Titel der Arbeit:

Three Essays on Unethical Behavior: The Role of Generalized Reciprocity, Discrimination and Norms

Schriftliche Promotionsleistung
zur Erlangung des akademischen Grades
Doctor rerum politicarum

vorgelegt und angenommen
an der Fakultät für Wirtschaftswissenschaft
der Otto-von-Guericke-Universität Magdeburg

Verfasser: Joschka Waibel
Geburtsdatum und -ort: 22.03.1991, Berlin
Arbeit eingereicht am: 16.05.2022

Gutachter der schriftlichen Promotionsleistung:
Jun.-Prof. Dr. Sabrina Jeworrek
Prof. Dr. Abdolkarim Sadrieh
Prof. Dr. Jeannette Brosig-Koch

Datum der Disputation: 27.02.2023

Contents

List of Tables

List of Figures

1. Introduction

Understanding human behavior in its entire complexity is an ambitious if not impossible challenge. It is however possible to study particular aspects of human behavior through experiments that allow us to isolate specific facets in the decision-making process, ultimately leading to a better understanding of human behavior as a whole. This thesis covers three experimental articles on unethical economic behavior and sheds light on the motives and circumstances that lead individuals to engage in these activities. Clearly, unethical behavior in all its different manifestations can pose great risk to society – both at the large (e.g. corporate tax evasion) and small (e.g. shoplifting) scale – making it a relevant topic to be studied in economic research.

Trying to understand unethical behavior through the lenses of traditional economic theory is problematic. Standard economic theory usually oversimplifies decision-making by assuming that people always take rational choices in a selfish-manner, irrespective of any normative considerations. However, this one-sided perspective on human behavior does not fully live up to what we observe and experience in daily life. Behavioral economic research enriched our understanding by acknowledging the relevance of apparent unselfish choices. Inspired by insights derived from behavioral science, economic experiments show that other-regarding preferences play a decisive role in the decision-making process (Cooper & Kagel, 2016). We observe that people deliberately repay kindness in a reciprocal manner and willingly help those in need as they donate money to charity. One way to conceptualize this apparent unselfishness is the idea that individuals tend to follow normative conventions, as they incorporate beliefs about what ought to be the right thing and act accordingly (Bicchieri, 2006).

Overall, economic decision-making should be understood as a process in which both selfish (and often unethical) as well as other-regarding motives are balanced. Further, it is key to recognize this individual weighting of opposing motives not necessarily as a set of fixed preferences but as a dynamic procedure that can be shaped by situational factors or time. In particular, people may exploit given information to readjust the normative evaluation of their actions. For instance, it is a well-known empirical finding that people show a lower hesitation to litter in public when the environment around them is already filled with litter (Bateson et al. 2013). One may further think about the role of emotions in decision-making (Drouvelis & Grosskopf, 2016). Building on own experiences gathered in life, we know that acting in fierce

anger or deep frustration can drive us towards decisions we would not take otherwise. These two simple examples illustrate nicely how malleable behavior is to the explicit situation we find ourselves in. In the following three chapters, I explore distinct decision-environments that open up the chance for decision-makers to justify – whether consciously or unconsciously – unethical behavior for their personal benefit.

Chapter 2 of this thesis covers joint work with Sabrina Jeworrek and is based on the research article Jeworrek and Waibel (2021a). Here, we investigate whether individuals are willing to blindly pass on experienced unkindness to innocent others. In contrast to a standard gift exchange setting, in which unkindness would be directly reciprocated to the perpetrator, we build on a concept known as "generalized negative reciprocity" (Strang et al, 2016). This concept looks at the specific situation in which the original source of unkindness is not a feasible target for retaliation. Instead, it examines whether individuals are willing to displace their negative gift on to an unrelated, innocent party. Given the steep hierarchal structures that characterize many workplaces and the naturally high hesitation to reciprocate back to superiors, we causally test the relevance of generalized reciprocity in an artificial workplace setting. In a lab experiment, we assigned subjects into groups of three, promoting one to be the group leader. In the first stage of the experiment, the group leader had the chance to treat her subordinates in an unkind way by openly reassigning them to work on a tedious instead of a joyful task. In the second stage and independent from their former group leader, both group members worked on a new real effort task and competed for a monetary bonus. Similar to Charness et al. (2014), both competitors could freely choose to behave unethically by either manipulating their own or by sabotaging their opponent's task score in order to increase their chance of winning. Our findings, however, do not support our main behavioral prediction. In comparison to the control group, subjects who experienced unkind leader treatment had a similar probability to cheat against their opponents in the subsequent tournament. With the help of a follow-up survey experiment, building on the norm-elicitation technique introduced by Krupka and Weber (2013), we confirm our lab findings. It shows that the moral appropriateness of cheating against someone who has suffered from the same superior mistreatment remained unaffected by the prior unkindness. However, in a second and slightly reframed workplace scenario, in which the counterpart was introduced as a new and unfamiliar coworker, the analysis shows that experienced mistreatment opened up some moral wiggle room to justify mild forms of cheating. Taken together, both experimental studies find limited support for the relevance of generalized negative reciprocity in our workplace setting while pronouncing the importance of the explicit social context for the emergence of the behavioral pattern.

Chapter 3 also deals with the behavioral consequences of mistreatment on subsequent decision-making. Here, I openly provide participants with an excuse that can be utilized as a justification to disobey a clear request to comply with a given rule. Clearly, the willingness to comply with rules is an essential societal pillar, guaranteeing a peaceful cohabitation in life. Hence, understanding why and under which circumstances individuals refuse to follow rules is of great interest. One basic legal principle in modern societies is the convention of "rule equality", ensuring that everyone plays by the same rules. In this chapter, I investigate whether an open violation of this convention can in turn undermine the willingness to comply. In the beginning of my online experiment, participants were first allocated into groups. In the main stage of the experiment, I introduced subjects to an unambiguous rule, prompting them to truthfully report private information. Inspired by the popular die rolling game (Fischbacher & Föllmi-Heusi, 2013), participants obtained their private information from three coin tosses executed in complete privacy. It was made clear that subjects who report three "heads" would receive a monetary price, while all other combinations resulted in zero payoff. Thus, everyone was free to disobey the rule and misreport in order to gain the monetary price. To causally test whether an unequal rule imposition within groups undermines the willingness to report truthfully, I randomly chose one participant in each group, who was publicly released from the original rule, while I still requested compliance from everyone else. In contradiction to my predictions, I find that unjustified rule discrimination had no impact on subsequent rule compliance. Furthermore, priorly induced group identities, aiming to increase the perception of equivalence among group members, did not change the results. Taken together, this project suggests that singular rule discrimination without severe monetary disadvantages seems to be insufficient to disrupt compliance across groups.

In contrast to Chapter 2 and 3, which analyze the effect of prior mistreatment on the decision to behave unethically, Chapter 4 takes a different perspective. In the last research article, which is again a joint project with Sabrina Jeworrek, we examine how large societal changes can affect the willingness to behave pro- or antisocial. In this paper, which is based on a revised version of Jeworrek and Waibel (2021b), we build on the well-established finding that social norms are important but also fragile drivers for normative behavior in society (Kimbrough & Vostroknutov, 2016; Bicchieri et al, 2022). Exploiting the rigorous restrictions introduced during the Covid-19 pandemic, we examine whether social distancing and the resulting experience of extensive social isolation, influenced people's perception of social norms and the willingness to comply with them. The paper draws on two online experiments. In the first, we elicited the normative beliefs on socially appropriate behavior in a Take-or-Give donation

game, a game in which subjects could freely take (=unethical behavior) or give (=ethical behavior) money from or to a popular charity organization. In the second experiment, we observed actual decision-making within the same game but with new participants, allowing us to draw conclusions on the willingness to comply with the underlying social norms. To derive causal insights on the question whether persistent social isolation during the pandemic shifted norm perceptions and compliance, we used the priming method (Cohn & Maréchal, 2016) to make subjects' personal lockdown and isolation memories salient. First, our results show that the normative expectations of appropriate behavior in the donation game did not change after recollecting individuals' social isolation memories. The second experiment on norm compliance, however, revealed a clear difference. Here, our findings indicate that salient social isolation memories led to a decline in prosocial choices as primed participants took larger amounts away from charity in order to increase their personal payoff. Two additional information treatments, in which we simply reminded participants about the applying norms in the game, successfully increased prosocial behavior once again.

Chapter 5 concludes this thesis and highlights the most important insights that can be drawn from the presented research articles.

References

Bateson, M., **Callow**, L., **Holmes**, J. R., **Redmond Roche**, M. L., & **Nettle**, D. (2013). Do images of 'watching eyes' induce behaviour that is more pro-social or more normative? A field experiment on littering. *PloS one*, 8(12), e82055.

Bicchieri, C. (2006). The grammar of society: The nature and dynamics of social norms. Cambridge University Press.

Bicchieri, C., **Dimant**, E., **Gächter**, S., & **Nosenzo**, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, 132, 59-72.

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

Cohn, A., & **Maréchal**, M. A. (2016). Priming in economics. *Current Opinion in Psychology*, 12, 17-21.

Cooper, D. J., & **Kagel**, J. H. (2016). Other-Regarding Preferences: A selective survey of experimental results. In J. H. Kagel and A. E. Roth (Eds), The Handbook of Experimental Economics, 217-290. Princeton University Press.

Drouvelis, M., & **Grosskopf**, B. (2016). The effects of induced emotions on pro-social behaviour. *Journal of Public Economics*, 134, 1-8.

Fischbacher, U., & **Föllmi-Heusi**, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3), 525-547.

