decided to organize the seminar, and we provided our research assistants with email blueprints to send out invitations. Another motive for us was to use the seminar setting for a follow-up experiment on information avoidance behavior.

The idea of the experiment was to make seminar participation more or less attractive depending on participants' preferences for inconvenient information. As the location was a ten-minute walk away from the main campus of the university, taking part entailed substantial opportunity costs. To add another benefit of showing up at the location, we offered the option to take a look into the data that each participant had generated for us. We randomized information about this option to manipulate the costs and benefits from participating in the seminar by slightly modifying the text in the invitation email:

[…] As announced, we offer all participants the opportunity to hear about the results of our study and cordially invite you to a public lecture as part of our IAAEU brownbag seminar. The lecture will take place next Thursday (May 11) in the institute's lecture hall (room H 714, 7th floor) on campus 2. The lecture starts at 12:15 and lasts an hour. Afterward, we will raffle off the 100 euro note in the lecture room, so that the winner can get the money immediately.

[Information treatment 'control':] *In case you are interested, there is also the option to take a look at the data we collected about you.*

[Information treatment 'performance':] *In case you are interested, there is also the option to take a look at the data we collected about you, which, for example, allows you to find out about your performance in the work task.*

[Information treatment 'performance rank':] *In case you are interested, there is also the option to take a look at the data we collected about you, which, for example, allows you to find out about your performance rank in the work task.*

We would be happy to hear from you. To participate in the lecture, sign on via email by May 8 (12:00).

In the control treatment, participants were told that there is an option to get information without any further specifics. We had no such statement in a baseline treatment. We considered two further information treatments, one mentioning the performance data and the other announcing that it was possible to find out about the performance rank. The latter makes it clear that those who underperformed in the task would potentially be exposed to inconvenient information, creating a decision context in which information preferences could become very relevant. Specifically, underperformers were expected to shy away from the seminar more often, the lower their preferences for inconvenient information were if (and only if) they have reason to be concerned about exposure to their underperformance, which we indicated with our performance rank treatment.

Our setting provides us with several indicators of interest in the seminar as outcomes of our information manipulation: Signing on to the seminar via email (immediately or after receiving a reminder) and actually participating. We were able to merge information on all three outcomes with the full dataset from our experimental investigation. For this purpose, each student received a special code after signing on via email, which had to be reported to our research assistants on location before entering the seminar. Thereby, we could check who showed up instead of just announcing interest in seminar participation.

The results from our information experiment provide support for the idea that preferences for inconvenient information matter for individual behavior. Table A.3.1 presents the results where we compare the interaction effect between the performance rank treatment and our information preferences measure from the online survey. We do this for all three outcome variables as well as for each of the two subsamples, which are divided into high versus low performance in the work task. For underperformers, we find robust and significant interaction effects across all outcomes (columns 2, 4, and 6), implying that in the information treatment, the likelihood of being interested in the seminar increases with preferences for inconvenient information. Here, seminar participation could lead to exposure of having performed poorly relative to others, and so the likelihood of signing on to the seminar as well as actually showing up increases with information preferences.

Table A.3.1 Information option, preferences for information and behavior

Dependent variable:	Signing on (early)		Signing on (any time)		Participation in the seminar	
	(1)	(2)	(3)	(4)	(5)	(6)
Information treatment: Performance rank	0.043	0.024	0.005	0.031	-0.014	0.035
	(0.039)	(0.046)	(0.062)	(0.055)	(0.052)	(0.049)
Information preferences	0.004	-0.025	-0.021	-0.023	-0.026	-0.026
	(0.007)	(0.017)	(0.019)	(0.019)	(0.017)	(0.018)
Information treatment X Information preferences	-0.069*	0.057**	-0.029	0.059**	-0.007	0.056**
	(0.039)	(0.026)	(0.049)	(0.027)	(0.051)	(0.025)
Task performance:	High	Low	High	Low	High	Low
N	197	211	197	211	197	211
R^2	0.065	0.023	0.015	0.017	0.017	0.022

Notes: The dependent variable in columns 1 and 2 is a dummy that is 1 if the participant signed on early for the seminar, i.e. before reminders were sent out, and 0 otherwise. The dependent variable in columns 3 and 4 is a dummy that is 1 if the participant signed on for the seminar, independent of the point in time (i.e. before or after reminders were sent out), and 0 otherwise. The dependent variable in columns 5 and 6 is a dummy that is 1 if the participant took part in the seminar, and 0 otherwise. 'Information treatment' is a dummy that is 1 if the participant had received an email invitation containing information regarding the option to find out about the performance rank in the work task, and 0 otherwise. 'Information preferences' is a variable with scores from 1 ('No, under no circumstances') to 10 ('Yes, under all circumstances'), as reported in response to the survey question on information preferences (wording: "Do you want to be informed about something even when the information could be uncomfortable?"), which was de-meaned for the analyses shown in this table. Based on a median split, the full dataset of 408 participants was divided into 'high' task performers (columns 1, 3, and 5) and 'low' task performers (columns 2, 4, and 6) in the work task, as measured by the individual number of output units according to the session records. Results from linear-probability estimations (without control variables) are shown. Robust standard errors are in parentheses. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In detail, we determine for each additional score on the information-preference scale that the likelihood of showing up increases by three percentage points when we consider the results in row 2 (-2.6 pp.) and row 3 (5.6 pp.) of column 6. This is a meaningful effect given a mean participation rate of roughly ten percent in the full sample. The picture differs for high performers, where we observe a weak negative interaction effect for the likelihood of signing on immediately (column 1). This suggests that information preferences could also matter in case of positive information about performance, but here the ostriches are more attracted by finding out about their (good) performance. Such a potential pleasure effect however does not seem to compare to the potential negative feeling of finding out about bad performance, given that for the two other outcomes, the effect becomes insignificant (columns 3 and 5).

We conducted several additional analyses and checks of the results. First, both the performance treatment as well as the control treatment do not significantly interact with information preferences, suggesting that the threat of being exposed to potentially inconvenient information has to be made clear, as we did with our performance rank treatment. Second, we also tested an alternative explanation of our main finding, which is that mentioning the term 'rank' in our treatment text could manipulate people's beliefs on the number of participants in the experimental project. We thus added the actual number of participants randomly in the email text to find out whether this interacts with the information treatment. While our main finding for the low performers is not sensitive in this respect, we observe some heterogeneity for high performers, suggesting that for the pleasure of finding out about good performance, the number of others who are participating could be relevant. Finally, we checked our main results along several lines: i) considering control variables in the regression model, ii) splitting the data into high and low performers by using an alternative performance indicator (the total number of articles), and iii) using a fixed threshold of 25 output units, instead of using median splits, which we also did for the alternative performance indicator. All these checks do not lead to different results and thereby confirm our main conclusion from the information experiment, which is that self-reported information preferences are behaviorally valid, as they explain variation in avoidance of potentially inconvenient information.

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

We experimentally analyze the behavior in a workplace environment, in which either two or three study participants of a university research project perform a real-effort task that is compensated by a team piece rate. We examine data on individual task performance and workplace perceptions via surveys to analyze the behavioral implications of manipulating team size. We find that an increase in team size is not harmful to team performance, which is robust across performance dimensions and robust towards the introduction of pay inequality between two work periods. Peer effects appear to alleviate the free-rider problem when team size increases. Contrary to discussions of mechanisms in previous research, our survey evidence speaks against the idea of peer pressure leading to positive performance spill-overs.

*This chapter is joint work with Adrian Chadi. We are grateful for comments by Laszlo Goerke and participants of the AKempor conference in Salzburg, the team dynamics and peer effects workshop in Trier 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.

3.1 Introduction

From numerous perspectives, be it that of a personnel department or an educational institution, the determinants of team outcomes are of interest. Research in various fields such as economics, management and applied psychology has led to a large number of practical guides and books published on this topic of how to improve team performance (Salas et al. 2013, Chiocchio et al. 2015, Griffith and Dunham 2015, Ben-Hafaïedh and Cooney 2017). Although the number of team members is an important quantitative characteristic, the effect of team size on performance has not been analyzed comprehensively so far. One explanation for this could be the difficulty to randomly assign teams to identify empirically the causal effect for outcomes, such as worker productivity. In a workplace context, the random assignment appears necessary for identification, given that the selection into teams is a relevant phenomenon (Dohmen and Falk 2011, Bäker and Mertins 2013, Kuhn and Villeval 2013, Kölle 2017). A second 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 and Demsetz 1972, Newhouse 1973, Holmström 1982). Accordingly, if the number of team members rises, the incentive to free-ride increases as well.

Personnel economists discuss task complementarities, specialization advantages and knowledge transfers between team members, as reasons to deploy teams (Lazear 1995). However, even if one abstracts from these factors, there are still good reasons 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 may be countered by psychological mechanisms, such as peer pressure. Experimental research in the field (Babcock et al. 2015) and in the laboratory (Corgnet et al. 2015a) provides some evidence in support of the idea that peer effects, which may result from mutual monitoring, social pressure or other aspects related to teamwork, have the potential to mitigate the free-rider problem and to thereby foster performance at work. Arguably, such behavioral phenomena may be able to also compensate economic disincentives that result from increasing the size of the team. Having said this, psychological mechanisms could in principle play the opposite role, as, for example, perceived unfairness might foster free-riding in larger teams and thus impair the willingness to put in effort into the teamwork. In general, there is a lack of research on this idea of a behavioral compensation of free-rider

incentives, which particularly concerns the consequences for performance when team size increases and the role of peer effects in this context.

The aim of this paper is to investigate how team size affects worker productivity and whether peer effects can solve the free-rider problem that could occur when the number of team members increases. By combining exact data on performance with rich survey data, we are particularly interested in the psychological motives underlying individual decisions to contribute to teamwork. To do so, we examine a workplace setting of a computer office at a German university with either two or three study participants performing a real-effort task, similar to a routine job of student assistants. Thanks to a team piece rate, individual payments varied depending on the average individual output level across all group members. While this emphasizes the idea of teamwork, it also implies stronger incentives to free-ride when working together with more beneficiaries of an individual's effort levels.[24]

We experimentally manipulate whether two or three team members have to accomplish the task while working together in the same office. We use the terms 'smaller teams' and 'larger teams' for these two conditions while noting that the participants were not aware of the manipulation of team size. The opportunity to study behavior in a controlled environment allows us to establish credible evidence on the effect of team size for task performance, to compare the occurrence of peer effects across smaller and larger teams, and to explore survey data to find out why those effects may or may not occur. Thanks to a bonus game between the two work periods, we can furthermore shed light on whether the team-size effect is robust over time towards randomly introduced inequality. Another feature of the workplace setting is the availability of precise information on performance in different dimensions, as we observe the total number of output units as well as the quality of work. This allows examining the possibility that free-riding takes place in the form of individuals reducing the quality of work while keeping the level of total output constant.

Our study yields the following results. Average performance levels are almost exactly the same across both team size conditions, which counters the idea of more free-riding in larger teams, compared to smaller teams. This remains when we i) analyze alternative performance

[24] 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.

indicators, ii) compare the results across two work periods, and iii) differentiate between equal and unequal pay in consequence of random bonuses that lower the perceived fairness in larger teams. In line with our main finding of robust performance levels across team size, the evidence from the survey confirms that on average there is no manipulation of any effort mechanism, such as willingness to free-ride. However, several of our survey findings indicate a potential for heterogeneous behavioral effects due to changes in team size. For example, an increase in the number of team members goes along with less transparency of peer performance as well as more positive perceptions of peers with respect to their role for one's own performance. This suggests heterogeneity in peer pressure, raising the question of whether peer effects play a different role across team size and thereby explain the lack of free-riding in larger teams.

By taking effect heterogeneity across different levels of peer performance into account, we find evidence that psychological motives for putting in the effort 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'). This phenomenon interacts with team size and is limited to larger teams only. 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 the routine computer task. A split between high-skilled peers and low-skilled peers reveals that individual performance in larger teams is lower than in smaller teams if peer skills are low. This underlines the idea that the free-rider problem in larger teams exists but can be mitigated thanks to positive peer effects, induced by high-skilled and hence high-performing team members.

To better understand the heterogeneity in peer effects across team size, we further exploit the survey data by investigating potential mechanisms. Peer pressure could be very relevant in this context, given that it depends on the team's size. Yet, the evidence apparently runs counter to expectations, concerning a positive role of peer pressure, as suggested by previous research on peer effects. In fact, we find that individuals in larger teams perceive less pressure when peer skills are high, as compared to individuals in smaller teams. In consequence, it appears that high peer pressure prevents positive peer effects to occur and rather triggers adverse effects in the performance of smaller teams, which is reflected in the performance data on the quality of work. By considering different performance dimensions, we find that the amount of incorrect output units increases in smaller teams, when peer skills are high, while the amount of correct work output declines. All this is inconsistent with the idea of peer pressure improving team

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.

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 in larger teams, free-riding is a possibility, leading to worse outcomes compared to smaller teams. Therefore, it makes sense for personnel management to complement larger teams with high-performers to ensure that free-riding can be prevented by positive peer effects. As another lesson learnt, 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.

With regard to the academic discussion on teamwork, we contribute to several strands of research. First, 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 and Oyer 2013) is rather silent on the role of team size and its impact on productivity so far. A reason for that may be that exogenously manipulating team size is hardly practical when investigating team performance, especially in the field. Hence, team size is often solely analyzed in a robustness check in order to reinforce the main outcome (Hamilton et al. 2003, Friebel et al. (2017), Englmaier et al. 2018, Delfgaauw et al. 2020)[25]. Second, we contribute with our findings to research on peer effects in workplace performance (for a review, see Herbst and Mas 2015), which has not provided a consistent picture regarding the underlying mechanisms yet[26] 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.

[25] Hamilton et al. (2003) mention that despite other expectations increasing team size does not decrease worker productivity in their field experiment. Friebel et al. (2017) predict in their model that individual efforts decrease with team size. However, find some evidence in the experiment that bonus effects increase with team size. Englmaier et al. (2018) observe that group size has a positive effect on performance. Noteworthy, they complement their field experiment with a laboratory experiment but decided to keep team size constant at three individuals in this additional investigation. Delfgaauw et al. (2020) do not report a negative or positive effect of team size on performance.

[26] Additionally, the research on peer effects, which in the laboratory focuses mainly on the comparison between peer groups of identical size (e. g. Bonein 2018) or the comparison between one individuum working alone and two peers working together (e.g. Falk and Ichino 2006), faces the identical problem regarding the endogeneity of group size when analyzing peer effects in the field (e.g. Cornelissen et al. 2017).

In the following Section 3.2, we describe our experimental setup in more detail, whereas supplementary material and information is provided in Appendix B.1. In Section 3.3, we present the results of our empirical investigation, which we complement with additional evidence provided in Appendix B.2. Based on the summary of results, we provide a discussion of the relevant literature and our contributions to ongoing research on teams and peer effects in the concluding Section 3.4.

3.2 Experiment

3.2.1 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. This flexible earnings component aligned with usual hourly wage levels for a job as a student assistant at the university during that time. 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 B.1.1 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 were randomly assigned to session participants and 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 and waiting rooms were separated from each other. Prior to the start of the session, participants were guided from the waiting room to the computer workplace by a research assistant. All sessions took place in the same computer office (see Appendix Figure B.1.1 for pictures).

To investigate the effect of team size on performance, 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, for which a sizable number of observations from three-person sessions was needed. As part of the procedure (see Appendix B.1.1 for more details), three-person sessions were not cancelled 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 non-randomly assigned two-person sessions. Note that we refer to those as 'transformed' two-person sessions in the remainder of the paper.[27]

The session host was a research assistant employed at the institute who guided the participants through the session by reading out 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 (see Appendix B.1.2 for the full translated script).[28] 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 in the office were connected to the internet. This was necessary to allow participants to work on their task.

The work task was to prepare scientific papers for research at the university by downloading and labelling journal articles from an online library, which was done in similar fashion by

[27] As the only difference between regularly assigned and transformed two-person session, we implemented a verbal treatment in the latter case to find out more about possible psychological effects. Arguably, individual behavior in transformed sessions might differ not only because of selectivity but also because of psychological effects. For example, the behavior of research assistants could be different in a situation in which a person was missing unintendedly, which in turn may affect perceptions and hence behavior of participants. Therefore, 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 (see instructions in Appendix B 1.2). Analyses of this manipulation do not clearly support the idea of psychological effects, as we do not find statistically significant differences in task performance. Yet, given a rather weak manipulation with just one statement, we cannot fully rule out that psychological effects could have occurred in transformed two-person sessions, which adds to possible selectivity in regard of individual characteristics.

[28] 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 Appendix Figure B.1.1, the computer screen of the third workplace was turned off and equipment was removed in case of a two-person session. To check whether this was in any way revealing regarding our experimental manipulation, we asked research assistants helping with organizing the sessions to carefully and comprehensively take notes on any striking features regarding the participants' behavior. 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.

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 Appendix Figure B.1.1), while also reading the instructions out loud. The task itself consisted of two work periods, lasting ten minutes each, while the clock of work period one started only after all session participants had finished a test run of successfully preparing one article. Figure 3.1 displays the timeline of a session.

Figure 3.1 Timeline of the work session

	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

After the session, participants received their payments separately in the waiting rooms in which they had been welcomed beforehand. 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 entitled 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 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, which allows us to investigate the role of pay inequality for the effect of team size.[29] Randomly, 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).[30]

3.2.2 Data

3.2.2.1 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 the performance-related pay. The session host followed clear rules to determine the performance, including a spot check on possible mistakes in the work output (see instructions in Appendix B.1.2).[31]

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 labelling. We then exploit the opportunity to analyze more deeply the qualitative dimension of performance by making use of computer software that was developed by the IT department of the research institute. The software conducts precise inspections of files produced in the sessions and identifies possible deviations from fully correct article labels. Thereby, it is

[29] Appendix B.1.2 and B.1.3 provide detailed information on the procedure of the bonus game. Note that the game was included in the setting as an important feature of the above-mentioned parallel experiment, which deals with the implications of dishonesty at work (Chadi and Homolka 2021a). Accordingly, there are 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.

[30] 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.

[31] The procedure was as follows. From all the articles downloaded by a participant in a work period, the session host checked a randomly chosen output unit for 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 rare case of three erroneous articles being detected, the individual work output of this team member was reduced to zero (to avoid spending too much time on this spot check), which happened only a few times (see Appendix Figure B.2.1).

possible to detect even small mistakes that a participant made when labelling 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') in further steps of our analysis.

3.2.2.2 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 3.1).

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 about how they perceived teamwork, which included free-rider intentions, among other things (Appendix Table B.2.1).[32] This module was repeated to allow for a before-and-after comparison regarding the effects of the bonus game. For this purpose, we included two questions on unfairness: perceiving the situation as unfair and feeling treated unfairly. 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 (Appendix Table B.2.2). Third, we asked participants about perceptions regarding their peers and channels of peer effects, such as psychological pressure (Appendix Table B.2.3).

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 five factors) as well as reciprocity and trust attitudes. Finally, we included questions on self-estimated skill levels in different areas, including mathematics, probability estimation and working with

[32] As an example, the module included one item to test whether members of larger teams are more likely than members of smaller teams to identify themselves with their team, which refers to the discussion on the role of team identity in preventing shirking (Eckel and Grossman 2005).

computers, all of which were obtained on a 5-point scale. To ease the interpretation of the effects in our empirical analysis, we redefine these skills variables to range from -2 (very lowly skilled) to 2 (very highly skilled), so that 0 always reflects mid-level skills.

3.2.2.3 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 3.1 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 3.1 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 clearly 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 comes 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 (columns 3), Kruskal-Wallis tests reveal significant gender differences, as the only statistically significant result.[33] Finally, by comparing data from transformed sessions (column 2) separately with the data from three-

[33] 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. When we check for measurement errors in this variable, which in principle could differ across team size, we cannot find evidence in support of this in deeper inspections of the data. As we know the first name of most participants, a measurement error is very unlikely.

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.[34]

Table 3.1 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, 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.

We conclude that there is reason to double-check any empirical result of our analysis regarding robustness across datasets. We hence proceed in the following analysis by first

[34] To learn more about possible selection effects, we make use of the available data to find out whether there is evidence of selectivity regarding individuals not showing up. While this could contribute to a potential selection bias among the remaining individuals of transformed sessions, our deeper analysis provides no evidence for this. We nevertheless cannot fully rule selection bias out, given that we can only use limited data from the webpage registration when analyzing individuals who did not participate. Similarly, by using the same information from the webpage registration, we conduct another analysis by comparing individuals as originally assigned to either two-person sessions or three-person sessions. This analysis reveals no evidence indicating imperfect randomization.

making use of the full sample with 271 observations and then reporting on whether results based on the dataset with 195 observation differ. To ease the flow of reading, we do so only if results indeed turn out to be different in terms of statistical significance. Furthermore, we check whether our main results are robust when we consider control variables in regression analyses.

3.2.3 Expectations

Our expectations rely on the previous literature in personnel economics on teams. First, team size is expected to reduce performance levels, because of increased incentives to free-ride on others' performance (Alchian and Demsetz 1972, Newhouse 1973, Holmström 1982). Second, peer effects could help to overcome the free-rider problem when the size of a team increases, and peer pressure could be a relevant mechanism to tackle disincentives (Kandel and Lazear 1992, Backes-Gellner et al. 2015).

In this literature on free-riding in teams, various psychological mechanisms have been proposed as potential explanations of why incentives to free-ride could be compensated, and peer pressure has received particular attention so far. For example, Corgnet et al. (2015a) 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 another experiment on peer pressure, Bonein (2018) investigates how the display of co-worker effort matters for social comparisons with other workers.

Regarding the role of team size, one could argue based on the previous literature that larger numbers of hard-working peers are able to put more pressure on others, thereby fostering the occurrence of positive performance spillovers. In contrast to a single peer, several peers working together might be more likely to establish a strong group norm of hard work that pressures all co-workers to exert high efforts. In this context, Carpenter et al. (2009) discuss the idea of establishing a norm of contributing to production as a means to prevent shirking in large teams. Furthermore, using a public-goods scenario, Carpenter (2007) shows that monitoring could be of particular importance to tackle the free-rider problem when group size increases. Arguably, having more peers watching their team members implies more mutual monitoring, which could increase the perceived pressure in those team members who may consider to shirk. Yet, on the other hand, one could argue that peer pressure might be mitigated in larger teams, given that individuals are always in two roles: As emphasized in Buechel et al. (2018),

individuals are not only monitored by their peers and thus pressured to work, they are also monitoring others, which generally takes time that could rather be spent productively on performing a task. Hence, if there are more individuals to be monitored, this makes it less likely that peers will actually identify a potential shirker with low willingness to contribute to the team's success. This is a particular issue if individual contributions to team performance become less transparent in larger teams, which might further reduce perceived pressure when comparisons with others become more difficult. In summary, the role of peer pressure is an open question in our research context.

There are other psychological mechanisms of team performance which also could mitigate free-riding but have not been investigated regarding interactions with team size so far. For example, based on the idea that team identity may deter free-riding (Eckel and Grossman 2005), one could expect positive performance effects if working in larger teams is perceived more positively. Another example is the perceived level of competitiveness, which could play a role for the occurrence of peer effects (Beugnot et al 2019). Increasing team size may modify perceived competitiveness if individual performance becomes less transparent and thus comparisons with others become more difficult. As another part of our investigation on psychological mechanisms, our setup allows us to inspect the role of perceived unfairness induced by pay inequality, which has been shown to affect worker effort (Cohn et al. 2014) and which could play a different role across team size.

To comprehensively investigate possible free-riding in teams, we also pay attention to different dimensions of performance, given that 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). The quality of work is also of particular interest in our analysis of peer pressure, given that individuals might produce more mistakes in performing a task under pressure (Ariely et al. 2009).

3.3 Empirical analysis

3.3.1 Team size and task performance

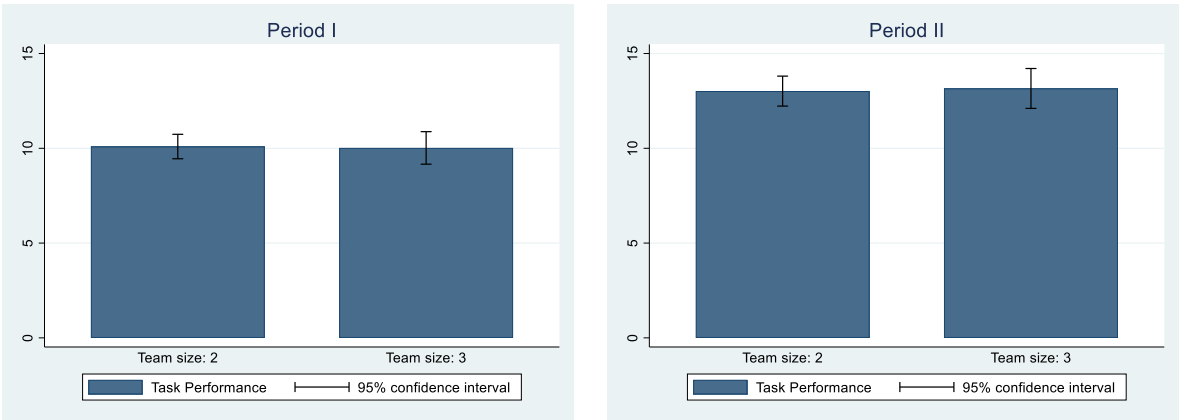
3.3.1.1 Main result

We start the presentation of our results by comparing average task performance in larger and smaller teams across work periods. To provide another perspective on the data, we present histograms of task performance across team size and work periods in Appendix Figure B.2.1.

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 routinely conduct two-sided tests throughout the empirical analysis.

Figure 3.2 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 of similar size for both smaller and larger teams, meaning that there is no significant difference across team size on average (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 only on the change in performance over time do not reveal significant differences between smaller and larger teams either (2.92 vs. 3.14 output units, p=0.554, t-test).

Figure 3.2 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.

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). Note that in regression analyses, we routinely cluster the

standard errors at the session level. In line with the visual evidence, the results in Appendix Table B.2.4 show that for both work periods one (Panel A) and two (Panel B), the team-size effect is close to zero. We make the same observation when inspecting the aggregated performance level across both work periods combined (Panel C).

3.3.1.2 Performance dimensions

In the next step, we analyze possible team-size effects across different performance dimensions. For this purpose, we differentiate between the total number of output units without any quality check and software-based quality-adjusted work output.[35]

Appendix Figure B.2.2 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 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 team-size effects, all six illustrations show no differences in performance levels, which appears to be robust across dimensions and work periods, conforming to the earlier result for task performance.

3.3.1.3 Unequal pay

Using a feature of our setup, we analyze the consequences of exogenously induced pay inequality for team performance across team size. Between the two work periods, unequal pay results of random bonus payments that are performance-independent and 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 examine responses regarding survey items on fairness perceptions, as potential channels of team performance.

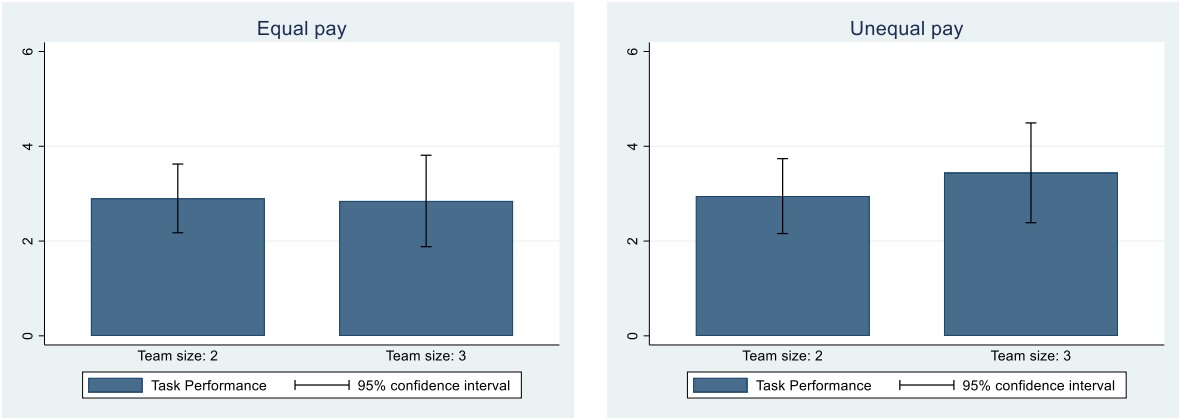
Appendix Table B.2.1 provides two pieces of evidence for manipulation of fairness perceptions, as shown in Panel B for data collected after the bonus game. First, we observe an increase in average scores reported for the item ‘unfair situation‘ (item 3) when comparing the responses to those from before the bonus game (Panel A). This effect is stronger in larger teams, as can be seen when comparing columns 1 and 2. Accordingly, perceived unfairness is 1.75 in smaller teams, compared to 2.24 in larger teams, which implies a statistically significant team-

[35] 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.

size effect (p=0.017, rank-sum test).[36] This gap is somewhat larger in those sessions when bonuses were paid out (columns 5 and 6). The second piece of evidence for an effective psychological manipulation comes from a survey question on the perception of being treated unfairly (item 4). Responses show the highest average score in larger teams when bonuses are paid out (column 6 in Panel B). Further analyses reveal that this is driven by those members of larger teams who did not get the bonus in the unequal pay scenario. They report much stronger perceptions of being treated unfairly compared to those in smaller teams (1.69 vs. 2.80, p=0.015 rank-sum test).

To test for effects in team performance related to differences in fairness perceptions, Figure 3.3 shows average changes in performance across work periods by team size and separated by the outcome of the bonus game. Despite the effective psychological manipulation of fairness perceptions through unequal pay, there are no relevant differences in team performance across team size. If anything, the increase in performance tends to be even stronger in larger teams when payments are unequal, but this is not statistically significant (2.95 vs. 3.44 output units, p=0.357, t-test).

Figure 3.3 Team size, unequal pay and changes in task performance



Notes: Bars show average changes in task performance, i.e. the difference in performance between periods two and one, by team size and by treatment. In the case of the 'Equal pay' treatment, no bonus was paid out between work periods one and two. In the case of the 'Unequal pay' treatment, a bonus was paid out between work periods one and two to all but one 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. 95% confidence intervals are displayed. Total sample size is 271 observations.

[36] While we conduct two-sided Mann-Whitney rank-sum tests to inspect significance of ordinal variables, we check robustness regarding clustering of standard errors at the session level, using Somers' D tests. We report on results from this alternative test in case the finding differs, when considering common thresholds for p-values.

Findings: *Members of larger teams on average do not underperform in comparison to smaller teams, which holds i) across different performance dimensions, ii) across work periods and iii) even after manipulating fairness perceptions by changing the payment structure. While pay inequality is perceived as less fair in larger teams, this does not go along with changes in team performance.*

3.3.2 Team size and subjective perceptions

3.3.2.1 Performance channels

To better understand our main result and to find out whether there is a potential for peer effects that could explain the lack of free-riding in larger teams, we analyze the data from our survey with its rich information on workplace perceptions and start by inspecting potential 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 effort being at play but cancelling each other out. For example, a free-rider motive to shirk in larger teams might be compensated by a stronger team identity, compared to smaller teams. However, when analyzing the data from the question battery on possible channels of team performance, shown in Appendix Table B.2.1, we cannot find any evidence for different channels responding to the manipulation of team size. Apart from the above-discussed changes in fairness perceptions after work period two (Panel B), we find no significant differences across team size (column 1 vs. column 2), which is also true for perceptions reported after work period one (Panel A). Average scores are very similar, including channels describing free-rider motives and team identity.

In addition to the question battery on channels that could affect performance, the participants also reported directly on potential reasons for providing more or less effort after finishing both work periods. Appendix Table B.2.2 shows average scores as reported by the participants for various items, such as perceived monitoring. Corresponding to the average performance levels, however, the data from this battery on possible effort channels also reveals no disparity across smaller and larger teams in general.

3.3.2.2 Satisfaction and attitudes towards teamwork

While our results show that team size as such does neither affect average performance levels nor its drivers in general, there could be important heterogeneity unrevealed in our analysis so far. Before we tackle the question of whether team size actually modifies the role of peer effects in performance, as an explanation for the lack of free-riding in larger teams, we first inspect

whether participants perceive working with peers in larger teams differently than in smaller teams. As differences in the perception of teamwork across team size could be indicative of a potential for heterogeneous peer effects, we exploit data from subjective assessments, based on questions regarding satisfaction with the teamwork and regarding future work motivation, in the following analysis. Table 3.2 details the wording of these questions and presents average scores as reported for all items, separated by team size, as well as *p*-values from rank-sum tests.

When comparing the two columns in Table 3.2, there are no significant differences in perceptions between smaller teams and larger teams regarding the task or the team as such. Yet, individuals report higher preferences for doing the task again in a larger team (last item) compared to members of a smaller team. This gap of 2.88 vs. 3.35 (on a 5-point scale) is statistically significant. We observe a very similar result in the satisfaction data when we focus on the team size item. Here, average satisfaction increases from 4.95 to 5.44 (on a 7-point scale) when comparing average responses from members of smaller and larger teams.[37]

Table 3.2 Team size, satisfaction with teamwork, and future work motivation

	(1) Mean	(2) Mean	(4) *p*-value
I.) Satisfaction: team communication	4.06	3.56	0.013
II.) Satisfaction: team size	4.95	5.44	0.020
III.) Satisfaction: team leader	5.47	5.51	0.713
IV.) Satisfaction: team in general	5.25	5.25	0.853
V.) Future work motivation: this task	3.23	3.37	0.283
VI.) Future work motivation: this team	3.58	3.59	0.810
VII.) Future work motivation: larger team	3.35	2.88	0.002
Number of team members:	2	3	
N	118	153	

Notes: Average scores of responses to various survey questions are shown in columns 1 and 2. Column 3 shows *p*-values from two-sided Mann-Whitney rank-sum tests. The wording of the questions is as follows:

- "How satisfied were you with the team communication?" (I.)
- "How satisfied were you with team size?" (II.)
- "How satisfied were you with the team leader?" (III.)
- "How satisfied were you with your team in general?" (IV.)
- "Would you want to do this task again?" (V.)
- "Would you want to work with this team constellation again?" (VI.)
- "Would you want to complete this task again with a larger team?" (VII.)

The response scale ranges from "Totally dissatisfied" (score: 1) to "Very satisfied" (score: 7) for items (I.) to (IV.) and from "No, under no circumstances" (score: 1) to "Yes, with pleasure" (score: 5) for items (V.) to (VII.).

[37] In the dataset without transformed sessions (N=195), the average score for satisfaction with team size increases from 4.90 in smaller teams to 5.44 in larger teams (p=0.068, rank-sum test).

Furthermore, members of larger teams 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 given a lack of robustness, we are cautious with interpretations.[38]

A separate survey question on perceptions of working with peers offers another option to study team-size related differences in perceptions of teamwork. 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 statistically significant increase in the likelihood of a positive assessment of working with peers (42.37% vs. 54.90%, p=0.041, rank-sum test).[39] In additional regression analyses, we confirm an increase of more than twelve percentage points in the likelihood of perceiving peers positively (see Appendix Table B.2.5). Arguably, more positive perceptions of peers could set the stage for more positive performance spillovers when the size of the team increases.

3.3.2.3 Performance estimations

Another indicator for a potential of heterogeneity in peer effects across team size comes from subjective assessments of performance. The idea here is, that estimating peer performance should become more difficult the larger the number of team members, implying potential differences in peer pressure and thus in the way that peer effects occur across team size. Before participants were informed about the team performance in work period one, they had 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.

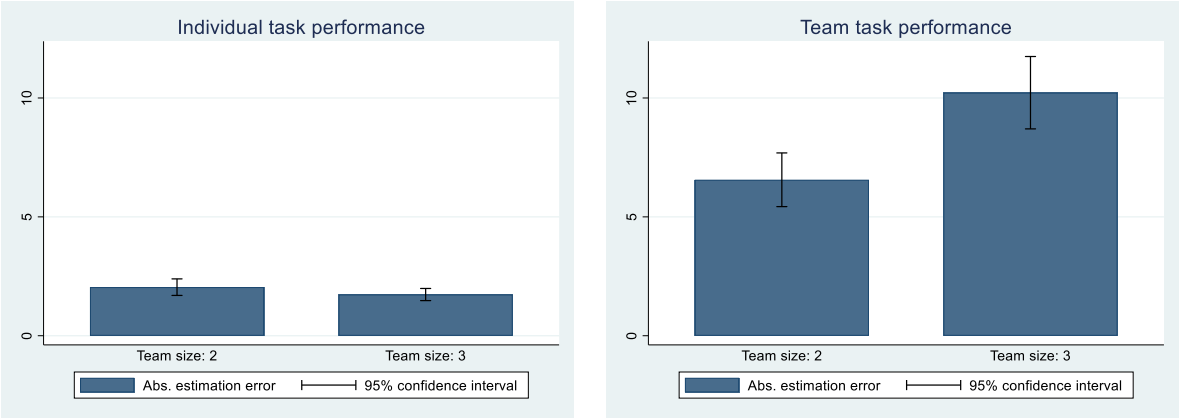
Figure 3.4 illustrates via bars how performance estimation errors differ across team size. In the left illustration, the focus is on performance estimation errors at the individual level, which reveals no significant differences in errors when estimating one's own performance across team size (2.04 vs. 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 larger teams compared to smaller teams, as the estimation error goes up from 6.56 to 10.22 (p=0.000,

[38] In the dataset without transformed sessions (N=195), the average score for satisfaction with team communication decreases from 3.86 in smaller teams to 3.56 in larger teams (p=0.337, rank-sum test).

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

t-test).[40] Accordingly, peer performance levels appear less clear in larger teams, which could be interpreted in light of the idea that less transparency regarding the performance of others induces less peer pressure. This would lead to the expectation of stronger peer effects in smaller teams, if we are willing to assume that peer pressure is a key driver of such performance spillovers.

Figure 3.4 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 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.

Findings: *Psychological motives to perform well or to shirk, such as free-rider intentions, do not differ across team size on average, contrary to satisfaction with the team size and perceptions regarding peers and their performance, indicating a potential for heterogeneity in peer effects across team size.*

[40] 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.

3.3.3 Peer effects, team size and task performance

3.3.3.1 Performance scatterplots

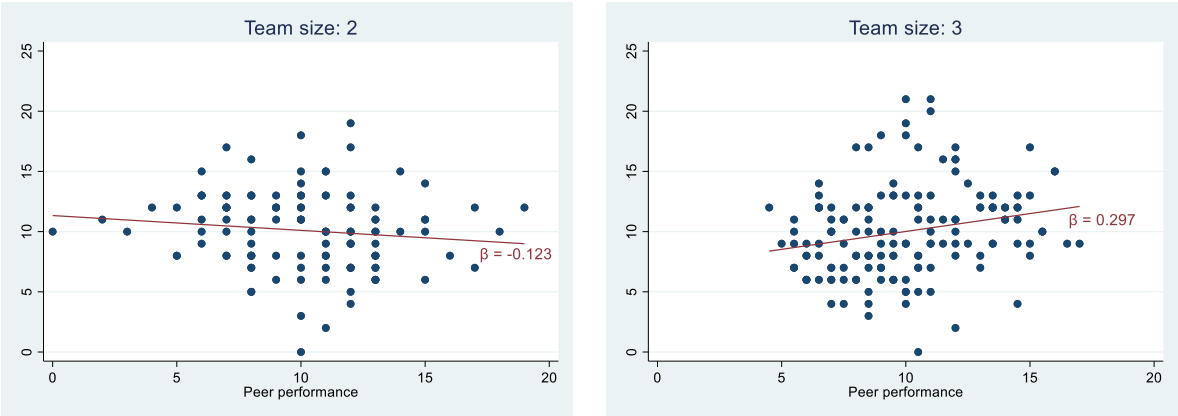
In a first step to investigate the occurrence of peer effects in our workplace setting, we show the raw performance data by means of scatterplots. A positive link between peer performance and own performance could be indicative of possible performance spillovers. In this context, we can then obtain first evidence on whether the behavior of peers could be relevant for one's own performance depending on the team's size.