Jeworrek, S., & **Waibel**, J. (2021a). Unethical employee behavior against coworkers following unkind management treatment: An experimental analysis. *Managerial and Decision Economics*, 42(5), 1220-1234.

Jeworrek, S., & **Waibel**, J. (2021b). Alone at home: The impact of social distancing on norm-consistent behavior. *IWH Discussion Papers* (8).

Kimbrough, E. O., & **Vostroknutov**, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3), 608-638.

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

Strang S., **Grote**, X., **Kuss**, K., **Park**, S. Q. & **Weber**, B. (2016). Generalized negative reciprocity in the dictator game – How to interrupt the chain of unfairness. *Scientific Reports*, 6:22316.

2. Unethical Employee Behavior Against Coworkers Following Unkind Management Treatment: An Experimental Analysis

We study unethical behavior toward unrelated coworkers as a response to managerial unkindness with two experiments. In our lab experiment, we do not find that subjects who experienced unkindness are more likely to cheat in a subsequent competition against another coworker who simultaneously experienced mistreatment. A subsequent survey experiment suggests that behavior in the lab can be explained by individuals' preferences for norm adherence, because unkind management behavior does not alter the perceived moral appropriateness of cheating. However, having no shared experience of managerial unkindness opens up some moral wiggle room for employees to misbehave at the costs of others.

This chapter is joint work with Sabrina Jeworrek and is based on the same-titled paper published in *Managerial and Decision Economics* in February 2021.

2.1 Introduction

Counterproductive work behavior hurts or is intended to hurt the organization or members of the organization, and its origin has been looked at from different theoretical perspectives (see Spector and Fox, 2002). One of these perspectives is the organizational justice approach. Skarlicki and Folger (1997) show that the response to perceived organizational unfairness goes beyond the classical employee satisfaction-performance relationship, focusing on retaliatory behaviors such as purposely damaged equipment. Thereby, organizational justice is decomposed into distributive, procedural and interactional justice. The latter is a subset of procedural justice, and it is of special interest since it deals with the question how supervisors treat their employees —which is relevant for all organizational levels and day-to-day work. However, given the hierarchical system of organizations and the thereof resulting fear of losing the job, directly or indirectly reciprocal behaviors may not be a feasible option for unkindly treated employees, which raises the question whether these disgruntlements can disrupt workplace behavior in a less obvious way. Bennett and Robinson (2000) refer to counterproductive work behaviors as workplace deviance and argue that an employee will choose from a set of deviant behaviors e.g. the most feasibly or least costly, depending on the situational context. If retaliation of the supervisor's behavior seems too costly but emotions caused by the supervisor's treatment motivate employees to engage in deviant behavior, they might choose an alternative target that is less powerful than the supervisor, even if this target is not responsible for the experienced mistreatment. Then, managerial unkindness may shape deviant behavior far beyond the known channel of direct reciprocity and impact decision-making in less obvious and indirect ways. Especially at workplaces where cooperation and personal interactions are of utmost importance, a contagious spread of unethical behavior from one employee to the other would become an imminent threat for the firm. A better understanding of how these behaviors move across personnel may inform employers and policy makers about regulatory needs and the adequate extent of costly administrative measures (i.e. employee monitoring) in such workplace environments. Hence, this study seeks to answer the question whether perceived interactional injustice will be passed on to unrelated coworkers if there is no possibility to hurt the original perpetrator.

To test whether unkind management behavior triggers counterproductive work behavior that purposely makes coworkers worse off, we first conducted a laboratory experiment in which participants were assigned to working groups of three, with one group leader. For each group, the computer allocated two distinct real effort tasks (counting zeros vs. rating jokes), which

differed regarding their perceived pleasantness. The computer allocated the unpleasant counting task either to the group leader or to the two remaining group members. Afterwards, the group leader was allowed to overrule the task allocation by switching e.g. from the counting to the more pleasant rating task. Since this particular reallocation is expected to trigger feelings of unkindness, we refer to it as the unkind treatment. After the first period, no further interactions between the group leader and the other two group members took place. The latter, however, received a new working task for which they could receive an additional monetary bonus if they performed better than the other group member. Here, subjects had the possibility to act unethical by manipulating the tournament. Our analysis reveals that more than half of the subjects (60%) chose to cheat. Comparing the treatment and control group, we do not observe a statistically significant difference, neither at the extensive nor the intensive margin.

The tendency to engage in counterproductive work behavior towards coworkers, however, might heavily depend on the personal relationship between those involved. The social identity theory (see, e.g., Tajfel & Turner, 1979; Akerlof & Kranton, 2000) suggests that individuals who belong to the same group develop preferences for one another which may alter decision making. Given that only knowing about identical painting preferences (i.e. Kandinsky vs. Klee, see Chen and Li, 2009) induces group identity, the shared experience of interactional injustice should establish an even stronger feeling of belonging to the same group but the empirical evidence for group identity and unethical behavior are rather mixed. Kato and Shu (2016), for example, show that Chinese workers tend to engage more frequently in aggressive competitive behavior when they face coworkers with a distinct group identity (urban vs. rural migrant workers). Benistant and Villeval (2019), however, revisit this question in a lab experiment and do not find evidence for an influence of shared group identity on sabotage in tournaments. Meub et al. (2016) test whether a shared group identity has an impact on participants' truth telling. Lying increased their own payoff while reducing the payoff of another participant. The authors find more honest reports when the deprived party shared the same group identity (fellow student vs experimenter). Similar results were obtained by Hermann and Ostermaier (2018) and Soraperra et al. (2019).

Given that people care about the social appropriateness of their actions (Krupka & Weber, 2013) but a sin does not always seem to be seen as a sin when motive and opportunity exist to cross social norms (Liu et al., 2014), one might hypothesize that supervisors' unkindness is more likely to affect the social appropriateness of counterproductive work behavior in the absence of a shared experience. Therefore, we conducted an online survey experiment with new

participants to evaluate the appropriateness of cheating in multiple hypothetical workplace scenarios closely related to our lab experiment. The scenarios build upon two employees in a fast food restaurant who compete for a monthly bonus. The winner is determined by the number of satisfied customers. For each customer who did not rate her service experience, the employee has the opportunity to leave a positive evaluation (i.e. the opportunity to cheat). Survey respondents were then asked to judge the moral appropriateness of cheating. Two scenarios replicated the lab experiment, framing a situation in which both employees previously suffered (or not) from unkind supervisor treatment. In line with our findings from the lab, we do not observe any differences between the control and treatment group so that the perceived social norm not to hurt a fellow in-group member is a likely candidate to explain subjects' observed behavior. In two additional scenarios, we varied the competition in that the competitor for the bonus is not the former colleague. In case of having no common history, moderate levels of cheating are indeed assessed to be less morally inappropriate after suffering from interactional injustice.

Taken together, by conducting two experimental studies that rule out both imitation behavior and retaliation motives, we show that employees do not blindly engage in counterproductive work behavior towards their coworkers after experiencing interactional injustice. Nevertheless, supervisors should not interpret this finding as a justification for not thinking about interactional justice since our data also indicate that experienced unkindness opens some moral wiggle room for unethical employee behavior, which might especially spread in workplaces characterized by loose and anonymous employee structures.

2.2 Related Literature

2.2.1 Generalized Reciprocity

One way to conceptualize the behavioral pattern of passing on perceived (un)kindness to an unrelated third party is called generalized reciprocity (Herne et al., 2013; Strang et al., 2016).[1] In recent years, an increasing number of scholars employing experimental methods to establish causal relationships started to take an interest in the concept of both positive and negative generalized reciprocity across different settings. Regarding positive generalized reciprocity, Stanca (2009) shows that people have a tendency to pass on perceived prosocial behavior in a gift-exchange lab experiment. Mujcic and Leibbrandt (2018) replicated this prosocial tendency in a large field experiment: People who were granted way in a big parking lot had a higher

[1] Krebs (1982) also refers to this pattern as "displaced reciprocity".

probability to grant way to other cars in a subsequent traffic interaction. Nowak and Sigmund (2005) describe generalized positive reciprocity as an important evolutionary prerequisite for human development. In contrast, Gray et al. (2014) examine the tendency to pass on unkind experiences to innocent others. Across multiple allocation experiments, the authors find that participants who received unkind shares are more likely to choose unkind allocations themselves. Whereas Schnedler and Stephan (2020) do not observe that unkindly treated individuals share less of their endowment with an anonymous receiver in a subsequent dictator game, Strang et al. (2016) highlight the role of negative emotions triggered in unkindly treated individuals. Offering individuals the chance to write a letter to the source of unkindness helped them to regulate their emotional state and significantly increased shares allocated to third parties. More closely related to the workplace is the laboratory experiment by Zdaniuk and Bobocel (2013) who show that only being confronted with the face of a fair or unfair leader had a significant impact on the interactional fairness of communicating a dismissal decision afterwards to an unrelated other. Somehow related to the idea of generalized reciprocity, Houser et al. (2012) find that after receiving an unfair share in a dictator game, participants have a higher probability to lie when they are asked for private information in a subsequent unrelated die-roll game. Della Valle and Ploner (2017) replicate this finding and argue that this behavior aims at restoring one's own subjective fairness balance. This suggests, however, that generalized negative reciprocity is not driven by the intention to hurt someone and, hence, perceived interactional injustice might not trigger counterproductive work behavior towards coworkers —especially since employees seem to care for the well-being of even unknown coworkers by punishing their employer with reduced performance after layoffs (Heinz et al., 2020).

2.2.2 Trickle Effects

Closely related to the concept of generalized reciprocity is the literature on the so-called trickle effects. In the context of organizational behavior, trickle effects describe how behavioral patterns, feelings or emotions are passed down or across hierarchal layers causing a contagious spread throughout entire companies (Masterson, 2001). Most studies examined how decisions originating at the top management level may ultimately interact with behavioral choices at lower hierarchical levels through the transmitting channel of mid-management personnel or intermediate supervisors (for a comprehensive review of such trickle effects in the workplace, see Wo et al., 2019). Multiple theoretical mechanisms have been proposed to conceptualize the decision-making process behind trickle effects in organizational behavior, with "Social Learning Theory" (Bandura, 1977; 1986), "Social Exchange Theory" (Blau, 1964) and

"Displaced Aggression" (Tedeschi and Norman, 1985; Marcus-Newhall et al., 2000) being the most popular ones.

The concept of social learning theory builds on the idea that individuals imitate the observed behavior of their role models (e.g. supervisors) and re-use these experiences as a guide for own supervising practices in future interactions. The social exchange theory is based on normative beliefs. Focusing on the concept of reciprocity, social exchange theory predicts that individuals who experience kind treatment from their supervisors may feel obliged to pass on kindness to others. The concept, however, faces some limitations in conceptualizing trickle effects given the absence of normative beliefs that stress the moral obligation to displace negative experiences towards innocent others. Whereas both approaches describe a cognitive-based behavioral mechanism, displaced aggression conceptualizes a more affect-driven behavior moderated by negative emotions (Wo et al., 2015, 2019). In particular, displaced aggression describes the tendency in human behavior to re-channel experienced frustration, anger or harmful behavior away from the originating source in fear of future consequences. Instead, victimized individuals may choose to unleash their negative emotions on unrelated others such as lower-ranked subordinates, fellow coworkers, or even close family members (Restubog et al., 2011; Hoobler and Hu, 2013), triggering a contagious spread of unkind behavior in interpersonal exchange. In a study based on survey data obtained from Chinese employees, Aryee et al. (2007) show that supervisors' perception of experienced interactional injustice trickles down to shape their own subordinates' perception of interactional injustice and in turn, undermined subordinates' workplace behavior in regard to both organization- and individual-oriented organizational citizenship. Their findings are in line with the concept of displaced aggression, noting that supervisors who are treated unkindly by their direct superiors, refrain from fighting back but rather re-channel their frustrations on to their own subordinates. Closely related to our research question is the study by Mawritz et al. (2012) which bases its empirical foundation on U.S. survey data. The authors' analysis confirms the findings by Aryee et al. (2007), showing that abusive leadership at the top-management level is positively associated with abusive supervisor behavior at the mid-management level. Moreover, their data reveal that subordinates who suffered from abuse are more likely to show deviant interpersonal behavior against other coworkers within their own work groups. The relationship between abusive supervision and interpersonal deviance, however, is moderated by the underlying work atmosphere, highlighting that personal ties among coworkers can serve as an effective buffer against displaced unkindness. Hence, their results emphasize the sensitivity of trickle effects with respect to social preferences and personal characteristics. By conducting two experimental

studies building one on another, we want to derive causal insights on the assumed relationship and therefore to address a frequently mentioned limitation discussed in the above-mentioned literature, whose empirical conclusions build almost exclusively on cross-sectional survey data (Wo et al., 2019). Furthermore, our controlled laboratory environment allows us to cleanly isolate behavioral spillovers from one individual to another, while eliminating concerns that decision-making is confounded by the individual desire to retaliate back the perpetrator through the interaction with third parties, a concern that cannot be neglected when relying on survey data.

2.3 Experimental Design

We conducted the laboratory experiment with a total of 249 student participants within nine sessions at the Magdeburg Experimental Laboratory of Economic Research (MaxLab) located at University of Magdeburg using "hroot" (Bock et al., 2014) for subject recruitment. The experimental design was executed in z-Tree (Fischbacher, 2007)[2]. On average, participants earned 11.60 EUR (including a 5 EUR show-up fee) for a one hour session. The experiment consisted of two stages and a follow-up questionnaire. Subjects were randomly allocated into groups of three with one subject as the group leader and two regular members.

Stage 1.— Each group had to work on two different real effort tasks but the allocation of tasks to the subjects was done by the computer, which was common knowledge. Participants were paid a fixed wage of 200 taler (with 100 taler = 1 EUR) and the working time was 10 minutes. Note that both the payment and the working time were identical for both tasks. We designed the tasks in such a way that they differed with regard to their perceived pleasantness (unpleasant vs. pleasant task), an idea similarly used by Gray et al. (2014) and Schnedler and Stephan (2020). To make the difference in pleasantness between the two tasks salient, all subjects received the instructions for both tasks and took part in 2-minute trial rounds.

One task was the *counting task* used by Abeler et al. (2011), in which subjects had to count the number of zeros within 15*15 matrices filled with randomly ordered zeros and ones.[3] It was crucial for our design that the counting tasks was perceived as being unpleasant. For this reason, we modified the version used by Abeler et al. (2011) and extended their original matrix size together with a reduction in font size. The task demands no specific skill-set but a substantial

[2] We are grateful for the provision of the z-Tree codes used in Gill and Prowse (2019), Benndorf et al. (2019), Murphy et al. (2011) and Abeler et al. (2011), which were publicly available.
[3] Screenshots and instructions (translated from German) are provided in the Appendix A.2

level of mental focus. As mentioned above we offered a fixed wage but in order to ensure effort exertion, we asked for a minimum number of solved matrices necessary to receive the offered fixed wage. In the second task, the so-called *rating task*, subjects were asked to read and rate jokes on a slider from 0 (not funny) to 100 (very funny). Contrary to the counting task, we intended subjects to enjoy reading the jokes so that we did not ask for a minimum number of rated jokes, the fixed wage was guaranteed. Comparing the characteristics of the two tasks, we assumed that the overwhelming majority of people would prefer to read jokes instead of engaging in a tedious counting activity.

After finishing the second trial period, all participants were informed about the task allocation within their group. We designed the program so that only two task allocation settings were realized.[4] In setting 1, the tedious counting task was assigned to the group leader while the two remaining group members were allocated the rating task. In setting 2, the allocation was reversed. Independent of the setting, the group leader learned that the allocated task would be her working task for stage 2 as well. Before starting the 10-minute working phase, we asked the group leader whether she wants to overrule the task allocation by exchanging her allocated tasks with the ones allocated to her group members. It was made clear to the group leader that there will be no further interactions between her and her group members in stage 2. This ensured free decision making, avoiding false believes about future dependencies and potential retaliation by the group members. Note that we allowed only for a swap between the settings 1 and 2, the leader could not exchange only one task with a particular group member. If the leader decided to switch task allocations, the group members were informed about the intervention, otherwise no information was shared. Since reallocating the group members' pleasant for the unpleasant task is expected to trigger feelings of unkindness, we refer to the task exchange in setting 1 as being our unkind treatment. Consequently, affected group members in setting 1 will represent the treatment group. The control group consists of members assigned to setting 2 if no task exchange took place. In short, our analysis compares group members who both worked on the tedious counting task but differ in the way they received this unpleasant task, either by the neutral computer program or an (unkind) personal decision made by the group leader.

Afterwards, the 10-minute working period started. At the end of this first stage, we asked subjects to self-report their emotions using an extended[5] version of the Positive and Negative

[4] To exclude deception concerns, we avoided the term randomization and told participants only that the computer automatically determined the distribution of tasks.

[5] We added the affects angry, happy and irritated to the original questionnaire since these emotions are likely to be affected by perceived unkind behavior.

Affect Schedule (PANAS) questionnaire (Watson et al., 1988) to assess potential differences in the emotional state of mind between the treatment and control group.

Stage 2.— In order to analyze generalized (instead of direct or indirect) negative reciprocity, it was clearly stressed to all subjects that the former group structure had been broken apart. The group members learned that their former group leader would now be working independently on an unrelated real effort task, without any further interactions or dependencies. As already mentioned, the group leader had to work on the previously allocated task once again, this time for 5 minutes and a fixed wage of 250 taler.