Figure 3.5 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 periods. Based on simple 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 larger teams (coefficient: 0.297, p=0.005). As the link is statistically insignificant and negative (coefficient: -0.123, p=0.184) for data on smaller 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).[41] For larger teams, 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), implying that the heterogeneity in peer effects across team size remains intact over time.

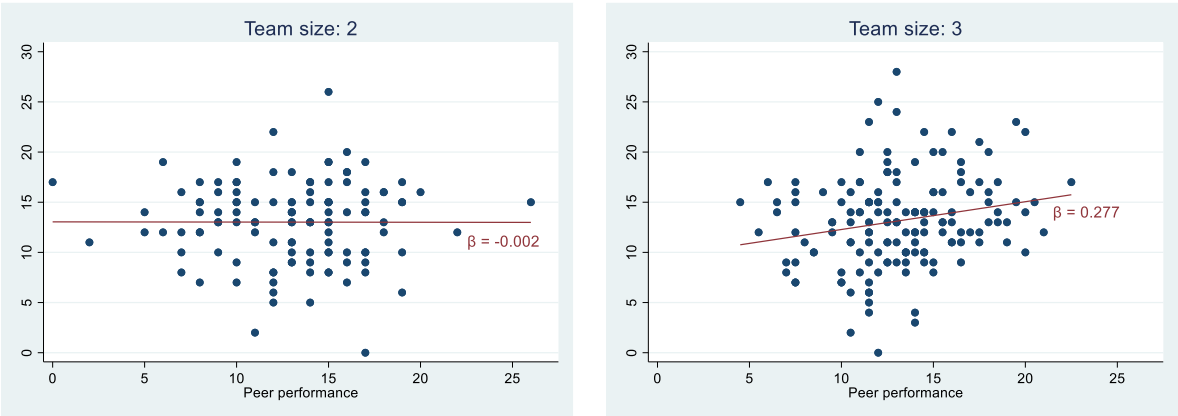
[41] In the dataset without transformed sessions, we even find a statistically significant negative relationship between peer performance and own performance in smaller teams in work period one (coefficient: -0.446, p=0.003), which indicates negative peer effects. In work period two, the correlation remains negative but becomes insignificant (coefficient: -0.054, p=0.736), in line with the result based on the full sample.

Figure 3.5 Team size and peer effects in task performance

A) Work period I



B) Work period II



Notes: Panel A (B) shows scatterplots of individual performance and peer performance in the first (second) work period by team size. 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. The β-coefficients shown are from linear regressions without control variables. Total sample size in each panel is 271 observations.

To learn more about the nature of potential performance spillovers between peers in larger teams, we modify the analysis of Figure 3.5 by distinguishing between 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 additional 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 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 are in line with the notion of low performers encouraging others to underperform.

3.3.3.2 Peer skills and task performance

There are several methodological challenges that researchers face when analyzing peer effects in groups. Our setup addresses common problems of identification like self-selection; yet, the reflection problem that occurs when regressing outcomes on outcomes (Manski 1993) deserves attention. In the following, we address this issue by using information from the survey on computer skills, as a separate indicator for the performance potential of peers. Since own computer skills are a strong determinant of own performance in this task of downloading and renaming scientific articles from the internet, we consider the computer skills of peers ('peer skills') as a determinant of peer effects.[42]

Table 3.3 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: the main task performance indicator and the total number of output units. In all columns, the results are consistent with the visual evidence in Figure 3.5 regarding heterogeneity in peer effects across team size. While better peer skills do not go along with higher task performance in smaller teams, there is evidence for effect heterogeneity, as the role of peer skills is more positive in larger teams. This is demonstrated by the interaction effect between team size and peer skills, whereas, in a separate row,

[42] 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.

Table 3.3 Team size and heterogeneous peer effects

	(1)	(2)	(3)	(4)
Dependent variable:	Task performance		Total work output	
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	1.070**	1.131*	1.043**	1.195*
team size 3	(0.529)	(0.666)	(0.486)	(0.635)
Total effect of peer skills	0.878*	1.051*	1.005**	1.111**
in teams of 3	(0.451)	(0.568)	(0.413)	(0.531)
R^2	0.015	0.013	0.020	0.016

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 is 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. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.3 also shows the total peer effect in larger teams (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 Appendix Table B.2.6).[43]

Based on the results in Table 3.3 with its interaction-term analysis, we can draw conclusions if and when free-riding in teams could occur. According to the estimates, the consequences of variations in peer skills are quite severe in larger teams, as for each score point on the scale the work output increases by roughly one output unit in each work period. Vice versa the lower the peer skills, the more likely there will be a negative performance effect of increasing team size. In fact, for mid-level peer skills (i.e., a value of zero, see Table 3.1), all results show a negative coefficient for the team size variable. One of our robustness checks even reveals a weakly significant negative effect of a team size increase (see column 1 in Appendix Table B.2.7). This

[43] We confirm our findings in several robustness checks, which include regression analyses based on the dataset without transformed sessions (see Appendix Table B.2.7). For another check, we exclude 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 B.1.1).

suggests that our first main finding of similar average performance levels across team size (see Section 3.3.1) could be due to a relatively large share of participants with sufficient skills to perform well in a simple computer task, whereas free-riding only occurs in larger teams if peer skills are low. To find out more, we focus on the small number of individuals (N=49) who are 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).

Further evidence on the idea of free-riding in larger teams being prevented by high peer skills comes from the survey data on potential transmission channels shown in Appendix Table B.2.1. For an analysis based on similarly sized subsamples, we split the data by distinguishing between settings with "low" peer skills and settings with "high" peer skills, which we define as positive values in the following. As an indicator of free-rider intentions, the item 'I lean back and let others do the work for me' shows higher scores for larger teams (1.45), compared to smaller teams (1.28), when we condition on individuals working together with low-skilled peers only, as can be seen in columns 3 and 4 of Panel A. This difference in average scores is weakly significant (p=0.096, rank-sum test).[44] While this finding is not robust across work periods, we consider this evidence as indicative of the idea that free-riding is more likely to occur when team size increases in a scenario without high-skilled peers.

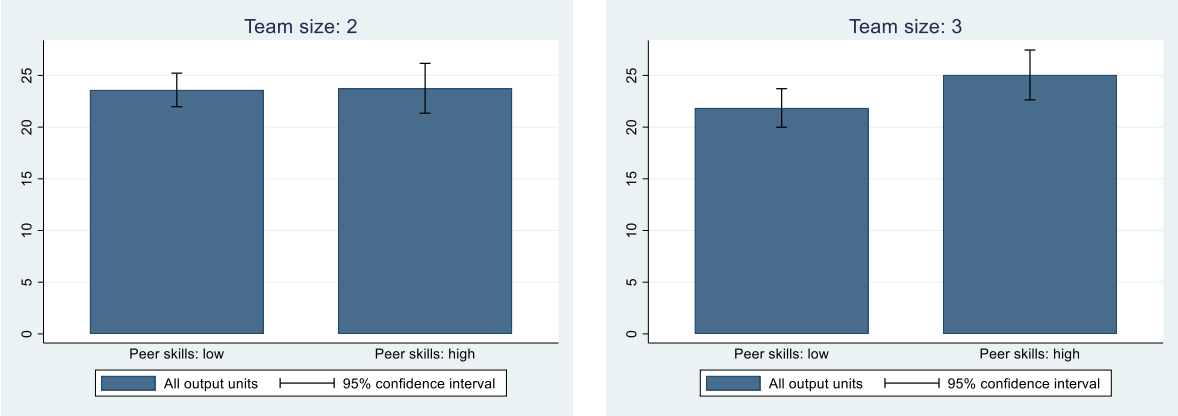
3.3.3.3 Performance dimensions

To learn more about the nature of performance-spillovers and why those seem to occur only in larger teams, we again turn to the software-generated performance data and analyze whether peer effects vary across performance dimensions. For a visual analysis (presented in Figure 3.6), we again differentiate between settings with "high" and those with "low" peer skills, as in the previous subsection, and consider three different outcome variables. Accordingly, we distinguish between numbers of high-quality work output units (Panel B) and low-quality work output units (Panel C), whereas the sum of both is the total number of work output units, independent of correctness (Panel A). In all three cases, the figure displays the combined work output which is aggregated across the two work periods.

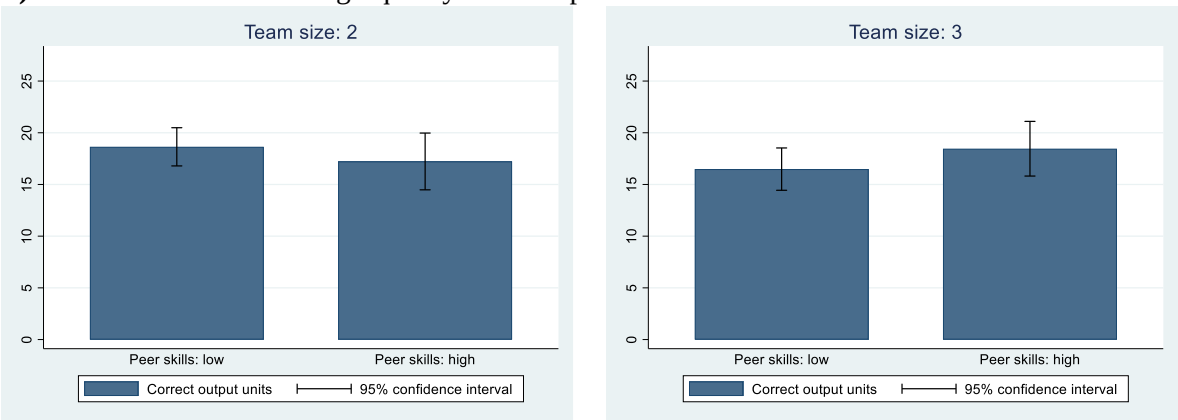
[44] The effect of team size becomes more pronounced (p=0.028, rank-sum test) when we use the dataset without transformed sessions, as the average score of self-reported free-rider intentions is 1.11 in smaller teams.

Figure 3.6 Peer effects and performance dimensions

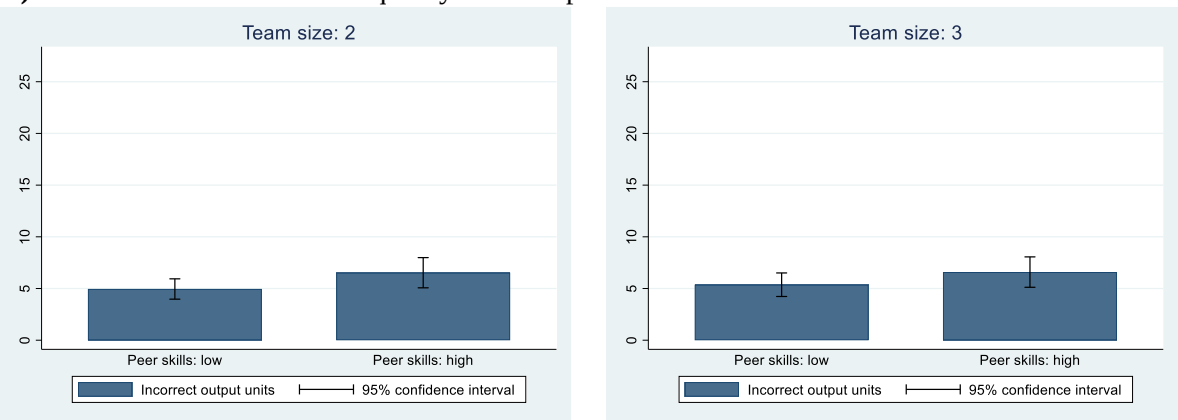
A) Performance measure: Total work output



B) Performance measure: High-quality work output



C) Performance measure: Low-quality work output



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. Panel B (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.

Several findings emerge out of the analysis presented in Figure 3.6. First and foremost, the heterogeneity in the effect of peer skills becomes even more striking when using a binary distinction into high and low peer skills, instead of using a linear peer skills variable as in the analysis of Table 3.3. For the total work output (Panel A), we find a significantly positive peer effect in larger teams when comparing high-skilled and low-skilled peers (21.9 vs. 25.0, p=0.010, t-test). No such effect is observed in smaller teams (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 larger teams, 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 smaller teams when we split the work output into high-quality vs. low-quality output units. On the one hand, there is an increase in the number of mistakes, which leads to 1.6 low-quality output units more when peers are highly skilled compared to when they are not (p=0.035, t-test). On the other hand, the number of high-quality output units decreases by 1.4 output units (p=0.310, t-test). Taken together, this explains the absence of any peer effects on performance, as measured in total output units. Strikingly, while members of smaller teams do not increase their total work output quantitatively when confronted with a high-skilled peer, our analysis of performance dimensions reveals that they produce the same quantitative amount of output, but in lower quality. In the following subsection, we explore possible explanations for these apparently negative peer effects in smaller teams by shedding light on the potential mechanisms behind these findings.

Findings: *Peer effects on task performance interact positively with team size. If peer skills are low, members of larger teams perform less well compared to smaller teams, and free-riding becomes more likely. If peer skills are high, an increase in team size is not detrimental to performance, whereas members of smaller teams produce more work output in low quality.*

3.3.4 Peer effects and subjective perceptions

Our finding of heterogeneous peer effects fits well to the above discussed survey evidence in Section 3.3.2, according to which working with others is perceived differently when team size increases. More positive perceptions of peers in larger teams could also explain why negative peer effects are only observed in smaller teams. By contrast, the survey evidence does not conform to the idea of performance-enhancing peer pressure as the dominant mechanism, given that in larger teams, peer performance was less transparent and hence the pressure to perform was probably weaker, in comparison to smaller teams.

To learn more about the mechanisms underlying peer effects in performance, we continue to analyze survey data. We start this final part of our analysis by examining responses from a question battery on workplace perceptions related to the occurrence of peer effects, which was included in the survey after work period one and is shown in Appendix Table B.2.3. For our analysis of heterogeneous peer effects, we compare average scores, as reported for various items, across team size and peer skills. The results show that 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, columns 5 and 6) when the team size increases and peer skills are high (p=0.059, rank-sum test), which speaks against the idea of well-performing peers exerting pressure on other members of a larger team to also perform well. Second, for the same subgroup comparison of high-skilled peers only, we observe a perception of a less competitive situation when team size increases (from 3.64 to 3.10, p=0.071, rank-sum test).[45]

To clarify which possible channel of peer effects explains best our finding on effect heterogeneity, we conduct a regression analysis using the data from the same survey battery (Table B.2.3) and employing interaction terms between team size and peer skills. The results in Table 3.4 again show that peer skills increase the perception of a competitive environment in smaller teams, but not so in larger teams (which can be seen in a separate row for the total effect).

[45] Both findings become significant at the 5% level when we check our results by employing the Somers‘ D test, which allows clustering the standard errors at the session level. If we make use of the dataset without transformed sessions, statistical significance is at the 10% level, but this could be due to smaller sample size, since both gaps in averages actually become larger.

Table 3.4 Channels of peer effects and team size

Dependent variable:	(1) Performance	(2) Monitored	(3) Observing	(4) Distracted
Team size 3	-0.492	0.222	-0.115	0.261*
	(0.454)	(0.226)	(0.211)	(0.154)
Peer Skills	-0.193	0.066	0.065	-0.133
	(0.276)	(0.175)	(0.141)	(0.097)
Interaction: Peer skills X	1.070**	-0.034	-0.044	0.023
Team size 3	(0.529)	(0.266)	(0.229)	(0.182)
Total effect of peer skills	0.878*	0.032	0.021	-0.110
in teams of 3	(0.451)	(0.201)	(0.180)	(0.155)
R^2	0.015	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	-0.365	-0.518**	-0.091	-0.232
Team size 3	(0.282)	(0.258)	(0.253)	(0.232)
Total effect of peer skills	-0.040	-0.431**	-0.071	-0.050
in teams of 3	(0.241)	(0.215)	(0.209)	(0.172)
R^2	0.019	0.015	0.002	0.006

Notes: Linear regressions without control variables are used. The dependent variables in specifications (2) to (8) are based on responses to survey questions regarding teamwork. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The question battery starts with: "What is your view regarding this statement 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 is 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 performance measure in the specification (1) is the number of output units used for performance pay after a spot check on correctness was conducted during sessions in work period one. The full sample (N=271) is used. Clustered standard errors (at the session level) are in parentheses.

As a second finding, peer skills significantly interact with team size when we turn to an item measuring the perceived pressure to succeed. This underlines the idea that peer pressure could play a negative role in the emergence of peer effects, as it seems that less pressure goes along with more positive spillovers in performance from peer to peer (as shown for illustration purposes in the first specification of Table 3.4). 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, contrary to expectations based on previous literature.[46]

As a final step, we use the survey data on reasons for putting in effort (Table B.2.2) to inspect differences in perceptions when peers are either highly skilled or not. We focus on larger teams (columns 4 and 6) to find out more about the occurrence of positive peer effects. By doing so, we find a highly significant negative effect of high peer skills on 'competition for best performance' as a reason to provide effort in larger teams (decline from 4.50 to 3.51, p=0.001, rank-sum test). In smaller teams (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 less peer pressure in larger teams could help prevent competitive situations from occurring.

Findings: *Peer skills affect workplace perceptions differently across team size. Peer pressure is a possible mechanism underlying the occurrence of peer effects, as pressure induced by high-skilled peers differs when the number of team members increases. Contrary to expectations, however, positive peer effects in larger teams could be due to the lower pressure exerted by high-skilled peers, in comparison to smaller teams, which could play a role in determining whether or not the work environment is perceived as a competition.*

[46] 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).

3.4 Discussion

3.4.1 Summary and interpretation of results

Our study on team size and task performance reveals several findings. We show that despite economic disincentives, larger teams are not necessarily outperformed by smaller teams. This appears to be a robust observation in our study, regarding time and performance dimensions and even regarding an introduction of pay inequality that creates a situation perceived as unfair. Psychological mechanisms seem to play a role in explaining relatively high work motivation in larger teams, as peer effects interact with team size and thereby contribute to preventing free-riding when team size increases. Thanks to well-performing peers, individuals in larger teams are complying in their willingness to put in effort. However, if peers are not high-skilled, free-riding becomes more likely when team size increases.

From a practical perspective, the concern of free-riding in larger teams is highly relevant for contexts in which the task is difficult and skills are limited. In such a scenario of a low-skilled workforce, the positive link between own and others' performance implies that larger teams could be outperformed by smaller teams. Adding high-skilled workers to larger teams or introducing economic incentives could be options to prevent free-riding. Another issue with practical relevance could be our finding of a negatively perceived pressure in smaller teams when peers are highly skilled. A higher number of mistakes among pressured individuals speaks to the idea of possible side effects when adding a high performer to the team and thereby increasing pressure. The idea of a negative role of peer pressure and hence the occurrence of negative peer effects deserves further attention in research on teams. Before we continue discussing how our paper adds to ongoing research in this context, we first examine possible limitations of our study in the following discussion.

3.4.2 Discussion of limitations

There are some possible limitations of our findings in regard to their practical applicability and generalizability, the first of which relates to stakes. As monetary incentives are not very high in our study, one could argue that differences in free-riding could be more severe across team size if individuals were more strongly incentivized. However, a payment structure with a rather small incentive component aligns with the situation in many companies where team incentive pay is added on top of a relatively high and fixed baseline pay to affect performance.[47]

[47] See Friebel et al. (2017) for a study on team incentives that increased employee performance in a retail chain.

Arguably, contributing one Euro more to the team output or not makes no big difference for actual income levels, but individuals may get pleasure from earning a prize so that monetary incentives entail powerful effects on their motivation. Furthermore, the evidence on psychological pressure observed in smaller teams supports the idea that individuals in our setting took the task very seriously, whereas higher stakes probably could have stimulated this particular channel even more, suggesting that we may even underestimate the importance of psychological effects in our setting.

Second, our workplace setup allows for a short-run investigation, which per se might limit the occurrence of free-riding. However, as we do observe evidence for free-riding in the short-run, though limited to the low-skilled-peers scenario, we believe that our finding of free-riding in larger teams being limited via peer effects could be very generalizable to other long-run settings. The issue of generalizability rather seems to be whether peer effects are also effective in the long-run, but there is research based on field data that provides support for this (Mas and Moretti 2009, Battisti 2017).

Third, in our study based on a controlled work environment, we are limited to teams consisting of either two or three co-workers. The question arises whether our findings are also 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 in larger teams the phenomenon of peer effects could become more relevant. Which one of the two phenomena dominates depends on various factors, which not only include the average level of peer skills but also other factors, such as the nature of the work task, making this question increasingly context-specific and subject to the characteristics of the workplace setup.

Fourth, with regard to the specific characteristics of our workplace setting, we decided to abstract from various features of teams, such as task complementarities. This allows for more generalizable insights but could be seen as another limitation. Hence, one may think about changing workplace characteristics to study the role of aspects neglected in our setup, such as social ties as an example of a factor that could modify the role of peer effects (Bandiera et al. 2010). While factors like that in our view are as interesting as they are difficult to manipulate in the context of teamwork, we leave this to future research, which we discuss next.

3.4.3 Contributions to the literature and future research

With our findings, 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 comparing team incentives with individual incentives (Hamilton et al. 2003, Babcock et al. 2015, Corgnet et al. 2015b) but pays less attention to the role of team size for workplace performance. We contribute to this research by providing 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 to recent results in the research on team size and peer effects, building upon the seminal contribution by Kandel and Lazear (1992).[48]

In a recent and comprehensive investigation into teams, Steinbach and Tatsi (2018) make use of data from a large transportation company to study peer effects. In a setting different from ours, they also find that peer effects interact with team size. Hence, our contribution is to confirm this finding in data 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 inspect theory-inspired ideas about the role of team size in explaining differences in effort levels. In line with our paper, they argue that peer effects may offset disincentives in larger teams, and they point out both social ties and peer pressure as important aspects in this context. Apart from confirming the basic idea of peer effects mitigating free-riding with experimental evidence, another contribution of ours is to provide evidence on mechanisms underlying peer effects in a workplace context, as we show that the role of peer pressure might be different than expected in theory. In regards to future research, it appears promising to further inspect the role of increases in team size, going beyond the number of team members investigated by us, given that both free-riding and psychological mechanisms might play a different role in very large teams.

Finally, we also contribute to the growing research on peer effects at work, a field in which researchers are still looking for explanations regarding the question why such psychological phenomena do occur in the first place. In a recent paper on the mechanisms behind peer effects,

[48] There are related strands of research, such as in economics where researchers study the role of contributions in groups, arguing that group size can affect individuals' willingness to contribute positively (Isaac and Walker 1988, Isaac et al. 1994). While contributing to a public good constitutes a similar incentive problem as putting in 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 contributes more, not less, which could be a consequence of greater social benefits. 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. In contrast to our setting, team members receive individual incentives, instead of team performance pay, and peer effects through presence of others are ruled out by design.

Reyniers (2018) investigates several hypotheses to solve what she calls a “puzzle” of what is bringing peer effects to work. Interestingly, she views competitiveness as an important aspect in the occurrence of peer effects, whereas our evidence suggests a potential for negative peer effects due to perceived competition and corresponding pressure to succeed. Notably, the idea of a ‘choking-under-pressure’ phenomenon (Baumeister 1984, Dohmen 2008) is discussed as a possibility in the context of peer pressure (Georganas et al. 2015). Furthermore, there is some evidence on negative peer effects in specific settings, be it in sports (Emerson and Hill 2018) and in an educational context (Brady et al. 2017). While our contribution to this research is to shed light on the possibility of adverse effects through high peer pressure in teams, it is certainly a very worthwhile objective for future research to continue the discussion on negative peer effects and the underlying mechanisms.

3.5 Appendix B.1: Additional background information

Appendix B.1.1: Recruitment

In this appendix section, we illustrate how the recruitment of study participants for the team experiment took place. To take part, individuals first had to fill in an online webpage interface to be considered for study participation. On this webpage, it was necessary to sign a data agreement and to provide basic information. An email address was required to be contacted by the organizers of the sessions, whereas a name and a mobile phone number could be entered on a voluntary basis. The research assistants who organized the sessions screened all registered persons via the information provided, according to the following three selection criteria: First, German language skills were required, as this was the language used during sessions. Second, student status was necessary for legal reasons, given that downloading scientific articles from a subscription-based online library is only allowed for individuals with university access. Third, to prevent multiple participations of the same person, available information on name, email address, and field of subject was used. The same information was also used to identify possible social ties between students in order to minimize the chances of individuals knowing each other being assigned for the same session. In the case of two individuals providing similar information, the assistant assigned the second individual to the subsequent session.

Registered students considered as suitable after the checks entered a list of soon-to-be-invited participants, ordered according to the time of registration. To foster positive response behavior and to minimize cancellations, students could use the registration webpage to inform the organizers about when they were not available. Depending on the experimental condition, sessions were filled with either two or three individuals. To do so, the next eligible individuals according to the list received invitation emails containing the start time of the session and a note on the importance of a timely response within 24 hours. All email communication was standardized using email templates. In the absence of a positive response, the next person from the list was invited for the empty slot. After confirming the invitation email positively, the invitee received a second standardized email on how to get to one of three waiting rooms. The room was determined by a random procedure.

In cases when slots had not been filled two days in advance, sessions were cancelled early on. In cases when all slots had been filled first, but then an early withdrawal took place on the day prior to the session day (e.g. due to illness), the organizing assistants were instructed to recruit a replacement (e.g. by asking students on the campus to register and take part spontaneously), as an alternative to cancelling the session on the day before. In case of a last-minute withdrawal (‘non-show-up’) on the day of the session, a three-person session was carried out as a ’transformed’ two-person session, whereas a two-person session was cancelled in such a case.

To avoid individuals showing up for a cancelled session, the organizer used all available contact information, including email address and mobile phone number. If unsuccessful, show-ups received 5 Euro and were told to still be eligible for study participation. Non-show-ups also were considered eligible for study participation and were invited again, but only if they provided a reason (e.g. traffic jam). If not, they were taken from the list of to-be-invited participants, just like those individuals were who did not respond to three invitation emails in a row.

Appendix B.1.2: Instructions

[All the text in quotation marks is to be read aloud by the session host. Supplementary instructions for the session host are in italics, which include demonstrations of procedures using the available big screen. The instructions for the session host also include references to material, which can be found in this Appendix (marked by '➔').]

Welcome participants and ask them to read the general information document [➔ Figure B.1.4], *switch mobile phones to flight mode as well as put them aside together with their bags. When all participants have been guided to the office, assign them to their places by drawing lots.*

"Welcome to this scientific study, which is moderated by me, the team leader. You will be the work team, consisting of ..."

Introduce every person with the correct team player number according to the lot drawn before.

In case of a scheduled three-person session with only two participants, randomly either say the following or not: "A third person was invited but apparently did not show up."

"First of all, the most important rule that you must follow is: Apart from what I say, verbal communication is prohibited until the end of the session. Misconduct leads cuts in payouts. We will communicate via email. In the following, we will test the email program installed on each computer. You can see the email addresses on the tables for everyone to read. We assigned an email address to each one of you."

Refer to each sign by hand.

"My address, as the team leader, is chef@iaaeu.de. I will send a test email to everyone now using the mailing list team@iaaeu.de."

Show the preparation of the first test email on the big screen and send it [➔ Email template 1 in Figure B.1.5].

"I will separately send out a second test email to myself and all of you by putting team@iaaeu.de in blind copy (BCC)."

Show this on the big screen using the mouse cursor.

"By putting you in BCC, you will receive an email without being able to see the other recipients of the email."

Send the second test email [➔ Email template 1 in Figure B.1.5].

"To send an email, click on 'New email' in the upper left corner of your email program. Now please write me a test email to chef@iaaeu.de and set the team mailing list, which is team@iaaeu.de, to BCC."

Wait until all emails are received.

“You can use the email function from now on. If you have any questions, you can always send me an email to chef@iaaeu.de.
In this study, the aim is to collect as many points as possible. For every point you score, you will be paid one Cent at the end. Thus, for 100 points you will receive 1 Euro. You will begin with 250 points for your attendance, so you are guaranteed to earn 2 Euros and 50 Cents.”

Send separate emails with a link to Survey 1 [➔ Email template 2 in Figure B.1.5].

“By filling out an online questionnaire completely and conscientiously, you will earn an additional 200 points. I have sent you a link via email. You have to click on the link to get to the questionnaire. Once everyone has completed the questionnaire, I will receive a confirmation and then we can continue. Please complete the questionnaire now.”

Wait until all confirmation emails for survey 1 are received.

“Everyone has completed the online questionnaire. You can close your browser window. I will now explain your work task. The following is about supporting research at our institute by procuring literature. As you may know, the university has limited access to scientific journals, which are very important for research. The limited availability impedes the research at the institute. In order to make studies of certain journals directly available for research purposes, your task is now to download as many scientific articles as possible and label them correctly. On every computer desktop you find an open Explorer window that shows a folder.”

Show folder using the Explorer window on the big screen.

“Behind me, you can see my folder. In this folder, the first two articles of a volume from a specific journal have been downloaded and correctly labeled. Please note that the journal’s volume year is different on each computer. If you accidentally close your folder, you can find a shortcut on your desktop. Alternatively, you can use a bookmark via the ‘Favorites’ function in the upper left corner of the Explorer window.”

Show the ‘Favorites’ function using the mouse cursor.

“I will now show you the website of the journal.”

Show the journal website in the browser window and show the relevant information using the mouse cursor in the following.

“Your task is to continue downloading articles and labeling them correctly. As you can see on the big screen behind me, the articles are ordered from top to bottom according to the page numbers. It is important to download the articles in order, which means you should download the articles one by one, moving from top to bottom. When you have finished downloading all of the articles from a volume’s issue, click on ‘next issue’ located at the top right of the page, and continue with downloading the articles from top to bottom. When you are finished will all of the issues from your volume, you can download the following volume’s articles.
Note that you do not have to download the documents named 'Front Matter' and 'Back Matter'. The standard procedure is summarized on the big screen behind me.”

Show information slide ‘task instructions’ on the big screen [➔ Figure B.1.3].

"Before informing you about the remuneration for the work task, I would like you all to practice the task I have just described. We will go through the process of downloading the respective third article together. Please keep in mind that we all work on different volumes and that the third article in my case is, therefore, different from yours."

Show all steps on the big screen in the following.

"Now we will all go to the journal's website. You will find the link as a bookmark in the upper left corner of your internet browser.

1. Click on the third article
2. Click on 'Download PDF'
3. Confirm the terms and conditions of use, if necessary
4. Click on 'Save document' on the top right side
5. Select 'Save File'
6. The folder to store the file should already be selected. If not, you can find the folder by using a bookmark via the 'Favorites' function in the upper left corner of your internet browser.
7. Assign the correct file name: Page number of the article's first page, underscore, surnames of the authors, each separated by a comma, underscore, the title of the study. You can use 'copy & paste' for this purpose.
8. Click on 'Save' to finish."

Show the information slide 'task instructions' on the big screen again [➔ Figure B.1.3].

"Note that special characters such as colons, question marks, or quotation marks can cause error messages. Therefore, you have to omit these when assigning the file names."

Check all folders to see if the files were successfully downloaded and help with problems. If a problem is not solvable via email communication, you can resort to verbal communication.

"Your remuneration is determined as follows. For each additional downloaded article, the team receives 10 points, as long as I cannot find mistakes in the labeling. The total amount of points at the end of the task is divided by the number of team players, so everybody receives the same earnings from this task. If all of you prepare three articles each, this means 30 points on average and thus 30 cents per team player. Note that the number of points you score affects your earnings but has otherwise no further consequences for you. Please keep in mind that talking is not allowed and is punished with point deductions. Of course, you can communicate at any time using the email program. If you want to ask me a question, the best way to do so is to reply to the email which I am sending you now. In this email, I am asking you to estimate the number of articles you will be able to download and label correctly during the following 10 minutes of working time as well as your estimation regarding the average performance of your team."

Send the email about performance estimations [➔ Email template 3 in Figure B.1.5].

"As soon as I have all your estimates, I will give the starting signal."

Ensure that the information slide 'task instructions' is on the big screen [➔ Figure B.1.3] *and wait until all emails are received. (If this is not the case after 30 seconds, say:* "I have not yet received an answer

from all team players." *If this is still not the case after another 15 seconds, contact the team player directly:* "I am still waiting for your email."*)*

Work period 1

"We will now start with the work period. I will now start the timer."

Start stopwatch. Prepare the next emails during the work period and begin checking articles of the participants while they work. Announce the end of the work period after 10 minutes.

"Time is up. Please finish the last process."

Wait a moment. Send separate emails with a link to Survey 2 [➔ Email template 2 in Figure B.1.5].

"While I count and check the articles to determine the remuneration, you can receive another 200 points by completing a second questionnaire conscientiously. Please click on the link again that you received via email and start filling out the questionnaire immediately."

Determine the results by using the calculation sheet and prepare the next email. Wait until the first confirmation email for survey 2 is received.

"As soon as everyone has completed the questionnaire, I can send you the results of the work period via email."

Wait until the second confirmation email for survey 2 is received.

"In the email, I will also ask a question that you should answer via email."

Send the email about the remuneration [➔ Email template 4 in Figure B.1.5] *and wait briefly for the responses.*

Bonus game

"We will now start with a bonus game. In front of you is a 20-sided die and a dice cup, which allows us to roll the die secretly so that no one else can see your number."

Show information slide 'Bonus game instructions' on the big screen [➔ Figure B.1.2].

"The game consists of three steps. In the first step, I, as the team leader, roll the die to determine which one of you participates in the game as bonus player A. In the second step, player A rolls the die and replies to my email with a number from 1 to 20. Based on a randomly chosen rule, the number player A has rolled will determine who bonus player B is. In the third step, player B rolls the die and replies to my email with a number of 1 to 20. This roll determines whether or not a bonus is paid out. You can see the rule on the big screen behind me."

Show the rule by using the mouse cursor and read it aloud. Wait briefly.

"Note that a different distribution of points is not possible."

Practice round

"We will go through the bonus game one time to practice. In the first step, I will now determine who is player A."

Roll secretly and simultaneously both 6-sided and 20-sided dice to determine the correct row and the correct column in the player selection sheet for the 'Selection of A' [➔ Figure B.1.6]. *Send the email to A* [➔ Email template 5 in Figure B.1.5].

"According to the number I rolled, I have determined who bonus player A is in this practice round. It is team player ...
Now step two: I have sent player A an email. As soon as my email is received, player A rolls the die and replies with a number from 1 to 20."

Wait for A's reply and continue after receiving it.

"I have received a reply. Based on the number sent to me by player A, I can now determine bonus player B. There is a 50% probability for either team player ... or ...."

Choose the same column as before in the player selection sheet and consider the reported number of A to determine the correct row in the player selection sheet for the 'Selection of B'. [➔ Figure B.1.6].

"According to the number sent by player A, I have determined who takes the role of player B in this practice round. It is team player ...
In the same way, I would now ask player B to reply with an email including a number from 1 to 20. This number then decides whether a bonus is paid out or not, according to the rule."

Read the rule aloud again [➔ Figure B.1.2].

Main round

"Now we will play for money. In the first step, I will determine who is player A."

Roll secretly and simultaneously both 6-sided and 20-sided dice to determine the correct row and the correct column in the player selection sheet for the 'Selection of A' [➔ Figure B.1.6]. *Send the email to A* [➔ Email template 5 in Figure B.1.5].

"According to the number I rolled, I determined who will be player A in this main round. As soon as my email is received, player A rolls the die and replies with a number from 1 to 20."

Wait for A's reply and continue after receiving it.

"I have received a reply. Based on the number sent to me by player A, I will determine now who will be player B in this main round of the bonus game."

Choose the same column as before in the player selection sheet and consider the reported number of A to determine the correct row in the player selection sheet for the 'Selection of B'. [➔ Figure B.1.6]. *Send an email to player B by forwarding the email reply from A and adding the sentence: 'Please reply to this email with a number from 1 to 20'.*

"As soon as my email is received, player B rolls the die and replies to my email with a number from 1 to 20."

Wait for B's reply and continue after receiving it.

"I have received an answer from player B. I will now announce the result of the bonus game via email."

Send the email with the result [➔ Email template 6 in Figure B.1.5].

<u>Work period 2</u>

"In the email, I will also ask you a question about the second work period that is about to start. Once you have replied to the question via email, you may continue with the work task. You will start from where you finished in the first period. This means that the next article according to the number of pages must be downloaded and labeled correctly. I will start the timer now."

Start stopwatch. Show the information slide 'task instructions' on the big screen [➔ Figure B.1.3]. *Prepare the next emails during the work period and begin checking articles of the participants while they work. Announce the end of the work period after 10 minutes.*

"Time is up. Please finish the last process."

Wait a moment. Send separate emails with a link to Survey 3 [➔ Email template 2 in Figure B.1.5].

"While I count and check the articles to determine the remuneration, you can receive another 200 points by completing the third questionnaire conscientiously. Please click on the link again that you received via email and start filling out the questionnaire immediately. It is important to fill out this third questionnaire completely. The study will only be finished after you have completed the questionnaire. This is necessary for receiving the payout."

Determine the results by using the calculation sheet and prepare the receipts for the payout. Wait until the first confirmation email for survey 3 is received.

"To be on the safe side, we ask you to delete all emails in your email program, in your inbox, in the sent folder, and the Recycle Bin."

Wait until all confirmation emails for survey 3 are received.

"The third and last online questionnaire has been completed by everyone. I have created the receipts with the amounts of money. With these you will receive your payout at the meeting room where you first arrived today. Thank you very much for participating in this study."

Appendix B.1.3: Supplementary figures

Figure B.1.1 The computer workstation

A. Setup: team of three

B. Setup: team of two

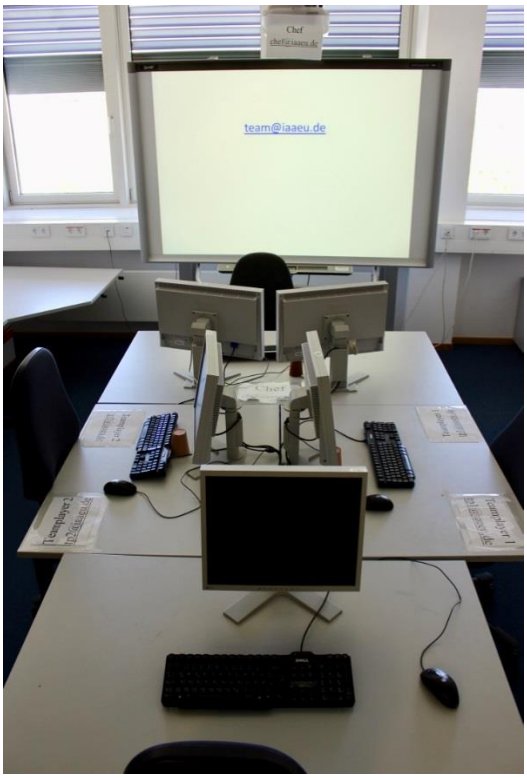


Figure B.1.2 Bonus game instructions

Step 1: **Who is A?**
The team leader rolls the die.
This number decides who A is, based on a pre-determined rule.

Step 2: **Who is B?**
A rolls the die and replies to the team leader's email with a number from 1-20.
This number decides who B is, based on a pre-determined rule.

Step 3: **Will there be a bonus?**
B rolls the die and replies to the team leader's email with a number from 1-20.

Rule:
Only if the numbers of B and the team leader match
(i.e., both numbers are even or both numbers are odd),
is a bonus is paid out (otherwise not).

Bonus:
<u>300 points for B</u>
{300 points for A, 300 points for B}

Notes: This figure shows the instructions for both two-person and three-person sessions. Differences between the treatment conditions are marked as follows. In two-person sessions, the text in curly brackets is left out, whereas, in three-person sessions, the underlined text is replaced by the text in curly brackets.

Figure B.1.3 Task instructions

Step 1: Download PDF

Step 2: Enter PDF Title -> **Page_author1, author2, author3, etc._title**

(Example: 1471_Jolls,Sunstein,Thaler_A behavioral approach to law and economics)

Rules:

1. NO ABBREVIATIONS when listing the authors

➔ "et al." is a taboo!

2. UPPERCASE can be retained, in general, there are no specifications of upper and lower case

➔ copy & paste the title of the study retrieved from the journal's website is generally allowed

3. SPECIAL CHARACTERS partly have to be left out when entering titles

➔ This refers particularly to colons (": "), question marks (" ? ") and quotation marks (" " ")

4. SUBTITLES can be left out

➔ convert "Norm-Based Trade Union Membership: Evidence for Germany" into

"Norm-Based Trade Union Membership"

Figure B.1.4 Information document

Important information about this experiment

First of all, we would like to thank all who are participating in this study and thus supporting the research at the Institute for Labour Law and Industrial Relations in the European Union (IAAEU). We would like to point out two important aspects.

Data protection

With the registration for this study, consent was given that individual data can be generated and subsequently analyzed anonymously. All data collected will be treated in strict confidence and the dataset generated in the course of this project will be stored anonymously. All observations will differ only in terms of identification numbers. It is therefore impossible to draw conclusions about specific individuals. In addition, the rules and laws of communication customary in Germany will also apply to this experiment at the computer workstation. If, for example, communication between study participants takes place via email or in chat programs without the explicit involvement of the session host, neither we as researchers nor third parties are permitted to read any such communication. This does not include emails sent to the session host. We will delete all other emails sent via the installed email program after the experiment. For the sake of data protection, study participants can of course also delete the emails themselves.

Ethical principles

This is a study conducted by economists. In contrast to experimental research in other fields such as psychology, the ethical principle in experimental economics is that the researchers do not deceive participants. In concrete terms, this means that all statements made by the session host are truthful. In accordance with this ethical principle, it also applies that no participant in the study has received instructions in advance, for example to act as an actor. Finally, we would like to point out that the concept for this experimental study was presented to the ethics committee of the University of Trier and was approved.

Figure B.1.5 Email templates used during sessions

1

Subject:	*Test*
Content:	*Test*

2

Subject:	*Survey*
Content:	*Please click on this link* _____

3

Subject:	*Estimation*
Content:	*Reply to this email with an estimate of how many articles your team will download and correctly label in 10 minutes. Additionally estimate, in brackets after the first estimate, how many articles you will download and label correctly in the work period that is about to start?*

4

Subject:	*Remuneration*
Content:	*Your team has created the following number of items: XXX* *You have thus earned the following points: YYY* *How satisfied are you with this remuneration?* *[Please answer with a value from 1 (for "completely dissatisfied") to 7 (for "completely satisfied").].*

5

Subject:	*Bonus game*
Content:	*Please reply to this email with a number from 1 to 20.*

6

Subject:	*Bonus game - Result*
Content:	*The number determined first: XXX* *The number determined by B: XXX* *The numbers add up./ The numbers do not match.* *There are 300 additional points <u>for B</u> {for A and for B}. / There are no extra points.* *Question: How many articles will you download and label? correctly in the upcoming work period?*

Notes: This figure shows the text for both two-person and three-person sessions. Differences between the treatment conditions are marked as follows. In two-person sessions, the text in curly brackets is left out, whereas, in three-person sessions, the underlined text is replaced by the text in curly brackets.

Figure B.1.6 Player selection sheet

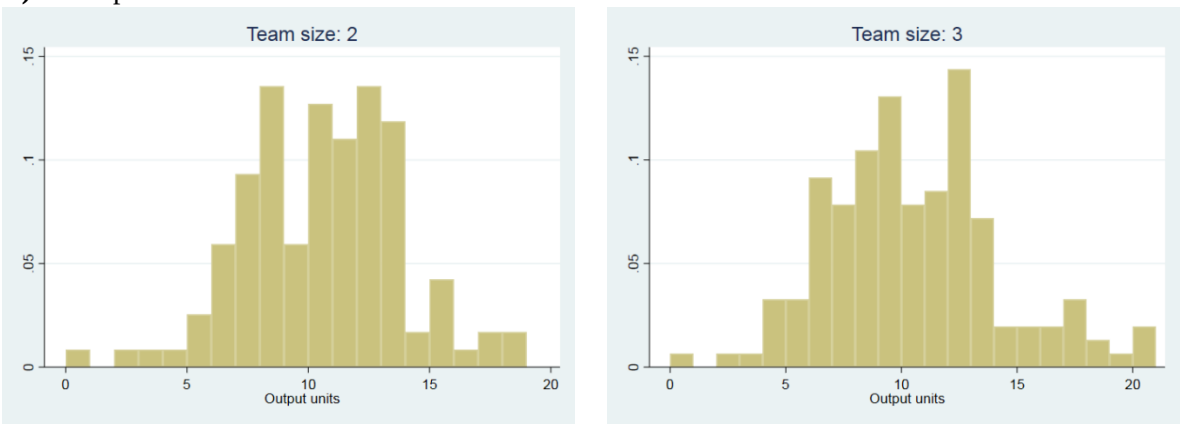
	Column selection (20-sided die of session host)			
	1 to 5	6 to 10	11 to 15	16 to 20
Selection of A (6-sided die of session host)	1,2,3 → team player 1 {1,2 → team player 1} 4,5,6 → team player 2 {3,4 → team player 3} {5,6 → team player 2}		1,2,3 → team player 2 {1,2 → team player 2} 4,5,6 → team player 1 {3,4 → team player 3} {5,6 → team player 1}	
Selection of B (20-sided die of A)	even number → team player on the left odd number → team player on the right	even number →team player on the right odd number →team player on the left	even number →team player on the left odd number →team player on the right	even number → team player on the right odd number → team player on the left

Notes: This figure shows the sheet for both two-person and three-person sessions. Differences between the treatment conditions are marked as follows. In two-person sessions, the text in curly brackets is left out, whereas, in three-person sessions, the underlined text is replaced by the text in curly brackets.

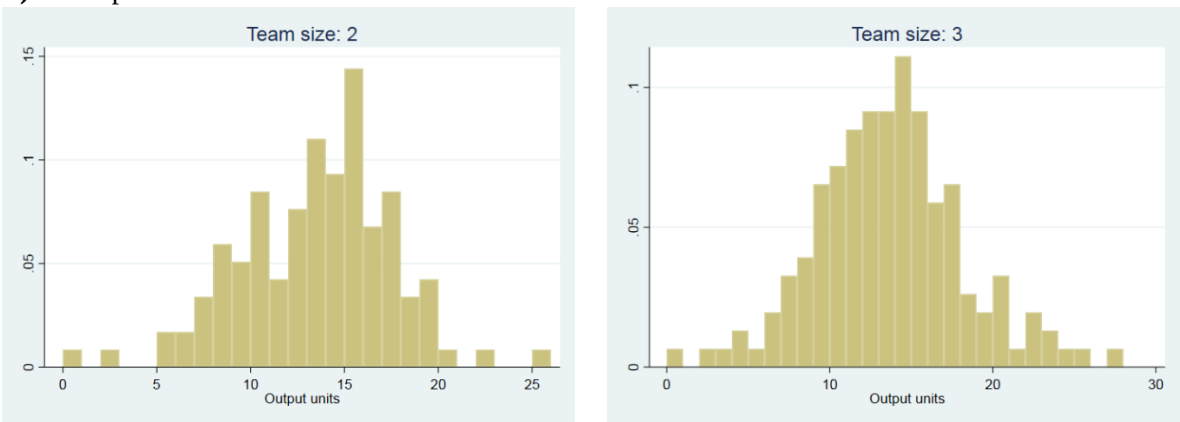
3.6 Appendix B.2: Supplementary statistics and analyses

Figure B.2.1 Histograms of task performance

A) Work period I



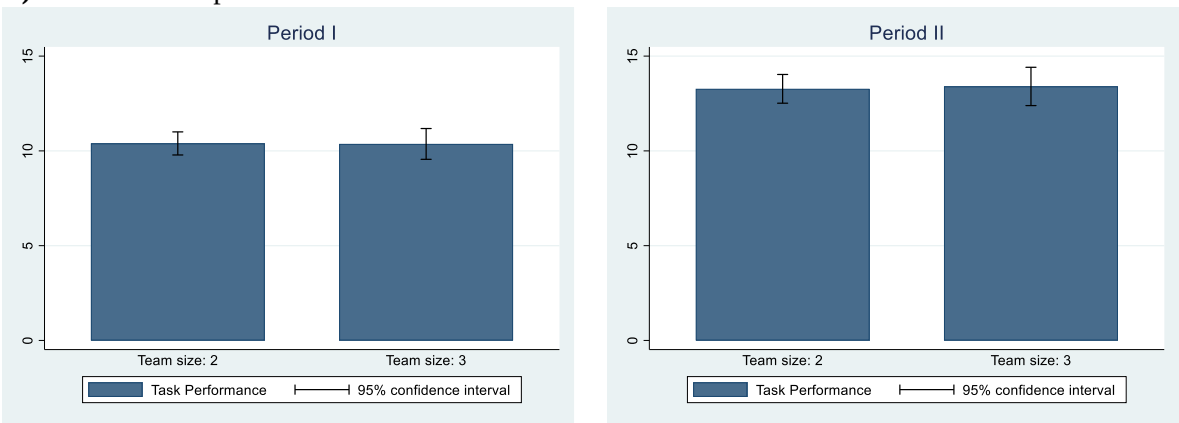
B) Work period II



Notes: Histograms show task performance by team size. The left (right) illustration displays task performance in teams of 2 (3). Panel A (B) shows task performance for work period one (two). 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.

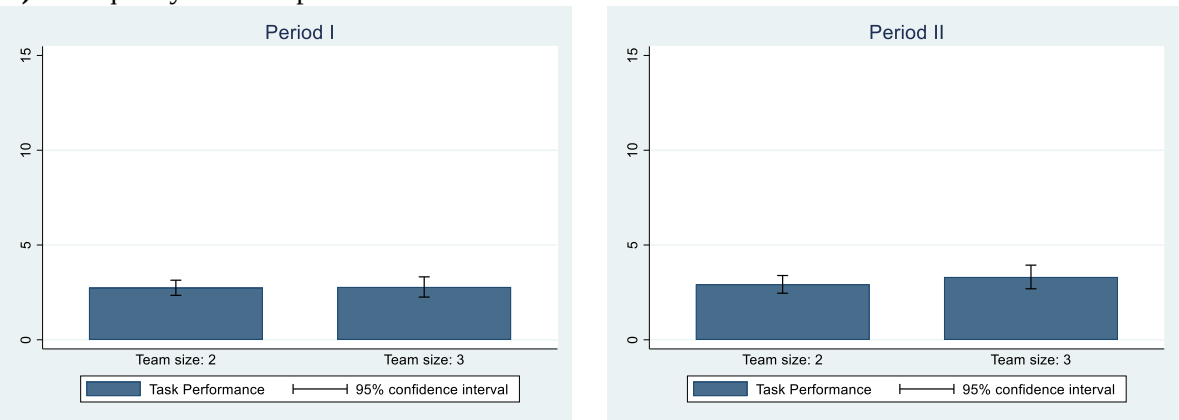
Figure B.2.2 Team size and performance dimensions

A) Total work output



B) High-quality work output

C) Low-quality work output



Notes: Bars show average performance by team size and by work period. Panel A displays the average number of output units, independent of correctness. Panel B (C) displays the average number of correct (incorrect) output units, according to software-based analyses. The left (right) illustration displays the first (second) work period. Total sample size in each panel is 267 observations.

Table B.2.1 Channels of team performance

A) Work period I	(1) Mean	(2) Mean	(3) Mean	(4) Mean	(5) Mean	(6) Mean
I.) Trust in the team	4.92	4.75	4.94	4.72	4.91	4.76
II.) Free-rider intention	1.37	1.41	1.28	1.45	1.49	1.38
III.) Unfair situation	1.58	1.60	1.62	1.63	1.53	1.58
IV.) Treated unfairly	1.38	1.39	1.37	1.47	1.40	1.33
V.) Assess my performance	3.54	3.38	3.71	3.40	3.34	3.37
VI.) Team identification	3.41	3.54	3.49	3.55	3.30	3.53
VII.) Not let down the team	5.35	5.35	5.25	5.53	5.47	5.24
VIII.) Work atmosphere	4.86	4.69	4.74	4.73	5.00	4.66
Number of team members:	2	3	2	3	2	3
Peer skills restriction:	-	-	Low	Low	High	High
N	118	153	65	60	53	93

B) Work period II	(1) Mean	(2) Mean	(3) Mean	(4) Mean	(5) Mean	(6) Mean
I.) Trust in the team	5.02	4.86	4.78	4.87	5.26	4.85
II.) Free-rider intention	1.38	1.35	1.52	1.35	1.24	1.36
III.) Unfair situation	1.75	2.24	1.62	2.06	1.88	2.43
IV.) Treated unfairly	1.50	1.70	1.47	1.50	1.53	1.91
V.) Assess my performance	3.77	3.63	3.88	3.74	3.66	3.51
VI.) Team identification	3.68	3.73	3.82	3.65	3.53	3.81
VII.) Not let down the team	5.31	5.24	5.22	5.33	5.40	5.15
VIII.) Work atmosphere	4.80	4.72	4.85	4.65	4.74	4.79
Number of team members:	2	3	2	3	2	3
(Un)equal pay treatment:	-	-	EP	EP	UEP	UEP
N	118	153	60	78	58	75

Notes: 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. In the case of the 'Equal pay' (EP) treatment, no bonus was paid out between work periods one and two. In the case of the 'Unequal pay' (UEP) treatment, a bonus was paid out between work periods one and two to all but one team members. Average scores of responses to various survey questions are shown in columns 1 to 6. The question battery starts with: "What is your view regarding this statement about the previous teamwork?" The statements are:

- "I have great trust in the other persons in this room." (I)
- "I lean back and let others do the work for me." (II)
- "I think what is happening here is unfair." (III)
- "Personally, I feel treated unfairly." (IV)
- "I can assess my performance within the team well." (V)
- "I identify myself with my team." (VI)
- "I do not want to let down my team." (VII)
- "The work atmosphere is good." (VIII)

The response scale ranges from "Totally disagree" (score: 1) to "Totally agree" (score: 7).

Table B.2.2 Reasons for work effort

	(1) Mean	(2) Mean	(3) Mean	(4) Mean	(5) Mean	(6) Mean
I.) Important contribution	5.76	5.54	5.83	5.22	5.68	5.74
II.) Felt monitored	2.59	2.78	2.74	2.58	2.42	2.90
III.) Image concerns	3.09	3.22	3.34	3.27	2.79	3.18
IV.) Competition	3.92	3.90	3.89	4.50	3.94	3.51
V.) Role model	3.90	3.75	3.91	3.77	3.89	3.74
VI.) Inspired	4.40	4.44	4.23	4.27	4.60	4.56
VII.) Exclusion fear	2.98	3.18	2.94	3.55	3.04	2.94
VIII.) Negative experience	2.36	2.41	2.54	2.38	2.13	2.42
IX.) Help others	4.59	4.67	4.71	4.78	4.45	4.59
Number of team members:	2	3	2	3	2	3
Peer skills restriction:	-	-	Low	Low	High	High
N	118	153	65	60	53	93

Notes: 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. Average scores of responses to various survey questions are shown in columns 1 to 6. The question battery starts with "Now we want to find out about why you have put in more or less effort. I have put in effort, because ..." The statements are:

- "...my performance was of great importance to the team results." (I)
- "...I felt like I was observed." (II)
- "... it was important to me what others might think about me." (III)
- "... the competition for the best performance motivated me." (IV)
- "...I want to be the team's role model through my performance." (V)
- "...the team's performance level motivated me." (VI)
- "... I did not want to be excluded from the team." (VII)
- "...it helps me better deal with negative experiences." (VIII)
- "... I like to do something for others." (IX)

The response scale ranges from "Totally disagree" (score: 1) to "Totally agree" (score: 7).

Table B.2.3 Peer skills and workplace perceptions

	(1) Mean	(2) Mean	(3) Mean	(4) Mean	(5) Mean	(6) Mean
I.) Monitored	3.00	3.22	3.05	3.12	2.94	3.28
II.) Observing	3.03	2.91	3.09	2.70	2.96	3.04
III.) Distracted	2.00	2.25	2.18	2.23	1.77	2.27
IV.) Competition	3.34	3.12	3.09	3.15	3.64	3.10
V.) Pressure	3.58	3.69	3.57	3.95	3.60	3.53
VI.) Uncomfortableness	2.56	2.69	2.54	2.70	2.58	2.68
VII.) Shirking others	2.52	2.56	2.28	2.60	2.81	2.53
Number of team members:	2	3	2	3	2	3
Peer skills restriction:	-	-	Low	Low	High	High
N	118	153	65	60	53	93

Notes: 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. Average scores of responses to various survey questions are shown in columns 1 to 6. The question battery starts with: "What is your view regarding this statement about the previous teamwork?" The statements are:

- "I feel monitored when working on the task." (I.)
- "I pay attention to the others while working on the task." (II.)
- "I feel distracted from working on my task." (III.)
- "I experience a competitive situation." (IV.)
- "I feel the pressure to succeed." (V.)
- "I feel uncomfortable when I have to work with strangers." (VI.)
- "I think it's problematic that not everyone is making an effort." (VII.)

The response scale ranges from "Totally disagree" (score: 1) to "Totally agree" (score: 7).

Table B.2.4 Team size and task performance (regression analysis)

Dependent variable:	Task performance			
A) Work period I	(1)	(2)	(3)	(4)
Team size 3	-0.074	-0.088	-0.338	-0.316
	(0.453)	(0.395)	(0.486)	(0.543)
Control variables	No	Yes	No	Yes
N	271	271	195	195
R^2	0.000	0.255	0.001	0.252
B) Work period II	(1)	(2)	(3)	(4)
Team size 3	0.140	0.176	-0.272	-0.125
	(0.568)	(0.485)	(0.697)	(0.699)
Control variables	No	Yes	No	Yes
N	271	271	195	195
R^2	0.000	0.273	0.001	0.277
C) Aggregated across periods	(1)	(2)	(3)	(4)
Team size 3	0.066	0.088	-0.609	-0.440
	(0.966)	(0.816)	(1.060)	(1.116)
Control variables	No	Yes	No	Yes
N	271	271	195	195
R^2	0.000	0.296	0.001	0.296

Notes: Linear regressions are used. The dependent variable is task performance, which is the number of output units used for performance pay after a spot check on correctness was conducted during sessions. Columns 1 and 2 show results for the full sample. Columns 3 and 4 show results for the dataset without transformed sessions. Team size 3 is a variable that is 1 if the team size is three and 0 if the team size is two. Clustered standard errors (at the session level) are in parentheses. See Table 3.1 for a list of the control variables used. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.2.5 Team size and the perception of working with peers

Dependent variable:	Positive peer perception			
	(1)	(2)	(3)	(4)
Team size 3	0.125**	0.131**	0.144*	0.181**
	(0.062)	(0.063)	(0.082)	(0.086)
Control variables	No	Yes	No	Yes
N	271	271	195	195
R^2	0.015	0.105	0.014	0.129

Notes: Linear regressions are used. Columns 1 and 2 show results for the full sample. Columns 3 and 4 show results for the dataset without transformed sessions. Team size 3 is a variable that is 1 if the team size is three and 0 if the team size is two. Dependent variable is a binary indicator for above-median responses to the question "How does it affect my motivation when others do the same work with me in a team?" with responses ranging from "Negative" (score: 1) to "Positive" (score: 50). Clustered standard errors (at the session level) are in parentheses. See Table 3.1 for a list of the control variables used. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.2.6 Team size and heterogeneous peer effects: Checks using control variables

	(1)	(2)	(3)	(4)
Dependent variable:	Task performance (aggregated)		Total work output (aggregated)	
Team size 3	-0.809	-0.778	-0.820	-0.764
	(0.962)	(0.891)	(0.925)	(0.863)
Peer skills	-0.272	-0.402	-0.121	-0.252
	(0.554)	(0.504)	(0.543)	(0.519)
Interaction: peer skills X team size 3	2.201*	2.148**	2.238**	2.127**
	(1.114)	(1.035)	(1.039)	(1.011)
Total effect of peer skills in teams of 3	1.929**	1.746*	2.116**	1.875**
	(0.966)	(0.881)	(0.886)	(0.835)
Control variables	No	Yes	No	Yes
R^2	0.016	0.309	0.020	0.286

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. Both performance measures are aggregated across the two work periods. 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 is 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. See Table 3.1 for a list of the control variables used. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.2.7 Team size and heterogeneous peer effects: Sample checks

	(1)	(2)	(3)	(4)
Dependent variable:	Task performance		Total work output	
Work period:	I	II	I	II
Team size 3	-1.024*	-1.039	-0.853	-0.968
	(0.554)	(0.785)	(0.542)	(0.781)
Peer skills	-0.595	-0.610	-0.588	-0.522
	(0.379)	(0.638)	(0.360)	(0.643)
Interaction: peer skills X team size 3	1.472**	1.661*	1.593***	1.633*
	(0.591)	(0.856)	(0.550)	(0.836)
Total effect of peer skills in teams of 3	0.878*	1.051*	1.005**	1.111**
	(0.454)	(0.570)	(0.415)	(0.533)
R^2	0.026	0.022	0.033	0.025

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 is 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 dataset without transformed sessions (N=195) is used. Clustered standard errors (at the session level) are in parentheses. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4 Perceived co-worker attractiveness and task performance: Beware of the opposite sex![*]

We investigate the role of perceived co-worker attractiveness for performance by analyzing a rich dataset on more than a hundred work groups. The results show that task performance is lower, the higher the perceived attractiveness of co-workers, but only in opposite-sex constellations. The performance drop is noticeable in two-person scenarios where the individual performance level is reduced by roughly 15% if the opposite-sex co-worker is perceived as above-average attractive. Our evidence on potential mechanisms speaks against distraction through mating activities but instead supports the idea of underperformance due to cognitively demanding self-regulation, which is reflected in increased emotional arousal.

[*]This chapter is joint work with Adrian Chadi. We are grateful for comments by Laszlo Goerke, Gari Walkowitz as well as participants of the IAAEU seminar, the Lüneburg workshop on microeconomics and the Trier workshop on team dynamics and peer effects. 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.

"The Lord prefers common-looking people. That is why he made so many of them." (Abraham Lincoln)

4.1 Introduction

Human interactions are strongly affected by the way people perceive each other. Being perceived as attractive or not can have substantial economic implications, according to research on earnings (e.g. Hamermesh and Biddle 1994, Mobius and Rosenblat 2006, Doorley and Sierminska 2015, Parrett 2015) and hiring chances (e.g. Rooth 2009, López Bóo et al. 2013, Baert and Decuypere 2014, Ruffle and Shtudiner 2015). While attractiveness appears to be relevant to one's own labor market outcomes, little is known about the behavioral and economic effects of working with someone who is subjectively perceived as more or less attractive. This question could be particularly relevant, given that recent research emphasizes the role of gender constellations in economic interactions (e.g. Bodenschatz and Walkowitz 2020, Gerhards and Kosfeld 2020), whereas organizations increasingly foster gender diversity in their workforces.

There are many reasons why perceived co-worker attractiveness could trigger behavioral responses and thereby potentially affect workplace performance. According to Karremans et al. (2009), men's cognitive function declines in opposite-sex interactions compared to same-sex interactions, especially if the opposite-female is perceived as attractive. The authors discuss several ideas, including mating goals, impaired self-regulation and impression management to explain their findings. For a workplace context, the first two explanations constitute clear threats to task performance if workers are distracted by mating activities or if they cannot focus on their work due to cognitively demanding self-regulation. In contrast, the role of impression management in this context is not so clear. Assuming that individuals may try to impress co-workers by putting in extra effort to be seen as particularly hard-working (Corgnet et al. 2015c), this motive could play a positive role for workplace performance if such efforts increase in perceived co-worker attractiveness. Another strand of research on group behavior also suggests potentially beneficial implications of perceived attractiveness, which might serve as a signal of willingness to cooperate with others (Andreoni and Petrie 2008).

In this paper, we provide first evidence on the role of perceived co-worker attractiveness in workplace performance by using a rich dataset on work groups with more than a hundred students participating in a large research project. We thereby aim to contribute to the ongoing discussion on the role of attractiveness in labor market outcomes and to shed light on possible side-effects of promoting gender diversity at the workplace. In contrast to previous research on

attractiveness, we are studying behavioral effects in individuals who are working together with others who are subjectively seen as more or less attractive by the very individuals whose behavior we analyze. Hence, we are not interested in the effects of objectively attractive co-workers, but rather we analyze the possible implications of being confronted with someone subjectively perceived as attractive from an individual's viewpoint. For this purpose, we can use single attractiveness assessments gathered from all project participants before performing a task at a computer workstation and being incentivized to put in effort via performance-related pay. Besides subjective ratings on others' appearance, which typically is difficult to obtain in authentic workplace contexts, our dataset enables us to study individual performance levels and underlying mechanisms when others are perceived as more or less attractive. Another feature of our dataset is the experimental manipulation of team size.

In the following Section 4.2, we provide more information on the background and the data. We present the empirical analysis in Section 4.3, and discuss conclusions from our findings in Section 4.4.

4.2 Empirical framework

4.2.1 Setting

Our dataset comes from a large research project on work groups, which took place in late 2016 and early 2017 at a German university. Hundreds of students were recruited as short-term employees to perform a routine task at a computer workstation in groups of two or three individuals. Data from this setting were used to address several research questions regarding teamwork, including the behavioral effects of cheating (Chadi and Homolka 2021a) and the role of team size in workplace performance (Chadi and Homolka 2021b). Recruitment of study participants was organized by our assistants and followed pre-defined rules with the objective to minimize both chances of social ties between session participants and selectivity in individual characteristics across work sessions. The determination of the co-workers and the resulting group composition was exogenous from an individual's point of view, as it was unclear who and how many other participants were assigned to attend the same session.

All sessions took place in the same office with several computer workplaces (for a picture, see Appendix Figure C.1.1). The task was to download and label journal articles from an online library. Given that this was done similarly by actual research assistants at that time, the task can be seen as authentic. Each participant had to work on a volume of a pre-selected journal and

prepare as many articles as possible within the working time. Journal volumes were pre-selected, while both the journal (for the session) and the volumes (within the session) were randomly drawn before each session. A male research assistant was in the role of the session host, guiding through the procedure by reading out instructions. While oral communication was prohibited for participants, they could communicate via email. The work task consisted of two periods, lasting ten minutes each, preceded by a short test run.

Payments were, on average, about 12 Euro for an hour-long participation. In addition to fixed payments (of 8.50 Euro) for showing-up and completing online surveys, participants could increase their payoffs based on a team piece rate scheme, which incentivized performance in the work task by rewarding the team's total average work output (0.1 Euro for each prepared article). To emphasize the idea of teamwork, participants were only informed about the team result. Furthermore, a dice game took place between the two work periods, which allowed for a bonus win (of 3 Euro).[49] After the session, participants received their payments in separate rooms.

4.2.2 Sample

The workplace setup provides us with a dataset that includes information on individual participants from various sources (including webpage registration, surveys, and session records). First and foremost, the dataset informs us about the perceived attractiveness of others, as our independent variable, thanks to subjective ratings regarding the appearance of other participants. Ratings were obtained on a seven-point Likert scale and were provided via a survey conducted prior to the start of the work (see Appendix Figure C.1.2).[50] As our main dependent variable, we focus on the number of scientific articles prepared in the work task, referred to as 'task performance' in the following, which comes from the session records. This performance indicator was determined by the session host after a spot-check procedure of the

[49] The bonus game plays an important role in Chadi and Homolka (2021a). We checked whether factors related to the bonus game are relevant for the findings presented in this paper, but could not find any evidence that this is the case (e.g. when we control for selection of bonus players or bonus outcomes in additional regression analyses). Similarly, we also checked whether aspects related to the second parallel experiment might be relevant for our study here. Of potential relevance is the fact that not all of the two-person sessions were randomly scheduled as such, but some were initially scheduled as three-person sessions with one invitee not showing up. As discussed in more detail in Chadi and Homolka (2021b), using data from those non-randomly assigned two-person sessions has no implications for the empirical analysis other than to increase the sample size.

[50] Note that subjective attractiveness ratings typically differ between males and females, the latter of which are often rated a bit more positively (Andreoni and Petrie 2008, Pfeifer 2012). This is also true for our data (see Appendix Figure C.1.3), as we observe an average female attractiveness that is almost one score point higher on the Likert scale compared to males. To address this, we employ a gender-conditional attractiveness indicator that considers this gender gap in ratings by reducing co-worker attractiveness by the gender-specific average.

work output's quality and served as the basis for the performance-related pay. For our main analysis, we aggregate individual task performance across both work periods (for histograms, see Appendix Figure C.1.4).

Various pieces of information allow us to learn more about the mechanisms through which perceived attractiveness of others could matter for one's own task performance, including data on the occurrence of email communication during sessions as an indicator of mating activities. To inspect a possible need for self-regulation, we analyze self-reported emotions gathered through a variant of the positive-affect-negative-affect scale, which was part of a survey conducted after the second work period. Furthermore, our dataset includes information on skills and individual characteristics, including Big Five personality, reciprocity, and trust, all serving as control variables in our analysis.

The workplace setup allows us to distinguish between opposite-sex and same-sex constellations in teams of two. In teams of three, it is possible to make this distinction as long as the two remaining co-workers were of the same gender. In this case, we determine the average co-worker attractiveness.[51] Table C.1.1 illustrates the dataset that emerges out of these considerations and contains information on 342 participants, of which ca. 60% are female.[52]

4.3 Empirical analysis

4.3.1 Perceived co-worker attractiveness and task performance

We start our empirical analysis by running regressions with task performance across work periods as the dependent variable and perceived co-worker attractiveness as the independent variable. For ease of interpretation, we run linear regressions and interpret perceived attractiveness as a continuous measure. In a first step, we consider a parsimonious set of control variables (team size, own gender, opposite-sex constellation) before we expand the set of controls by using further variables, as shown in Table C.1.1.

[51] We do not use observations from individuals in three-person teams when the two co-workers are of different gender, as perceived average co-worker attractiveness would not be clear in its meaning, if attractiveness of others matters dependent on gender. Additional analyses are in line with this expectation, as we find no evidence for a link between individual performance and others' average attractiveness if the two others have different gender. Note that the information on gender of participants comes from session records and is part of the observable data collected during the session. We checked for measurement errors in this variable, but could not find any evidence pointing in this direction based on additional inspections of the available information on the participants.

[52] Since no intervention in the natural selection process was made via gender-based sample stratification, there is also a similar gap in the likelihood of same-sex vs. opposite-sex constellations (see Table C.1.1).

Panel A of Table 4.1 shows that there is no empirical link between perceived co-worker attractiveness and one's own task performance when we do not distinguish between opposite-sex and same-sex constellations, which is independent of the set of control variables (columns 1 and 2). When we inspect gender constellations separately using sample-split interaction variables, we find a performance decline with increasing co-worker attractiveness in opposite-sex constellations (columns 3 and 4).[53] For same-sex constellations, we observe a weakly positive performance effect of others' attractiveness, but this is not robust across specifications. We exclude same-sex constellations in the following, to find out more about the performance decline in the case of opposite-sex constellations.

Panel B of Table 4.1 shows that across specifications, the role of perceived co-worker attractiveness is particularly strong in teams of two individuals. This indicates that co-worker attractiveness matters less the larger the number of co-workers is.[54] Regarding possible gender differences, we find that individual performance responds similarly negative to perceived co-worker attractiveness in smaller teams, independent of one's own gender. While the effects are statistically insignificant in larger teams, we observe that the direction of the effect for females is reversed from negative to positive.

As part of our sensitivity analyses, we inspect performance effects separately for each work period. Appendix Table C.1.2 shows a similar picture across work periods and when using an alternative performance indicator, that is, the total number of output units (independent of correctness). As an alternative to regression analyses with an independent variable assumed to be continuous, we generate a high-attractiveness dummy by separating between above-average and below-average attractiveness scores in order to conduct non-parametric significance testing. A two-sided ranksum-test (p=0.023) provides further evidence that co-worker attractiveness matters for performance in small teams (two-sided t-test: p=0.034).

[53] For interpretation purposes, we prefer employing sample-split interactions terms, which inform us about the total effect of attractiveness for each subsample, rather than employing differential interaction terms, which would inform us about the difference in the effects between subsamples. For example, the differential interaction effect across sex constellations in column 4 of Panel A amounts to -1.490 output units per score point, which is highly significant (p=0.002) but the meaning of this gap in the effects between sex constellations is not clear.

[54] As an alternative interpretation, it might technically be more difficult to capture co-worker attractiveness if there is more than one co-worker. Instead of using average co-worker attractiveness in teams of three, as we do in our baseline specification, we hence use only the rating of the more attractive (less attractive) co-worker among the two others in additional analyses, which does not reveal significant effects either.

Table 4.1 Task performance

A) Full sample (N=342)	(1)	(2)	(3)	(4)
Team size 3	0.439	0.364	0.377	0.268
	(0.826)	(0.764)	(0.822)	(0.751)
Female gender	-2.305**	-1.293	-2.279**	-1.295
	(0.904)	(1.055)	(0.897)	(1.053)
Opposite-sex constellation	-0.489	-0.397	-0.518	-0.460
	(0.866)	(0.815)	(0.863)	(0.816)
Co-worker attractiveness	-0.227	0.032		
	(0.306)	(0.273)		
Interactions: Co-worker attractiveness…				
X Same-sex constellation			0.219	0.606*
			(0.381)	(0.317)
X Opposite-sex constellation			-0.940**	-0.884**
			(0.455)	(0.399)
Further control variables	No	Yes	No	Yes
R^2	0.023	0.252	0.031	0.265
B) Opposite-sex sample (N=140)	(1)	(2)	(3)	(4)
Team size 3	-0.078	1.069	0.071	1.054
	(1.073)	(1.053)	(1.077)	(1.061)
Female gender	-1.786	-1.952	-1.326	-1.564
	(1.256)	(1.441)	(1.278)	(1.393)
Interactions: Co-worker attractiveness…				
X Team size 2	-1.751***	-1.644***		
	(0.362)	(0.594)		
X Team size 3	-0.300	-0.020		
	(0.763)	(0.629)		
X Male gender X Team size 2			-1.907***	-1.727**
			(0.530)	(0.830)
X Female gender X Team size 2			-1.639***	-1.664**
			(0.362)	(0.781)
X Male gender X Team size 3			-1.393	-0.634
			(1.146)	(0.815)
X Female gender X Team size 3			1.108	0.696
			(0.783)	(0.922)
Further control variables	No	Yes	No	Yes
R^2	0.054	0.428	0.074	0.433

Notes: The sample in Panel A (B) includes 342 (140) individual observations. The dependent variable is the number of output units prepared across both work periods. Linear regressions are used. Throughout the table's specifications, split-interactions are used, distinguishing between the effects of perceived co-worker attractiveness on task performance across subsamples. Further controls include variables for age, semester, degree, Big5, reciprocity, trust, and skills. For more information on the variables used, see Appendix Table C.1.1. Clustered standard errors (at the session level) are in parentheses. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Strikingly, the number of output units is 21.81 when the opposite-sex co-worker is seen as above-average attractive, which is a substantial drop of 15.3% compared to 25.75 output units when perceived co-worker attractiveness is below-average.

In further analyses, we inspect whether co-worker attractiveness serves as a proxy for other factors instead of being directly relevant for performance. As can be seen in Appendix Table C.1.3, our main results are unaffected when we consider skills of co-workers as additional control variables. Furthermore, in Appendix Table C.1.4, we consider all available ratings from the survey module on co-worker appearance, including other subjective items besides attractiveness. The results reveal a clear picture, as only the attractiveness variable is significantly related to performance. Among the insignificant items are perceived fairness and helpfulness, which counters the idea that others' appearance could serve as a performance-increasing signal of cooperativeness in a work group.[55] In the following, we discuss explanations of why perceived co-worker attractiveness may decrease workplace performance.

4.3.2 Mechanisms

One reason why perceived attractiveness of co-workers could be performance-decreasing relates to mating activities. Individuals may invest time in communicating with others seen as attractive instead of focusing on the task and being productive. In order to socialize with others during the work session, participants in our setting had to send emails. Panel A of Table 4.2 shows the results from regression analyses using a continuous indicator (columns 1 and 2) and a binary indicator (columns 3 and 4) of email communication with other participants as the dependent variables. If anything, email communication becomes less likely in small teams the more attractive the other participant is perceived, which does not conform to the idea of distraction through mating activities via email as an explanation for the performance decline.[56]

[55] Results also hold if we consider the co-worker ratings analyzed in Table C.1.4 as additional covariates. A possible reason to conduct such a check is that subjective perceptions in general could be affected by external factors, e.g. current weather conditions, which may separately affect the performance at work, though in the opposite way. We find however no empirical evidence in support of this idea. In fact, our dataset also includes ratings about oneself, which we can alternatively add as control variables without changing the findings.

[56] Further evidence against a role of mating in our setting comes from a check, for which we use survey responses about partnership status. We find no significant interaction with co-worker attractiveness, suggesting that our main performance result is not driven by individuals who are more likely than others searching for a partner.