Given that the present study aims at investigating unethical behavior towards coworkers, we introduced a new real effort task. Similar to Dato and Nieken (2014) and Erkal et al. (2011), subjects were asked to encode words into numbers[6] for 5 minutes (after a 2-minute trial round) with a fixed wage of 250 taler. It was made clear that both former group members would be working on the encoding task, competing against each other for a bonus payment of another 250 taler. The winner of this tournament was determined by the number of points subjects had after finishing the task, with one point for each word that was correctly encoded. However, subjects had the possibility to modify the tournament outcome before its start in order to increase their probability of winning the monetary bonus. Similar to Charness et al. (2014), two modification options existed: subjects could either increase their own point score (redemption) or decrease the point score of their opponent (sabotage). [7] In contrast to Charness et al. (2014), in which redemption and sabotage were implemented in separate treatments, we allowed contestants to freely choose between both options. If subjects chose to increase their own score, they could buy up to 20 points at a cost of 10 taler for each point bought. If the sabotage option was chosen, subjects could reduce the point score of their opponent in the same size and at the same cost. Of course, subjects could refrain from any modification. Even though there is no payoff relevant difference between sabotage and redemption, we intentionally introduced both options simultaneously since different forms of unethical behavior are likely to bear different moral costs for individuals (Hermann & Mußhoff, 2019). Compared to cheating on one's own outcome, choosing to sabotage within the tournament clearly represents the intention to hurt the counterpart. By comparing modification choices within the treatment and control group, our

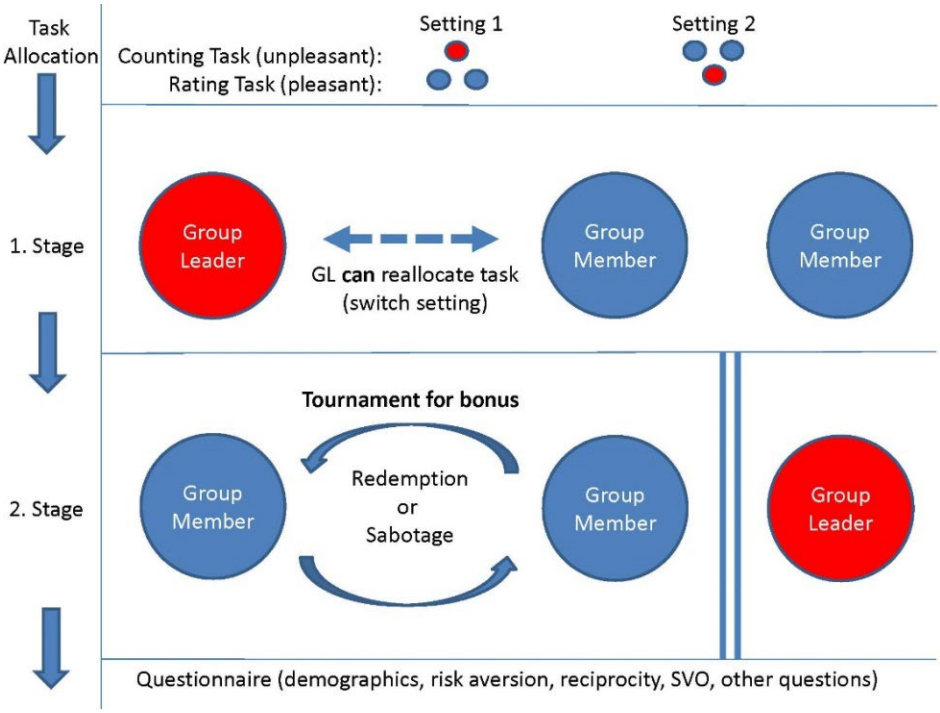
[6] For each word a two-digit number was assigned to a respective letter of the alphabet, displayed in an encryption table at the bottom of the screen. Each word and each specific encryption table was randomly created by the computer program (Benndorf, et al., 2019). The real effort task was chosen as it requires no specific skill set.
[7] We used a neutral wording for both sabotage and redemption in order to avoid any behavioral adjustments. See Appendix A.3 for exact wording.

design allows us to investigate whether this relative cost structure changes after experiencing the unkind treatment and subjects become more prone to hurt someone else directly. Following the decision on whether to cheat and if so, to what extent, the 5-minute working period started. Figure 1 provides an overview of the main experimental design features. It is important to highlight that our experiment examines workplace behavior across two distinct task dimensions. This ensures that modification choices in stage 2 are unlikely to be driven by imitation or learning experiences, derived from engaging in the allocation task of stage 1. Furthermore, by strictly separating group manager and group members after entering stage 2, we guarantee that differences in modification choices between treatment and control group cannot be explained by retaliation motives against the manager.

After finishing the working period in stage 2, we elicited subjects' beliefs regarding their opponent's modification behavior. If they correctly assessed their opponent's type of modification (or no modification at all) in a first step, subjects received a fixed payment of 150 taler. In a second step, they were asked about the extent of the opponent's modification.[8] To elicit beliefs as accurately as possible, we employed an incentive structure similar to Gächter and Renner (2010) where the payoff decreases quickly with less accurate beliefs. For a correct guess, subjects could earn 500 taler.[9] For all other guesses, subjects received 250 taler divided by the absolute estimation error. Finally, participants were asked to answer a short questionnaire which contained questions regarding the experiment (manipulation checks), some socio-demographic characteristics, risk attitudes and reciprocal behavior using the reduced 6-item version of the original scale developed by Perugini et al. (2003). Additionally, we elicited subject's social value orientation (SVO) (Murphy et al., 2011) and personality traits within a reduced form of the big five personality framework (Costa & McCrae, 1989).

[8] This step was omitted if they believed that there was no modification at all, and the second belief elicitation was not mentioned while selecting the first one in order to avoid hedging.
[9] As in Gächter and Renner (2010), we allowed for an error tolerance of +/-1 point.

Figure 1: The Stages of the Experimental Design



2.4 Results

2.4.1 Modification Choices

Due to the fact that our treatment and control groups were realized endogenously through choices made by the group leaders, we ended up with 60 regular group member subjects in the treatment group and 80 subjects in the control group. More than half of these subjects (60%) chose to cheat in the tournament to increase their winning chances. Comparing treatment and control, we do not observe a statistically significant difference, the share of subjects who decided to cheat is even slightly higher in the control than in the treatment group (61.25 vs. 58.33%, χ^2 (2) = 0.21, p = 0.899).[10][11] Similarly, the average number of modification points bought does not differ between control and treatment group (3.59 vs. 3.47, p = 0.737) [12][13]. Figure 2 illustrates subjects' modification choices including the differentiation between

[10] Table 3 in Appendix A.1 provides summary statistics for subjects' socio-demographic characteristics and personality traits. We observe a statistically significant difference as regards subjects' gender between control and treatment group. This difference, however, cannot explain the null result since we neither observe a treatment effect for females nor for males (both $p > 0.8$).

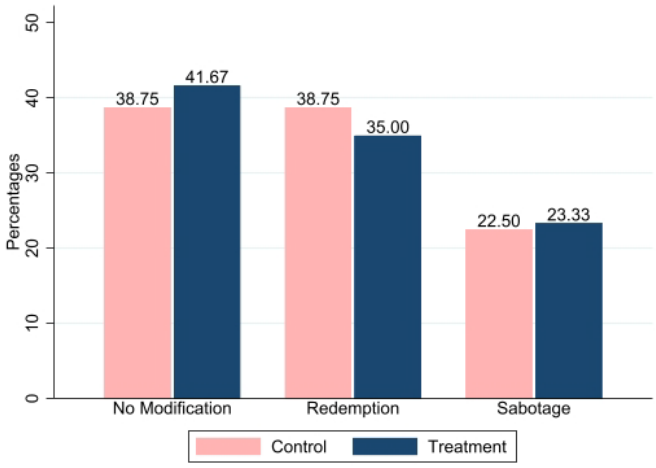
[11] To compare: 68.18% of the subjects who were allocated the rating task (N=22) and allowed to keep this task decided to modify their tournament outcome, whereas all subjects who were allocated the counting task and received the rating task from the group leader (N = 4) refrained from buying modification points.

[12] If not stated otherwise, the p-values are obtained from a two-sided Wilcoxon rank-sum test.

[13] To assess the significance of this null results, we calculated the statistical power given our sample size for three different hypothetical effect sizes (equal to a Cohen's d of 0.2, 0.5 and 0.8). We used the software G*Power (Faul et al., 2009) and set the p-value equal to 5%. Considering a low, a medium and a large treatment effect, the statistical power would be equal to 20%, 81% and 99%, respectively.

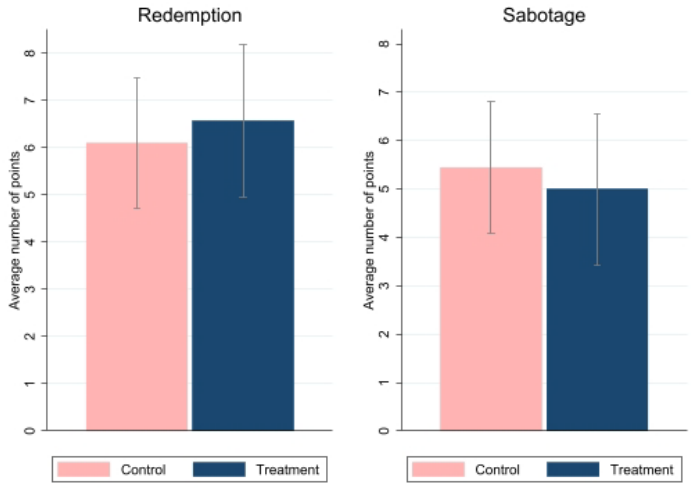
redemption and sabotage. In line with our suggestion that sabotage might induce higher moral costs, we find that subjects chose the redemption option more often[14] but again, there is no difference between treated and untreated subjects.

Figure 2: Modification Choices (Shares)



Looking at the intensive margin (see Figure 3), a similar picture emerges: conditioned on modifying the tournament via the redemption option, the control group bought on average 6.10 points while the treatment group invested only slightly more with an average of 6.57 points (p = 0.646). This difference is even smaller in case of the sabotage option (on average 5.44 vs. 5.00 points, p = 0.494).

Figure 3: Size of Redemption and Sabotage (Intensive Margin)



[14] A two-sided binomial test reveals a significant difference between sabotage choices (39.39%) and the 50% threshold (p = 0.038), indicating non-random modification choices.