Table 4.2 Communication and emotions

A) Email communication	(1)	(2)	(3)	(4)
Dependent variable:	Number of emails		Any email communication	
Team size 3	0.806***	0.746**	0.075	0.081
	(0.303)	(0.322)	(0.101)	(0.105)
Female gender	-0.281	0.078	-0.057	0.050
	(0.327)	(0.303)	(0.083)	(0.105)
Interactions: Co-worker attractiveness…				
X Male gender X Team size 2	0.004	-0.102	-0.091	-0.103*
	(0.152)	(0.213)	(0.067)	(0.059)
X Female gender X Team size 2	-0.252	-0.359	-0.104**	-0.143*
	(0.156)	(0.239)	(0.052)	(0.069)
X Male gender X Team size 3	0.264	0.098	0.012	-0.023
	(0.304)	(0.231)	(0.071)	(0.065)
X Female gender X Team size 3	-0.072	0.146	0.012	0.034
	(0.228)	(0.334)	(0.076)	(0.082)
R^2	0.057	0.212	0.038	0.149
Further control variables	No	Yes	No	Yes
B) Emotional arousal	(1)	(2)	(3)	(4)
Dependent variable:	Average emotions score		Share of emotional responses	
Team size 3	0.009	0.018	0.016	-0.002
	(0.067)	(0.071)	(0.031)	(0.032)
Female gender	-0.049	-0.072	-0.026	-0.032
	(0.075)	(0.095)	(0.035)	(0.039)
Interactions: Co-worker attractiveness…				
X Male gender X Team size 2	0.143***	0.176***	0.086***	0.099***
	(0.040)	(0.055)	(0.022)	(0.028)
X Female gender X Team size 2	-0.003	-0.001	0.057*	0.061**
	(0.035)	(0.048)	(0.033)	(0.028)
X Male gender X Team size 3	0.095	0.123*	0.043	0.052**
	(0.061)	(0.063)	(0.026)	(0.026)
X Female gender X Team size 3	-0.056	-0.065	0.031	0.026
	(0.075)	(0.066)	(0.034)	(0.027)
R^2	0.060	0.175	0.107	0.271
Further control variables	No	Yes	No	Yes

Notes: The sample includes 140 individual observations. The dependent variable in columns 1 and 2 of panel A is the number of emails sent to other co-worker (columns 3 and 4: a dummy reflecting the incidence that any email was sent to another co-worker). The dependent variable in columns 1 and 2 of panel B is the average emotional score established based on all 20 items reflecting positive and negative emotions on the 5-point scale (columns 3 and 4: the share of items from the emotions survey module for which a score of at least 2 on the 5-point scale was reported). For an analysis using all 20 emotion items, see Appendix Table C.1.5. Linear regressions are used. Throughout the table's specifications, split-interactions are used, distinguishing between the effects of perceived co-worker attractiveness on task performance across subsamples. Further controls include variables for age, semester, degree, Big5, reciprocity, trust, and skills. For more information on the variables used, see Appendix Table C.1.1. Clustered standard errors (at the session level) are in parentheses. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In a second step, we analyze whether perceived attractiveness of others could trigger emotional arousal and thereby lower performance due to impaired self-regulation. We first use the average score across all emotions from the 20-item survey module as our dependent variable. The results in Panel B of Table 4.2 (columns 1 and 2) indicate that males report more emotional arousal, especially in small teams, the more they perceive the female co-worker as attractive. Appendix Table C.1.5 presents the full results separately for all 20 items, showing that for males, the psychological effects are often significant and are sometimes positive (e.g. excited) sometimes negative (e.g. sad). The full results also show that females appear to be emotionally responsive in some ways, albeit solely negatively (e.g. upset). When we employ an alternative indicator by using the share of emotional responses across all survey items as the dependent variable in Panel B in Table 4.2 (columns 3 and 4), we find more evidence for emotional arousal, especially in small teams and also for both genders, which supports the idea that performance-decreases could be due to lacking self-regulation.

Overall, our findings do not entirely rule out possible gender differences in the mechanisms of how performance declines when co-worker attractiveness is perceived as high. Results from further analyses of the available data suggest that females could be engaged in a form of impression management and want to impress male co-workers perceived as attractive by putting in extra efforts. Yet, in a face-to-face environment seen as particularly competitive, their performance might then rather decline under high pressure to perform well. Evidence for this comes from the results in Table 4.1, as the effect of male co-worker attractiveness on performance of females reverses from negative to positive when team size increases. Arguably, the psychological pressure to perform well is lower in larger teams with less transparency of individual contributions to the team's overall work output. Further evidence comes from our rich survey dataset, which includes individual workplace perceptions, according to which females perceive the situation as more competitive the higher the perceived co-worker attractiveness is (see Appendix Table C.1.6).[57] Nonetheless, we are cautious with conclusions at this point, given that various unique workplace features could affect gender differences in behavior, including, for example, the nature of the task (Günther et al. 2010).

[57] Inspecting other subjective items from the survey dataset, such as task enjoyment, we observe further indication that females in smaller teams are reporting more negatively about the work environment, suggesting a certain feeling of discomfort, when they work with a male co-worker perceived as more attractive. This might also be reflected in responses to survey questions on future work motivation (see Appendix Table C.1.7). Here, we find that males tend to report higher motivation the more female co-workers are perceived as attractive, which indicates a potential for positive performance effects in the long-run, while no such evidence is observed in females working together with male co-workers perceived as attractive.

4.4 Discussion

With this paper, we present first evidence for potential performance reductions when working together with opposite-sex co-workers perceived as attractive. A likely explanation is impaired self-regulation when emotional arousal prevents individuals from performing well. Having said this, while performance effects in small teams are robust across the sexes, our evidence on mechanisms leaves room for gender differences.

In an effort to draw implications concerning the discussion on gender diversity, our results are informative about possible short-run side-effects when working together with opposite-sex co-workers. While research on team composition typically emphasizes the benefits of gender diversity for performance (e.g. Apesteguia et al. 2012, Hoogendoorn et al. 2013), the evidence from our setting does not conform to this notion and hence should caution personnel departments pursuing diversity policies without considering all the possible impacts on worker behavior. However, assuming that emotional arousal and lack of self-regulation could result from transitory problems of working together with opposite-sex co-workers perceived as attractive, one could also argue in favor of possible long-run benefits that are not revealed in our analysis of task performance. If so, the question of how co-worker attractiveness might affect workplace outcomes becomes increasingly complex; but, in our view, this underlines even more the attractiveness of the topic and the need for future research.

4.5 Appendix C.1: Supplementary statistics and analyses

Figure C.1.1 The computer workstation

Notes: The picture shows the computer office with three workplaces for the team members and the seat of the session host in the background. The computer of the host was connected to several screens, including a big screen that was used to display important procedural information, such as task instructions. Note that in case of a session with only two team members, the equipment of the third workplace was removed prior to the start of the session.

Figure C.1.2 Timeline of the work session

Welcome
Introduction to the rules regarding communication and pay

Survey 1
Including items for reciprocity, trust levels, risk attitudes and co-worker ratings, such as attractiveness

Instruction phase
Including test run

Work period 1
10 minutes

Survey 2
Including items for Big 5 personality, skills, and workplace perceptions

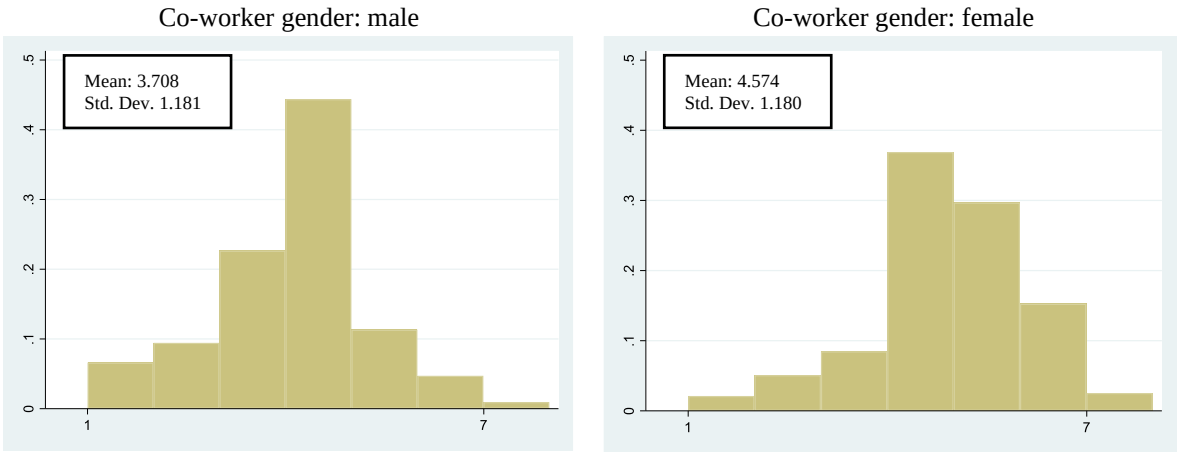
Bonus game
Possible win for randomly selected bonus player(s)

Work period 2
10 minutes

Survey 3
Including items for emotions and future work motivation

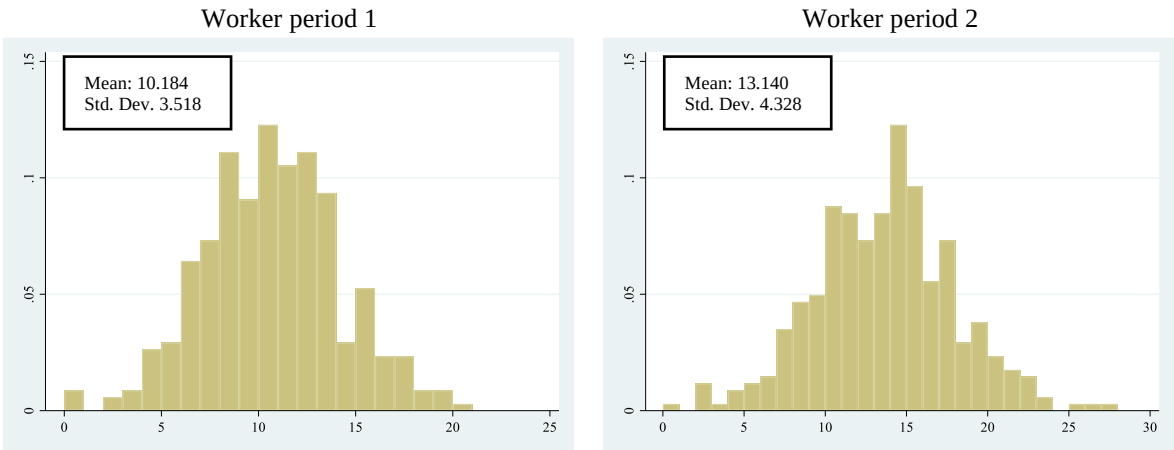
Payout
Pay and final feedback in separate rooms

Figure C.1.3 Perceived co-worker attractiveness by co-worker gender



Notes: The sample used in the left-side (right-side) illustration includes 106 (236) individual observations. The left-side illustration shows a histogram of attractiveness ratings reported by participants of two-person teams with a male co-worker and by participants of three-person teams with two male co-workers. The right-side illustration shows a histogram of attractiveness ratings reported by participants of two-person teams with a female co-worker and by participants of three-person teams with two female co-workers.

Figure C.1.4 Task performance by work period



Notes: The sample includes 342 individual observations. The left-side (right-side) illustration shows a histogram of the number of output units prepared in work period 1 (2).

Table C.1.1 Descriptive statistics for the full sample

	(1) Mean	(2) Std. dev.	(3) Min.	(4) Max.
Task performance (across both work periods)	23.325	7.304	2	47
Co-worker attractiveness (de-meaned)	0.000	1.179	-3.574	3.292
Number of emails	1.032	2.011	0	13
Incidence of email communication	0.652	0.477	0	1
Average emotional score	1.915	0.423	1	3.9
Share of emotional responses	0.487	0.193	0	1
Basic control variables				
Team size 3	0.655	0.476	0	1
Female gender	0.614	0.488	0	1
Opposite-sex constellation	0.409	0.492	0	1
Further control variables				
Age	24.898	3.607	18	53
Number of semester	5.234	4.118	1	25
Bachelor's degree	0.295	0.457	0	1
Big5: extraversion	4.909	1.295	1.333	7
Big5: agreeableness	5.355	0.975	1.667	7
Big5: openness	4.942	1.188	1.333	7
Big5: neuroticism	4.407	1.249	1	7
Big5: conscientiousness	5.161	1.071	1.333	7
Reciprocity	6.080	0.775	2	7
Trust	3.512	0.820	1	5
Math skills	3.187	1.056	1	5
Probability skills	3.003	1.009	1	5
Computer skills	3.380	0.982	1	5

Notes: The sample includes 342 individual observations. Column 1 display means. Column 2 displays standard deviations. Columns 3 and 4 display minimum and maximum values. Task performance is the number of output units prepared taken from the session records. Co-worker attractiveness is a survey-based rating of the other co-worker in a team of two, respectively the average rating for the two other co-workers in teams of three, regarding attractiveness. This variable is based on responses given on a scale ranging from 1 ("does not apply at all") to 7 ("does fully apply") for the "attractive" item and is de-meaned (by using gender-specific averages regarding co-worker attractiveness). Number of emails are all emails sent to co-workers except for procedural emails that were sent as a result of the instructions. For information on average emotional score and share of emotional responses, see the notes in Tables 4.2 and C.1.5. Team size 3 is a dummy that is 1 if the individual is part of a three-person team or 0 if the individual is part of a two-person team. Female gender is a dummy variable taken from the session records. Opposite-sex constellation is a dummy that is 1 if the other individual in a team of two has a different gender, respectively the other two individuals in a team of three have a different gender, and 0 otherwise. Age in years, number of semesters, and bachelor's degree (yes/no) are variables self-reported by participants during webpage registration. The wording of 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. Variables on skills are measured on a 5-point ordinal scale, ranging from "Very bad" (score: 1) to "Very good" (score: 5). The wording of the survey questions on math / probability / computer skills is "How good are you…at math?" / "…at judging probabilities?" / "…with computers?".

Table C.1.2 Task performance: Sensitivity analyses

	(1)	(2)	(3)	(4)
Dependent variable:	Main performance indicator		Alternative performance indicator	
Work period:	1	2	1	2
Team size 3	-0.238	0.160	-0.113	0.046
	(0.498)	(0.710)	(0.474)	(0.687)
Female gender	-1.078*	-0.708	-1.138**	-0.614
	(0.591)	(0.801)	(0.567)	(0.744)
Interactions: Co-worker attractiveness…				
X Team size 2	-0.913***	-0.838***	-0.757***	-0.778**
	(0.233)	(0.302)	(0.244)	(0.298)
X Team size 3	-0.158	-0.142	-0.164	-0.223
	(0.384)	(0.437)	(0.377)	(0.414)
R^2	0.069	0.032	0.067	0.030

Notes: The sample includes 140 individual observations. The dependent variable in columns 1 and 2 is the number of output units prepared in work periods 1 and 2. The dependent variable in columns 3 and 4 is the total number of output units (independent of correctness) in work periods 1 and 2. Linear regressions are used. For more information on the variables used, see Appendix Table C.1.1. Clustered standard errors (at the session level) are in parentheses. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.1.3 Task performance: Co-worker skills

	(1)	(2)	(3)	(4)
Dependent variable:	Main performance indicator		Alternative performance indicator	
Work period:	1	2	1	2
Team size 3	-0.085	1.210	-0.114	1.143
	(1.099)	(1.095)	(1.040)	(1.063)
Female gender	-1.662	-1.907	-1.698	-1.757
	(1.365)	(1.483)	(1.370)	(1.521)
Co-worker math skills	0.093	-0.379	0.202	-0.293
	(0.747)	(0.724)	(0.665)	(0.680)
Co-worker probability skills	-0.475	-0.084	-0.429	-0.067
	(0.724)	(0.719)	(0.734)	(0.724)
Co-worker computer skills	0.569	0.182	0.616	0.309
	(0.847)	(0.729)	(0.818)	(0.730)
Interactions: Co-worker attractiveness…				
X Team size 2	-1.788***	-1.570**	-1.607***	-1.450**
	(0.390)	(0.609)	(0.410)	(0.589)
X Team size 3	-0.286	0.017	-0.385	-0.118
	(0.769)	(0.640)	(0.734)	(0.640)
R^2	0.058	0.430	0.056	0.403

Notes: The sample includes 140 individual observations. The dependent variable in columns 1 and 2 is the number of output units prepared in work periods 1 and 2. The dependent variable in columns 3 and 4 is the total number of output units (independent of correctness) in work periods 1 and 2. Linear regressions are used. Co-worker skills are (average) skills as reported by other co-workers in the same session on a 5-point ordinal scale, ranging from "Very bad" (score: 1) to "Very good" (score: 5). For more information on the variables used, see Appendix Table C.1.1. Clustered standard errors (at the session level) are in parentheses. Levels of significance are: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C.1.4 Performance and different ratings of co-workers

	(1)	(2)	(3)	(4)	(5)	(6)
Team size 3	0.384	0.422	-0.078	0.327	0.687	0.508
	(1.141)	(1.106)	(1.073)	(1.153)	(1.038)	(1.095)
Female gender	-1.822	-1.996	-1.786	-1.927	-2.130	-2.205*
	(1.294)	(1.273)	(1.256)	(1.297)	(1.291)	(1.289)
Interactions: Co-worker…	motivated	shy	attractive	nervous	fair	helpful
X Team size 2	0.827	-0.616	-1.751***	0.161	0.824	-0.207
	(0.941)	(0.682)	(0.362)	(0.578)	(0.610)	(0.739)
X Team Size 3	-0.897	-0.392	-0.300	0.096	0.556	1.078
	(0.919)	(0.965)	(0.763)	(0.706)	(0.916)	(0.825)
R^2	0.033	0.026	0.054	0.019	0.027	0.035

Notes: The sample includes 140 individual observations. The dependent variable is the number of output units prepared across both work periods. The independent variables include a co-worker rating (de-meaned using gender-specific averages), interacted with gender constellation. The ratings of other participants are captured on a 1 to 7 Likert scale and refer to six different terms: motivated (column 1), shy (column 2), attractive (column 3), nervous (column 4), fair (column 5), and helpful (column 6). Linear regressions are used. For more information on the variables used, see Appendix Table C.1.1. Clustered standard errors (at the session level) are in parentheses. Levels of significance are: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.1.5 All emotions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable:	Active	Distressed	Excited	Upset	Guilty	Scared	Hostile	Proud	Irritable	Enthusiastic
Team size 3	-0.036	0.045	-0.094	0.006	0.070	-0.078	0.079	0.053	0.089	0.100
	(0.033)	(0.082)	(0.080)	(0.071)	(0.086)	(0.057)	(0.052)	(0.090)	(0.080)	(0.077)
Female gender	-0.034	-0.176**	0.001	-0.104	0.016	-0.087	-0.009	0.084	-0.072	0.015
	(0.039)	(0.082)	(0.086)	(0.088)	(0.079)	(0.060)	(0.049)	(0.083)	(0.091)	(0.084)
Interactions: Co-worker attractiveness…										
X Male gender X Team size 2	-0.034	-0.015	0.139**	0.210***	0.126**	0.140**	-0.063	0.016	0.061	0.107*
	(0.035)	(0.080)	(0.056)	(0.055)	(0.051)	(0.053)	(0.057)	(0.077)	(0.073)	(0.058)
X Female gender X Team size 2	0.008	0.111*	-0.014	0.132**	0.110*	0.123**	0.097	0.060	0.071	0.029
	(0.007)	(0.057)	(0.070)	(0.057)	(0.064)	(0.057)	(0.059)	(0.066)	(0.066)	(0.085)
X Male gender X Team size 3	0.040	0.040	0.096	-0.044	0.010	0.000	-0.115**	0.116*	-0.023	0.119*
	(0.042)	(0.069)	(0.064)	(0.070)	(0.070)	(0.042)	(0.050)	(0.066)	(0.070)	(0.064)
X Female gender X Team size 3	0.053	-0.005	-0.054	0.006	-0.014	-0.031	0.065	0.064	0.132**	0.026
	(0.049)	(0.085)	(0.073)	(0.082)	(0.081)	(0.038)	(0.046)	(0.071)	(0.061)	(0.064)

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Dependent variable:	Ashamed	Nervous	Determined	Attentive	Jittery	Afraid	Esteemed	Sad	Uneasy	Unsettled
Team size 3	0.031	0.096	0.004	-0.060*	-0.045	0.003	0.053	0.067	-0.043	-0.028
	(0.078)	(0.094)	(0.055)	(0.031)	(0.093)	(0.080)	(0.076)	(0.067)	(0.088)	(0.090)
Female gender	-0.047	0.050	-0.081	-0.061	0.029	0.002	-0.004	-0.021	-0.013	-0.003
	(0.078)	(0.089)	(0.061)	(0.041)	(0.088)	(0.077)	(0.079)	(0.080)	(0.094)	(0.092)
Interactions: Co-worker attractiveness...										
X Male gender X Team size 2	0.159***	0.087	0.054	-0.000	0.155***	0.116**	0.067	0.152**	0.108	0.140
	(0.055)	(0.091)	(0.076)	(0.002)	(0.056)	(0.049)	(0.070)	(0.060)	(0.075)	(0.094)
X Female gender X Team size 2	0.086	0.089	0.035	-0.020	0.003	0.070	0.059	0.032	0.026	0.028
	(0.063)	(0.068)	(0.044)	(0.027)	(0.075)	(0.064)	(0.073)	(0.066)	(0.079)	(0.073)
X Male gender X Team size 3	-0.019	0.071	0.122**	0.050	-0.008	0.039	0.169***	-0.023	0.092	0.121*
	(0.060)	(0.067)	(0.052)	(0.036)	(0.069)	(0.046)	(0.049)	(0.062)	(0.060)	(0.061)
X Female gender X Team size 3	0.059	0.107	-0.019	0.010	0.076	0.102*	-0.001	0.003	0.040	-0.001
	(0.075)	(0.072)	(0.059)	(0.023)	(0.072)	(0.059)	(0.056)	(0.086)	(0.079)	(0.081)

Notes: The sample includes 140 individual observations. The dependent variable is a dummy for each emotion item that is 1 for scores of at least 2 or 0 for score 1 on the 5-point scale. Each emotion item is interacted with team size and gender constellations. Linear regressions are used. For more information on the variables used, see Appendix Table C.1.1. Clustered standard errors (at the session level) are in parentheses. Levels of significance are: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.1.6 Workplace perceptions

	(1)	(2)	(3)	(4)
Dependent variable:	Performance	Monitored	Observing	Distracted
Interactions: Co-worker attractiveness…				
X Male gender X Team size 2	-1.907***	-0.046	0.207	-0.022
	(0.530)	(0.242)	(0.312)	(0.240)
X Female gender X Team size 2	-1.639***	0.186	0.306	0.201
	(0.362)	(0.259)	(0.239)	(0.141)
X Male gender X Team size 3	-1.393	-0.152	0.207	0.068
	(1.146)	(0.228)	(0.267)	(0.184)
X Female gender X Team size 3	1.108	0.579***	0.028	0.563***
	(0.783)	(0.189)	(0.294)	(0.190)
R^2	0.074	0.055	0.019	0.073

	(5)	(6)	(7)	(8)
Dependent variable:	Competition	Pressure	Strangers	Shirking
Interactions: Co-worker attractiveness…				
X Male gender X Team size 2	-0.011	0.240	-0.202	0.010
	(0.205)	(0.275)	(0.170)	(0.324)
X Female gender X Team size 2	0.533**	0.067	0.158	0.551**
	(0.220)	(0.276)	(0.305)	(0.231)
X Male gender X Team size 3	-0.260	0.044	0.181	-0.104
	(0.247)	(0.245)	(0.229)	(0.216)
X Female gender X Team size 3	0.467	0.376	0.382	-0.118
	(0.300)	(0.379)	(0.273)	(0.273)
R^2	0.054	0.020	0.034	0.052

Notes: The sample includes 140 individual observations. Linear regressions are used. For more information on the variables used, see Appendix Table C.1.1. Clustered standard errors (at the session level) are in parentheses. Levels of significance are: * p < 0.1, ** p < 0.05, *** p < 0.01. The dependent variables in specifications (2) to (8) are based on responses to survey questions regarding workplace perceptions. The question battery starts with: "What is your view regarding the following statements about the work atmosphere?" 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 "Does not apply at all" (score: 1) to "Does fully apply" (score: 7).

Table C.1.7 Future work motivation

Dependent variable:	(1)	(2)	(3)	(4)
	Future work motivation:			
	This task	This team	Larger team	Teamwork
Interactions: Co-worker attractiveness…				
X Male gender X Team size 2	0.096	0.285***	0.013	0.271*
	(0.144)	(0.100)	(0.161)	(0.162)
X Female gender X Team size 2	-0.156	0.151	-0.205	-0.360*
	(0.167)	(0.152)	(0.217)	(0.189)
X Male gender X Team size 3	0.384**	0.490***	0.130	0.001
	(0.159)	(0.123)	(0.167)	(0.125)
X Female gender X Team size 3	-0.237	0.005	-0.286	-0.116
	(0.174)	(0.145)	(0.210)	(0.133)
R^2	0.068	0.142	0.048	0.065

Notes: The sample includes 140 individual observations. Linear regressions are used. For more information on the variables used, see Appendix Table C.1.1. Clustered standard errors (at the session level) are in parentheses. Levels of significance are: * p < 0.1, ** p < 0.05, *** p < 0.01. The dependent variables in specifications (1) to (4) are based on responses to the following survey questions: "Would you want to…"

- "…do this task again?" (1)
- "…work with this team constellation again?" (2)
- "…complete this task again with a larger team?" (3)
- "…do your work in a team or alone in your later professional life?" (4)

The response scale ranges from "No, under no circumstances" (score: 1) to "Yes, definitely" (score: 5) for items (1) to (3) and from "Definitely alone" (score: 1) to "Definitely in a team" (score: 5) for item (4).

5 The impact of cashless payment options on face-to-face fundraising – Evidence from a natural field experiment*

Offering a cashless payment option (CPO) in a charity context may deter individuals from donating. We conduct a field experiment to investigate this phenomenon. After taking a guided tour of an art exhibition, visitors are asked to support the exhibition financially. As a first treatment, we randomly manipulate the opportunity to donate, either offering a CPO or not. To check whether the findings are special to donations, we additionally implement two pay-what-you-want (PWYW) treatments (with and without CPO) by asking to compensate for the service of the guided tour. Our results show that donors shy away from using the CPO in all treatment conditions. Despite that, there is no negative effect of the CPO, neither in the PWYW nor in the donation condition, on the frequency of financial support and its magnitude.

*This chapter is joint work with Adrian Chadi, Sabrina Jeworrek, and Laszlo Goerke. We are grateful for the support by Trier University and the chair of modern art history, Stephan Brakensiek, Andrea Diederichs, Ulrike Gehring and Andrea Günther as well as our research assistants Miguel da Fonseca, Benjamin Hattemer, Justina Heinz, Tom Hitzler, Lisa Nagel, Joana Pestana, Ruth Regnauer, Kristina Reznik, Manuel Riede and Lorenz Schlotter. We are further grateful for comments by participants of the SABE/IAREP conference 2019 in Dublin, the European ESA meeting 2019 in Dijon, the Vechta workshop on social behavior, the workshop on labour economics in Trier as well as seminar participants in Konstanz and Trier.

5.1 Introduction

Cashless payment instruments are on the rise worldwide and card payments are already ubiquitous in some regions to the extent that face-to-face transactions are almost cash free (Sveriges Riksbank 2018). Moreover, many charitable organizations and other institutions such as museums that rely on donations observe a decline in cash donations and therefore look out for a future fundraising method (Institute of Fundraising 2018, Straughan 2019). Intuitively, cash-affine persons should not be deterred from providing financial support if a cashless payment option (CPO) is available additionally. Besides, card-affine persons may be inclined to use the CPO. Hence, the share of donors should increase due to a CPO. Furthermore, contributions are no longer limited to the amount of cash that donors are carrying at the time of the request.

While cash as the norm in face-to-face fundraising seems paradox at first, this can be reconciled with empirical evidence from a pioneering study on cashless payments in door-to-door fundraising. Soetevent (2011) finds that donors prefer cash to card payments. When restricting donations to card payments only, participation rates drop dramatically. This finding reveals that, even though cashless payments are frequently used in point of sales transactions, people abstain from them in charitable giving. Further, when comparing a cash-only and a card-or-cash option condition, the share of donations goes down as a result of giving donors the opportunity to pay by card. This finding points to a strong role of psychological mechanisms, which may undermine the theoretically positive impact of implementing a CPO in face-to-face fundraising. Given the lack of additional research, one could also argue that it is not conclusively clarified whether a CPO in charitable giving is per se a bad idea. Assuming that the norm in face-to-face fundraisings is to donate via cash – and deviations from the norm are punished with refusing to contribute –, it seems worthwhile to change the framing of giving from donating to paying for a received service. Building upon the literature on charitable giving and pay-what-you-want (PWYW) (Gneezy et al. 2010, Gneezy et al. 2012, Park et al. 2017), a CPO might be perceived differently and, hence, could be more or less successful in fundraisers.

In this paper, we exploit the background of an art exhibition at a German university to conduct a natural field experiment on the effects of a CPO on the frequency of financial support and its magnitude. We cooperated with the organizers, who offered tours to visitors of the exhibition on pre-arranged dates. While this service was provided for free, each tour ended with a request to financially support the exhibition's continuation and complete a questionnaire. At this point, we intervened with our modifications of the fundraising. We implemented a CPO as one experimental treatment by randomly putting a card terminal on the desk (*CPO* condition)

or not (*Cash* condition). As part of a 2x2 design, we randomized the framing of the request for financial support by distinguishing between a *donation* condition, in which visitors were asked to donate for the exhibition, and a pay-what-you-want (*PWYW*) condition, in which visitors were invited to pay for the provision of the tour. The condition helps to bridge the gap between donations for charity or arts, which are usually made in cash in face-to-face encounters, and purchases of private goods often paid for by card. Our second manipulation, hence, gives us the opportunity to compare possible implications for the adoption of modern payment technologies based on the idea that cashless payments may be more common when paying for clearly defined services. In order to get some further insights, we asked visitors to answer a feedback questionnaire that also included some queries on socio-demographics. To merge individual payment data with data from the survey, all cash donations were made in marked envelopes.

We find strong evidence for psychological barriers to technology adoption, as our results confirm the notion of cash as the dominant payment choice (almost 97%) in face-to-face fundraising. This holds for both the *donation* as well as the *PWYW* condition. For both the share and the amount of contributions, however, our evidence contrasts with previous findings portraying CPO as harmful for people's willingness to contribute. When we focus on the core group of exhibition visitors by excluding students from the sample, we observe that the willingness to donate increases when offered the option to pay cashless. Given that almost all donations are still made in cash, the positive effect of the CPO is likely to be a psychological one.

So far, the literature is rather silent regarding the mechanisms that drive donors' willingness to give (or not) when confronted with a CPO in face-to-face fundraising. Drawing on additional collected survey data, we observe that individuals who report using card payments regularly are driving the increase in donations in the *CPO* condition, although they do not use the CPO. As long as visitors are asked to donate in cash only, those with strong preferences for card payments have the opportunity to regulate their emotions and possible feelings of guilt by relying on the fact that they do not use cash in other contexts. By putting a card terminal right in front of them, we take away their psychological excuse not to give and eliminate any moral wiggle room for card-affine individuals. Hence, our data suggest that a CPO could pressure some people to give who otherwise would have abstained from giving. Furthermore, we find a negative effect of the CPO on visitor satisfaction, which again is driven by card-affine individuals. This suggests discomfort among frequent card users in the presence of a card terminal and a resulting increase in the pressure to donate. Student visitors, however, behave differently not only in that they give significantly less money, the CPO also seems to have a

rather negative impact (even though statistically insignificant) which might be related to a decline in the perceived awareness of need caused by the CPO (Bekkers and Wiepking 2011).

Our conclusion of increased pressure as the main channel to increase payments fits well with the observation that individuals tend to make more contributions under the *PWYW* condition than in the *donation* condition. Arguably, PWYW generally makes solicitation efforts of fundraisers more salient and thus pressures more people to give. In line with that suggestion, social pressure is already high even without the card terminal's existence so that the CPO does not increase contributions any further in the PWYW context.

With our paper, we contribute to the field of charitable giving as we show how technological devices modify economic decision-making. We thereby expand the research on the role of cashless payments in face-to-face fundraising (Soetevent 2011) and provide evidence on underlying mechanisms of individual choices to give or not to give. Our conclusion that social pressure and norms play a strong role in people's motivation to give adds to related discussions on voluntary contributions to charities or in PWYW settings (Landry et al. 2010, DellaVigna et al. 2012, Andreoni et al. 2017, Feldhaus et al. 2019). Finally, our results inform about the role of PWYW in charitable giving (Gneezy et al. 2010, Gneezy et al. 2012, Park et al. 2017), as we discuss how framing the act of giving, namely as paying for a specific service rather than donating, can modify individual choices to participate in a face-to-face fundraiser.

The further paper is structured as follows. In Section 5.2, we outline the setting and the experimental design. Section 5.3 presents results and offers possible explanations for our findings. In Section 5.4, we offer some concluding remarks. The Appendix D contains supplementary information.

5.2 Experiment

5.2.1 Background

We set up a natural field experiment in cooperation with the organizers of an art exhibition named *"generator"* at Trier University in Germany.[58] The exhibition is located in the university's cellar rooms and its focus is on light art as well as art from recent modern artists interconnected through one main motive (for pictures, see Appendix Figure D.1.1). The generator was initiated in 2015 by the department of the History of Arts at Trier University. The interventions presented in our paper occurred in the third edition of the exhibition, which took place in winter 2016/2017.

[58] The permanent website of the art exhibition can be found here: http://generator.uni-trier.de/ (in German).

For the organizers, an important issue is how to collect more money to ensure the continuation of the exhibition. Fundraising in previous editions was not very successful. Hence, there was a desire for a cooperation with economists, which allowed us to conduct research on economic choices in an authentic context.

Whereas in earlier editions of the exhibition visitors could walk around freely, in the time span of the experiment regular tours were offered in order to attract more visitors. Thereby, standardized one-hour tours were implemented, including the allocated experimental treatment at the end of each tour. For the purpose of identifying the effects of our interventions and possible mechanisms, we also collected data about individual payments as well as self-reported information from a visitor survey. Besides asking for some socio-demographics, we solicited feedback from visitors. This did not only serve the purpose of improving future editions of the exhibition but also distracting respondents from our main intentions so that we were able to conduct a natural field experiment closely following the definition proposed by Harrison and List (2004). All interventions were agreed upon with the artists, but our assistants carried out the implementation. To ensure the ethical appropriateness of our intervention, we asked for and received approval from the ethics committee of Trier University.

On a regular basis, tours of the exhibition took place on Thursdays and Saturdays, as soon as at least five individuals had registered via the generator's own webpage. In an effort to further promote the exhibition and ensure sufficiently large group sizes we distributed flyers on campus. In consequence, a large number of students were attracted to the exhibition. In our analysis, students have to be considered a special group, given that they most likely differ from the core visitors regarding opportunity costs (to get to the exhibition), financial capacities (to give to charity) as well as actual interest in arts.

5.2.2 Experimental design

Immediately after taking the tour, visitors were invited by the tour guide to answer the feedback survey and contribute financially for the purpose of continuing the exhibition. Subsequent to this invitation, the guide referred to an assistant who provided more information about the survey and how to support the exhibition. The announcement of the assistant took the experimental condition into account, which was also induced in clear and visible fashion using posters (see Appendix Figure D.1.2). Our assistants handed out a clipboard to all visitors, equipped with the questionnaire, a pencil, and an envelope for cash payments (see Appendix Figure D.1.3 for the arrangement and Figure D.1.4 for the content of the questionnaire). Visitors

could fold the questionnaire and throw it in one box, while the envelope, including cash, could be placed in another one.

As a first treatment, we randomly manipulated the opportunity to make a cashless donation. For that purpose, we either placed a terminal on the desk, or we refrained from doing. In the latter case, the *cash* condition, the terminal was hidden behind the desk and was neither visible nor mentioned. However, visitors could use the card terminal if they asked for it. In the CPO condition at the end of a tour, the university-owned card terminal was placed visibly on a desk, which stood next to the spot where the tour ended and the fundraiser started (see Appendix Figure D.1.5). The possibility to use a debit or credit card was further mentioned during the announcements. A CPO transaction was proceeded as in German stores at the time of our experimental conduct. Participants had to inform one research assistant about the amount they wanted to give and afterward type in their PIN in private into the device. Cash payments were always collected in anonymized envelopes. To match individual payments afterward with the survey responses, we marked those envelopes in the following way: The logo of the exhibition was printed on each questionnaire, but in a certain coloring, while the coloring could be found in reversed order on a corresponding envelope in which cash could be placed. In case of a card payment, the card transaction's receipt was placed in the envelope and matched in the same manner with the questionnaire later on.

As part of our 2x2 design, which is illustrated in Table 5.1, we also implemented the concept of *PWYW* (with and without the cashless option) by asking visitors to compensate the service of the tour. They were told (identical to the *donation* condition) that the purpose was to collect money to ensure the exhibition's continuation. Here, we benefitted from the fact that the tour guides offered their service for free, allowing us to analyze whether the preference for or against cash may depend on what people believe is the most common choice.

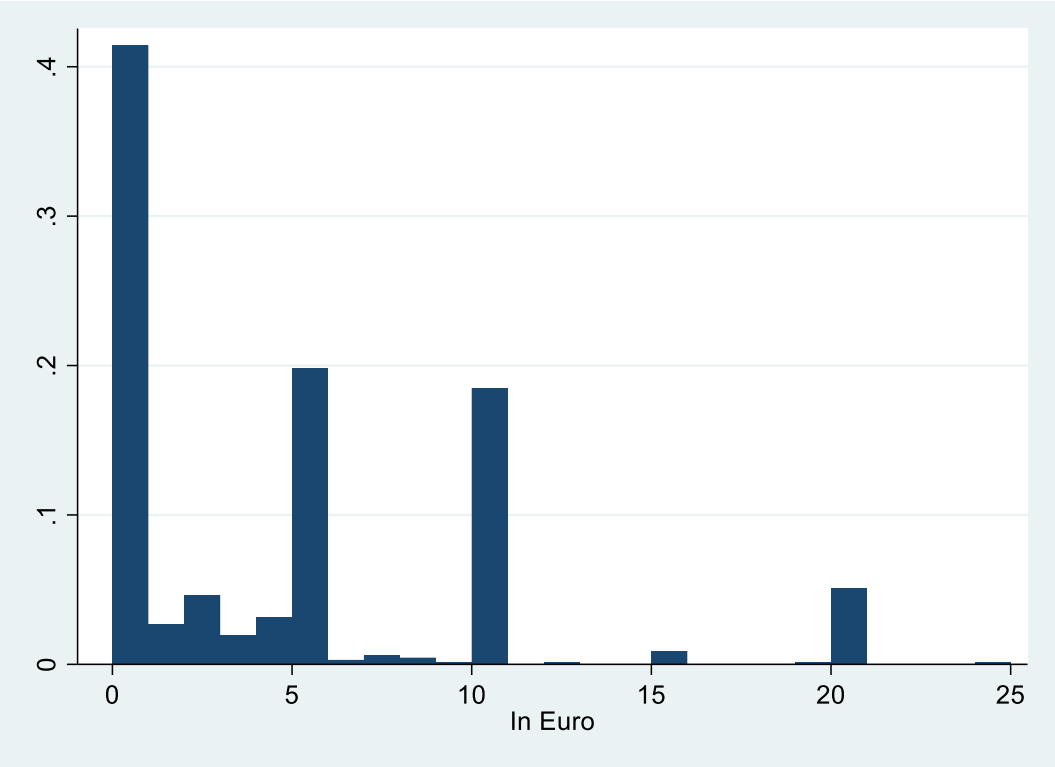
Table 5.1 Experimental design and observations numbers

	Treatment conditions	
	Cash	**Card**
Donation	13 tours (N=156)	14 tours (N=186)
PWYW	15 tours (N=167)	14 tours (N=165)

5.2.3 Sample description

In total, our dataset covers 674 exhibition visits distributed over 56 tours. 409 visitors gave money. In the vast majority, the amount was less than 25 Euro, as shown in Figure 5.1, which depicts the frequency of payments across all treatments. On top of that, there was one payment of 40 Euro, one of 50 Euro, and another one of 100 Euro.

Figure 5.1 Frequency of payments across all treatments



Notes: Distribution of amounts given are shown regardless of the treatment. Additionally, there was one payment of 40 Euro, one of 50 Euro, and another one of 100 Euro.

523 feedback survey questionnaires were submitted, which is a response rate of 77.6%. Looking at both tour-level and individual-level information, we find that the randomization of tours into the different treatment interventions worked very well with all *p*-values[59] above 0.1 (for details, see Appendix Table D.2.1). Notably, the survey participation was nearly identical across all four groups. Based on this survey, roughly 60% of visitors were female (which matches with the observational data for the whole group of visitors obtained by our assistants), 31% were students, slightly less than 10% were visiting the generator individually, and the majority of visitors (i.e. almost 70%) were frequent users of credit and debit cards. Given the

[59] If not reported otherwise, *p*-values are obtained from Wilcoxon rank sum or chi² tests, respectively.

sizeable share of student visitors, we conduct a sample split to compare the core group of visitors with the student visitors. Besides the rather obvious fact of being much younger, the data also indicates that individuals in the student sample are generally less interested in arts and less often financially support arts and culture (see Appendix Table D.2.2). Adding to the fact that our advertising efforts encouraged university students' attendance, as discussed above, these differences provide empirical evidence in support of considering the groups of students and core visitors separately.

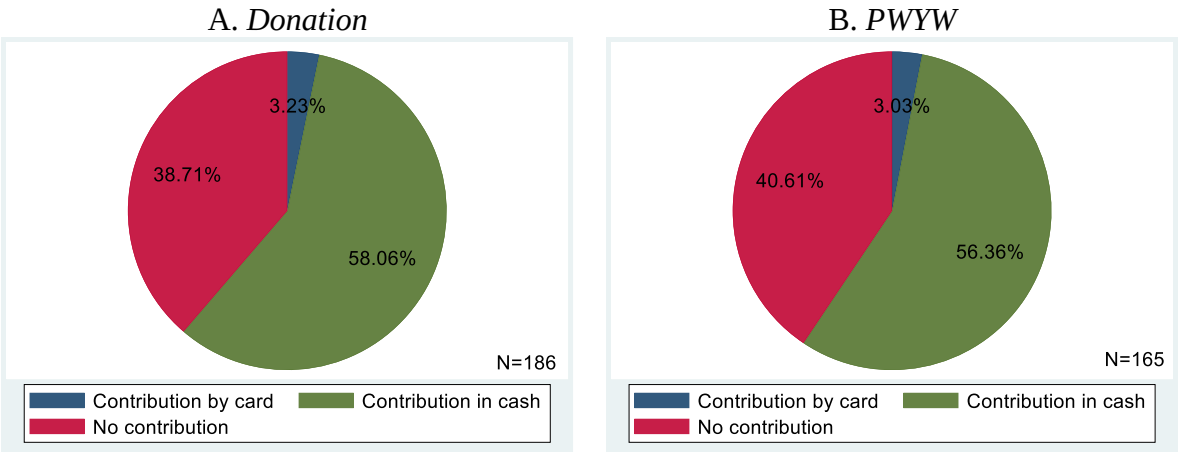
5.3 Results

In this section, we, first, focus on the use of debit or credit cards in the CPO treatment. Subsequently, we analyze how the CPO affects the frequency and level of financial support, both for the donation and the PWYW setting. We support the according descriptive findings by regression evidence. Finally, we provide possible explanations for our results using additional survey data.

5.3.1 Payment choice: Cash or card

As shown in Figure 5.2, visitors shy away from using a debit or credit card in the CPO setting, irrespective of whether a donation was asked for or visitors were invited to pay for the tour (PWYW). 3.23% of all visitors in the *Donation* treatment paid by card, that is, about 5.3% of those who provided financial support (Panel A) and 3.03% (5.1%) of all (paying) visitors in the *PWYW* treatment (Panel B). In absolute terms, only eleven persons contributed by using debit or credit cards. In the *Cash* treatment, no visitor asked whether it would be possible to give by card. Hence, we consider these numbers as clear evidence that individuals prefer cash in face-to-face fundraising.

Figure 5.2 Contributions and payment choices in the CPO conditions



Notes: In Panel A percentage of payment choices are shown for the CPO donation condition. In absolute numbers, 6 participants contributed by card, 108 contributed in cash and 72 did not contribute anything. Supplementary, in the cash donation condition (not shown here) 88 contributed in cash and 68 nothing at all. In Panel B percentage of payment choices are shown for the CPO PWYW condition. Here, 5 participants contributed by card, 93 contributed in cash and 67 did not contribute anything. Supplementary, in the cash PWYW condition (not shown here) 109 contributed in cash and 58 nothing at all.

Our findings on people's payment choices reject two ideas in the context of cashless payments and voluntary giving. First, as suggested by Soetevent (2011), social image concerns are unlikely to explain people's payment choices. Given the anonymity of cash payments due to putting the money in an envelope which was then thrown in a box, we would expect especially the higher amounts of money to be paid cashless in order to benefit from a positive social image. All payments that were equal to or bigger than 20 Euro were given in cash, whereas the highest amount given cashless was 15 Euro (only once). Second, PWYW which is related to the idea that in combination with a CPO it may foster payments (Natter and Kaufmann 2015) also did not modify payment choices. It seems that the norm to pay by cash was too strong to be modified with this manipulation of the act of giving. Our findings are, however, in line with evidence provided by Lynn et al. (2013) in that individuals' decisions to give are most likely influenced by their preference for round prices instead of other potential explanatory processes, 71.15% of giving tour visitors decided to give either five, ten or twenty Euro.

Besides a strong social norm, other explanations for people's reluctance to use their cards are less likely to play a role in our setting. First, people might be concerned about data protection. However, as the majority of people use cards frequently in other contexts like profit-orientated stores, it is unlikely to expect that the same individuals become concerned about their data when being asked to give by a non-profit organization of artists at a public German

university. Second, people might shy away from using the card terminal because it takes more time than putting money into an envelope. While it is true that signature- and PIN-required digital payments are less quick than cash payments at the checkout of the grocery store (Klee 2008), this fact does not prevent individuals from using cards in exactly those other contexts (and 70% of our survey respondents claimed to be frequent card users). Furthermore, there was no particular benefit from being quick in the art exhibition, given that visitors typically stayed after the fundraiser and kept talking to others or went back to the exhibits for a second look. One explanation for the norm to pay by cash might be found in the costs that are related to cashless payments. Gneezy et al. (2014) have shown that the willingness to donate increases sizably if the organization's overhead costs are already covered. This finding suggests that donors desire their donation to support the social cause in full. Given that most people should be aware of the fact that a fee is associated with the CPO, they might shy away from using this option.

In summary, the idea of introducing a CPO in the fundraiser could be seen as an economic failure, given that many visitors did not use it. However, there could also be psychological effects due to the CPO, which influence the frequency and magnitude of contributions and thereby affect the fundraiser's success. Both the impact of the CPO on visitors' contributions as well as potential mechanisms will be investigated in the next two chapters.

5.3.2 CPO and giving

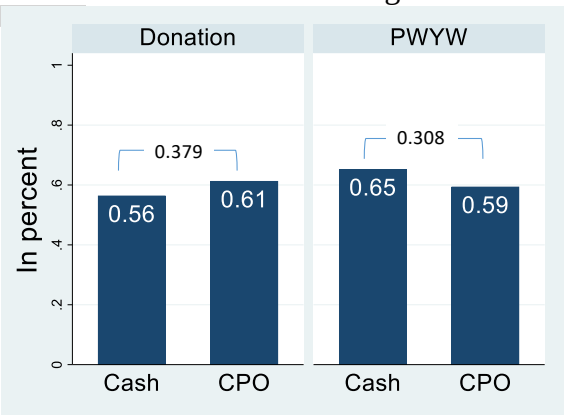
To analyze the impact of a CPO, we look at the share of contributing visitors and visitors' average contributions. Figure 5.3 summarizes our results for the full sample, the core group of visitors excluding students, and the student sample. Looking at the bar charts A1. and A2., which show the full sample, we do not find any effect of the CPO on the share of visitors who provided financial support as well as on the average amount contributed, neither in the *Donation* nor the *PWYW* framework.

When splitting the full sample, differences become discernible.[60] The bar charts B1. and B2. show that the core group of visitors is significantly (p=0.014) more likely to give when confronted with the option to donate by card. The higher share of donors then also translates into an increase of about 23% in the average amount contributed (with p=0.011). This result remains unchanged if we exclude cases of card payments, which suggests that the CPO effect is a purely psychological one.

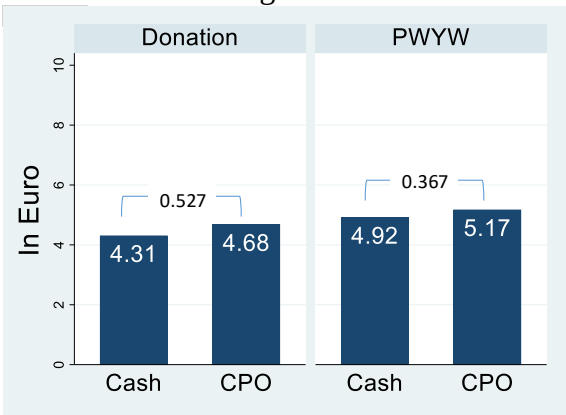
[60] These results are, however, not driven by the reduced sample of visitors who responded to the questionnaire. Again, we do not find any significant difference between the *Cash* and the *CPO* treatments.

Figure 5.3 Contributions by treatment

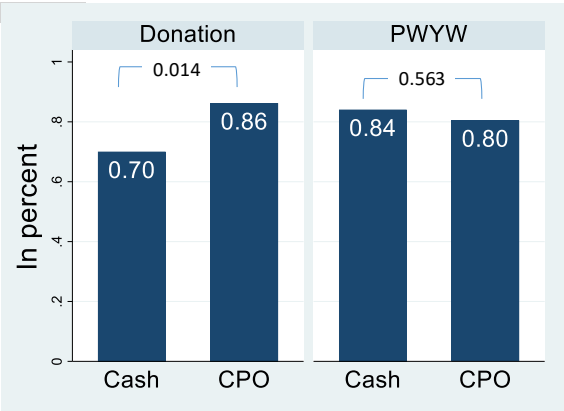
A1. Share of contributing visitors



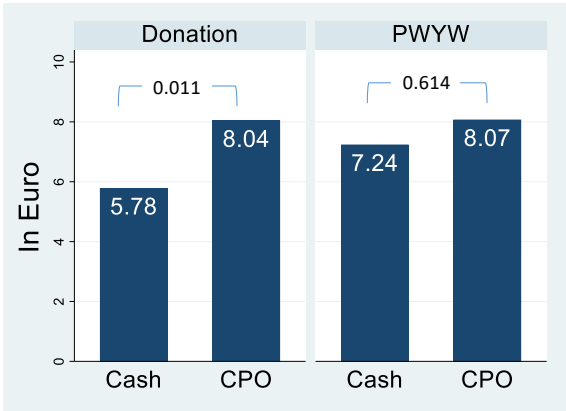
A2. Average contribution



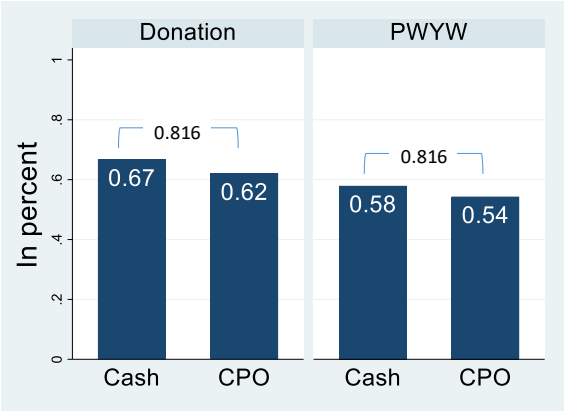
B1. Share of contributing core group visitors



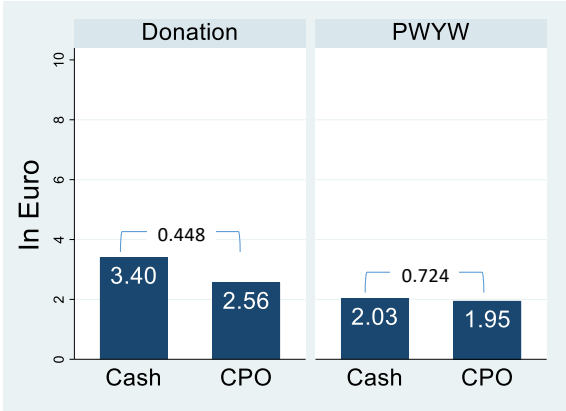
B2. Average contribution by core group visitors



C1. Share of contributing student visitors



C2. Average contribution by student visitors



Note: Panel A refers to the full sample, whereas Panel B and C show the subgroup results for the core group of visitors and student visitors.

In the *PWYW* treatment, the CPO neither affects the share of contributing visitors nor the average contribution. In the *cash* treatments, however, the share of contributing visitors is already quite high in the *PWYW* compared to the *donation* setting (84% vs. 70%, *p*=0.027) so that the *card* and *PWYW* treatments combined do not further increase people's willingness to give.

Bar charts C1. and C2. present the findings for students. They differ substantially in their payment behavior from members of the core group, as students contribute significantly smaller amounts (2.57 Euro vs. 8.43 Euro, *p*=0.000) and also give less frequently (60.3% vs. 80.5%, *p*=0.000). Moreover, the CPO lowers students' contribution, however, statistically insignificant.

In order to check the robustness of our results with regard to the characteristics of the tour (i.e., the tour guide, group size, and the weekday of the tour), we conducted an additional regression analysis depicted in Table 5.2. Specifications (1) and (2) show the results for the full sample. In line with our non-parametric test results, we do not find any treatment effect; the inclusion of tour controls changes the point estimates only slightly.

Table 5.2 OLS results

	(1) *Full sample*	(2)	(3)	(4) *Student sample*	(5) *Core group*	(6) *Donation only*
CPO	0.369	0.580	1.836	-1.248	3.067**	0.015
	(0.855)	(0.759)	(1.112)	(0.829)	(1.342)	(1.430)
PWYW	0.608	1.133	1.239	-1.459*	2.274**	
	(0.836)	(0.742)	(0.857)	(0.741)	(0.972)	
CPO X PWYW	-0.110	-0.570	-1.792	0.711	-2.569	
	(1.287)	(1.285)	(1.664)	(0.986)	(2.284)	
Female			-0.969	-0.249	-1.464	-1.964
			(0.684)	(0.566)	(0.997)	(1.309)
Student			-4.828***			
			(0.628)			
Frequent card user						-1.866
						(1.662)
CPO X freq. card user						4.870**
						(2.355)
Tour controls	No	Yes	Yes	Yes	Yes	Yes
R^2	0.002	0.042	0.158	0.069	0.110	0.243
N	674	674	498	154	344	161

Note: The dependent variable is visitors' amount given (in Euro). Robust standard errors, clustered on the tour-level, in parentheses. The tour controls include the tour guide, the weekday of the tour and a dummy for large groups (according to a median split). Levels of significance are: * p < 0.1, ** p < 0.05, *** p < 0.01.

Focusing on those visitors for whom we have the additional information from the feedback questionnaire in the specification (3), we still do not find a treatment effect, but we observe a large negative and highly significant point estimate for students, showing that they give on average almost 5 Euro less compared to the core group of visitors. By splitting the sample into the two subsamples of students and core group (see specifications (4) and (5)), the observation that students behave differently from other visitors can be confirmed. Even though not or only marginally significant, the point estimates for both the *CPO* and the *PWYW* setting are negative for student visitors but positive and significant at the 5% level for our core group.

5.3.3 Channels of giving

Given that we find no overall effect of the CPO on the frequency of payments and its magnitude but strong heterogeneities among the two visitor subgroups, we next aim to shed light on the possible underlying channels.

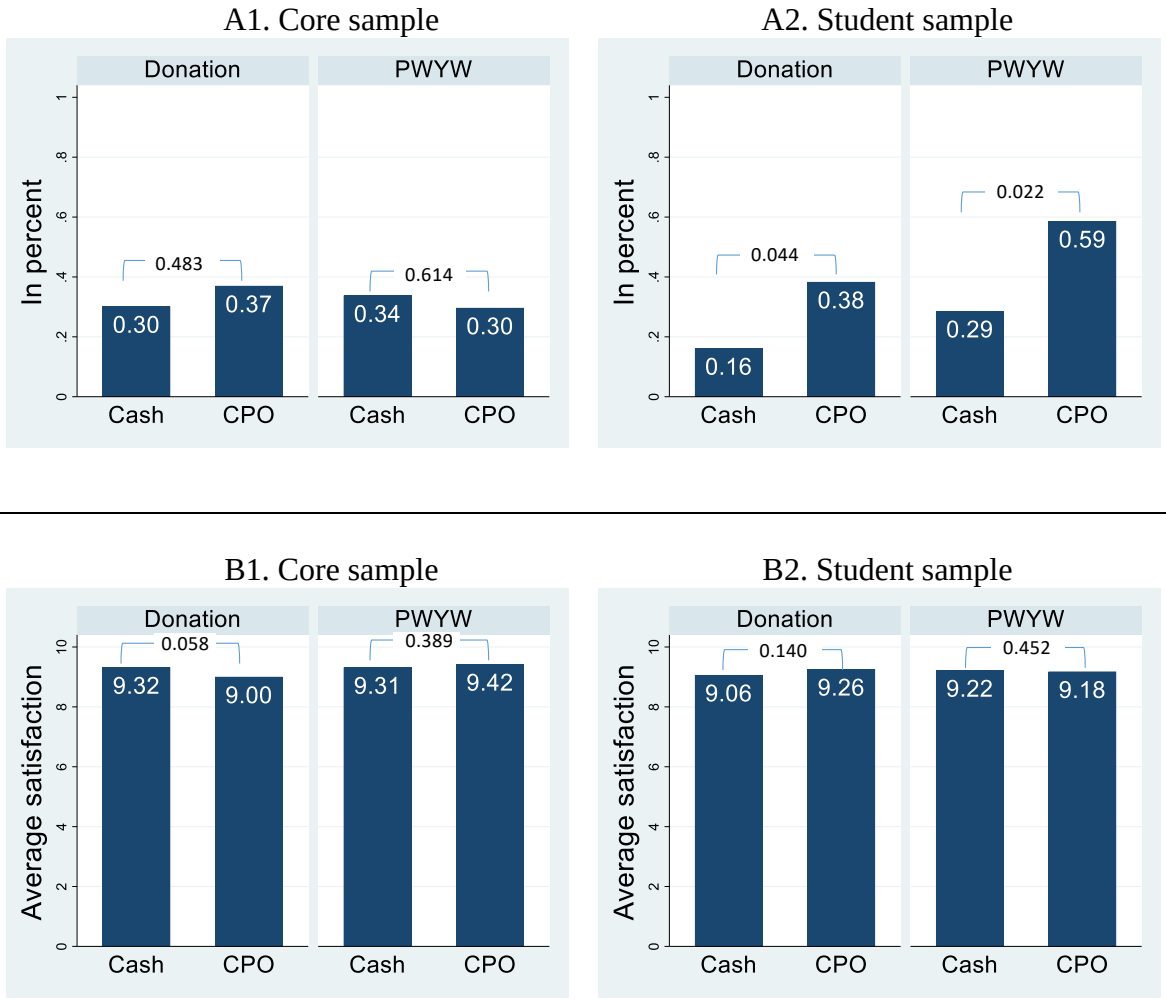
One psychological mechanism to affect charitable giving might be the perceived awareness of need (Bekkers and Wiepking 2011). The idea is that the perception of how much the solicitor needs money can be changed by the CPO. In Figure 5.4, bar charts A1. and A2. show the responses to the yes-or-no survey question *"Do you think the generator can be financed through voluntary payments?"* for the same subsamples as before.

Whereas there is no effect of the CPO on the perceived awareness of need for the core group, the figure reveals that students exposed to a CPO are roughly twice as likely to report that the art exhibition can be financed through voluntary payments. Combined with the observation that average contributions tend to decrease in the student sample if a CPO exists, the reduced awareness of need is likely to explain why people refrain from giving (more).

Since previous studies suggest that happier people are also more generous (see, e.g., Boenigk and Mayr 2016), we asked visitors about their satisfaction with the exhibition, the tour, and with the team. Whereas the CPO is unlikely to have a direct effect on visitor satisfaction, it would be conceivable that visitors perceive the exhibition as more professional and, hence, more satisfactory. All three satisfaction categories are on a high level (averages of more than 8 on a 10-point Likert scale), but the bar charts B1. and B2. in Figure 5.4 show that there is a slightly negative CPO effect under the *donation* condition for the satisfaction with the team (all other treatment comparisons are statistically insignificant). Intriguingly, this negative CPO effect is in contrast to the increase of both share and amount of contributions observed among the core group of visitors, they seem to feel uncomfortable in the presence of the card terminal and the option to pay cashless. This suggestion can be backed up by another piece of evidence obtained

from the survey data: Visitors were asked whether they would prefer to pay in cash or by card for future exhibitions and we find that visitors in the *card* treatments are significantly more likely to report *"in cash"* (83% vs. 76.3% with p=0.064 for the full sample, and 86.2% vs. 73.9% with p=0.005 for the core group of visitors).

Figure 5.4 Awareness of affordability (Panel A) and satisfaction with the team (Panel B)



Notes: Panel A refers to the survey question: *„Do you think the generator can be financed through voluntary payments?"* Panel B refers to the survey question: *"On a scale of 1 to 10, how satisfied are you with the Generator-Team?"*

This latter finding points us to discuss the role of social pressure regarding the core group of visitors. Arguably, a mechanism of increased pressure, as induced by a CPO, should be particularly effective in manipulating the decision-making of those visitors with strong preferences to pay cashless in other contexts. Hence, the increase of donations within the core group of visitors due to the CPO should be driven by those individuals who also stated to use

their credit/debit cards frequently. Indeed, the CPO has no effect for visitors who are no frequent card users (p=0.980). On the contrary, the average donation rises by about 55% in the subgroup of frequent card users (5.60 Euro vs. 8.71 Euro, p=0.020). We interpret this result in the way that the CPO impairs people's moral wiggle room to abstain from giving, as the option to pay cashless takes away frequent card users' psychological excuse to not give. As a consequence of this, card-affine individuals are pressured to take a look into their wallets. Again, the regression analysis confirms the finding that the increase in donations within the core group of visitors is completely driven by frequent card users. Focusing on the *donation* treatments only (see specification (6) in Table 5.2), the inclusion of an interaction term for frequent card users and the *CPO* condition results in a point estimate for the CPO that is close to zero.[61]

Our finding that the CPO has no effect in the *PWYW* framing further supports the idea of pressure playing a key role in face-to-face fundraising. In that setting, the pressure to give might already be a result of the request to pay for a service that was provided for free, which cannot be further increased by the introduction of a CPO. Moreover, compassion and moral feelings of guilt may be more relevant in the *donation* setting compared to a situation in which people decide on how to remunerate a clearly defined service. In the latter case, one could expect that people would give more if they liked the tour, while satisfaction with this service might play no or only a minor role in donation decisions. In line with this suggestion, we observe that visitors who reported to be fully satisfied with the tour give significantly more money (6.88 Euro vs. 4.24 Euro, p=0.002) in the *PWYW* setting, whereas this is not true in the *donation* setting (p=0.681).

In summary, our survey data suggest that neither the awareness of need for the art exhibition itself nor the professionalism of the tour are affected in a positive manner by the CPO. Intriguingly, our analyses indicate that the CPO increases psychological pressure on people to donate rather than affecting their willingness to pay for a service. This underlines the role of the different drivers of giving when comparing *PWYW* and *donation* scenarios.

5.4 Conclusion

We conducted a natural field experiment to investigate whether the implementation of a cashless payment option is worthwhile in a fundraising setting and whether a PWYW frame can ease the implementation of this technology. Our results show that donors shy away from the CPO across all treatment conditions. However, we do not find strong evidence in support

[61] When applying the same specification for the student sample, the point estimate for the interaction term is 0.228 with p=0.909.

of a negative psychological effect of the CPO. On the contrary, we find a positive effect of the CPO on people's willingness to donate for our core group of visitors. As a matter of fact, we find a decrease in reported satisfaction rates which is in line with the idea that some visitors felt an unpleasant pressure to give in the presence of the CPO. Subsample analyses show that frequent card users are driving the positive CPO effect in donations. We argue that this result can be seen as evidence that a CPO takes away the moral wiggle room of card users not to give and thereby increases the pressure on them, just as social pressure to give has been discussed as an essential channel for giving (Landry et al. 2010, DellaVigna et al. 2012, Andreoni et al. 2017).

Implications from our study are straightforward: The replacement of cash by card in face-to-face fundraising may require particular efforts, but as regards offering a cashless payment *option*, our evidence does not suggest any economic harm, contrary to previous research. Interestingly, the CPO has no effect at all in the PWYW fundraiser even though one might expect a cashless payment option to be more important for an actual payment situation. Nevertheless, this finding is very well in line with the suggestion that social pressure is driving the increase in donations since individuals might in general feel more pressure when being asked to pay for a service they just have received, so that an increase in pressure due to the CPO is not necessary.

What needs to be taken into account is that the provision of this payment option is not for free. As a result, there is a trade-off between the costs and benefits of using a card terminal to offer possible givers a CPO. In our case, the implementation of an additional payment choice was not worth the costs since the share of positively affected individuals was too small. Thus, one should not be surprised when in face-to-face fundraisers not only the donors but also the solicitors may continue to prefer cash. Moreover, our findings also express the need for future research. Given that the positive CPO effect is likely to be based on psychological pressure, it would be worth investigating whether affected individuals would be willing to visit another exhibition or, more generally, be willing to get in touch with the fundraising organization once again.

5.5 Appendix D.1: Additional background information

Figure D.1.1 Pictures of the exhibition

A. Entrance

B. Computerized light animation

Figure D.1.2 Posters

A. Setup

B. Donation

SPENDEN

(*DONATE*)

Bitte unterstützen Sie die Fortführung des Projekts, indem Sie eine Spende in beliebiger Höhe tätigen! Zahlen ist keine Pflicht, aber jeder Euro hilft!

(***Please support the continuation of this project by making a donation of any amount! Paying is not a duty, but every Euro helps!***)

C. Pay what you want

ZAHLEN SIE, WAS SIE WOLLEN

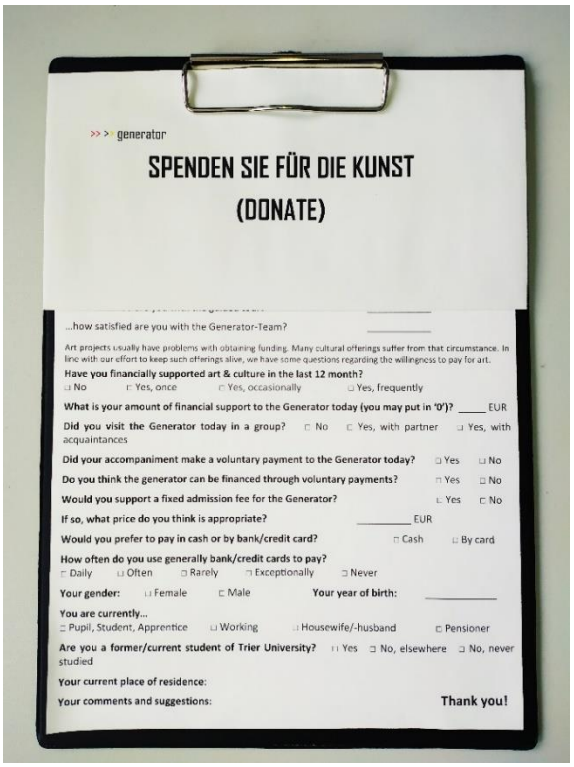
(*PAY WHAT YOU WANT*)

Bitte unterstützen Sie die Fortführung des Projekts, indem Sie die Führung durch einen Betrag in beliebiger Höhe entlohnen! Zahlen ist keine Pflicht, aber jeder Euro hilft!

(***Please support the continuation of this project by remunerating the guided tour with a payment of any amount! Paying is not a duty, but every Euro helps!***)

Figure D.1.3 Clipboard setup

A. Donate



>> > generator

SPENDEN SIE FÜR DIE KUNST

(DONATE)

...how satisfied are you with the Generator-Team?

Art projects usually have problems with obtaining funding. Many cultural offerings suffer from that circumstance. In line with our effort to keep such offerings alive, we have some questions regarding the willingness to pay for art.

Have you financially supported art & culture in the last 12 month?
☐ No ☐ Yes, once ☐ Yes, occasionally ☐ Yes, frequently

What is your amount of financial support to the Generator today (you may put in '0')? _____ EUR

Did you visit the Generator today in a group? ☐ No ☐ Yes, with partner ☐ Yes, with acquaintances

Did your accompaniment make a voluntary payment to the Generator today? ☐ Yes ☐ No

Do you think the generator can be financed through voluntary payments? ☐ Yes ☐ No

Would you support a fixed admission fee for the Generator? ☐ Yes ☐ No

If so, what price do you think is appropriate? _____ EUR

Would you prefer to pay in cash or by bank/credit card? ☐ Cash ☐ By card

How often do you use generally bank/credit cards to pay?
☐ Daily ☐ Often ☐ Rarely ☐ Exceptionally ☐ Never

Your gender: ☐ Female ☐ Male **Your year of birth:** _____

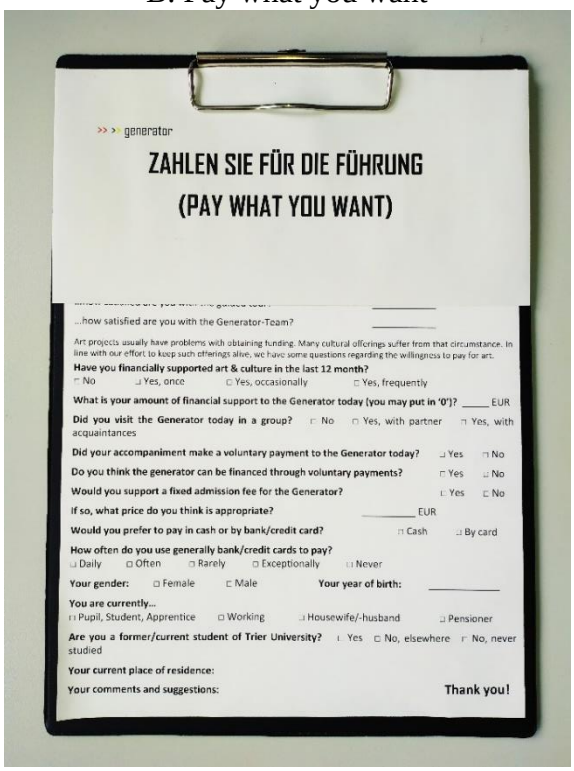
You are currently...
☐ Pupil, Student, Apprentice ☐ Working ☐ Housewife/-husband ☐ Pensioner

Are you a former/current student of Trier University? ☐ Yes ☐ No, elsewhere ☐ No, never studied

Your current place of residence:

Your comments and suggestions: **Thank you!**

B. Pay what you want



>> > generator

ZAHLEN SIE FÜR DIE FÜHRUNG

(PAY WHAT YOU WANT)

...how satisfied are you with the Generator-Team?

Art projects usually have problems with obtaining funding. Many cultural offerings suffer from that circumstance. In line with our effort to keep such offerings alive, we have some questions regarding the willingness to pay for art.

Have you financially supported art & culture in the last 12 month?
☐ No ☐ Yes, once ☐ Yes, occasionally ☐ Yes, frequently

What is your amount of financial support to the Generator today (you may put in '0')? _____ EUR

Did you visit the Generator today in a group? ☐ No ☐ Yes, with partner ☐ Yes, with acquaintances

Did your accompaniment make a voluntary payment to the Generator today? ☐ Yes ☐ No

Do you think the generator can be financed through voluntary payments? ☐ Yes ☐ No

Would you support a fixed admission fee for the Generator? ☐ Yes ☐ No

If so, what price do you think is appropriate? _____ EUR

Would you prefer to pay in cash or by bank/credit card? ☐ Cash ☐ By card

How often do you use generally bank/credit cards to pay?
☐ Daily ☐ Often ☐ Rarely ☐ Exceptionally ☐ Never

Your gender: ☐ Female ☐ Male **Your year of birth:** _____

You are currently...
☐ Pupil, Student, Apprentice ☐ Working ☐ Housewife/-husband ☐ Pensioner

Are you a former/current student of Trier University? ☐ Yes ☐ No, elsewhere ☐ No, never studied

Your current place of residence:

Your comments and suggestions: **Thank you!**

Figure D.1.4 Questionnaire

>> >> generator

The following survey is very important for us because we want to improve the generator's portfolio. Thank you very much for taking the time to answer this questionnaire. Your data will be captured and processed anonymously.

How have you heard about the Generator? □ Acquaintances □ Press □ Online Media □ Others

Would you recommend the Generator/this exhibition? □ Yes □ No

Is this the first time you have visited Campus II? □ Yes □ No

On a scale from 1 to 10,… *(1=barely, rarely and 10=very often, very much)*

…how often do you visit exhibitions? _____

…would you call yourself an art lover? _____

…how interested are you in other local cultural offerings? _____

On a scale from 1 to 10,… *(1= very unsatisfied, negative and 10= very satisfied, positive)*

…how did you like this exhibition? _____

…how satisfied are you with the guided tour? _____

…how satisfied are you with the Generator-Team? _____

Art projects usually have problems with obtaining funding. Many cultural offerings suffer from that circumstance. In line with our effort to keep such offerings alive, we have some questions regarding the willingness to pay for art.

Have you financially supported art & culture in the last 12 month?
□ No □ Yes, once □ Yes, occasionally □ Yes, frequently

What is your amount of financial support to the Generator today (you may put in '0')? _____ EUR

Did you visit the Generator today in a group? □ No □ Yes, with partner □ Yes, with acquaintances

Did your accompaniment make a voluntary payment to the Generator today? □ Yes □ No

Do you think the generator can be financed through voluntary payments? □ Yes □ No

Would you support a fixed admission fee for the Generator? □ Yes □ No

If so, what price do you think is appropriate? _____ EUR

Would you prefer to pay in cash or by bank/credit card? □ Cash □ By card

How often do you use generally bank/credit cards to pay?
□ Daily □ Often □ Rarely □ Exceptionally □ Never

Your gender: □ Female □ Male **Your year of birth:** _____

You are currently…
□ Pupil, Student, Apprentice □ Working □ Pensioner □ Housewife/-husband □ Other

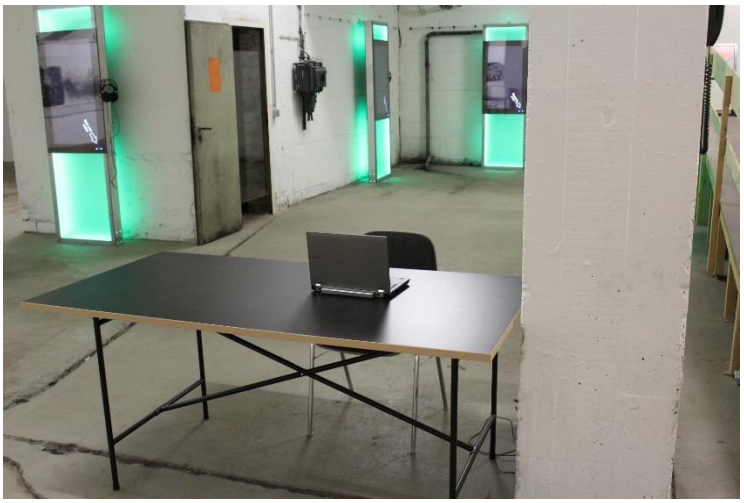
Are you a former/current student of Trier University? □ Yes □ No, elsewhere □ No, never studied

Your current place of residence:

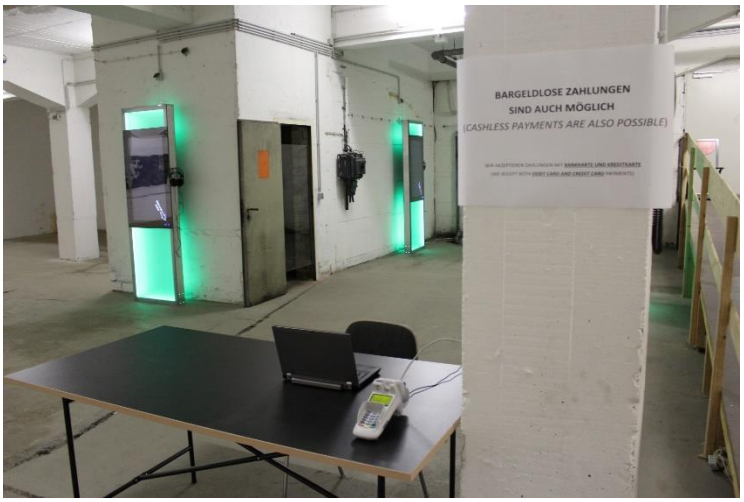
Your comments and suggestions: **Thank you!**

Figure D.1.5 Treatment manipulation

A.



B.



C.

BARGELDLOSE ZAHLUNGEN SIND AUCH MÖGLICH

(CASHLESS PAYMENTS ARE ALSO POSSIBLE)

WIR AKZEPTIEREN ZAHLUNGEN MIT **BANKKARTE UND KREDITKARTE**

(*WE ACCEPT BOTH **DEBIT CARD AND CREDIT CARD** PAYMENTS*)

5.6 Appendix D.2: Supplementary statistics and analyses

Table D.2.1 Balance check across treatment conditions

	Cash / Donation	*Card / Donation*	*Cash / PWYW*	*Card / PWYW*	*p*-value
Panel A) Tour-level information					
	N = 13	*N* = 14	*N* = 15	*N* = 14	
Tour size in persons (actual)	12.00	13.29	11.20	12.14	0.709
Tour size in persons (scheduled)	12.54	15.36	11.33	13.50	0.367
Tour on Thursday	0.54	0.57	0.40	0.50	0.810
Tour on Saturday	0.23	0.29	0.33	0.29	0.949
Tour on any other day	0.23	0.14	0.27	0.21	0.876
Panel B) Individual-level information					
	N = 156	*N* = 186	*N* = 167	*N* = 165	
Survey participation	0.76	0.76	0.81	0.77	0.582
Based on survey information:					
Female	0.64	0.61	0.59	0.54	0.488
Student	0.29	0.36	0.29	0.29	0.438
Unaccompanied	0.08	0.10	0.10	0.11	0.830
Card use often	0.75	0.73	0.70	0.62	0.171

Note: *p*-values are obtained from Kruskal-Wallis or chi^2 tests, respectively.

Table D.2.2 Sample comparison based on survey data

	Core visitors	Student visitors	*p*-value	*N*
How have you heard about the generator?				
Acquaintances	0.55	0.45	0.033	496
Press	0.21	0.04	0.000	496
Online Media	0.17	0.27	0.012	496
Others	0.21	0.32	0.008	496
Recommend the exhibition	1.00	0.98	0.058	497
First time visiting this campus	0.41	0.20	0.000	499
On a scale from 1 to 10…				
…how often visiting exhibitions	5.30	3.97	0.000	489
…calling yourself an art lover	6.34	5.26	0.000	492
…interested in cultural offerings	6.63	5.90	0.002	488
Financially support art & culture	0.64	0.49	0.003	491
Age	50.65	26.34	0.000	409

Note: *p*-values are obtained from Kruskal-Wallis or chi^2 tests, respectively. In total, 523 visitors responded to the questionnaire, but 19 individuals refrained from stating their current labor market status so that we end up with a maximum of 504 observations consisting of 156 students and 348 core visitors.