Next, we address potential heterogeneity concerns. To do so, we pooled both modification options for this analysis since the relative moral cost for the two modification types do not seem to have changed. Moreover, observation numbers within the subgroups become rather small if we refrain from pooling the two modification types. Table 1 summarizes the average modification size[15] for different subgroups and differentiated according to control and treatment group. We observe a pattern that supports the suggestion that social preferences play a major role in the decision-making process with regard to unethical behavior (see, e.g. Grosch and Rau, 2020). First, based on the SVO questionnaire (Murphy et al., 2011), we classified subjects into pro-social and individualistic types. Subjects characterized by a higher tendency to act selfish (individualistic type) buy on average more modification points than more pro-social individuals ($p = 0.006$). Similarly, subjects splitted according to their negative reciprocal inclination, the average number of modification points bought is higher for subjects with a stronger tendency to respond to unkind behavior ($p = 0.005$). Risk preferences ($p = 0.437$) and the fact whether subjects passed the threshold in stage 1 to earn the fixed wage ($p = 0.891$) are not related to subjects' cheating behavior. Remarkably, there is consistently no treatment effect across all subgroups.

Table 1: Average Size of Modification for Different Subgroups

	All	*N*	Control	Treatment	*p*-value
1) Pro-Social-Type	2.78	86	2.74	2.83	0.527
Individualistic-Type	4.74	54	5.00	4.42	0.922
2) Low negative reciprocal inclination	2.62	69	2.82	2.37	0.520
High negative reciprocal inclination	4.42	71	4.32	4.57	0.839
3) Risk-loving	3.97	70	4.06	3.89	0.956
Risk-averse	3.10	70	3.26	2.78	0.380
4) Threshold passed	3.31	74	3.28	3.35	0.752
Threshold not passed	3.79	66	3.90	3.62	0.902

Note: The subgroups for risk preference and reciprocal inclination were determined through a median split.

2.4.2 Manipulation Checks

In order to rule out that this null effect is not specifically due to our experimental design, we have to check whether the group leader's behavior was perceived as unkind in first instance. Therefore, we have two pieces of evidence.

[15] Results separated into extensive and intensive margin can be found in Table 4 (Appendix A.1).

First, it is crucial that the two working tasks in stage 1 were indeed perceived as being very distinct with regard to their degree of pleasantness. We therefore asked participants how entertaining both tasks have been on a scale from 1 to 7 (with 7 equal to "very entertaining"). Due to the fact that every subject experienced both tasks at least through the 2-minute trail round, we feel confident to ask for the assessment of both tasks simultaneously. Subjects clearly assessed the rating task to be more entertaining than the counting task. For both the control (4.63 vs 2.13) and the treatment group (4.40 vs 2.30), this difference is statistically highly significant ($p < 0.001$, Wilcoxon sign rank test), whereas the assessment of the single task types does not differ across treatment and control group (rating task: $p = 0.425$; counting task: $p = 0.482$). Consequently, subjects should have preferred to work on the rating task. Hence, the prerequisite for triggering a feeling of unkindness through the group leader's reallocation decision is fulfilled.

Second, we directly asked subjects of the treatment group about their fairness[16] perceptions for the task allocations in stage 1. The two questions read "How fair do you think was the computer's task allocation in stage 1?" and "How fair do you think was the final task allocation?", both again on a scale from 1 to 7 (with 7 equal to "very fair"). Even though we cannot completely eliminate this concern, we placed the second question on a successive screen in order to reduce the experimenter demand effect. Taking the obtained results therefore with some care, subjects claim the final task allocation to be more unfair (2.85 vs. 4.25 with $p < 0.001$, Wilcoxon signed-rank test), which supports the previous finding on the difference in the perceived pleasantness of the two tasks.

The third piece of evidence is obtained from the modified PANAS questionnaire that was implemented between the stages 1 and 2 to evaluate subjects' emotional state of mind. On a scale from 1 (not at all) to 5 (extremely), subjects had to state to which extend the 23 items fitted to their current mood. Comparing outcomes between control and treatment group, we find a (weakly) significant difference for the negative trait *hostile* ($p = 0.065$).[17] Notably, when analyzing this item for the subgroups we have been looking at before, we find that there is no difference for individualistic types and subjects with a low reciprocal inclination ($p = 0.537$ and $p = 0.530$, respectively). The two groups, however, that are most likely to be sensitive to the

[16] Due to the fact that there is no fitting translation for "unkindness" in German, we asked for participant's assessment of perceived unfairness in the questionnaire. We claim that especially from a victim's perspective, unkindness and unfairness can be seen as sufficiently close.
[17] For the full list of the PANAS trait comparisons, please see Table 5 in Appendix A.1.

unkind reallocation of tasks, the pro-social and highly reciprocal individuals, show significantly higher values in case of being treated (1.64 vs. 1.30 with $p = 0.069$, and 1.93 vs. 1.56 with $p = 0.065$, respectively). These numbers may serve as further evidence that the task reallocation has been perceived as unkind.

2.4.3 Regression Analysis

Given that the group leader's reallocation decision did not trigger unethical behavior towards coworkers, we conducted a detailed regression analysis to control for potential differences between control and treatment group and to investigate other important determinants that might explain subjects' cheating behavior. We look both at subjects' binary decision to engage in modification activities using probit and the actual modification size using tobit models. As before, we continue to pool both modification options for the regression analysis and look at unethical behavior in general. Results are summarized in Table 2. With the exception of one specification only, the estimated treatment effect is slightly negative, but statistically far from being significant. Hence, using different models and an extensive set of controls, the regression analysis confirms our results obtained from the preceding non-parametric testing.

Looking at further possible determinants of subjects' cheating behavior, it obviously does not matter whether subjects passed the threshold in the counting task and earned the fixed wage. This suggests that even though subjects are competing for a monetary bonus, cheating behavior is more likely to be driven by non-monetary aspects. One of these aspects is the belief about the opponent's behavior. As observable in specifications (2) to (4), the more modification points a subject expects from her opponent (0 to 20), the more likely she is to engage in cheating, too. The same holds for the number of modification points bought, as shown in specifications (6) to (8). Given that the point estimate is below 1, however, we find that subjects do not perfectly adjust their own behavior towards their beliefs about their opponent's behavior. As already indicated in the context of the subgroup analysis, we find that individuals with a higher negative reciprocal inclination are about 20 percentage points more likely to cheat. However this effect almost disappears for the average number of modification points using the tobit model. We also tested for an interaction effect between being highly reciprocal and modification beliefs (detailed results are available upon request) but both groups behave nearly identical. Notably, it does not only matter whether subjects expected their opponents to cheat in general, they also adjusted the type of modification to their expectations: 69.62% of all subjects chose the modification option they also expected from their opponent. One might argue that subjects tried to justify their own behavior by stating that they expected the same behavior from their

opponent but due to the incentivized belief elicitation, this should be true, if at all, only for a minority of subjects.

Table 2: Main Regression Results

	Probit, dep. var.: Modification choice (0/1)			
	(1)	(2)	(3)	(4)
Treated	-0.029	-0.030	-0.009	0.007
	(0.042)	(0.043)	(0.047)	(0.047)
Threshold passed		0.042	0.024	0.035
		(0.085)	(0.084)	(0.085)
Rating task more fun		0.024*	0.022	0.025**
		(0.014)	(0.014)	(0.013)
Mod. belief (size)		0.024***	0.018**	0.019**
		(0.008)	(0.008)	(0.009)
Highly neg. reciprocal			0.184**	0.225**
			(0.086)	(0.101)
Individualistic type			0.111	0.072
			(0.093)	(0.089)
Risk loving			-0.097	-0.124*
			(0.082)	(0.069)
Constant				
Observations	140	140	140	140
Pseudo R^2	0.001	0.059	0.106	0.159
Additional Controls	NO	NO	NO	YES
	Tobit, dep. var.: Modification size (0-20)			
	(5)	(6)	(7)	(8)
Treated	-0.317	-0.357	-0.332	-0.168
	(0.884)	(0.744)	(0.837)	(0.829)
Threshold passed		0.269	-0.083	-0.012
		(0.935)	(0.955)	(1.060)
Rating task more fun		0.255	0.182	0.259
		(0.218)	(0.219)	(0.204)
Mod. belief (size)		0.647***	0.581***	0.628***
		(0.166)	(0.161)	(0.142)
Highly neg. reciprocal			2.077	2.229*
			(1.449)	(1.318)
Individualistic type			1.750*	1.082
			(1.054)	(1.185)
Risk loving			-0.328	-0.798
			(1.228)	(1.233)
Constant	1.808***	-2.367*	-3.261**	-1.305
	(0.653)	(1.268)	(1.325)	(4.794)
Observations	140	140	140	140
Pseudo R^2	0.0001	0.052	0.063	0.085
Additional Controls	NO	NO	NO	YES

Note: Additional controls include socio-demographic characteristics (gender, age, field of study, number of friends in current session) and the Big 5 personality traits. For the probit specifications, the table reports average marginal effects. Standard errors clustered on the session level in parentheses. Significance levels are denoted as follows: *** p<0.01, ** p<0.05, * p<0.1.

The remaining control variables have the expected signs but they are less robust across the different specifications. Additionally, controlling for some basic socio-demographic characteristics and the Big 5 personality traits in specifications (4) and (8) does not alter our main results.

2.4.4 Group Leader Behavior

Even though the group leaders' behavior is not of our main interest, their decision making was crucial for obtaining the observations of interest. Hence, we shortly also look at group leaders' behavior to finalize the data analysis. Out of 83 groups, 41 groups faced the computer generated task allocation of setting 1 (i.e. the group leader received the unfavorable counting task). 11 leaders refrained from exchanging tasks but only one assessed the counting task to be more entertaining than the rating task. Given that it was made clear that no further interactions with the other two group members would take place in stage 2, this behavior suggests altruistic preferences —nine of these group leaders are also classified as pro-social individuals based on the SVO questionnaire. In allocation setting 2, only 2 out of 42 individuals actively decided to undertake the unpleasant task, both individuals were also classified as pro-social individuals.

Within the final questionnaire, we additionally asked all group leaders "Please guess, how many group leaders would decide to exchange their counting for the rating tasks / their rating for the counting task". Even though the belief elicitation was not incentivized, the observed group leader behavior corresponds quite well to the predicted one. Leaders in setting 1 predicted a 75.27% exchange probability (vs. 73.17% actual exchange rate) and a share of 17.74% (vs. 4.76%) was predicted in setting 2. More importantly, we wanted to examine whether leaders have been aware of the impact that exchanging the tasks might have. Therefore we asked "Please put yourself into the position of your group member. Do you think your intervention in the task allocation was perceived as being fair or unfair?", on a scale ranging from 1 (very unfair) to 7 (very fair). We find that group leaders who exchanged the unpleasant for the pleasant tasks stated an average of 2.67. This is very close to the perceived level of unfairness stated by the treated subjects (i.e. 2.85) and hence, leaders seemed to be aware of their behavioral impact and still intentionally engaged in unkind behavior for their own benefit.

Taken together, the group leaders' behavior and questionnaire responses support the evidence obtained from the regular group members that subjects indeed perceived the rating task to be more favorable and that the exchange of tasks in setting 1 was an unkind action which was beneficial for the group leader but adverse for the group members. Hence, the question arises whether generalized negative reciprocity simply does not play any role in the underlying

context or whether our result is driven by the fact that both competing individuals were victims of the same unkind treatment. Hence, we invited a new group of university students to participate in an online survey experiment, which will be described in the next section.

2.5 Incentivized Online Survey Experiment

2.5.1 Experimental Design

In total, 206 survey participants were allocated to one out of four workplace scenarios which are closely related to our lab experiment. We used hroot (Bock et al., 2014) for recruitment and the survey was executed with SoSci Survey (Leiner, 2019). All scenarios build upon two fictional employees in a fast food restaurant who compete for a monthly bonus. The winner is determined by the number of satisfied customers. Given that both employees work their shifts separately and without any supervision, they could cheat at the cost of their absent coworker by leaving a positive evaluation for each customer who did not rate her service experience. The two *colleague* scenarios replicate the lab experiment, framing a situation in which both employees previously suffered (or not) from an unkind supervisor who has just left the company. Additionally, we conducted two *unknown coworker* scenarios in which we varied the competition in that the competitor for the bonus is not the former colleague but an unknown coworker who just joined the restaurant because the former colleague has also left.[18] Given our primary interest in the *unknown coworker* setting, we oversampled these two scenarios so that we end up with 81 observations for the *colleague* (in treatment, $N = 42$ and in control, $N = 39$) and with 125 observations for the *unknown coworker* (in treatment, $N = 64$ and in control, $N = 61$) scenarios.

After reading their randomly allocated scenario, survey respondents should judge the moral appropriateness of cheating. In order to prevent all participants from stating that cheating is morally completely inappropriate since they personally would never behave in such a way, we closely follow the experimental design proposed by Krupka and Weber (2013) and asked them about their beliefs on how the other participants judge the moral appropriateness of cheating. Given that the intensity of cheating might play a crucial role, participants had to judge different options the fictional employee had, such as no cheating at all, submitting a positive rating for every tenth, every fifth, every third, every other, or for every customer. For each possible option, the employee's behavior can be perceived as being "very morally inappropriate", "somewhat morally inappropriate", "somewhat morally appropriate", and "very morally appropriate". To

[18] Please note that we never used the term "colleague" or "unknown coworker" to avoid framing effects. For all details, see Appendix A.4.

incentivize participants to reveal their true beliefs, they received a payoff of 5 EUR in case of choosing the mode value for one randomly picked option, and no payoff otherwise.

At the end of the survey, we collected socio-demographic information on age, gender and the course of study. Additionally, we included a question regarding the participants' personal engagement in any kind of voluntary work, which will serve as a proxy for highly pro-social individuals (summary statistics can be found in the Appendix A. Table 6). To answer the survey completely, it took on average about 5 minutes.

2.5.2 Results

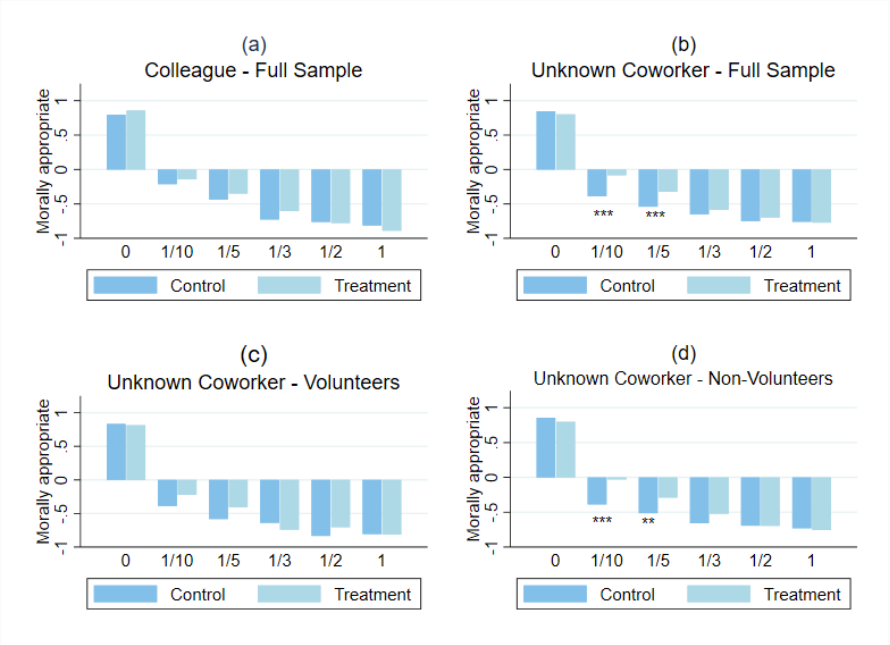
In general, our results (see Figure 4) show that the decision to refrain from any cheating is always assessed to be highly moral (high positive moral score) while the opposite is true for all five cheating options (negative moral score). In line with our findings from the lab experiment, Figure 4A indicates no difference between the treatment and control scenario across any cheating intensity within the *colleague* setting. Hence, experienced unkindness does not seem to justify cheating at the cost of a colleague who suffered from the same experience.

Figure 4B shows the results for the *unknown coworker* setting. Whereas there is again no difference for high levels of cheating, the experienced unkindness seems to justify at least minor forms of cheating against an unknown coworker. In case of submitting a positive rating for every tenth possible customer, the average scores of -0.39 in the control and -0.08 in the treatment scenario already express a clear difference ($p = 0.002$). That difference becomes even more obvious when comparing shares of participants who judged this kind of cheating to be morally appropriate since this was true for only 18.03% of participants in the control scenario and this number more than doubles (40.63%) in the treatment scenario. For the option of submitting a positive rating for every fifth customer, the difference is still highly significant ($p = 0.008$) but less pronounced, the shares of participants who judge that behavior to be morally appropriate goes down to 11.48% in the control and to 18.75% in the treatment scenario.

Given that moral perceptions might differ between individualistic and pro-social individuals, we split the sample into two subgroups, see Figure 4C and Figure 4D. For highly pro-social survey respondents, as identified by regular volunteering activities, we do not find any treatment effect. For non-volunteers, however, the pattern remains nearly unchanged compared to the pooled sample, with significant differences between treatment and control for the two lowest cheating intensities. This finding is also in line with the suggestion that individuals prone to a specific type of deviant workplace behavior, such as organizational deviance, are likely to

be different from individuals who are prone to interpersonal deviance (Bennett and Robinson, 2000).

Figure 4: Mean Moral Appropriateness Rating for Cheating



Note: Participants' answers were converted into numerical scores with "very morally inappropriate" corresponding to -1, "somewhat morally inappropriate" to -1/3, "somewhat morally appropriate" to +1/3 and "very morally appropriate" to +1. Significance levels are denoted as follows: *** p<0.01, ** p<0.05, * p<0.1.

2.6 Conclusion

By conducting a laboratory and an online survey experiment, we investigated whether individuals are (more) likely to engage in counterproductive work behavior towards coworkers after experiencing unkind supervisor treatment if the supervisor is no feasible target for retaliation. Our laboratory experiment reveals that workers who suffered from interactional managerial injustice do not show a higher probability in cheating activities at the cost of their coworkers. In our survey experiment, we implemented a setting comparable to our lab experiment to investigate the perceived moral appropriateness of cheating. In the case of having a joint experience of unkind group leader behavior, there is no difference between control and treatment group. Given previous evidence showing that shared (traumatic) experiences can bond and shift preferences between involved parties quite intensively (see, e.g., Jong et al., 2015; Whitehouse et al., 2017), we additionally analyze whether the common experience of victimization might drive individuals' inclination for unethical behavior against others. When introducing a previously unknown and unencumbered coworker into the hypothetical workplace scenarios, we indeed find that cheating is perceived to be less morally inappropriate after experiencing unkind supervisor behavior.