6 All right, then. Keep your incentives. How increased transparency of data processing changes the willingness to disclose personal data in a research project*

I conduct an experiment to test whether increased transparency of data processing affects data disclosure and whether the results change if it is indicated that the implementation of the GDPR happened involuntarily. In the online experiment, participants consent to one out of four data protection pages, fill out a questionnaire, and decide whether they want to participate in an incentivized lottery and therefore have to disclose personal data. I find that increased transparency raises the number of participants who do not disclose personal data by 21 percent. However, this is not the case in the involuntary-signal treatment, where the share of non-disclosures remains relatively high. Moreover, I show that, when transparency is low, participants base their disclosure decision on time invested in the survey.

*I am grateful for comments by Laszlo Goerke, Gabriel Schultze as well as participants of the workshop on digitalization in economics and management in Heilbronn and seminar participants in Trier. Jessica Halle, Johanna Kluge, Simon Kleinert, and Dominic Reichert provided excellent research assistance.

"Data protection is both a central issue for research ethics in Europe and a fundamental human right. It is intimately linked to autonomy and human dignity, and the principle that everyone should be valued and respected. For this principle to guide the development of today's information society, data protection must be rigorously applied by the research community."
(European Commission, Ethics and data protection (2018a))

6.1 Introduction

In May 2018, the European Commission implemented a set of supranational general data protection rules (GDPR) to strengthen people's control over their personal data and create a level playing field for businesses (European Commission 2018b). By addressing businesses operating within the EU as well as serving most likely as a blueprint for other democratic nations and states associated with the EU, research on the direct and indirect effects of this regulation is highly required. Currently, the focus lays in studying whether the GDPR has indeed led to reduced data collection activities (Momen et al. 2019, Goldberg et al. 2019) as well as declining third-party server appearances (Sørensen and Kosta 2019, Batikas et al. 2020). Further, also unintended changes due to the GDPR are analyzed, such as an increasing market concentration among web technology vendors and apps available on the Google Play Store (Johnson et al. 2020, Kesler et al. 2020), as well as a decrease in EU venture investment (Jia et al. 2020).

Moreover, research itself is also directly targeted by the GDPR (GDPR 2016, Article 33), which is of great importance, especially in a time of growing empirical online research fostered by the ongoing digitalization and the restriction on face-to-face data gathering due to the covid-19 outbreak. As key elements for personal data processing, the GDPR proposes that "*it should be transparent to natural persons that personal data concerning them are collected, used, consulted or otherwise processed and to what extent the personal data are or will be processed"* (GDPR 2016, Article 39) as well as that *"consent should be given by a clear affirmative act"* (GDPR 2016, Article 32). It is evident that in the field of social and economic science, those mandated instruments do not qualify as new. Although, the extent to which the GDPR enforces transparency and the degree of control a natural person remains to possess over their personal data, does present a new approach. Nevertheless, the question that comes to mind is whether the implementation of higher data protection standards may change data subjects' behavior in research. According to the findings in the field of economics of privacy (for reviews see Cecere et al. 2017, Acquisti et al. 2016) as well as in the field of digital economics (for reviews see Goldfarb and Tucker 2019, Tucker 2016) regarding transparency and control mechanisms, there are arguments why this could be the case.

Initiated by research on the privacy paradox (for reviews see Gerber et al. 2018, Kokolakis 2017), describing the gap between high stated preferences for privacy and rather low observed actual privacy-protecting behavior, studies have shown that framing the act of data protection differently can affect the willingness to share personal data in various ways (e.g. Adjerid et al. 2013, Tucker 2014). On the one hand, Brandimarte et al. (2013) show that a higher perception of control leads to even more data sharing. On the other hand, Marreiros et al. (2017) show that privacy concerns are raised regardless of whether negative or positive aspects of online privacy policies are highlighted. To draw a conclusion from this research, it seems that even changing small details regarding data protection standards can affect data subject behavior critically.

Having said this, there are also valid arguments why participant behavior should remain unaffected. First, empirical research in social and economic science relies on a random sample selection of survey and experiment participants. This is why researchers always signal high data processing standards, in order to also attract participants with privacy concerns and avoid a, in this regard, biased sample selection (Diekmann 2007). Second, truthful responses of data subjects are essential for gathering reliable data. This would be at stake if improper data storage and use, harm the researcher and participants' relationship. This is why scientific researchers always aim to improve data sharing and safeness (Van Den Eynden 2008, Sieber 1988), and it is a common goal of European and German institutions to harmonize current practice with the GDPR (EDPS 2020, Schaar 2016). Third, even when abstracting from the argument that a trusting relationship between participants and researchers is vital, one could think of ethical reasons to abstain from exploiting the subject's personal data. Recapitulating these arguments, data standards in social and economic science are already high and thus, implementing the GDPR may change research settings only in a negligible way and not interfere with data subject behavior.

In addition to that, not only the framing of the data processing itself could impact data subject behavior, but also the communication process on how the GDPR is implemented. Research in the field of trust and reciprocity claims that if an action is perceived as voluntarily executed, it is rewarded higher than if the act is perceived as involuntary (McCabe et al. 2003, Toussaert 2017, Orhun 2018). This idea regarding reciprocal behavior and perceived intentions is taken up in the field of personal information disclosure. It has been shown that trust is also essential for predicting careless online behavior (Aïmeur and Sahnoune 2020), information sharing in gift-giving networks (Teubner et al. 2013), as well as for knowledge sharing of online bloggers (Chai et al. 2011). Transferring this idea, signaling voluntary or involuntary intentions when implementing the GDPR may also affect an organization's perception. According to the

research in the field of reciprocity, a voluntary signal could help to build trust towards participants, whereas signaling that the GDPR was implemented involuntarily may trigger a reverse effect.

Finally, while many studies in the field of economics of privacy assess the monetary value of personal data and which factors affect the willingness to share them (e.g. Cloos et al. 2019, Benndorf and Normann 2018, Plesch and Wolff 2018, Schudy and Utikal 2017), there is no consent so far regarding what describes the decision process behind those outcomes best.[62] Adjerid et al. (2018) address this issue by comparing the predictive power of rational modeling through a cost-benefit approach with behavioral modeling, focusing on factors such as cognitive heuristics and biases. As a result, they find that in actual choice settings, the influence of expected cost-benefit thinking plays a minor role in comparison to hypothetical choice contexts, whereas behavioral factors are highly relevant when it comes to real privacy decisions. As data disclosure for online research settings combines elements of actual choice settings and hypothetical ones, it seems interesting to investigate which approach describes outcomes in the following best.

In result of these considerations, I conduct an experiment to investigate whether increased transparency of the data protection rule of a research institute changes data subjects' willingness to share personal data. Second, whether highlighting that the GDPR was implemented involuntarily compared to voluntarily, increases privacy concerns. Third, whether privacy decisions-making is predicable through differences in the stated perceived value of data and time.

In the investigation, I exploit data from an online survey containing questions about deviant behavior and behavioral characteristics. The treatment manipulation is initiated on the first survey page, informing about the research institute's data protection rule. In the increased-transparency treatments, the wording of the data processing is displayed, whereas, in the no-information treatments, a text about the research institute that is unrelated to data protection is displayed. Since there is no difference in actual data processing between treatments and since in all treatments the entire data protection rule can be examined via a link, I focus on framing effects by varying the information shown and not changing the actual size of the threat for privacy. Functioning as the second involuntary-signal treatment, it is displayed that the research institute was forced to implement the GDPR. In the baseline voluntary-signal treatments, I

[62] For instance, Hermstrüwer and Dickert (2017) describe as reason of increased private data sharing after introducing higher government surveillance the increased need to comply with the expected social norm to give up privacy.

utilize a standard introduction sentence about data protection rules, signaling intrinsic motivation to implement the GDPR. The combination of both treatments results in a 2 x 2 design.

In order to measure privacy concerns, I ask participants at the end of the survey to disclose personal information or to abstain. In exchange for revealing their email addresses, names, and student-IDs, participants receive a lottery ticket with the opportunity to win one of two 50 Euro vouchers. Since it is announced that survey data is only collected for research, I focus on a clean tangible benefit and exclude the possible influence of more intangible benefits, such as image gains through sharing personal information or giving in to peer pressure by joining a social network (Ackfeld and Güth, 2019).

In addition, the value of data is indicated through the perceived value of reported data and the stated sensitivity of reported data. The value of time is measured by the perceived value of time spent on the survey as well as the actual time investment into the survey.

The results are as follows. In the voluntary treatment, I find, by comparing the no-information and increased-transparency treatment, a significant increase of 21 percent in the share of participants who do not disclose personal information. Although in both involuntary-signal treatments, the share of participants who abstain from the lottery is significantly higher than in the voluntary-signal-no-information treatment, there is no change in the share between those two involuntary-signal treatments. Further, by analyzing the distribution of the value of data and time stated in Euro, it becomes clear that participants struggle to assign a specific monetary value to their data, whereas, when it comes to transforming the value of time into a monetary one, this seems to be more feasible. Consequently, there is no link between the share of participants who abstain from the lottery and reported data value. However, recognizable for the value of time, there is a significant difference in the no-information treatments between participants who reveal personal information and participants who do not. This difference vanishes again in the increased-transparency treatments, which yields an intriguing piece of the puzzle to understand privacy decision-making: Since estimating the value of private data and the risk of potential data misuse presents a challenge, people tend to use the value of time as a cognitive proxy to make a decision about revealing personal information when transparency is rather low. Nevertheless, if data processing transparency is more transparent, potentially raised privacy concerns overrule the influence of this heuristic. To reinforce these outcomes, results remain robust after analyzing the stated deviant behavior, the stated sensitivity of data, as well as the actual time spent on filling out the survey.

The paper continues with the experimental setup in Section 6.2, followed by the results in Section 6.3. In the conclusion in Section 6.4, it is briefly discussed whether the GDPR functions as intended or whether it makes sense to adjust the regulation for scientific research in the future.

6.2 Experiment

6.2.1 Survey

From January 10, 2019, until January 30, 2019, students of a German University were invited via general university announcements to participate in an online survey (see an English version of the original announcement in Appendix Figure E.1.1). The organization responsible for data processing and survey content was a research institute related to the university, which was clarified as accountable regarding data protection issues.[63] The survey was announced to contain questions about personality and deviant behavior. Furthermore, it was advertised that two 50 Euro vouchers of an e-commerce company were to be drawn out in a lottery among all participants. Based on a pretest, it was signaled to participants that completing the survey would last approximately 10 minutes.

The experiment was created with the "Enterprise Feedback Suite Survey" software in German language and was built up in the following manner (see Figure E.1.2): After an introductory welcome page (see Appendix Figure E.1.3), the survey began with questions on age, gender, highest acquired degree, study major, number of semesters, and place of residence of the parents in 1989. The survey continued with questions taken from the German Socio-Economic Panel (GSOEP) about personal characteristics, including a group of questions capturing the Big Five, risk aversion, trust, and locus of control. A question about information preferences in this section, was constructed in the manner of Chadi and Homolka (2021a). At its core, the survey contained three question-batteries about deviant behavior recreated from the German General Social Survey (GGSS 2000). These questions focused on fare dodging in public transport, stealing, and drunk driving and were supplemented by two questions about streaming content illegally and cheating during an exam. Each topic was depicted in three different variants in order to receive more truthful responses and reduce social desirability (Wasmer et al. 1991). In this regard, participants were first interrogated about their general attitude towards the particular kind of behavior; second, participants were asked hypothetically to consider behaving in the outlined manner. Subsequently, participants were asked whether they had engaged in this specific kind of deviant behavior within the preceding year and, if so,

[63] Noteworthy, this was the first survey conducted by the research institute after the introduction of the GDPR.

how frequently, starting with "never" up to "more often than four times". In order to evaluate the sensitivity of the given answers so far, the items were followed up by a question about the general sensitivity towards the reported data. Subsequently, there were questions about the motivation to participate in the experiment; be it to gain money or due to an interest in, or the wish to support research. Lastly, before the lottery decision, participants were asked to assess the length of the survey and put a hypothetical value in terms of Euro on the time they had spent filling out the survey and the personal information they had listed. Since the hypothetical value of data and time spent are crucial for the subsequent analysis of possible mechanisms, the exact wording of these two questions is reported here: *"Assuming you received compensation for the data you provided as part of this study (your time spent on this study), how much would you consider to be fair?"*

The survey ended with the decision to participate in the lottery or to abstain, for which participants were prepared, due to the first introductory page of the survey. If they decided to join, participants had to report twice their university email addresses, their student-IDs, and their first names.[64]

6.2.2 Treatments

As obliged since May 25, 2018, before beginning with the survey, the research institute's data protection rule was pointed out. As an affirmative act, participants had to consent to the rule in order to proceed.[65] Whereas the first introduction sentence always remained the same, and the survey page always concluded with a link to the actual data processing rule[66], the two text modules in between were used to implement treatment manipulations portrayed in the following (All four treatments in their original web environment can be seen in Appendix Figure E.1.4).

In order to increase the transparency of data processing to its fullest, the main section of the first survey page was used. Therefore, the actual rule from the data protection website was displayed, providing complete information concisely: *"When studies are carried out, data is generated from the answers given by the participants. This data is scientifically evaluated by the IAAEU. The input is anonymized and cannot be assigned to any person. In this sense, participation in these studies is anonymous. The generated, anonymized data is used for the preparation of scientific research papers and presentations. These papers will be published."*

[64] These three pieces of personal information were necessary to ask for in order to ensure that lottery winners were students of the university and vouchers could ultimately be funded by the university.

[65] When participants did not agree with the legal conditions, they automatically dropped out of the experiment.

[66] Paragraph accessible in German at: https://www.iaaeu.de/de/datenschutz#3-welche-daten-wir-verwenden-und-warum.

Here, it is explicitly stated that data is going to be processed without any link to the participants' personal information, which is also legally binding for the research institute and its employees. Whereas this legal text was also accessible in the other treatment manipulations via the link to the survey page itself, the text in those treatments was replaced with rather general information about the nature of the research institute: *"The employees of the IAAEU research the underlying economic and legal conditions for labour in a constantly changing European society. Our task is to conduct research that has interdisciplinary approaches and is socially relevant. The IAAEU was founded as a public law foundation in 1983. The institute is financed by the German Federal State of Rhineland-Palatinate and is also a scientific institution attached to Trier University. It is located on Campus II of the University."* [67] indicating information about the research institute's general intentions and history, the paragraph is not related to data processing and is referred to as no-information treatment in the following.

In order to vary the signal of intentions revealed by the organization, the first introductory sentence was exploited as a second treatment. To implement a voluntary-signal treatment, the following wording was used: *"The IAAEU has committed itself to high standards of data protection."* This sentence is currently used by many online companies operating within the EU.[68] It suggests that the organization is voluntarily motivated to fulfill high standards of data protection. However, there is no factual information on how actual data processing is taking place. In contrast to that, in the involuntary-signal treatment, participants were confronted with the following: *"Due to the new EU General Data Protection Regulation, we are forced to inform you about our new privacy policy."*[69] Here, the implementation of the data protection rule appears to be exogenously forced by the lawmaker.

By combining the variation of "increased-transparency" versus "no-information" and displaying a "voluntary" versus an "involuntary-signal", four treatments are created. The following chapter discusses the impact those treatments may have regarding the personal information disclosure for the raffle.

6.2.3 Expectations

This paper aims to shed light on the effect of increased transparency of data processing on privacy concerns as well as to analyze the underlying mechanisms. These research questions

[67] This content is accessible at the starting page of the research institute's homepage: https://www.iaaeu.de/en/.

[68] For example, Apple asserts "your privacy is important". Google emphasizes the big responsibility and hard work "to protect your information and put you in control". Similarly, Twitter affirms they believe "you should always know what data we collect from you and how we use it".

[69] A quick online search produces a warning of phishing emails beginning with this particular sentence in German, e.g. https://www.mimikama.at/allgemein/achtung-aufgrund-der-neuen-eu-datenschutz-grundverordnung/.

are of general interest since research on the implications of the GDPR is still scarce and the regulation most likely will serve as a role model for other legislative processes (Bradford 2020). Further, the controlled environment in which this online experiment takes place enables an analysis of any isolated effects an introductory page may trigger. An insight that is likely to be overlooked by projects and surveys focusing on broader issues, but which rely on an unbiased beginning on the first page. However, almost all previous studies regarding privacy concerns focus on data sharing not only with the researcher but also with a company (e.g. Benndorf and Normann 2018), other data subjects (e.g. Schudy and Utikal 2017) or a public internet website (e.g. Hermstrüwer and Dickert 2017). By allowing for such an audience-monitoring-effect, setups clearly are more transferrable to real-world scenarios; though, in these setups, it is difficult to disentangle between channels triggered by the presence of the researcher and the presence of the additional audience, such as the need to conform with the behavior of others (Ackfeld and Güth, 2019). In this experiment, it is clear that the researcher drives all effects being the only data receptor. Having said this, this comes accompanied with the disadvantage that results of previous studies are limited, comparable to this experiment, since the audience differs. Nevertheless, it still seems to be worth considering previous empirical outcomes for forming expectations since the scientific receptor is present in all studies and previous results still may indicate which expectations are plausible to consider.

In the experiment, privacy concerns are indicated first and foremost by the decision to reveal personal information in exchange for a lottery ticket or to abstain. From an economic perspective, the incentive to participate in the lottery is clear: Only the benefit of a possible lottery win can compensate the effort put into the survey by a participant. (and consequently, to receive a lottery ticket). A previous finding of Schudy and Utikal (2017) is that roughly 14 percent of experimental participants decide to reveal personal data consistently in exchange for an incentive regardless of the experimental condition. As a consequence, in this experiment participants with strong preferences for a monetary benefit should choose to join the lottery regardless of data processing rules being transparent or not.

Vice versa, Benndorf and Normann (2018) find that one-sixth of study participants have no interest in a monetary reward for personal data, regardless of the incentive. Thus, it is plausible to expect that a share of participants will also abstain from the lottery independently from the treatment variation in this experiment.[70] Besides those two groups of participants for whom decision-making is rather settled, the remaining participants may be affected by the monetary

[70] As reason for sharing information and spending time on a survey without receiving a reimbursement could be the pure joy of sharing information or supporting research or even due to boredom.

incentive and most importantly, by the treatment variation. Following Adjerid et al. (2018), I assume that the feeling of concern about one's own privacy regardless of whether this concern is objectively justified or not is inconvenient and, therefore, costly. However, receiving a signal of increased transparency of (secure) data processing reduces the perceived risk of disclosure by reassuring participants of their data safety and therefore minimizing subjective privacy concerns. In contrast to that, when no information about data processing is highlighted, the perceived risk of disclosure seems in comparison more realistic and therefore privacy concerns of participants remain high. Thus, participant behavior in the no-information and the increased-transparency treatment is expected to differ regarding the lottery participation decision. Whereas in the increased transparency treatment, the lottery ticket in exchange for personal data seems to be an adequate incentive for a greater share of participants, which exceeds the cost of a minimal data disclosure risk. In the no information treatment, a minor share of participants perceives the lottery ticket as an adequate incentive in exchange for personal data. These considerations result in the following first hypothesis:

> *Hypothesis 1: The share of participants who reveal personal data for the lottery ticket is higher in the voluntary-signal-increased-transparency treatment than in the voluntary-signal-no-information treatment*

When displaying at first, an involuntary signal from the organization within the no-information treatment, perceived trust in the organization from the participants should decrease, and therefore the perceived risk of a possible data disclosure should increase. Thus, data sharing becomes more costly in the involuntary-signal-no-information treatment.

> *Hypothesis 2: The share of participants who reveal personal data for the lottery ticket is lower in the involuntary-signal-no-information treatment than in the voluntary-signal-no-information treatment.*

Having proclaimed this second hypothesis, it becomes clear that in the involuntary-signal treatment, two opposing effects should be at work by inducing increased transparency. Therefore, it remains ex-ante unclear whether the first postulated effect overrules the second effect or vice versa.

Moving on to the deviant behavior stated by the participants within the survey, which may serve as a qualitative channel indicating the privacy threat participants perceive. Assuming that

disclosed data becomes more sensitive when participants reveal deviant behavior, the rate of admitting responses should decline when privacy concerns are perceived as high.

> *Hypothesis 3:* *When the share of participants who reveal personal data for the lottery ticket is lower in a treatment, the share of participants who reveal deviant behavior within the survey is also lower in this treatment, compared to other treatments where the share of participants who reveal personal data for the lottery ticket is higher.*

Shifting the focus to potentially underlying mechanisms for privacy decision-making, the above-mentioned behavioral approach, as well as the classical cost-benefit thinking, come to mind (Adjerid et al. 2018). Resulting from the latter, the group of participants with a high stated value of their time and data may opt for a reward through the lottery more frequently on average than the group of participants with a low stated value of time and data. This distinction should be particularly strong in hypothetical choice settings when privacy concerns are rather low. Here a participant's decision to reveal personal information for the lottery should depend on the given subjective value of their disclosed data as well as their value of time invested. In contrast to that, in line with the behavioral approach, when the threat of privacy seems to be rather realistic, participants should no longer base their decision on a cost-benefit model, but instead, be affected by a possible treatment manipulation. Thus, data and time value should not play a decisive role in decision-making in these treatments.

> *Hypothesis 4:*
>
> *a) When privacy concerns are low, the likelihood of the lottery participation decision can be predicted based on the perceived value of data and time participants state.*
>
> *b) When privacy concerns are high, the likelihood of the lottery participation decision cannot be predicted based on the perceived value of data and time participants state.*

These hypotheses will be tested in the following.

6.2.4 Sample

The data analyzed in the results chapter below stem from three different data sources: First, administrative data from the survey software, second, administrative data from the university, and third and foremost responses within the survey from participants. Based on the first and second data source in this chapter the analyzed sample is defined. Following that, relying on

the third data source, an overview of the sample statistics is presented as well as randomization checks of the general sample response distribution over treatments.

Beginning the overview with accesses to the survey, overall, 362 accesses to the survey have been registered by the administrative data from the survey software. Of those 362 accesses, 271 reached one of the multiple final survey pages. This can be decoded in a response quote of 74.86%, which is high compared to other online surveys (Nulty 2008, Yetter and Capacciolo 2010). From those 271 accesses, two dropped out in the beginning by not consenting to the data protection rule. Those two were confronted with the involuntary-signal-no-information treatment, which could already hint towards increased privacy concerns in this treatment.[71] Additionally, one more access had to be excluded due to technical reasons. Further, eight separate accesses included doubled personal information, most likely in order to raise the lottery-winning probability. From those eight data sets, the four latter reporting identical personal information were excluded. The experimental sample ultimately consisted of 264 data sets, which are in the following referred to as participants of the survey.

In the next step, the email address, student-ID, and name revealed by participants in order to take part in the lottery, were matched with data from the university to be verified. This was necessary to identify additional participants who did not reveal personal information truthfully. This was the case for two persons who revealed mismatching email addresses, names, and student-IDs. These two persons took part in the voluntary-signal-increased-transparency treatment and the involuntary-signal-increased-transparency treatment. Two additional persons typed in an eight-digit student-ID, whereas only six or seven digits weren given from the university at that time. These two persons took part in the involuntary-signal-no-information and involuntary-signal-increased-transparency treatment. In the following, all four persons are treated as if they did not reveal personnel data. In additional robustness analyses, I check whether all outcomes are robust excluding these four persons and always report when outcomes differ in the following analyses.

The third and main data source are the answers of participants within the survey. The overall sample statistics for all 264 participants are shown in Table 6.1. Means for socio-demographics, education and personal characteristics, as well as motivation are displayed by the treatment group in columns 1-4 and the minimum and maximum value for each item in columns 5 and 6. Treatment groups are roughly equally distributed, with the smallest containing 62 participants and the largest 69. In the next step Table E.2.1 displays the results for Kruskal-Wallis equality

[71]It would be interesting to analyze the treatment allocation of all 91 dropouts, which did not even reach a final survey page. Unfortunately, the software does not provide information on these dropouts.

of populations rank tests for all variables and treatment group differences. Looking for significant differences over all four treatment groups combined, the tests reveal no significant differences, which implies that all samples come from the same population. Going one step further by comparing only two treatment groups at the same time, confirms this result, except of the share of females and the information avoidance characteristic. Between the voluntary-signal-no-information and the involuntary-signal-no-information treatment, females' share is significantly different (chi-squared with ties = 3.872, p= .049). As shown in Table 6.1 the share of female participants in the voluntary-signal-no-information treatment is 62 percent, which indicates a lower, but overall still high share of females, compared to the other three treatments. For the information avoidance variable there is a significant difference for the voluntary-signal-increased-transparency treatment to the voluntary-signal-no-information treatment (chi-squared with ties = 3.953, p= .047) as well as to the involuntary-signal-increased transparency treatment (chi-squared with ties = 35.130, p= .024). This results from a high mean of 9 in the voluntary-signal-increased-transparency treatment compared to an overall mean of 8.5. Although the overall test is insignificant for both variables and the reported significance is rather weak, in the following analysis I control carefully whether the share of females or the information avoidance characteristic affects the results. Regardless, all variables shown in Table 1 are exploited as control variables in regression analyses below.

Table 6.1 Sample statistics – Control variables

	Treatments					
	(1)	(2)	(3)	(4)	(5)	(6)
Specification	Mean	Mean	Mean	Mean	Min	Max
Socio-Demographics						
Age	23.23	22.94	23.53	23.50	18	54
Female	0.62	0.75	0.76	0.79	0	1
Education						
Education level	0.32	0.27	0.41	0.32	0	2
Subject: Educational science	0.43	0.54	0.42	0.41	0	1
Subject: Empirical science	0.28	0.21	0.21	0.30	0	1
Subject: Humanities and others	0.29	0.25	0.37	0.29	0	1
Number of semester	3.88	4.03	3.90	3.73	0	11
Personal characteristics						
Big5: extraversion	4.51	4.68	4.57	4.67	3	7
Big5: agreeableness	4.86	4.94	4.78	4.97	2	6.67
Big5: openness	5.15	4.99	4.94	5.01	1.67	7
Big5: neuroticism	4.59	4.56	4.65	4.61	2.34	6.67
Big5: conscientiousness	4.85	4.96	4.87	4.90	2.34	7
Risk aversion	5.00	4.84	5.16	4.95	1	10
Locus of control	4.51	4.47	4.49	4.55	1.9	6.1
Trust	6.62	6.52	6.11	6.53	1	10
Information avoidance	8.48	9.00	8.53	8.50	3	10
Motivation						
Money	0.51	0.58	0.58	0.48	0	1
Research	1.23	1.09	1.15	1.02	0	2
N	69	67	62	66		

Notes: Means by treatments are displayed in columns 1, 2, 3, and 4. Minimum and maximum values are reported in columns 5 and 6. 1: voluntary-signal-no-information; 2: voluntary-signal-increased-transparency; 3: involuntary-signal-no-information; 4: involuntary-signal-increased-transparency. All variables are from the survey. Age is a metric variable. Female is a dummy, which is "1" when stated female and "0" when stated male or other. The "Education" variable consists of three categories: school advanced level, intermediate university degree, high university degree. One observation is excluded in the education row because of stated education as "other". Mean is calculated over all three categories Subject consist of three categories: Educational science includes teacher trainees and psychology students. Empirical science includes stem students and social science students. Humanities and others includes linguistic students, law students and remnant subjects. The number of semester is a metric variable. The wording of survey questions on Big5 personality dimensions (15 items), locus of control (10 items), and trust (1 item) are taken from the German Socio-Economic Panel study. The wording of the survey question on information avoidance (1 item), motivation for money (1 item), and motivation to support research (2 items) are taken from Chadi and Homolka (2021a). For sample comparisons based on Kruskal-Wallis equality of populations rank tests see Table E.2.1.

6.3 Empirical analysis

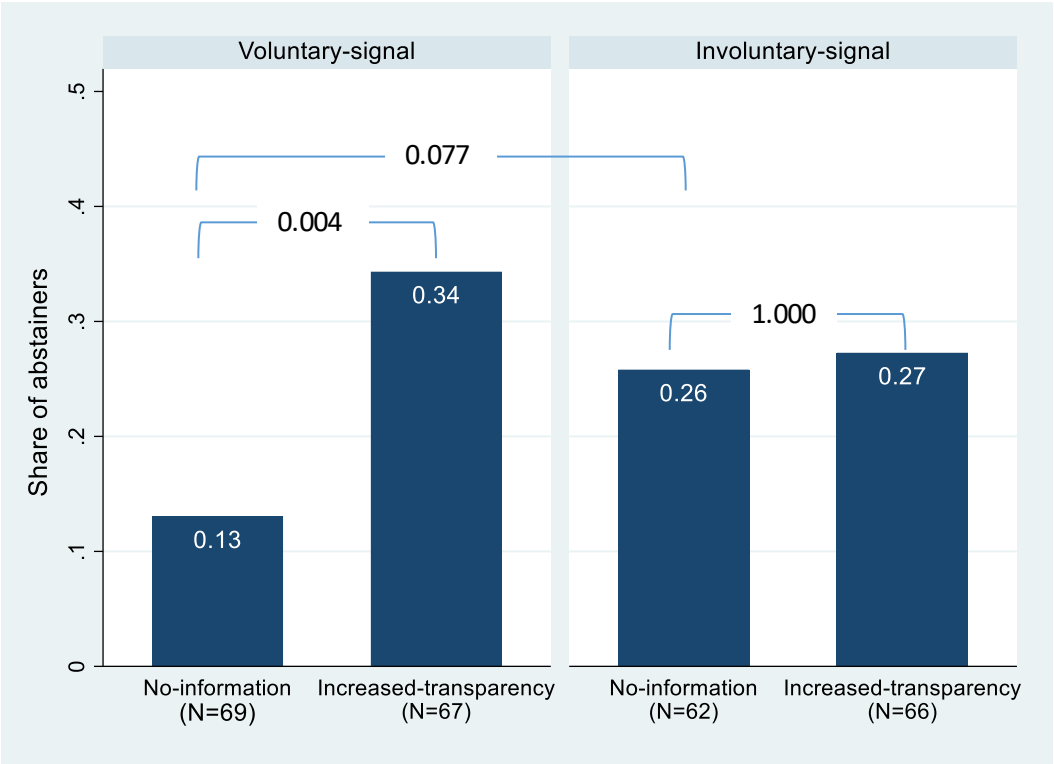
Section 6.3.1 initiates the analyses regarding the impact of the four treatments on privacy concerns, expressed through a participants' decision to take part in the lottery or not and therefore reveal personal information. In Section 6.3.2, the captured results are checked for robustness. Section 6.3.3 continues with the analysis of the possible mechanisms which may explain the results presented before.

6.3.1 Privacy concerns

Figure 6.1 presents the main outcomes of the study. Indicated in blue bars is the share of participants who decided to abstain from the lottery and thereby expressed their privacy concerns. The overall share of participants who abstained (henceforth called abstainers) over all treatments is precisely 25 percent and 75 percent for participants who revealed personal data (henceforth called revealers). In the first two bars, the percentage share of abstainers in the voluntary-signal-no-information treatment (N=69) and the voluntary-signal-increased-transparency treatment (N=67) is presented. In contrast to hypothesis 1, the share of abstainers in the voluntary-signal-no-information treatment is the lowest over all treatments with 13.04%, whereas in the voluntary-signal-increased-transparency treatment it is the highest with 34.33%. Estimating a *p*-value for a Fisher's exact test reveals that the treatment difference is also highly statistically significant (two-sided Fisher's exact test: p=.004). It seems that, contrary to the expectations expressed in hypothesis 1, when a voluntary signal is displayed at first, increased transparency of the data protection rule raises the share of participants who decide not to participate in the lottery and hence, privacy concerns increase. In the next step, focusing on the two bars on the right, I show the percentage share of abstainers for the involuntary-signal-no-information treatment (N=62), as well as the involuntary-signal-increased-transparency treatment (N=66). Although the percentage share of 25.81 in the involuntary-signal-no-information treatment is higher than in the voluntary-signal-no-information treatment (two-sided Fisher's exact test: p=.077), which is in line with hypothesis 2, there is no difference in the share of abstainers in comparison to the involuntary-signal-increased-transparency treatment. This is also true after calculating a Fisher's exact test (two-sided Fisher's exact test: p=1.000). These results are capturing first that in comparison to the voluntary-signal-no-information treatment, the involuntary-signal treatment results in a significantly higher share of abstainers, which is in line with hypothesis 2. Additionally, the substantial difference between no-information and increased-transparency treatment observed in the voluntary-signal condition cannot be found again in the involuntary-signal condition. Thus, the strong negative

effect on the share of revealers of increased transparency in the voluntary-signal treatment seems to differ in the involuntary-signal treatment.

Figure 6.1 Abstaining behavior and treatments



Notes: Percent of participants who decided to abstain from the lottery are shown by treatment. *p*-values of two-sided Fisher's exact test are displayed in brackets above bars.

6.3.2 Additional analyses

First, to support the main findings, I carry out additional probit-regression analyses, adjusting for the variables shown in Table 6.1. As shown in Table E.2.2, the increased transparency effect in the voluntary-signal conditions remains strong, and the effect size even increases, adjusting for variables of personal characteristics. Additionally, in Table E.2.3, the difference between the share of abstainers in the voluntary-signal-no-information treatment and the share in the involuntary-signal-no-information treatment remains stable. This can be seen as further evidence for the rejection of hypothesis 1 and the maintenance of hypothesis 2. Interestingly, in both cases, the parameter “neuroticism”, as part of the Big Five, is a predictor for being an abstainer or revealer, which makes sense considering that this variable should capture personal traits such as the tendency towards nervousness. Furthermore, in the voluntary-signal treatments in Table E.2.2, studying educational science and being in a higher semester makes

participants more likely to abstain, whereas in the no-information treatments in Table E.2.3, reporting a higher locus of control results in a lower probability of being an abstainer.

In the next step, I address hypothesis 3 and whether the results found above are also mirrored in the amount participants reporting deviant behavior. In expectancy of this, an additional sign of high privacy concerns could be indicated by participants underreporting deviant behavior compared to other treatments. Therefore, in Table E.2.4, the mean of the 15 variables regarding deviant behavior is shown by treatment in columns 1 to 4 as well as the minimum and maximum values in columns 5 and 6. Using Kruskal-Wallis equality of populations rank tests over response behavior, reveals almost no difference between treatments. Notably, there are a few exceptions. In the voluntary-signal conditions, in columns 1 and 2, there are significant differences for the question regarding drunk driving of others (chi-squared=3.303, *p*=.069) and the question regarding hypothetical cheating in an exam (chi-squared=2.926, *p*=.087). For both questions, the mean for the voluntary-signal-no-information treatment is higher, which translates to judging drunk driving as "not that bad" and instead to consider cheating in an exam compared to participants in the voluntary-signal-no-information treatment. Using rank-sum tests, the question regarding drunk driving (*p*=.017), as well as the question regarding cheating (*p*=.024), are also significantly different between treatments. When comparing the no-information conditions in column 1 and column 3, two means are different, first, for the question regarding hypothetical fare dodging (chi-squared=4.342, *p*=.037) and second, the question regarding committed fare dodging (chi-squared=3.375, *p*=.066). Again, in both cases, the mean of the voluntary-signal-no-information treatment is higher than in the involuntary-signal-no-information treatment, which translates to considering fare dodging in public transport more and also committing to it more often. Further, this is also significantly different when using rank-sum tests for the question regarding hypothetical fare dodging (*p*=.031) as well as actual fare dodging (p=.013). To conclude, there seems to be a small tendency for participants to report slightly more deviant behavior in the voluntary-signal-no-information treatment. However, overall differences appear only for a few questions and thus, hypothesis 3 must be rejected.[72]

In another effort, accesses to the institute's actual data protection rule website are analyzed to check whether participants, independent from treatment manipulation, have checked out the data protection rule and thereby received a similar treatment manipulation as in the increased-transparency conditions. To rule out that data processing transparency was affected by such an

[72] Bearing this in mind, also expanding the regression models shown in Table E.2.2 and E.2.3 and adjusting for the variables addressing deviant behavior does not change the significance of the two results reported in Figure 6.1 (not shown here). Concluding, both effects are robust regarding the set of possible control variables and are not driven by the questions regarding deviant behavior.

endogenous effect, possible correlations between the daily accesses to the survey and page views of the actual data protection rule are analyzed. As shown in Figure E.2.1, there is no visible consonance between the daily number of accesses to the survey, shown in red, and the accesses to the data protection website, shown in blue, regarding the experiment's time frame, marked in green. This picture is supported by the OLS-regression analysis shown in Table E.2.5. In column 1 of Table E.2.5, the daily accesses to the survey are not related to the daily accesses to the data protection website. This non-correlation is also observable in column 2 when using the deviation from the weekly mean of accesses to the data protection website as a dependent variable to adjust for outliers. In the next step in column 3 and 4, it becomes even more evident that no link between page views of the website and the days the survey was accessible when analyzing exists. The time frame of the experiment compared to the months before and after the experiment, has no effect on the daily accesses to the data protection website. Further, in columns 5 and 6, the regressions in columns 3 and 4 are adjusted by checking whether the outcome changes when the monthly dispatch of a newsletter mail shared by the research institute is added into the model as well.[73] Interestingly, those newsletter mails seem to increase the number of page views to the data protection rule on the day sent and the day afterwards strongly. However, this effect does not uncover a correlation between page views and the experimental time frame. Consequently, there is no evidence that participants extensively visited the actual data protection website, which could explain the presented results.

6.3.3 Mechanisms

6.3.3.1 Value of data

The analysis starts with the perceived value of data. According to hypothesis 4a, especially in the voluntary-signal-no-information treatment, where the share of abstainers is relatively low and therefore privacy concerns may be negligible, the lottery participation decision should be predictable based on the stated value of data. In contrast to that, the effect of such cost-benefit thinking may be rather low in the voluntary-signal-increased-transparency treatment, where the share of abstainers is relatively high, and therefore, privacy concerns prevail. Starting with the hypothetical stated value of reported data Table E.2.6, Panel A shows the mean per treatment in the first row. It is worthwhile to mention that participants used the maximum possible hypothetical value-range from 0 up to 9999 hypothetical Euro, which can be seen in columns 5 and 6. An explanation of this wide range of values may be that putting an exact price tag on the value of your own personal information is difficult for participants. Hence, in Panel B of

[73] This newsletter is not in any way related to the experiment and deals with new developments in labor law.