The latter finding is in line with Chang et al. (2019) who show that norm perceptions are sensitive across differently framed situations and may depend on individuals' social identity. Evidence stressing the importance of personal relationships among workers has been derived by Yang et al. (2013) who show that sharing a strong collective identity (e.g. within work groups) can moderate the positive relationship between experienced unfairness and counterproductive behavior against others. In regard to trickle effects of unethical behavior within organizations, Mawritz et al. (2012) support this narrative. Their results indicate that a sound and functioning work atmosphere among coworkers can buffer or even reverse the tendency to pass on frustration in the aftermath of unkind supervision. The data derived from our survey experiment clearly confirm this suggestion and further enrich our understanding of the complex relationship between other-regarding preferences and the emergence of unethical behavior in the workplace. In particular, our experimental approach allows us to contribute to the literature by offering causal insights on how personal ties among workforce members can moderate the tendency to displace frustrations against innocent others, while controlling for potential confounding factors such as indirect retaliation motives. In regard to the literature on generalized negative reciprocity, our results support the suggestion by Della Valle and Ploner (2017) who find that unethical choices in the aftermath of own unfair treatment seems to aim at restoring one's own subjective fairness balance rather than the explicit wish to hurt innocent others.

One might argue that the null result in the laboratory experiment might be explained by the fact that behavior in contests is simply driven by the intrinsic wish to win the contest price (Benistant & Villeval, 2019). Nevertheless, findings presented by Flory et al. (2016) and Harbring and Irlenbusch (2005) indicate that people take social preferences into account when they decide to cheat in a tournament. Even though the survey experiment is not directly comparable to the lab results, the difference between the colleague and unknown coworker setting support this suggestion. Given that many workplace structures rely on tournament environments to increase work effort (Lazear & Rosen, 1981; Bull et al., 1987; Harbring et al., 2007) and that competitive markets are known to be an ideal breeding ground for unethical behavior (Harbring & Irlenbusch, 2011; Cartwright & Menezes, 2014; Charness et al., 2014; Dato & Nieken, 2014, 2019), the investigation of this specific setting seems to be relevant. Nevertheless, our study definitely raises a number of questions for further research.

Besides calling for further studies that investigate non-tournament settings, it is important to point out that our analysis examines interactional injustice in a single one-shot scenario.

Whereas individuals might be able to deal quite well with managerial unkindness e.g. in case of having only a bad day, more severe or repeated mistreatment may increase one's personal frustration level and trigger an outburst of accumulated anger against third parties, following the notion of displaced aggression. Due to ethical concerns, however, it is unreasonable to push participants over their emotional edge and, hence, the experimental approach is not suited for investigating more severe forms of managerial unkindness. Blind and uncontrolled outbursts of accumulated frustrations, however, are probably also less likely than deliberately engaging in less obvious and more subtle forms of counterproductive behavior. In that regard, our survey experiment presents empirical evidence showing that experienced injustice on the job can indeed open up some moral wiggle room in which affected workers can justify moderate levels of cheating —which in turn could negatively affect perceived group identity and, hence, increase the danger of more serious forms of counterproductive work behaviors.

It is important to note that in contrast to our laboratory experiment, we do not to observe actual behavior in the survey study. However, Krupka and Weber (2013) and Chang et al. (2019) show that differences in social norm assessments, elicited through the incentivized approach used here, serve as a powerful predictor for actual decision making. Nevertheless, one might question whether subjects undergo the same emotional distress from reading a hypothetical workplace scenario than from actually experiencing it. Hence, it is an undisputed fact that more research on this topic is needed. Especially quasi-experimental environments could solve ethical issues while allowing scholars to derive deeper causal insights on the impact of perceived interactional unfairness on counterproductive work behavior towards coworkers.

Taken together, our study shows that even in a competitive workplace setting in which immoral behavior is incentivized, people seem to refrain from immediately lashing out against other group members. The moral appropriateness of deviant workplace behavior, however, seems to be dependent on the situational context. The risk of a contagious spread of unethical behavior might be especially prevalent in workplaces characterized by loose and anonymous employee structures such as warehouses with high employee turn-over (Min, 2007). Besides pointing to the relevance of interactional justice, our paper delivers an additional argument in favor of creating a functional working atmosphere beyond the traditional narrative of pushing labor productivity. Especially in contrast to costly monitoring measures which are known to bear the risk of triggering adverse employee reactions (Falk & Kosfeld, 2006), investments enhancing the personal relationship among employees (such as office remodeling or joint employee

activities) appear to be a powerful instrument to contain or at least buffer the spread of unkind behavior in the aftermath of (perceived) interactional injustice.

Undoubtedly, the displacement of personal frustrations and anger can also be observed outside the workplace. Card and Dahl (2011), for example, find a strong increase in domestic violence in areas in which the local football team unexpectedly lost. If displaced aggression is even more prevalent among individuals that are less closely connected, such trickle effects may help to explain broader societal upheavals such as the rise of xenophobia. Individuals may have a tendency to displace their own experienced frustrations on to minority groups and refrain from displaying the same degree of acrimony against otherwise similar fellow countrymen. Of course, the question of whether trickle effects indeed contribute to the rise of xenophobia is beyond the scope of this paper. However, we believe that our findings call for more extensive efforts to bring (apparently) dissimilar or unknown people closer together —whether on the job or in daily life.

References

Abeler, J., **Falk**, A., **Goette**, L. & **Huffman**, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2), 470-492.

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

Aryee, S., **Chen**, Z. X., **Sun**, L., & **Debrah**, Y. A. (2007). Antecedents and outcomes of abusive supervision: Test of a trickle-down model. *Journal of Applied Psychology*, 92, 191-201.

Bandura, A. (1977). Social learning theory. New York: General Learning Press.

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice Hall.

Benistant, J. & **Villeval**, M. C. (2019). Unethical behavior and group identity in contests. *Journal of Economic Psychology*, 72, 128-155.

Benndorf, V., **Rau**, H. A. & **Sölch**, C. (2019). Minimizing learning in repeated real-effort tasks. *Journal of Behavioral and Experimental Finance*, 22, 239-248.

Bennet, R. J. & **Robinson**, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology*, 85(3), 349-360.

Blau, P. (1964). Exchange and power in social life. New York: Wiley.

Bock, O.; **Nicklisch**, A.; **Baetge**, I. (2014). Hroot: Hamburg registration and organization online tool. *European Economic Review*, 71, 117–120.

Bull, C., **Schotter**, A. & **Weigelt**, K. (1987). Tournaments and piece rates: An experimental study. *Journal of Political Economy*, 95(1), 1-33.

Card, D. & **Dahl**, G. (2011). Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior. *Quarterly Journal of Economics*, 126(1), 103-143.

Cartwright, E. & **Menezes**, M. L. (2014). Cheating to win: Dishonesty and intensity of competition. *Economic Letters*, 122(1), 55-58.

Chang, D., **Chen**, R. & **Krupka**, E. (2019). Rhetoric matters: A social norms explanation for the anomaly of framing. *Games and Economic Behavior*, 166, 158-178.

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

Chen, Y. & **Li**, S. X. (2009). Group identity and social preferences. *The American Economic Review*, 99(1), 431-457.

Costa, P. T. & **McCrae**, R. R. (1989). The NEO-PI/NEO-FFI manual supplement. Odessa, FL.: Psychological Assessment Resources.

Dato, S. & **Nieken**, P. (2014). Gender differences in competition and sabotage. *Journal of Economic Behavior & Organization*, 100, 64-80.

Dato, S. & **Nieken**, P. (2019). Gender differences in sabotage: The role of uncertainty and beliefs. *Experimental Economics*,23, 353-391.

Della Valle, N. & **Ploner**, M. (2017). Reacting to unfairness: Group identity and dishonest behavior. *Games*, 8(3), 28.

Erkal, N., **Gangadhara**, L. & **Nikiforakis**, N. (2011). Relative earnings and giving in a real-effort experiment. *The American Economic Review*, 101(7), 3330-3348.

Falk, A., & **Kosfeld**, M. (2006). The hidden costs of control. *American Economic Review*, 96(5), 1611–1630.

Faul, F., **Erdfelder**, E., **Buchner**. A. & **Lang**, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149-1160.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 171–178.

Flory, J., **Leibbrandt**, A. & **List**, J. A. (2016). The effects of wage contracts on workplace misbehaviors: Evidence from a call center natural field experiment. *NBER Working Paper* 22342.

Gächter, S. & **Renner**, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, 13(3), 364–377.

Gill, D. & **Prowse**, V. (2019). Measuring costly effort using the slider task. *Journal of Behavioral and Experimental Finance*, 21, 1-9.

Gray, K., **Ward** , A. F. & **Norton**, M. I. (2014). Paying it forward: Generalized reciprocity and the limits of generosity. *Journal of Experimental Psychology: General*, 143(1), 247-254.

Grosch, K., & **Rau**, H. A. (2020). Procedural unfair wage differentials and their effects on unethical behavior. *Economic Inquiry*, 58(4), 1689-1706.

Harbring, C. & **Irlenbusch**, B. (2005). Incentives in tournaments with endogenous prize selection. *Journal of Institutional and Theoretical Economics*, 161(4), 636-663.

Harbring, C. & **Irlenbusch**, B. (2011). Sabotage in tournaments: Evidence from a laboratory experiment. *Management Science*, 57(4), 611-627.

Harbring, C., **Irlenbusch**, B., **Kräkel**, M. & **Selten**, R. (2007). Sabotage in corporate contests – an experimental analysis. *International Journal of the Economics of Business*, 14(3), 367-392.

Heinz, M., **Jeworrek**, S., **Mertins**, V., **Schumacher**, H., & **Sutter**, M. (2020). Measuring the indirect effects of adverse employer behaviour on worker productivity: A field experiment. The *Economic Journal*, 130(632), 2546-2568.

Hermann, D. & **Mußhoff**, O. (2019). I might be a liar, but I am not a thief: An experimental distinction between the moral costs of lying and stealing. *Journal of Economic Behavior & Organization*, 163, 135-139.