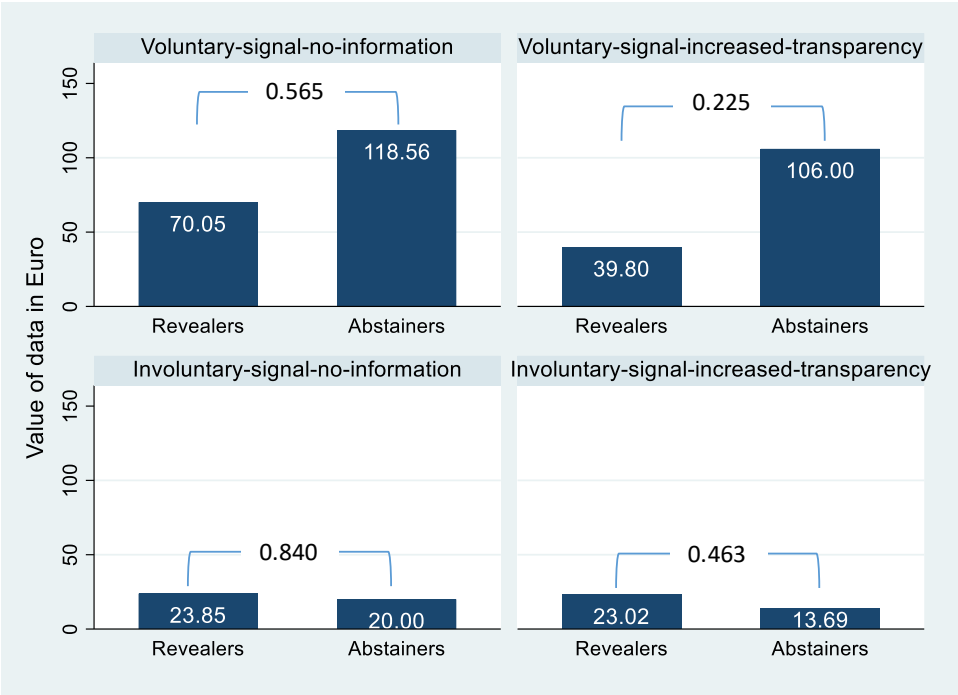
Table E.2.6, a version of this variable is created excluding the top 1 percent of all observations, which translates to dropping three observations.[74] Two participants considered their data to be of the value of 9999 and one the value of 3000. This results in the following variable named "Value of data 99%". All tests hereafter are executed using this version of the variable as well as the complete variable value of data and a log transformed version in order to check whether results also hold for the three outliers.

Building upon this, Figure 6.2 shows the variable "value of data 99%" by treatment differences between participants who decided to reveal personal information in order to participate in the lottery and abstainers who chose not to participate.

At first sight, it seems that in both voluntary-signal conditions, participants value their data more and especially abstainers in those treatments seem to have a very high value of their data. However, shifting the focus on to *p*-values of t-tests displayed in brackets above does not reveal any statistical significance. Further analyses reveal that using the complete version of the variable value of data, there is a slight tendency that in the voluntary (p=.071, two-sided t-test) as well as in the involuntary-signal (p=.020, two-sided t-test) condition, that the data value for abstainers is higher in comparison to revealers in the increased-transparency condition. However, this difference is again not statistically significant for the log-version (p=.119, two-sided t-test) (p=.244, two-sided t-test).

[74] It is common standard when excluding outliers to cut off the lower and upper bound. However, in this case there are good reasons against that. First of all, the hypothetical value “0” for the value for data was overall treatments stated 30 times. That means, that hypothetical zero Euro at the lower bound are not outliers, but an important part of the sample. Additionally, as tackled in Section 6.2.3 there are also good reasons for some participants to just share their data for free with the researcher.

Figure 6.2 Abstaining behavior, treatments, and stated value of data



Notes: Average stated value of data in hypothetical Euro is shown by treatment and lottery participation decision. *p*-value of t-tests are shown in brackets above bars. For the display as well as for the t-tests, three outliers representing the top 1% are excluded. For more information on the variable "Value of data 99%" see Table E.2.6, Panel B.

Additionally, another possible estimator of participant's concern towards their data in Figure E.2.2, is the average outcomes by treatment and abstainers versus revealers, shown by reported data's stated sensitivity. Captured using a 10-point Likert-scale, treatment differences between the stated sensitivity may be indicating another decision-mechanism for privacy concerns. However, checking whether results differ from the non-finding for the stated value of data, does not reveal different outcomes. Shown rank-sum tests and additionally carried out t-tests affirm this non-finding. Thus, according to both measurements, it must be concluded that in contrast to hypothesis 4, no matter whether the share of abstainers is low or high, perceived data sensitivity and value of data cannot predict the decision to participate in the lottery.

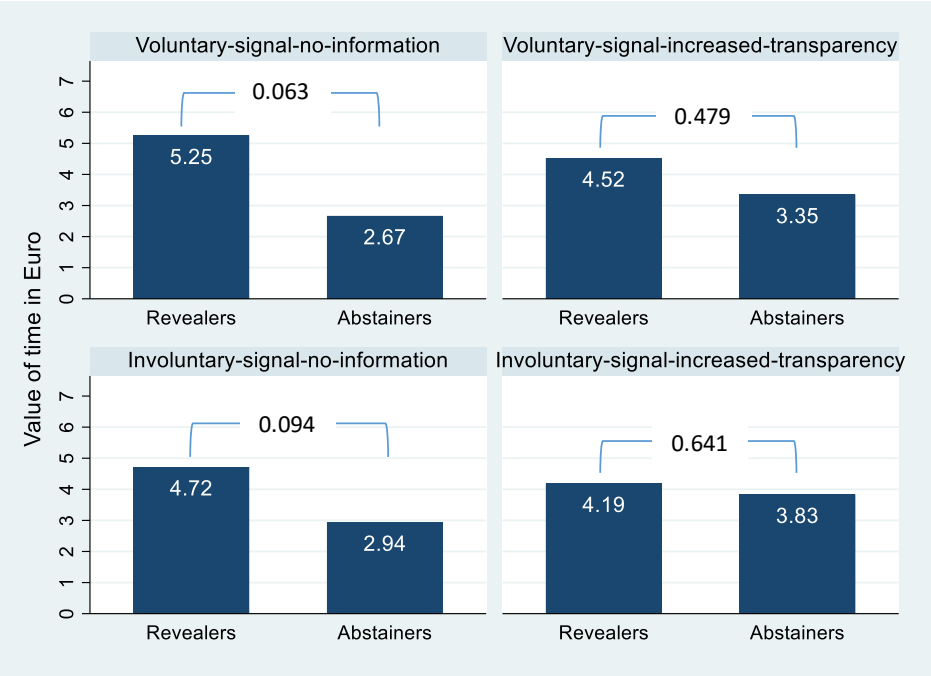
6.3.3.2 Value of time

In the next step, I shift the focus to the impact of the hypothetical stated value of time and time spent to fulfill the survey. For the hypothetical value of time, Table E.2.6, Panel A reveals that in contrast to the hypothetical value of data, participants have some consensus about what their time is worth. Mean values reported are between 4.91 Euro in the voluntary-signal-no-

information treatment and 4.10 Euro in the involuntary-signal-increased-transparency treatment.

In Figure 6.3, the differences between abstainers and revealers of the mean value of time are shown for each treatment. Beginning with the voluntary-signal-no-information treatment, there is a clear difference between revealers who value their time with 5.25 Euro the highest and abstainers who value their time with 2.67 Euro the lowest, compared to all other treatment groups. This is a significant difference as shown in Figure 6.3 in the reported t-test as well as when using a rank-sum test (p=.021). In the voluntary-signal-increased-transparency treatment, this difference shrinks and is no longer statistically significant. Interestingly, this finding supports hypothesis 4 and the impact of time on the decision-making when privacy concerns are low such as in the voluntary-signal-no-information treatment, which is mitigated within the voluntary-signal-increased-transparency treatment when privacy concerns are rather high. Additionally, in the involuntary-signal-no-information treatment, the perceived time value seems to have a weaker impact, significant using a t-test (p=.094, two-sided t-test), but insignificant using a rank-sum test (p=.166), when it comes to decision-making, whereas in the involuntary-signal-increased-transparency treatment, the effect vanishes again completely.

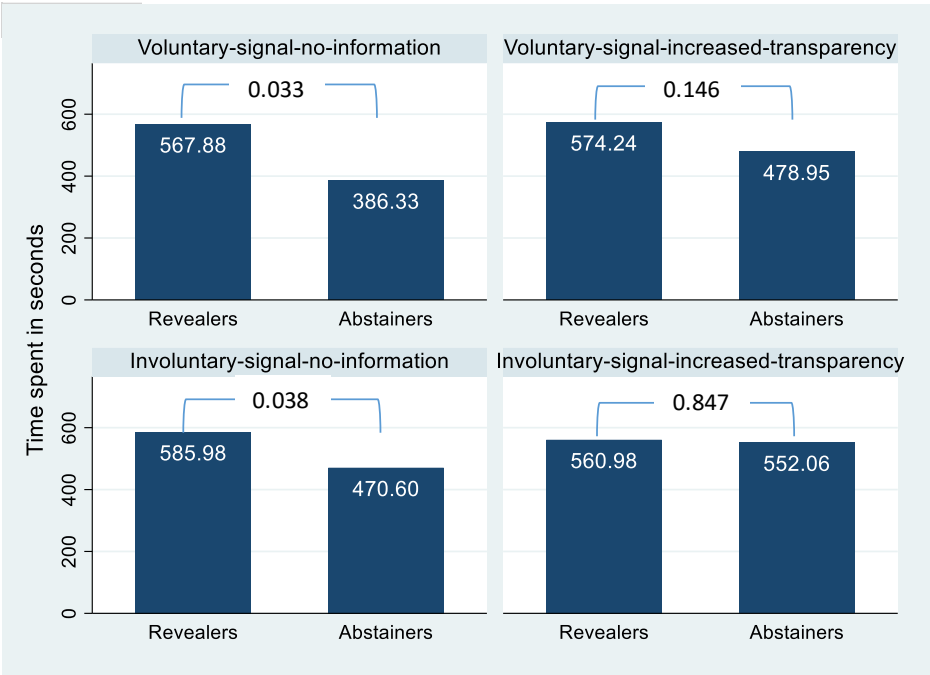
Figure 6.3 Abstaining behavior, treatments, and stated value of time



Notes: Average stated value of time in hypothetical Euro is shown by treatment and lottery participation decision. p-value of t-tests are shown in brackets above bars. For more information on the variable "Value of time" see Table E.2.6, Panel A.

Analyzing the duration in seconds spent on the survey supports this picture. As can be seen in Table E.2.6, Panel C, this variable includes some large outliers. Hence, for the visualization in Figure 6.4 a version of this variable is used excluding the top and bottom one percent shown in Table E.2.6, Panel D.[75] Using t-tests, it can be shown that in the voluntary-signal-no-information treatment for the lottery participation, time spent on the survey plays a significant role for being a revealer (more time spent on the survey) or an abstainer (less time spent on the survey). This difference of more than 3 minutes is also significant using a log-transformation (p=.008, two-sided t-test) and vanishes again in the voluntary-signal-increased-transparency treatment (p=.552, two-sided t-test).

Figure 6.4 Abstaining behavior, treatments, and time spent filling out the survey



Notes: Average values of time spent in seconds on the survey are shown by treatment and lottery participation decision. p-value of t-tests are shown in brackets above bars. For the display as well as for the t-tests six observations representing the top and bottom 1% are excluded. For more information on the variable "Time spent 98%" see Table E.2.6, Panel D.

Additionally, the same can be observed in the involuntary-signal condition, where in the no-information treatment, time spent plays a significant role predicting decision-making (p=.014, two-sided t-test). Further, this effect is not observable in the increased-transparency treatment (p=.338, two-sided t-test). When considering the results of the impact of perceived time value and the time spent on the lottery participation decision, it becomes clear that in both no-

[75] Also, all tests are repeated with the original variable as well as a log transformation version here.

information conditions, cost-benefit thinking regarding time costs plays a significant role. On the contrary, the impact is non-existent in the increased-transparency treatments, which speaks for a different decision process within the increased-transparency treatments and the maintaining of hypothesis 4 regarding the value of time.

In a final effort, I exploit the administrative data from the survey software once more and analyze second precise data on how much time participants have spent on each survey page. The goal is to find out whether the time difference between revealers and abstainers is caused by mainly one survey page, for instance, the treatment page, or if the result is based on an effect coming from the time difference over all pages combined. Table E.2.7 displays regression results for each survey page from 1 to 15 separately by treatment and distinguishing between being a revealer or abstainer. The variable shown in Table E.2.6, Panel D, excluding the upper and lower bound, is analyzed here again. Looking at the results, it is hard to identify one survey page as responsible for the amount of time invested less by abstainers in the voluntary-signal-no-information treatment in column 1 and the involuntary-signal-no-information treatment in column 3. In order to analyze also the outcomes for all observations shown in Table E.2.6, Panel C, this general trend can also be observed using a log transformation of the duration time spent on each survey page using linear regression models (not shown here). On all pages, abstainers spent less time than revealers in those two treatments except the lottery decision page, which can be explained by the fact that revealers had to complete one more page in order to disclose personal data, whereas abstainers were able to stay on this page even when they had completed the survey.

In conclusion, it is difficult for participants to evaluate their data reported in the survey in terms of monetary value over all treatments. Instead, participants tend to utilize the value of time as well as the actual time spent for the lottery decision-making in the voluntary-signal-no-information treatment. This effect is partly also observable in the involuntary-signal-no-information treatment and is not driven by a particular survey page. Hence, as presented in Figures 6.3, 6.4 and Table E.2.7, the amount of time invested seems to be used by participants with a cost-benefit approach in mind, to decide whether to participate in the lottery when there is no precise information about data processing and regardless of which motivation the organization signals.

6.4 Conclusion

This study sheds light on how data subjects' behavior is affected when implementing the GDPR and instruments, such as increased transparency of data processing or signaling a voluntary

versus involuntary motivation. Furthermore, potential heuristics for privacy decision-making are analyzed. The results firstly show that in contrast to the expectations, the rise of increased transparency of data processing raises privacy concerns regardless of perpetually high data protection standards. However, as a second finding, it can be shown that the effect is undermined when an involuntary signal is presented along with the data protection rule. Thus, more research is needed to uncover whether the first or the latter finding is the dominant pattern when it comes to privacy decision-making in scientific research. As expected, also signaling the involuntary implementation of the GDPR raises privacy concerns compared to a voluntary implementation. All findings are neither influenced by the stated deviant behavior, nor correlated with access to the actual data protection website. Furthermore, the analysis of the underlying mechanism reveals intriguing insights. First, for participants, the value of data seems challenging to estimate and therefore has no measurable impact on privacy decision-making independent of the treatment. This is in line with previous research (for an overview, see Wessels et al. 2019). Though regarding the perceived value of time as well as the actual time spent on the survey, participants tend to take those dimensions into account when it comes to decision-making and the organization signals voluntary intentions implementing the GDPR.

Further, this study offers interesting insights for researchers and policymakers. First, the partial link between incentives, the value of time, and the time spent on an online survey seem to be a predictable proxy for privacy decision-making when privacy concerns are low. This could be worthwhile to investigate further in real-world settings regarding data sharing with apps or in online social networks. Second, in the era of the COVID-19 outbreak, researchers increasingly rely on online research, which is required in various fields be it to support health research or advise policymakers on how to nudge citizens to keep social distancing. Hence, to gather truthful data in order to allow for appropriate political actions, online scientific research is of great importance. This study shows in line with other online research (Brandimarte et al. 2013, Marreiros et al. 2017) how small changes can affect participants' privacy concerns and, therefore, may cause a biased sample selection in this regard. As a consequence of that, one could argue that legislation should exempt scientific research, at least to a certain degree, from the GDPR in order to prevent biased sample selection. However, lawyers argue that Article 33 of the GDPR is necessary to secure high data standards and prevent loopholes for companies that conduct (market) studies declared as research (EDPS 2020). Thus, scientific research taking place in the European Union needs to present data processing rules with care in order to address privacy concerns adequately and fulfill the GDPR at the same time.

6.5 Appendix E.1: Additional background information

Figure E.1.1 Announcement flyer of the online survey

IAAEU

Institute for Labour Law and Industrial Relations in the European Union

Online survey:

Rebel or bourgeois?

No matter! We want to know how you tick.

The IAAEU conducts a 10-minute survey on the subject of personality and deviant behaviour. As a reimbursement for your participation, we raffle two 50€ Amazon vouchers among all participants.

Use the next bus ride, smoking break or learning phase to challenge your luck and support research! If you are interested and you have a solid knowledge of German, follow this link to participate:

https://www.unipark.de/uc/abweichendes_Verhalten/?a=

Thank you very much!

Your IAAEU Lab Team

Email: labor@iaaeu.de

Figure E.1.2 Design of the experiment

Page	Description
1	Treatments: Voluntary-signal-increased-transparency Voluntary-signal-no-information Involuntary-signal-increased-transparency Involuntary-signal-no-information
2	Welcome page
3	Socio-demographics
4	Big5
5	Risk aversion, information preferences and trust
6	Locus of control
7	Deviant behavior: fare dodging
8	Deviant behavior: stealing
9	Deviant behavior: streaming
10	Deviant behavior: cheating
11	Deviant behavior: drunk driving
12	Sensitivity
13	Motivation
14	Stated value of data and time
15	Lottery decision
16	Personal data (only if lottery decision is positive)

Figure E.1.3 Welcome page of the survey

Thank you very much for letting us evaluate your data!

The following study focuses on the connection between deviant behavior and personality. By deviant behavior, we mean behavior that violates existing norms or rules. Examples could be fraudulent income tax returns or disregard for traffic regulations.

In order to investigate the connection between deviant behavior and personality, we will ask you questions about this in the following survey. The study will take about ten minutes. Please answer them as truthfully as possible. As a thank you for your participation in this study, we are giving away two 50 Euro Amazon vouchers. On the last page of the survey, you have the possibility to decide whether you want to take part in the raffle. In order to participate, you must provide your first name, your matriculation number, and your university email address.

Figure E.1.4 English version of the treatment pages used in the online survey

A) Voluntary-signal-no-information treatment

Welcome to this study by the Institute for Labor Law and Industrial Relations in the European Union (IAAEU)!

We first need your consent for this, so that we can evaluate your private data for research purposes afterwards.

The IAAEU has committed itself to high standards of data protection.

The employees of the IAAEU research the underlying economic and legal conditions for labour in a constantly changing European society. Our task is to conduct research that is interdisciplinary approaches and socially relevant. The IAAEU was founded as a public law foundation in 1983. The institute is financed by the German Federal State of Rhineland-Palatinate and is also a scientific institution attached to Trier University. It is located on Campus II of the University.

You can find our privacy statement here: https://www.iaaeu.de/en/datenschutz

I agree that my data may be evaluated after sending the survey.

Agree ○ Disagree ○

B) Voluntary-signal-increased-transparency treatment

Welcome to this study by the Institute for Labor Law and Industrial Relations in the European Union (IAAEU)!

We first need your consent for this, so that we can evaluate your private data for research purposes afterwards.

The IAAEU has committed itself to high standards of data protection.

When studies are carried out, data is generated from the answers given by the participants. This data is scientifically evaluated by the IAAEU. The input is anonymized and cannot be assigned to any person. In this sense, participation in these studies is anonymous. The generated, anonymized data is used for the preparation of scientific research papers and presentations. These papers will be published.

You can find our privacy statement here: https://www.iaaeu.de/en/datenschutz

I agree that my data may be evaluated after sending the survey.

Agree ○ Disagree ○

C) Involuntary-signal-no-information treatment

Welcome to this study by the Institute for Labor Law and Industrial Relations in the European Union (IAAEU)!

We first need your consent for this, so that we can evaluate your private data for research purposes afterwards.

Due to the new EU General Data Protection Regulation, we are forced to inform you about our new privacy policy.

The employees of the IAAEU research the underlying economic and legal conditions for labour in a constantly changing European society. Our task is to conduct research that is interdisciplinary approaches and socially relevant. The IAAEU was founded as a public law foundation in 1983. The institute is financed by the German Federal State of Rhineland-Palatinate and is also a scientific institution attached to Trier University. It is located on Campus II of the University.

You can find our privacy statement here: https://www.iaaeu.de/en/datenschutz

I agree that my data may be evaluated after sending the survey.

Agree ○ Disagree ○

D) Involuntary-signal-increased-transparency treatment

Welcome to this study by the Institute for Labor Law and Industrial Relations in the European Union (IAAEU)!

We first need your consent for this, so that we can evaluate your private data for research purposes afterwards.

Due to the new EU General Data Protection Regulation, we are forced to inform you about our new privacy policy inform.

When studies are carried out, data is generated from the answers given by the participants. This data is scientifically evaluated by the IAAEU. The input is anonymized and cannot be assigned to any person. In this sense, participation in these studies is anonymous. The generated, anonymized data is used for the preparation of scientific research papers and presentations. These papers will be published.

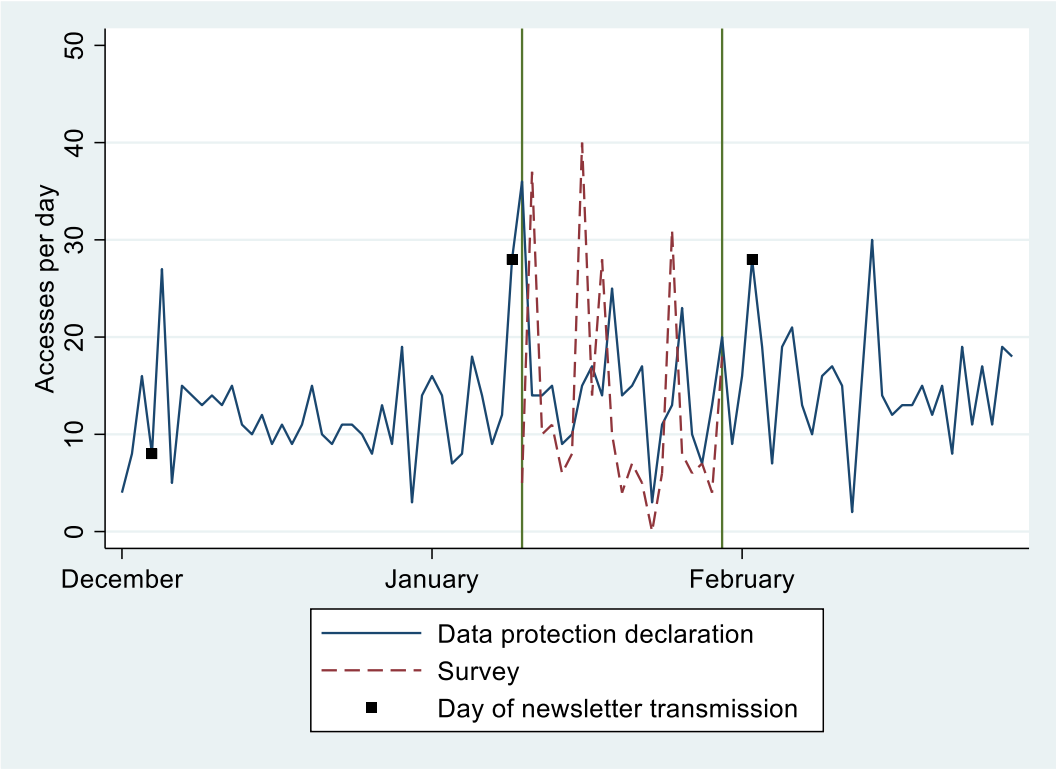
You can find our privacy statement here: https://www.iaaeu.de/en/datenschutz

I agree that my data may be evaluated after sending the survey.

Agree ○ Disagree ○

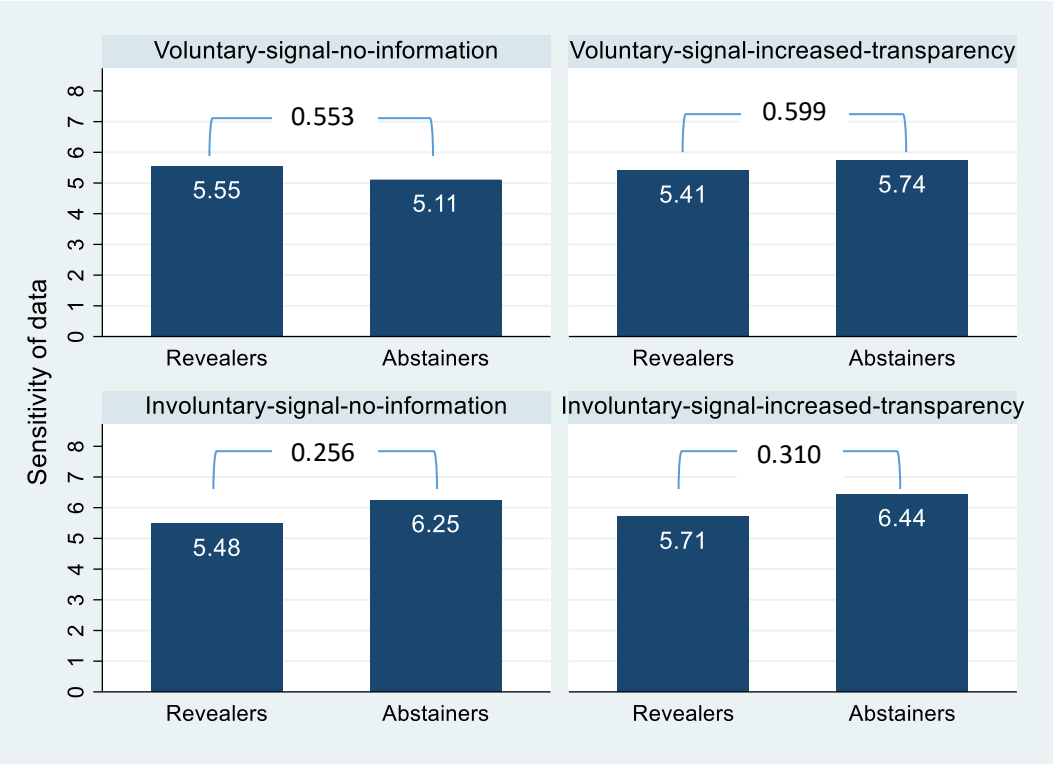
6.6 Appendix E.2: Supplementary statistics and analyses

Figure E.2.1 Experimental time frame and accesses to the data protection declaration



Notes: Timeline in months is displayed on the x-axis. On the y-axis, accesses to the data protection declaration per day are viewed in blue as well as accesses to the survey per day in red. The green frame marks the time period where the experiment took place. Three black squares mark the point in time on which the institute distributes a monthly newsletter not related to the experiment. See Table E.2.5 for more information.

Figure E.2.2 Abstaining behavior, treatments, and stated sensitivity of data



Notes: Values of the stated sensitivity are shown by treatment and lottery decision. p-value of rank-sum-tests are shown in brackets above bars. For more information on the variable "Sensitivity of data" see Table E.2.6, Panel A.

Table E.2.1 Randomization checks

	Treatment comparison				
	(1)	(2)	(3)	(4)	(5)
Specification	Overall Treatments	Treatment 1 versus 2	Treatment 3 versus 4	Treatment 1 versus 3	Treatment 2 versus 4
Socio-Demographics					
Age	1.637	0.951	0.091	0.006	1.501
	(0.651)	(0.329)	(0.763)	(0.939)	(0.221)
Female	6.078	2.364	0.004	3.872	0.320
	(0.108)	(0.124)	(0.951)	(0.049)	(0.572)
Education					
Education level	2.245	0.632	0.763	0.554	0.384
	(0.523)	(0.427)	(0.383)	(0.457)	(0.536)
Subject: Educational science	2.798	1.420	0.014	0.032	2.176
	(0.424)	(0.233)	(0.907)	(0.859)	(0.140)
Subject: Empirical science	2.357	0.810	1.445	0.757	1.535
	(0.507)	(0.368)	(0.229)	(0.384)	(0.215)
Subject: Humanities and others	2.244	0.222	0.993	0.967	0.195
	(0.523)	(0.637)	(0.319)	(0.325)	(0.659)
Number of semester	0.279	0.078	0.090	0.000	0.254
	(0.964)	(0.779)	(0.765)	(0.983)	(0.614)
Personal characteristics					
Big5: extraversion	2.580	1.830	0.423	0.262	0.001
	(0.461)	(0.176)	(0.516)	(0.609)	(0.973)
Big5: agreeableness	2.834	0.931	1.740	0.213	0.003
	(0.418)	(0.335)	(0.187)	(0.645)	(0.958)
Big5: openness	1.224	0.772	0.017	0.495	0.007
	(0.747)	(0.380)	(0.895)	(0.482)	(0.935)
Big5: neuroticism	0.438	0.026	0.109	0.325	0.084
	(0.932)	(0.873)	(0.741)	(0.569)	(0.771)
Big5: conscientiousness	1.082	0.928	0.034	0.165	0.297
	(0.781)	(0.335)	(0.854)	(0.684)	(0.586)
Risk aversion	0.682	0.257	0.258	0.167	0.169
	(0.877)	(0.612)	(0.612)	(0.683)	(0.681)
Locus of control	0.680	0.006	0.011	0.122	0.392
	(0.878)	(0.939)	(0.916)	(0.727)	(0.531)
Trust	2.022	0.049	0.836	1.838	0.057
	(0.568)	(0.825)	(0.361)	(0.175)	(0.811)
Information avoidance	6.132	3.953	0.180	0.061	5.130
	(0.105)	(0.047)	(0.671)	(0.806)	(0.024)
Motivation					
Money	1.983	0.762	1.169	0.703	1.254
	(0.575)	(0.383)	(0.280)	(0.402)	(0.263)
Research	4.536	1.829	1.452	0.899	0.481
	(0.209)	(0.176)	(0.228)	(0.343)	(0.488)

Notes: Chi-squared values with ties for Kruskal-Wallis tests are shown. *p*-values are displayed in the brackets below. In column 1 all treatments are taken into account 1 In columns 2, 3, 4 and 5 always two treatments are solely taken into account. Treatment 1: voluntary-signal-no-information; treatment 2: voluntary-signal-increased-transparency; treatment 3: involuntary-signal-no-information; and treatment 4: involuntary-signal-increased-transparency.

Table E.2.2 Abstaining behavior in the voluntary-signal treatments

	(1)	(2)	(3)
Dependent variable:		Abstain	
Increased transparency	0.721***	0.734***	0.932***
	(0.248)	(0.255)	(0.294)
Socio-Demographics			
Age		0.019	0.001
		(0.032)	(0.040)
Female		0.174	0.240
		(0.289)	(0.337)
Education			
Education level		0.264	0.408
		(0.244)	(0.294)
Subject: Empirical science		0.016	0.174
		(0.316)	(0.355)
Subject: Humanities and others		0.137	0.320
		(0.307)	(0.346)
Number of semester		0.009	0.013
		(0.048)	(0.053)
Personal characteristics			
Big5: extraversion			-0.258
			(0.215)
Big5: agreeableness			-0.037
			(0.222)
Big5: openness			0.022
			(0.129)
Big5: neuroticism			0.437**
			(0.220)
Big5: conscientiousness			-0.127
			(0.186)
Trust			0.042
			(0.077)
Locus of control			-0.605*
			(0.342)
Information preferences			-0.076
			(0.096)
Risk aversion			0.025
			(0.068)
Motivation			
Money			-0.180
			(0.285)
Research			-0.202
			(0.230)
Constant	-1.124***	-1.856**	1.376
	(0.191)	(0.778)	(2.638)
Observations	136	136	136
Pseudo R^2	0.059	0.087	0.162

Notes: Standard errors are in parentheses. Probability-likelihood models are used. For more information on the variables used, see Table 6.1. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E.2.3 Abstaining behavior in the no-information treatments

	(1)	(2)	(3)
Dependent variable:		Abstain	
Involuntary-signal	0.475*	0.511*	0.550*
	(0.257)	(0.276)	(0.314)
Socio-Demographics			
Age		0.043	0.040
		(0.040)	(0.044)
Female		0.102	-0.019
		(0.316)	(0.359)
Education			
Education level		0.047	0.057
		(0.266)	(0.327)
Subject: Empirical science		-0.587	-1.132**
		(0.376)	(0.526)
Subject: Humanities and others		-0.521	-0.885**
		(0.343)	(0.429)
Number of semester		0.067	0.136**
		(0.054)	(0.066)
Personal characteristics			
Big5: extraversion			-0.132
			(0.267)
Big5: agreeableness			-0.396
			(0.277)
Big5: openness			-0.086
			(0.134)
Big5: neuroticism			0.587***
			(0.216)
Big5: conscientiousness			0.154
			(0.253)
Trust			0.051
			(0.084)
Locus of control			-0.276
			(0.300)
Information preferences			0.043
			(0.091)
Risk aversion			-0.015
			(0.086)
Motivation			
Money			0.503
			(0.340)
Research			-0.204
			(0.259)
Constant	-1.124***	-2.284**	-2.204
	(0.191)	(0.923)	(2.552)
Observations	131	128	128
Pseudo R^2	0.027	0.107	0.187

Notes: Standard errors are in parentheses. Probability-likelihood models are used. For more information on the variables used, see Table 6.1. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E.2.4 Sample statistics – Deviant behavior

	Treatments					
	(1)	(2)	(3)	(4)	(5)	(6)
Specification	Mean	Mean	Mean	Mean	Min	Max
Deviant behavior: others						
Fare dodging	2.94	2.97	2.87	2.94	1	4
Stealing	1.87	1.79	1.94	1.88	1	4
Streaming	3.38	3.37	3.23	3.27	1	4
Cheating in exam	2.39	2.36	2.31	2.41	1	4
Drunk driving	1.52	1.30	1.37	1.59	1	4
Deviant behavior: hypothetical						
Fare dodging	0.49	0.48	0.31	0.41	0	1
Stealing	0.10	0.04	0.06	0.05	0	1
Streaming	0.64	0.70	0.63	0.62	0	1
Cheating in exam	0.30	0.13	0.27	0.23	0	1
Drunk driving	0.10	0.03	0.13	0.06	0	1
Deviant behavior: myself						
Fare dodging	1.09	0.75	0.44	0.74	0	5
Stealing	0.09	0.04	0.11	0.09	0	5
Streaming	1.51	1.66	1.26	1.59	0	5
Cheating in exam	0.06	0.09	0.16	0.11	0	5
Drunk driving	0.12	0.06	0.32	0.18	0	5
N	69	67	62	66		

Notes: Means by treatments are displayed in columns 1, 2, 3, and 4. Minimum and maximum values are reported in columns 5 and 6. 1: voluntary-signal-no-information; 2: voluntary-signal-increased-transparency; 3: involuntary-signal-no-information; 4: involuntary-signal-increased-transparency.
The wording of the questions is as follows:
I) "How bad do you think it is when someone …
- … uses public transportation without a valid ticket?"
- … in a department store or store takes goods without paying?"
- … illegally streams a series or a movie on the internet?"
- … cheats in an exam?"
- … drives a car under the influence of alcohol (blood alcohol level of 0.5 or above)?"
II) Would it be a viable option for you…
- … to use public transportation without a valid ticket?"
- … to pick up goods in a department store or store without paying?"
- … to stream a series or a movie illegally on the internet?"
- … to cheat in an exam?"
- … to drive a car under the influence of alcohol (blood alcohol level of 0.5 or above)?"
III) "Have you…
- … used public transportation in the last 12 months without a valid ticket?"
- … taken anything from a department store or store in the last 12 months without paying?"
- … illegally streamed a series or movie on the internet in the last 12 months?"
- … cheated on an exam in the last 12 months?"
- … been driving a car under the influence of alcohol (blood alcohol level of 0.5) in the last 12 months?"
The response scale for the questions listed under I) ranges from "Very bad" (score: 1) to "Not bad at all" (score: 4).
The response options for the questions listed under II) are "Yes" (score: 1) and "No" (score: 0).
The response scale for the questions listed under III) ranges from "No" (score: 0) to "Yes, more than four times" (score: 5).

Table E.2.5 Experimental time frame and page views to the data protection declaration

	(1) Page views	(2) Dev page views	(3) Page views	(4) Dev page views	(5) Page views	(6) Dev page views
Accesses survey	0.024	-0.005				
	(0.139)	(0.132)				
Experimental time frame			1.696	0.093	2.077	0.205
			(1.459)	(1.352)	(1.274)	(1.221)
Newsletter t-1					3.038	-0.730
					(2.604)	(2.494)
Newsletter t					8.872***	5.793**
					(2.982)	(2.857)
Newsletter t+1					14.179***	11.725***
					(2.968)	(2.843)
Newsletter t+2					-4.487	-6.513**
					(2.968)	(2.843)
Observations	21	21	90	90	90	90
R^2	0.002	0.000	0.015	0.000	0.303	0.244

Notes: Standard errors are in parentheses. Standard OLS Regressions are used. In columns 1, 3, and 5, the number of page views per day are used as a dependent metric variable. In columns 2, 4, and 6, the deviation from the weekly mean is used as a dependent metric variable. Page views of the data protection website of the research institute are counted via software. In column 1 and 2 the 21 days are analyzed when the survey was accessible. In column 3, 4, 5, and 6 the time span from December 2018 is analyzed, including the time span of the experimental data gathering. The Newsletter variable indicates the monthly day the institute shares a newsletter with new developments in the area of labor law. t-1 indicates the day before this newsletter mail. t+1(t+2) indicates the day (two days) after the newsletter. For a graphical presentation see Figure E.2.1. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E.2.6 Measurements for data and time

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment:	Voluntary-signal-no-information		Voluntary-signal-increased-transparency		Involuntary-signal-no-information		Involuntary-signal-increased-transparency			
Specification:	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Min	Max
Panel A										
Value of data	76.38	10	105.72	10	22.85	10	323.06	10	0	9999
Value of time	4.91	5	4.12	3	4.26	4.5	4.10	5	0	50
Sensitivity of data	5.49	6	5.52	6	5.68	6	5.91	6	1	10
N	69	69	67	67	62	62	66	66		
Panel B										
Value of data 99%	76.38	10	61.86	10	22.85	10	20.68	10	0	1000
N	69	69	66	66	62	62	64	64		
Panel C										
Time spent	570.35	505	2289.67	509.5	552.39	526	577.98	521	266	116196
N	68	68	66	66	61	61	65	65		
Panel D										
Time spent 98%	543.49	498	541.48	509.5	557.13	526	558.57	521	272	1960
N	67	67	64	64	60	60	63	63		

Notes: Means by treatments are displayed in columns 1, 3, 5, and 7. Medians by treatments are displayed in columns 2, 4, 6, and 8. Minimum and maximum values are reported in columns 9 and 10. The variables value of time and data in Panel A are from the survey and answers to the questions "Assuming you received compensation for the data you provided as part of this study (your time spent on this study), how much would you consider to be fair?". Answers are in hypothetical Euro. The values for the variable sensitivity of data come from responses to the question "How sensitive do you rate the data that you provided in the course of this questionnaire?" Responses ranging from 1 ("Very low sensitivity") to 10 ("Very high sensitivity"). For the variable value of data 99% in Panel B 3 three outliers are excluded from the variable value of data shown in Panel A. For the time spent metric variable measured in seconds in Panel C administrative data is used. 4 observations have been dropped due to technical reasons. For the variable time spent 98% in Panel D six observations representing the top and bottom 1% are excluded from the variable time spent shown in Panel C.

Table E.2.7 Abstaining behavior, treatments, and time spent on each survey page

	(1)	(2)	(3)	(4)
Time spent in seconds on each survey page:		Abstain		
1. Treatment	-4.33	-0.16	-4.58	-13.04
	(4.421)	(18.179)	(7.604)	(25.558)
2. Welcome page	-13.81*	-16.82	1.82	3.02
	(7.524)	(17.835)	(3.997)	(4.267)
3. Socio-demographics	-0.79	11.40**	-4.49	6.60
	(4.697)	(5.770)	(6.162)	(4.979)
4. Big5	-17.29	1.64	-8.29	13.82
	(16.924)	(7.167)	(6.873)	(9.935)
5. Risk aversion, information avoidance and trust	-8.63	-3.36	-3.60	-8.33
	(11.727)	(5.051)	(5.340)	(6.961)
6. Locus of control	-31.92	-0.07	-12.62	-0.15
	(37.860)	(12.759)	(12.618)	(6.787)
7. Deviant behavior: fare dodging	-9.38	-3.96	-13.67*	12.78*
	(11.998)	(7.064)	(8.392)	(7.016)
8. Deviant behavior: stealing	-5.30	7.05*	-2.20	-1.23
	(5.882)	(3.975)	(1.775)	(2.324)
9. Deviant behavior: streaming	-6.33	-1.87	-2.69	-0.58
	(4.301)	(7.468)	(3.194)	(2.324)
10. Deviant behavior: cheating	-3.66	0.74	-0.20	-2.47
	(2.904)	(1.585)	(2.736)	(2.977)
11. Deviant behavior: drunk driving	-6.19	-0.13	1.18	15.47**
	(10.225)	(1.259)	(3.582)	(7.620)
12. Sensitivity of data	-14.11	4.54	-2.16	7.17
	(16.877)	(2.779)	(4.583)	(8.932)
13. Motivation	-3.29	-4.45	-13.22*	-0.90
	(5.107)	(6.431)	(7.081)	(4.269)
14. Stated value of data and time	-1.42	-11.96	-11.73	5.50
	(10.775)	(16.775)	(27.316)	(6.833)
15. Lottery decision	2.96*	6.23*	23.31**	3.54
	(1.698)	(3.671)	(10.336)	(2.936)
Observations	67	64	60	63

Notes: Standard errors are in parentheses. Separate standard OLS regressions are used for each coefficient. In columns 1, 2, 3, and 4. differences between abstainers and revealers and time spent in seconds on each survey page are reported. 1: voluntary-signal-no-information; 2: voluntary-signal-increased-transparency; 3: involuntary-signal-no-information; 4: involuntary-signal-increased-transparency. Six observations representing the top and bottom 1% are excluded. For more information on the variable time spent 98% used as the independent variable for each survey page see Table E.2.6, Panel D. Levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7 Concluding remarks and outlook

This thesis contributes, in the first part, to the better understanding of teamwork and how different determinants in- or decrease team performance. The results presented in Chapters 2, 3, and 4 reveal that although the investigated determinants do not affect team performance at the aggregated level, substantial heterogeneity effects between individuals' contributions to the team are at work in all three studies. These three heterogeneities and the underlying mechanisms analyzed are essential for further research as well as real-world teams at the workplace.

In Chapter 2, two co-workers' dishonest behavior does not lead to a decline in performance either of the entire team nor the bystander in particular. However, distinguishing between bystanders with higher preferences for inconvenient information and bystanders with lower preferences toward inconvenient information reveals that there are two different response types on average. Whereas individuals with higher preferences for inconvenient information reduce performance after witnessing the potential fraudulent action, the more information avoiding types put significantly more effort into the task. Additional survey evidence hints towards the fact that this increase in performance results from a strategy to distract themselves from the cheating incidence by working harder. Thus, the bystander's contribution to the team output relies very much on her or his attitudes towards dealing with the behavior of the residual work team. This result adds an important finding to the literature on cheating and information (Benistant et al. 2021) as well as honesty at the workplace (Losey et al. 2007).

Chapter 3 starts with a similar non-finding on the aggregated level, where the increase in team size does not affect team performance. However, focusing on individual performance offers the insight that peer effects are able to counteract potential free-rider effects emerging in larger teams when high-skilled peers are a part of the team. Whereas vice versa in smaller teams, the possible negative peer pressure of high-skilled peers prevents a higher aggregated team performance here. This is an essential finding for team research in general and sheds light on

the more profound understanding of peer effects, a common phenomenon discussed in group research, but what is so far still puzzling (Reyniers 2018).

Additionally, in Chapter 4, when observing the effect of perceived attractiveness on performance in same and opposite-sex constellations on an aggregated level, the appeal of others within the team does not have an effect. Nevertheless, focusing on individuals with one opposite-sex co-worker reveals that perceived attractiveness has a negative impact when it comes to performance, which is still noticeable in teams with two opposite-sex co-workers. Our survey evidence implies that not mating activities are the reason for this effect, but especially for males, an increase in emotional arousal. Thus, we contribute to a better understanding of gender effects in a team context and the perception of other co-workers and possible effects on one's own performance.

Coming to limitations of this first part of this thesis, all studies share the short-run characteristic. This is especially important when aiming to distinguish between behavioral factors which persist long-term and others which have a more impulsive and short-term nature. We address this issue within surveys by asking participants about their future motivation, such as the willingness to work in the same team again at a later point in time. However, subjective expectations about the future are only reliable to a certain degree and may change when it comes to actual behavior in the future. Thus, to gather substantial evidence, long-term studies are required in further research projects to analyze long-term consequences, such as for the on average non-effect of dishonesty on team performance. Moreover, the relatively small incentives compared to real-world earnings lead to whether results are applicable for actual workplace behavior. Having said this, taking a closer look at the observed behavior of the participants at the workstation, there are good reasons to believe that participants put in real effort in order to receive the incentive and take the entire study seriously. Task behavior was not dominated by free-riding, which is also reflected in survey response on an item regarding putting effort into the work as well as the rules established at the beginning of each session, such as the prohibition of oral communication, were taken seriously in all sessions at all times. Nevertheless, to test whether studies with greater incentives validate this, thesis' findings would be essential.

Chapters 5 and 6 of this thesis contribute to a better understanding of technology implementation within fundraising and the impact of the GDPR on privacy concerns within a research context. Both main findings reveal avoidance behavior of individuals when facing the treatment changes. Besides these outcomes, especially the underlying mechanisms are intriguing.

In Chapter 5, visitors abstain from using the CPO for donations in all conditions. Though, for the subgroup of core visitors, the results hint that those are exploiting a moral wiggle room in the sense of Dana et al. (2007) in order to avoid giving in cash when no CPO is offered. This opportunity is taken away when the CPO is offered and the group of core visitors starts to contribute more, however still in cash. Nevertheless, a precise investigation of this hypothesis is required to allow a better understanding of this plausible mechanism. Further, especially for research on fundraising due to the covid-19 crisis and its still not foreseeable long-term effects, it must be questioned whether the acceptance of a CPO will rise due to a possible large boost of digital payment technologies in the time of the pandemic.

In Chapter 6, implementing a high transparency data processing declaration leads to increased privacy concerns expressed through the relatively low percentage of participants who share personal data with the researcher. However, this effect is only found when the motivation to implement the GDPR is signaled to be voluntary and not extrinsically forced. Thus, more research is needed to uncover whether the first or the latter case is the dominant pattern when it comes to privacy decision-making. Also intriguing are the mechanisms that drive decision-making when data processing transparency is relatively low. Here participants base their decision-making, to exchange personal data for a lottery ticket, on the time they invest into the survey. This observation adds an important finding to the literature, which aims to understand how individuals evaluate their data and privacy (Acquisti et al. 2016).

Furthermore, the results of all five studies and their future validity in real-world settings depend largely on the ongoing digitalization development. Questions, such as to which extent human behavior in a virtual environment remains unchanged compared with behavior in physical interactions, are key for the research field of (virtual) group behavior (e.g. Graff et al. 2021), as well as (online) fundraising and data protection. Nonetheless, to pave the way for further research, this thesis offers important findings.

References

Abeler, J., Becker, A., & Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics, 113*, 96-104.

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica, 87*(4), 1115-1153.

Ackfeld, V., & Güth, W. (2019). Personal information disclosure under competition for benefits: Is sharing caring?. *MPI Collective Goods Discussion Paper 2019/4*.

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature, 54*(2), 442-492.

Adjerid, I., Acquisti, A., Brandimarte, L., & Loewenstein, G. (2013). Sleights of privacy: Framing, disclosures, and the limits of transparency. In *Proceedings of the ninth Symposium on usable Privacy and Security*, 1-11.

Adjerid, I., Peer, E., & Acquisti, A. (2018). Beyond the privacy paradox: Objective versus relative risk in privacy decision making. *Management Information Systems Quarterly, 42*(2), 465-488.

Aïmeur, E., & Sahnoune, Z. (2020). Privacy, trust, and manipulation in online relationships. *Journal of Technology in Human Services, 38*(2), 159-183.

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

Andreoni, J., & Petrie, R. (2008). Beauty, gender and stereotypes: Evidence from laboratory experiments. *Journal of Economic Psychology, 29*(1), 73-93.

Andreoni, J., Rao, J., & Trachtman, H. (2017). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of Political Economy, 125*(3), 625-653.

Andries, M., & Haddad, V. (2020). Information aversion. *Journal of Political Economy, 128*(5), 1901-1939.

Apple Inc. (2019). *Privacy policy*. Retrieved August 21, 2020, from https://www.apple.com/legal/privacy/en-ww/.

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *Review of Economic Studies, 76*(2), 451-469.

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

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.

Backes-Gellner, U., Arndt, W., & Mohnen, A. (2015). Effort provision in entrepreneurial teams: Effects of team size, free-riding and peer pressure. *Journal of Business & Economic Statistics, 85*(3), 205-230.

Baert, S., & Decuypere, L. (2014). Better sexy than flexy? A lab experiment assessing the impact of perceived attractiveness and personality traits on hiring decisions. *Applied Economics Letters, 21*(9), 597-601.

Bäker, A., & Mechtel, M. (2019). The impact of peer presence on cheating. *Economic Inquiry, 57*(2), 792-812.

Bäker, A., & Mertins, V. (2013). Risk-sorting and preference for team piece rates. *Journal of Economic Psychology, 34*, 285-300.

Bandiera, O., Barankay, I., & Rasul, I. (2010). Social incentives in the workplace. *Review of Economic Studies, 77*(2), 417-458.

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

Barclays (2015). *The Barclays Way*. Retrieved April 06, 2020, from https://home.barclays/content/dam/home-barclays/documents/citizenship/our-reporting-and-policy-positions/policy-positions/The-Barclays-Way.pdf.

Bar-El, R., & Tobol, Y. (2017). Honesty toward the holy day. *Journal of Behavioral and Experimental Economics*, *68*, 13-17.

Bartling, B., & Von Siemens, F. (2011). Wage inequality and team production: An experimental analysis. *Journal of Economic Psychology*, *32*(1), 1-16.

Batikas, M., Bechtold, S., Kretschmer, T., & Peukert, C. (2020). European privacy law and global markets for data. *Available at SSRN No. 14475.*

Battisti, M. (2017). High wage workers and high wage peers. *Labour Economics*, *46*, 47-63.

Baumeister, R. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, *46*(3), 610-620.

Bazerman, M., & Sezer, O. (2016). Bounded awareness: Implications for ethical decision making. *Organizational Behavior and Human Decision Processes*, *136*, 95-105.

Bekkers, R., & Wiepking, P. (2011). A literature review of empirical studies of philanthropy: Eight mechanisms that drive charitable giving. *Nonprofit and Voluntary Sector Quarterly*, *40*(5), 924-973.

Bénabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, *117*(3), 871-915.

Ben-Hafaïedh, C., & Cooney, T. (Eds.). (2017). *Research handbook on entrepreneurial teams: Theory and practice*. Edward Elgar Publishing, Cheltenham.

Benistant, J., Galeotti, F., & Villeval, M. (2021). The distinct impact of information and incentives on cheating. *IZA Discussion Paper No. 14014.*

Benndorf, V., & Normann, H. (2018). The willingness to sell personal data. *The Scandinavian Journal of Economics*, *120*(4), 1260-1278.

Beugnot, J., Fortin, B., Lacroix, G., & Villeval, M. (2019). Gender and peer effects on performance in social networks. *European Economic Review*, *113*, 207-224.

Bodenschatz, A., & Walkowitz, G. (2020). Slipping on stereotypes – Gender effects in the erosion of ethical behavior. *Available at SSRN No. 3728551*.

Boenigk, S., & Mayr, M. (2016). The happiness of giving: Evidence from the German socioeconomic panel that happier people are more generous. *Journal of Happiness Studies*, *17*(5), 1825-1846.

Bonein, A. (2018). Peer pressure and social comparisons with heterogeneous ability. *Managerial and Decision Economics*, *39*(2), 142-157.

Bonhomme, S. (2021). Teams: Heterogeneity, sorting, and complementarity. *University of Chicago, Becker Friedman Institute for Economics Working Paper, (2021-15)*.

Bradford, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press, New York.

Bradler, C., Dur, R., Neckermann, S., & Non, A. (2016). Employee recognition and performance: A field experiment. *Management Science*, *62*(11), 3085-3099.

Brady, R., Insler, M., & Rahman, A. (2017). Bad company: Understanding negative peer effects in college achievement. *European Economic Review*, *98*, 144-168.

Brandimarte, L., Acquisti, A., & Loewenstein, G. (2013). Misplaced confidences: Privacy and the control paradox. *Social Psychological and Personality Science*, *4*(3), 340-347.

Brandts, J., & Charness, G. (2003). Truth or consequences: An experiment. *Management Science*, *49*(1), 116-130.

Breuer, C., Hüffmeier, J., & Hertel, G. (2016). Does trust matter more in virtual teams? A meta-analysis of trust and team effectiveness considering virtuality and documentation as moderators. *Journal of Applied Psychology*, *101*(8), 1151-1177.

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.

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

Cadsby, C., Du, N., & Song, F. (2016). In-group favoritism and moral decision-making. *Journal of Economic Behavior & Organization*, *128*, 59-71.

Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, *102*(6), 2981-3003.

Carpenter, J., & Gong, E. (2016). Motivating agents: How much does the mission matter?. *Journal of Labor Economics*, *34*(1), 211-236.

Carpenter, J., Bowles, S., Gintis, H., & Hwang, S. (2009). Strong reciprocity and team production: Theory and evidence. *Journal of Economic Behavior & Organization*, *71*(2), 221-232.

Carpenter, J. (2007). Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, *60*(1), 31-51.

Cecere, G., Le Guel, F., Manant, M., & Soulié, N. (2017). The economics of privacy. In Vernengo, M., Caldentey, E., & Rosser, B. (Eds.), *The New Palgrave Dictionary of Economics*, Palgrave Macmillan, Basingstoke, 1-11.

Chadi, A., & Homolka, K. (2020a). Little lies and blind eyes – Experimental evidence on cheating and task performance in work groups. *Unpublished manuscript.*

Chadi, A., & Homolka, K. (2020b). Under (peer) pressure: Experimental evidence on team size and task performance. *Unpublished manuscript.*

Chadi, A., Jeworrek, S., & Mertins, V. (2017). When the meaning of work has disappeared: Experimental evidence on employees' performance and emotions. *Management Science*, *63*(6), 1696-1707.

Chai, S., Das, S., & Rao, H. (2011). Factors affecting bloggers' knowledge sharing: An investigation across gender. *Journal of Management Information Systems*, *28*(3), 309-342.

Chen, R., & Gong, J. (2018). Can self selection create high-performing teams?. J*ournal of Economic Behavior & Organization*, *148*, 20-33.

Childs, J. (2012). Gender differences in lying. *Economics Letters*, *114*(2), 147-149.

Chiocchio, F., Kelloway, E., & Hobbs, B. (Eds.). (2015). *The psychology and management of project teams*. Oxford University Press, New York.

Clark, A., & Oswald, A. (1996). Satisfaction and comparison income. *Journal of Public Economics*, *61*(3), 359-381.

Clark, A., Masclet, D., & Villeval, M. (2010). Effort and comparison income: Experimental and survey evidence. *Industrial & Labor Relations Review*, *63*(3), 407-426.

Cloos, J., Frank, B., Kampenhuber, L., Karam, S., Luong, N., Möller, D., Monge-Larrain, M., Tan Dat, N., Nilgen, M., & Rössler, C. (2019). Is your privacy for sale? An experiment on the willingness to reveal sensitive information. *Games*, *10*(3), 28.

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.

Conrads, J., Irlenbusch, B., Rilke, R., & Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology*, *34*, 1-7.

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

Corgnet, B., Hernán-González, R., & Schniter, E. (2015b). Why real leisure really matters: Incentive effects on real effort in the laboratory. *Experimental Economics*, *18*(2), 284-301.

Corgnet, B., Hernán-González, R., & Rassenti, S. (2015c). Firing threats: Incentive effects and impression management. *Games and Economic Behavior*, *91*, 97-113.

Cornelissen, T., Dustmann, C., & Schönberg, U. (2017). Peer effects in the workplace. *American Economic Review, 107*(2), 425-456.

d'Adda, G., Gao, Y., Golman, R., & Tavoni, M. (2018). It's so hot in here: Information avoidance, moral wiggle room, and high air conditioning usage. *Working Paper*.

Daimler (2011). *Richtlinie für integres Verhalten. Was uns gemeinsam bei Daimler leitet. Unsere Verhaltensgrundsätze und Leitlinien für das Handeln*. Retrieved April 06, 2020, from https://docmaster.supplier.daimler.com/DMPublic/de/doc/ALD00000629.2017-04-03.DE.pdf.

Dana, J., Weber, R., & Xi Kuang, J. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory, 33*(1), 67-80.

De Jong, B., Dirks, K., & Gillespie, N. (2016). Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *Journal of Applied Psychology, 101*(8), 1134-1150.

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

Delfgaauw, J., Dur, R., & Souverijn, M. (2020). Team incentives, task assignment, and performance: A field experiment. *The Leadership Quarterly, 31*(3), 101241.

DellaVigna, S., List, J., & Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics, 127*(1), 1-56.

Diekmann, A. (2007). *Empirische Sozialforschung: Grundlagen, Methoden,* Anwendungen (*18th edition*). Rowohlts Enzyklopädie, Hamburg.

Dohmen, T. (2008). Do professionals choke under pressure? *Journal of Economic Behavior & Organization, 65*(3-4), 636-653.

Dohmen, T., & Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review, 101*(2), 556-590.

Doorley, K., & Sierminska, E. (2015). Myth or fact? The beauty premium across the wage distribution in Germany. *Economics Letters, 129*, 29-34.

Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters, 99*(1), 197-199.

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

Ederer, F., & Fehr, E. (2017). Deception and incentives: How dishonesty undermines effort provision. *Working Paper*.

EDPS – European Data Protection Supervisor (2020). *A preliminary opinion on data protection and scientific research*. Retrieved August 21, 2020, from https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf.

Eil, D., & Rao, J. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics, 3*(2), 114-138.

Eisenkopf, G., Gurtoviy, R., & Utikal, V. (2017). Punishment motives for small and big lies. *Journal of Economics & Management Strategy, 26*(2), 484-498.

Ekman, P. (2009). *Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised edition)*. WW Norton & Company, New York.

Emerson, J., & Hill, B. (2018). Peer effects in marathon racing: The role of pace setters. *Labour Economics, 52*, 74-82.

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 Series No. 6903*.

Erat, S., & Gneezy, U. (2012). White lies. *Management Science, 58*(4), 723-733.

European Commission (2018a). *Ethics and data protection*. Retrieved August 21, 2020, from https://ec.europa.eu/info/sites/info/files/5._h2020_ethics_and_data_protection_0.pdf.

European Commission (2018b). *Reform of EU data protection rules*. Retrieved August 21, 2020, from https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules/eu-data-protection-rules_en.

Ewers, M., & Zimmermann, F. (2015). Image and misreporting. *Journal of the European Economic Association*, *13*(2), 363-380.

Ezquerra, L., Kolev, G., & Rodriguez-Lara, I. (2018). Gender differences in cheating: Loss vs. gain framing. *Economics Letters*, *163*, 46-49.

Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, *24*(1), 39-57.

Falk, A., & Zimmermann, F. (2017). Beliefs and utility: Experimental evidence on preferences for information. *Working Paper*.

Fehrler, S., & Kosfeld, M. (2014). Pro-social missions and worker motivation: An experimental study. *Journal of Economic Behavior & Organization*, *100*, 99-110.

Feldhaus, C., Sobotta, T., & Werner, P. (2019). Norm uncertainty and voluntary payments in the field. *Management Science*, *65*(4), 1855-1866.

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise – An experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525-547.

Fischer, M., Rilke, R., & Yurtoglu, B. (2020). Two field experiments on self-selection, collaboration intensity, and team performance. *IZA Discussion Paper No. 13201*.

Fosgaard, T., Hansen, L., & Piovesan, M. (2013). Separating will from grace: An experiment on conformity and awareness in cheating. *Journal of Economic Behavior & Organization*, *93*, 279-284.

Freddi, E. (2020). Do people avoid morally relevant information? Evidence from the refugee crisis. *Review of Economics and Statistics*, 1-45.

Frick, B., Goetzen, U., & Simmons, R. (2013). The hidden costs of high-performance work practices: Evidence from a large German steel company. *Industrial & Labor Relations Review*, *66*(1), 198-224.

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.

Gaffney, S., & Francis, C. (2009). *Honesty Sells: How to Make More Money and Increase Business Profits*. John Wiley & Sons, Hoboken.

Ganguly, A., & Tasoff, J. (2017). Fantasy and dread: The demand for information and the consumption utility of the future. *Management Science*, *63*(12), 4037-4060.

Garbarino, E., Slonim, R., & Villeval, M. (2018). A method to estimate mean lying rates and their full distribution. *Journal of the Economic Science Association*, *4*(2), 136-150.

Garbarino, E., Slonim, R., & Villeval, M. (2019). Loss aversion and lying behavior. *Journal of Economic Behavior & Organization*, *158*, 379-393.

Gawn, G., & Innes, R. (2018). Do lies erode trust?. *International Economic Review*, *59*(1), 137-161.

GDPR (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).* Retrieved August 21, 2020, from https://eur-lex.europa.eu/eli/reg/2016/679/oj.

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.

Gerber, N., Gerber, P., & Volkamer, M. (2018). Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers & Security*, *77*, 226-261.

Gerhards, L. (2015). The incentive effects of missions – Evidence from experiments with NGO employees and students. *European Economic Review*, *79*, 252-262.
Gerhards, L., & Kosfeld, M. (2020). I (don't) like you! But who cares? Gender differences in same-sex and mixed-sex teams. *Economic Journal*, *130*(627), 716-739.
Gerhards, L., & Siemer, N. (2016). The impact of private and public feedback on worker performance – Evidence from the lab. *Economic Inquiry*, *54*(2), 1188-1201.
GGSS (2000). German general social survey. *GESIS – Leibniz-institute for the social sciences.*
Gill, D., Prowse, V., & Vlassopoulos, M. (2013). Cheating in the workplace: An experimental study of the impact of bonuses and productivity. *Journal of Economic Behavior & Organization*, *96*, 120-134.
Gino, F., & Pierce, L. (2009). Dishonesty in the name of equity. *Psychological Science*, *20*(9), 1153-1160.
Gino, F., Ayal, S., & Ariely, D. (2013a). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior & Organization*, *93*, 285-292.
Gino, F., Krupka, E., & Weber, R. (2013b). License to cheat: Voluntary regulation and ethical behavior. *Management Science*, *59*(10), 2187-2203.
Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, *95*(1), 384-394.
Gneezy, A., Gneezy, U., Nelson, L., & Brown, A. (2010). Shared social responsibility: A field experiment in pay-what-you-want pricing and charitable giving. *Science*, *329*(5989), 325-327.
Gneezy, A., Gneezy, U., Riener, G., & Nelson, L. (2012). Pay-what-you-want, identity, and self-signaling in markets. *Proceedings of the National Academy of Sciences*, *109*(19), 7236-7240.
Gneezy, U., Keenan, E., & Gneezy, A. (2014). Avoiding overhead aversion in charity. *Science*, *346*(6209), 632-635.
Goldberg, S., Johnson, G., & Shriver, S. (2019). Regulating privacy online: An economic evaluation of the GDPR. *Available at SSRN No. 3421731*.
Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, *57*(1), 3-43.
Goldman Sachs (2020). *Code of business conduct and ethics*. Retrieved April 06, 2020, from https://www.goldmansachs.com/investor-relations/corporate-governance/corporate-governance-documents/revise-code-of-conduct.pdf.
Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information avoidance. *Journal of Economic Literature*, *55*(1), 96-135.
Google LLC (2020). *Google Privacy Policy*. Retrieved August 21, 2020, from https://policies.google.com/privacy?hl=en.
Graff, F., Grund, C., & Harbring, C. (2021). Competing on the holodeck – The effect of virtual peers and heterogeneity in dynamic tournaments. *Journal of Behavioral and Experimental Economics*, *90*, 101596.
Griffith, B., & Dunham, E. (2015). *Working in teams. Moving from high potential to high performance*. Sage Publications, London.
Grosch, K., & Rau, H. (2017). Gender differences in honesty: The role of social value orientation. *Journal of Economic Psychology*, *62*, 258-267.
Günther, C., Ekinci, N., Schwieren, C., & Strobel, M. (2010). Women can't jump? – An experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior & Organization*, *75*(3), 395-401.
Grossman, Z., & Van Der Weele, J. (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association*, *15*(1), 173-217.
Gylfason, H., Arnardottir, A., & Kristinsson, K. (2013). More on gender differences in lying. *Economics Letters*, *119*(1), 94-96.

Hamermesh, D., & Biddle, J. (1994). Beauty and the labor market. *American Economic Review, 84*(5), 1174-1194.

Hamilton, B., Nickerson, J., & 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.

Harrison, G., & List, J. (2004). Field experiments. *Journal of Economic Literature, 42*(4), 1009-1055.

Herbst, D., & Mas, A. (2015). Peer effects on worker output in the laboratory generalize to the field. *Science, 350*(6260), 545-549.

Hermstrüwer, Y., & Dickert, S. (2017). Sharing is daring: An experiment on consent, chilling effects and a salient privacy nudge. *International Review of Law and Economics, 51*, 38-49.

Hertwig, R., & Engel, C. (2016). Homo ignorans: Deliberately choosing not to know. *Perspectives on Psychological Science, 11*(3), 359-372.

Ho, E., Hagmann, D., & Loewenstein, G. (2021). Measuring information preferences. *Management Science, 67*(1), 126-145.

Hofmann, J., Piele, A., & Piele, C. (2020). Arbeiten in der Corona-Pandemie – Auf dem Weg zum New Normal. *Frauenhofer-Institut für Arbeitswissenschaft und Organisation in Kooperation mit der Deutschen Gesellschaft für Personalführung (DGFP), Stuttgart*.

Hoogendoorn, S., Oosterbeek, H., & Van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science, 59*(7), 1514-1528.

Holmström, B. (1982). Moral hazard in teams. *The Bell Journal of Economics 13*(2), 324-340.

Houser, D., Vetter, S., & Winter, J. (2012). Fairness and cheating. *European Economic Review, 56*(8), 1645-1655.

Huck, S., Szech, N., & Wenner, L. (2017). More effort with less pay: On information avoidance, optimistic beliefs, and performance. *WZB Discussion Paper No. SP II 2015-304r2*.

Institute of Fundraising. (2018). *Cash and digital payments in the new economy: Call for evidence response by the Institute of Fundraising*. Retrieved November 03, 2019, from https://www.institute-of-fundraising.org.uk/library/hm-treasury-call-for-evidence-on-cash-and-digital-payments-in/.

Innes, R., & Mitra, A. (2013). Is dishonesty contagious?. *Economic Inquiry, 51*(1), 722-734.

Isaac, M., & Walker, J. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics, 103*(1), 179-199.

Isaac, M., Walker, J., & Williams, A. (1994). Group size and the voluntary provision of public goods. Experimental evidence utilizing large groups. *Journal of Public Economics, 54*(1), 1-36.

Jacobsen, C., Fosgaard, T., & Pascual-Ezama, D. (2018). Why do we lie? A practical guide to the dishonesty literature. *Journal of Economic Surveys, 32*(2), 357-387.

Jia, J., Jin, G., & Wagman, L. (2020). GDPR and the localness of venture investment. *Available at SSRN No. 3436535*.

Johnson, G., Shriver, S., & Goldberg, S. (2020). Privacy & market concentration: Intended & unintended consequences of the GDPR. *Available at SSRN No. 3477686*.

Kajackaite, A. (2015). If I close my eyes, nobody will get hurt: The effect of ignorance on performance in a real-effort experiment. *Journal of Economic Behavior & Organization, 116*, 518-524.

Kajackaite, A. (2018). Lying about luck versus lying about performance. *Journal of Economic Behavior & Organization, 153*, 194-199.

Kandel, E., & Lazear, E. (1992). Peer pressure and partnerships. *Journal of Political Economy, 100*(4), 801-817.

Karlsson, N., Loewenstein, G., & Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty, 38*(2), 95-115.

Karremans, J., Verwijmeren, T., Pronk, T., & Reitsma, M. (2009). Interacting with women can impair men's cognitive functioning. *Journal of Experimental Social Psychology, 45*(4), 1041-1044.

Kesler, R., Kummer, M., & Schulte, P. (2020). Competition and privacy in online markets: Evidence from the mobile app industry. In *Academy of Management Proceedings* (Vol. 2020, No. 1, p. 20978). New York: Academy of Management.

Klee, E. (2008). How people pay: Evidence from grocery store data. *Journal of Monetary Economics, 55*(3), 526-541.

Kocher, M., Praxmarer, M., & Sutter, M. (2020). Team decision-making. In Zimmermann, K. (Eds.), *Handbook of Labor, Human Resources and Population Economics* (1-25). Springer International Publishing.

Kocher, M., Schudy, S., & Spantig, L. (2018). I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science, 64*(9), 3995-4008.

Kocher, M., Strauß, S., & Sutter, M. (2006). Individual or team decision-making – Causes and consequences of self-selection. *Games and Economic Behavior, 56*(2), 259-270.

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security, 64*, 122-134.

Kölle, F. (2017). Affirmative action, cooperation, and the willingness to work in teams. *Journal of Economic Psychology, 62*, 50-62.

Korbel, V. (2017). Do we lie in groups? An experimental evidence. *Applied Economics Letters, 24*(15), 1107-1111.

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

Kube, S., Maréchal, M., & Puppe, C. (2012). The currency of reciprocity: Gift exchange in the workplace. *American Economic Review, 102*(4), 1644-1662.

Kuhn, P., & Villeval, M. (2013). Are women more attracted to co-operation than men?. *The Economic Journal, 125*(582), 115-140.

Kvaløy, O., Nieken, P., & Schöttner, A. (2015). Hidden benefits of reward: A field experiment on motivation and monetary incentives. *European Economic Review, 76*, 188-199.

Landry, C., Lange, A., List, J., Price, M., & Rupp, N. (2010). Is a donor in hand better than two in the bush? Evidence from a natural field experiment. *American Economic Review, 100*(3), 958-983.

Lane, T. (2017). How does happiness relate to economic behaviour? A review of the literature. *Journal of Behavioral and Experimental Economics, 68*, 62-78.

Lazear, E. (1995). *Personnel Economics*. MIT Press, Cambridge.

Lazear, E. (1999). Globalisation and the market for team-mates. *The Economic Journal, 109*(454), 15-40.

Lazear, E., & Oyer, P. (2013). Personnel Economics. In Gibbons, R., & Roberts, J. (Eds.), *The Handbook of Organizational Economics* (479-519). Princeton University Press, Princeton.

López Bóo, F., Rossi, M., & Urzúa, S. (2013). The labor market return to an attractive face: Evidence from a field experiment. *Economics Letters, 118*(1), 170-172.

Losey, M., Meisinger, S., & Ulrich, D. (Eds.). (2007). *The future of human resource management: 64 thought leaders explore the critical HR issues of today and tomorrow*. John Wiley & Sons, Hoboken.

Lynn, M., Flynn, S., & Helion, C. (2013). Do consumers prefer round prices? Evidence from pay-what-you-want decisions and self-pumped gasoline purchases. *Journal of Economic Psychology, 36*, 96-102.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies, 60*(3), 531-542.

Mao, A., Winter, M., Siddharth, S., & Watts D. (2016). An experimental study of team size and performance on a complex task. *PloS one, 11*(4), e0153048.

Marreiros, H., Tonin, M., Vlassopoulos, M., & Schraefel, M. (2017). Now that you mention it: A survey experiment on information, inattention and online privacy. *Journal of Economic Behavior & Organization, 140*, 1-17.

Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review 99*(1), 112-145.

McCabe, K., Rigdon, M., & Smith, V. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization, 52*(2), 267-275.

Miller, S. (1987). Monitoring and blunting: Validation of a questionnaire to assess styles of information seeking under threat. *Journal of Personality and Social Psychology, 52*(2), 345-353.

Mobius, M., & Rosenblat, T. (2006). Why beauty matters. *American Economic Review, 96*(1), 222-235.

Momen, N., Hatamian, M., & Fritsch, L. (2019). Did app privacy improve after the GDPR?. *IEEE Security & Privacy, 17*(6), 10-20.

Mol, J., van der Heijden, E., & Potters, J. (2020). (Not) alone in the world: Cheating in the presence of a virtual observer. *Experimental Economics, 23*(4), 961-978.

Muehlheusser, G., Roider, A., & Wallmeier, N. (2015). Gender differences in honesty: Groups versus individuals. *Economics Letters, 128*, 25-29.

Müller, J., & Upmann, T. (2017). Eigenvalue productivity: Measurement of individual contributions in teams. *CESifo Working Paper Series No. 6679*.

Natter, M., & Kaufmann, K. (2015). Voluntary market payments: Underlying motives, success drivers and success potentials. *Journal of Behavioral and Experimental Economics, 57*, 149-157.

Newhouse, J. (1973). The economics of group practice. *Journal of Human Resources, 8*(1), 37-56.

Nulty, D. (2008). The adequacy of response rates to online and paper surveys: What can be done?. *Assessment & Evaluation in Higher Education, 33*(3), 301-314.

Ockenfels, A., Sliwka, D., & Werner, P. (2015). Bonus payments and reference point violations. *Management Science, 61*(7), 1496-1513.

Orhun, A. (2018). Perceived motives and reciprocity. *Games and Economic Behavior, 109*, 436-451.

Oster, E., Shoulson, I., & Dorsey, E. (2013). Optimal expectations and limited medical testing: Evidence from Huntington disease. *American Economic Review, 103*(2), 804-830.

Park, S., Nam, S., & Lee, J. (2017). Charitable giving, suggestion, and learning from others: Pay-what-you-want experiments at a coffee shop. *Journal of Behavioral and Experimental Economics, 66*, 16-22.

Parrett, M. (2015). Beauty and the feast: Examining the effect of beauty on earnings using restaurant tipping data. *Journal of Economic Psychology, 49*, 34-46.

Pfeifer, C. (2012). Physical attractiveness, employment and earnings. *Applied Economics Letters, 19*(6), 505-510.

Pierce, L., & Snyder, J. (2008). Ethical spillovers in firms: Evidence from vehicle emissions testing. *Management Science, 54*(11), 1891-1903.

Plesch, J., & Wolff, I. (2018). Personal-data disclosure in a field experiment: Evidence on explicit prices, political attitudes, and privacy preferences. *Games, 9*(2), 24.

Reuben, E., & Stephenson, M. (2013). Nobody likes a rat: On the willingness to report lies and the consequences thereof. *Journal of Economic Behavior & Organization, 93*, 384-391.

Reyniers, D. (2018). Peers and productivity: Evidence from an experimental factory. *Working Paper*.

Rooth, D. (2009). Obesity, attractiveness, and differential treatment in hiring a field experiment. *Journal of Human Resources, 44*(3), 710-735.

Rosenbaum, S., Billinger, S., & Stieglitz, N. (2014). Let's be honest: A review of experimental evidence of honesty and truth-telling. *Journal of Economic Psychology*, *45*, 181-196.

Ruffle, B., & Shtudiner, Z. (2015). Are good-looking people more employable?. *Management Science*, *61*(8), 1760-1776.

Salas, E., Tannenbaum, S., Cohen, D., & Latham, G. (Eds.). (2013). *Developing and enhancing teamwork in organizations. Evidence-based best practices and guidelines*. Jossey-Bass Publishers, San Francisco.

Schaar, K. (2016). What is important for data protection in science in the future? General and specific changes in data protection for scientific use resulting from the EU general data protection regulation. *RatSWD Working Paper No. 258.*

Schudy, S., & Utikal, V. (2017). 'You must not know about me' – On the willingness to share personal data. *Journal of Economic Behavior & Organization*, *141*, 1-13.

Schweitzer, M., & Hsee, C. (2002). Stretching the truth: Elastic justification and motivated communication of uncertain information. *Journal of Risk and Uncertainty*, *25*(2), 185-201.

Schwieren, C., & Weichselbaumer, D. (2010). Does competition enhance performance or cheating? A laboratory experiment. *Journal of Economic Psychology*, *31*(3), 241-253.

Shalvi, S., Soraperra, I., van der Weele, J., & Villeval, M. (2019). Shooting the messenger? Supply and demand in markets for willful ignorance. *Tinbergen Institute Discussion Paper No. 2019-071/I.*

Sicherman, N., Loewenstein, G., Seppi, D., & Utkus, S. (2016). Financial attention. *The Review of Financial Studies*, *29*(4), 863-897.

Sieber, J. (1988). Data sharing. *Law and Human Behavior*, *12*(2), 199-206.

Soetevent, A. (2011). Payment choice, image motivation and contributions to charity: Evidence from a field experiment. *American Economic Journal: Economic Policy*, *3*(1), 180-205.

Soraperra, I., Weisel, O., Kochavi, S., Leib, M., Shalev, H., & Shalvi, S. (2017). The bad consequences of teamwork. *Economics Letters*, *160*, 12-15.

Sørensen, J., & Kosta, S. (2019). Before and after GDPR: The changes in third party presence at public and private European websites. In *The World Wide Web Conference*, 1590-1600.

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

Straughan, C. (2019). *What does the future of Museum fundraising look like?*. Retrieved November 03, 2019, from https://www.museumnext.com/article/the-future-of-museum-fundraising/.

Sveriges Riksbank (2018). *Payment patterns in Sweden 2018.* Stockholm: Sveriges Riksbank.

Teubner, T., Hawlitschek, F., Adam, M., & Weinhardt, C. (2013). Social identity and reciprocity in online gift. In *2013 46th Hawaii International Conference on System Sciences*, 708-717.

Toussaert, S. (2017). Intention-based reciprocity and signaling of intentions. *Journal of Economic Behavior & Organization*, *137*, 132-144.

Tucker, C. (2014). Social networks, personalized advertising, and privacy controls. *Journal of Marketing Research*, *51*(5), 546-562.

Tucker, C. (2016). Privacy and the internet. In Anderson, S., Waldfogel, J., & Strömberg, D. (Eds.), *Handbook of Media Economics* (547-562). Elsevier BV, Amsterdam.

Twitter Inc (2020). *Twitter Privacy Policy*. Retrieved August 21, 2020, from https://twitter.com/en/privacy.

Van de Ven, J., & Villeval, M. (2015). Dishonesty under scrutiny. *Journal of the Economic Science Association*, *1*(1), 86-99.

Van Den Eynden, V. (2008). Sharing research data and confidentiality: Restrictions caused by deficient consent forms. *Research Ethics*, *4*(1), 37-38.

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.

Ward, D., & Beck, W. (1990). Gender and dishonesty. *The Journal of Social Psychology*, *130*(3), 333-339.

Wasmer, M., Koch, A., & Wiedenbeck, M. (1991). *Methodenbericht zur "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften"* (ALLBUS) 1990.

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences*, *112*(34), 10651-10656.

Wessels, N., Wagner, A., Prakash Sarswat, J., & Buxmann, P. (2019). *What is your selfie worth? A field study on individuals' valuation of personal data*. In *14th International Conference on Wirtschaftsinformatik*, 1322-1336.

Wiltermuth, S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes*, *115*(2), 157-168.

Wolter, M., Mönnig, A., Maier, T., Schneemann, C., Steeg, S., Weber, E., & Zika, G. (2021). Langfristige Folgen der Covid-19-Pandemie für Wirtschaft, Branchen und Berufe. *Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg. No. 202102.*

Yetter, G., & Capaccioli, K. (2010). Differences in responses to web and paper surveys among school professionals. *Behavior Research Methods*, *42*(1), 266-272.

Zhang, X., & Zhu, F. (2011). Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review*, *101*(4), 1601-1615.