Hermann, D. & **Ostermaier**, A. (2018). Be close to me and I will be honest. How social distance influences honesty. *Center for European, Governance and Economic Development Research Discussion Papers* No. 340.

Herne, K., **Lappalainen**, O. & **Kestilä-Kekkonen**, E. (2013). Experimental comparison of direct, general, and indirect reciprocity. *The Journal of Socio-Economics*, 48, 38-46.

Hoobler, J., & **Hu**, J. (2013). A model of injustice, abusive supervision, and negative effect. *The Leadership Quarterly*, 24(1), 256-269.

Houser, D., **Vetter**, S. & **Winter**, J. (2012). Fairness and cheating. *European Economic Review*, 56(8), 1645-1655.

Jong, J., **Whitehouse**, H., **Kavanagh**, C. & **Lane**, J. (2015). Shared negative experiences lead to identity fusion via personal reflection. *PLoS ONE*, 10(12): e0145611.

Kato, T. & **Shu**, P. (2016). Competition and social identity in the workplace: Evidence from a Chinese textile firm. *Journal of Economic Behavior & Organization*, 131, 37-50.

Krebs, D. L. (1982). Prosocial behavior, equity, and justice. In J. Greenberg and R. L. Cohen (Eds) Equity and Justice in Social Behavior, 261-308. Academic Press.

Krupka, E. L., **Weber**, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3), 495-525.

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

Leiner, D. J. (2019). SoSci Survey (Version 3.1.06) [Computer software]. Available at https://www.soscisurvey.de.

Liu, Y., **Lu**, H., & **Veenstra**, K. (2014). Is sin always a sin? The interaction effect of social norms and financial incentives on market participants' behavior. *Accounting, Organizations and Society*, 39(4), 289-307.

Marcus-Newhall, A., **Pedersen**, W. C., **Carlson**, M. & **Miller**, N. (2000). Displaced aggression is alive and well: A meta-analytic review. *Journal of Personality and Social Psychology*, 78(4), 670-689.

Masterson, S. S. (2001). A trickle-down model of organizational justice: Relating employees' and customers' perceptions of and reactions to fairness. *Journal of Applied Psychology*, 86(4), 594-604.

Mawritz, M. B., **Mayer**, D. M., **Hoobler**, J. M., **Wayne**, S. J. & **Marinova**, S. V. (2012). A trickle-down model of abusive supervision. *Personnel Psychology*, 65(2), 325–357.

Meub, L., **Proeger**, T., **Schneider**, T. & **Bizer** K. (2016). The victim matters – experimental evidence on lying, moral costs and moral cleansing. *Applied Economics Letters,* 23(16), 1162–1167.

Min, H. (2007). Examining sources of warehouse employee turnover. *International Journal of Physical Distribution & Logistics Management,* 37(5), 375-388.

Mujcic, R. & **Leibbrandt**, A. (2018). Indirect reciprocity and prosocial behaviour: Evidence from a natural field experiment. *The Economic Journal*, 128(611), 1683-1699.

Murphy, R. O., **Ackermann**, K. A. & **Handgraaf**, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8), 771–781.

Nowak, M. A. & **Sigmund**, K. (2005). Evolution of indirect reciprocity. *Nature*, 423, 1291–1298.

Perugini, M., **Galucci**, M., **Presaghi**, F. & **Ercolani**, A. P. (2003). The personal norm of reciprocity. *European Journal of Personality*, 17(4), 251-283.

Restubog, S. L. D., **Scott**, K. L., & **Zagenczyk**, T. J. (2011). When distress hits home: The role of contextual factors and psychological distress in predicting employees' responses to abusive supervision. *Journal of Applied Psychology*, 96(4), 713-726.

Schnedler, W., & **Stephan**, N. L. (2020). Revisiting a remedy against chains of unkindness. *Schmalenbach Business Review*, 72(3), 347-364.

Skarlicki, D. P. & **Folger**, R. (1997). Retaliation in the workplace: The roles of distributive, procedural and interactional justice. *Journal of Applied Psychology*, 82(3), 434-443.

Soraperra, I., **Weisel**, O. & **Ploner**, M. (2019). Is the victim Max (Planck) or Moritz? How victim type and social value orientation affect dishonest behavior. *Journal of Behavioral Decision Making*, 32(2), 168–178.

Spector, P. E., & **Fox**, S. (2002). An emotion-centered model of voluntary work behavior: Some parallels between counterproductive work behavior and organizational citizenship behavior. *Human Resource Management Review*, 12(2), 269-292.

Stanca, L. (2009). Measuring indirect reciprocity: Whose back do we scratch? *Journal of Economic Psychology*, 30(2), 190–202.

Strang S., **Grote**, X., **Kuss**, K., **Park**, S. Q. & **Weber**, B. (2016). Generalized negative reciprocity in the dictator game – How to interrupt the chain of unfairness. *Scientific Reports*, 6:22316.

Tajfel, H. & **Turner**, J. (1979). An integrative theory of intergroup conflict. Monterey, CA: Brooks-Cole.

Tedeschi, J. T., & **Norman**, N. M. (1985). A social psychological interpretation of displaced aggression. *Advances in Group Processes*, 2, 29-56.

Watson, D., **Clark**, L. A. & **Tellegen**, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070.

Whitehouse, H., **Jong**, J., **Buhrmester**, M. D., **Gomez**, A., **Bastian**, B., **Kavanagh**, C. M., **Newson**, M., **Matthews**, M., **Lanman**, J., **McKay**, R. & **Gavrilets**, S. (2017). The evolution of extreme cooperation via shared dysphoria experiences. *Scientific Reports*, 7:44292.

Wo, D.X., **Schminke**, M. & **Ambrose**, M. L. (2019). Trickle-down, trickle-out, trickle-up, trickle-in, and trickle-around effects: An integrative perspective on indirect social influence phenomena. *Journal of Management*, 45(6), 2263-2292.

Wo, X. H. D., **Ambrose**, L. M., & **Schminke**, M. (2015). What drives trickle-down effects? A test of multiple mediation processes. *Academy of Management Journal*, 58(6), 1848-1868.

Yang, L., **Johnson**, R.E., **Zhang**, X., P.E., **Spector** & **Xu**, S. (2013). Relations of interpersonal unfairness with counterproductive work behavior: The moderating role of employee self-identity. *Journal of Business and Psychology*, 28, 189–202.

Zdaniuk, A. & **Bobocel**, D. R. (2013). The automatic activation of (un)fairness behavior in organizations. *Human Resource Management Review*, 23, 254-265.

Appendix A

A.1 Tables

Table 3: Subjects' Socio-demographics and Personality Traits

	Treatment	Control	*p*- value
Age	23.55	23.55	0.529
Female	0.433	0.700	0.024
First Degree	0.467	0.400	0.432
Friends	0.483	0.613	0.594
Risk	4.900	4.088	0.037
Individualistic Type	0.400	0.375	0.764
Reciprocal inclination (neg.)	2.994	2.950	0.893
Reciprocal inclination (pos.)	5.944	5.867	0.757
Openness	4.850	5.013	0.265
Conscientiousness	4.911	4.854	0.852
Extraversion	4.989	4.700	0.221
Agreeableness	5.322	5.329	0.790
Neuroticism	4.022	4.238	0.253
N	60	80	

Note: p-values are obtained from Pearson's χ^2 or two-sided Wilcoxon rank-sum tests, respectively.

Table 4: Modification Choices within Subgroups (Extensive and Intensive Margin)

	Extensive Margin				*Intensive Margin*			
	Control	Treat	*p*	*N*	Control	Treat	*p*	*N*
1) Pro-Social-Type	0.56	0.47	0.421	86	4.89	6.00	0.791	45
Individualistic-Type	0.70	0.75	0.684	54	7.14	5.89	0.488	39
2) Low neg. reciprocal inclination	0.51	0.43	0.512	69	5.50	5.46	0.924	33
High neg. reciprocal inclination	0.71	0.73	0.810	71	6.10	6.23	0.984	51
3) Risk-loving	0.58	0.59	0.873	70	7.05	6.55	0.850	41
Risk-averse	0.64	0.56	0.555	70	5.10	4.92	0.442	43
4) Threshold passed	0.63	0.56	0.563	74	5.24	6.00	0.677	44
Threshold not passed	0.60	0.62	0.901	66	6.50	5.88	0.626	40

Note: p-values are obtained from Pearson's χ^2 or two-sided Wilcoxon rank-sum test, respectively.

Table 5: PANAS Results

	Control	Treatment	*p*-value
Proud	2.48	2.72	0.224
Happy	2.54	2.58	0.729
Irritable	2.45	2.40	0.705
Enthusiastic	1.78	1.95	0.405
Ashamed	1.30	1.32	0.731
Angry	2.10	2.08	0.805
Alert	2.93	2.97	0.788
Nervous	1.61	1.68	0.733
Determined	2.85	2.92	0.788
Attentive	3.31	3.32	0.979
Jittery	1.73	1.63	0.789
Afraid	1.14	1.15	0.933
Distressed	1.59	1.45	0.458
Interested	2.58	2.55	0.877
Irritated	1.85	1.87	0.693
Excited	1.84	1.90	0.927
Strong	2.21	2.23	0.846
Envious	1.74	1.77	0.735
Guilty	1.19	1.10	0.236
Scared	1.26	1.25	0.724
Hostile	1.40	1.57	0.065
Inspired	2.09	2.07	0.935

Note: p-values are obtained from two-sided Wilcoxon rank-sum tests.

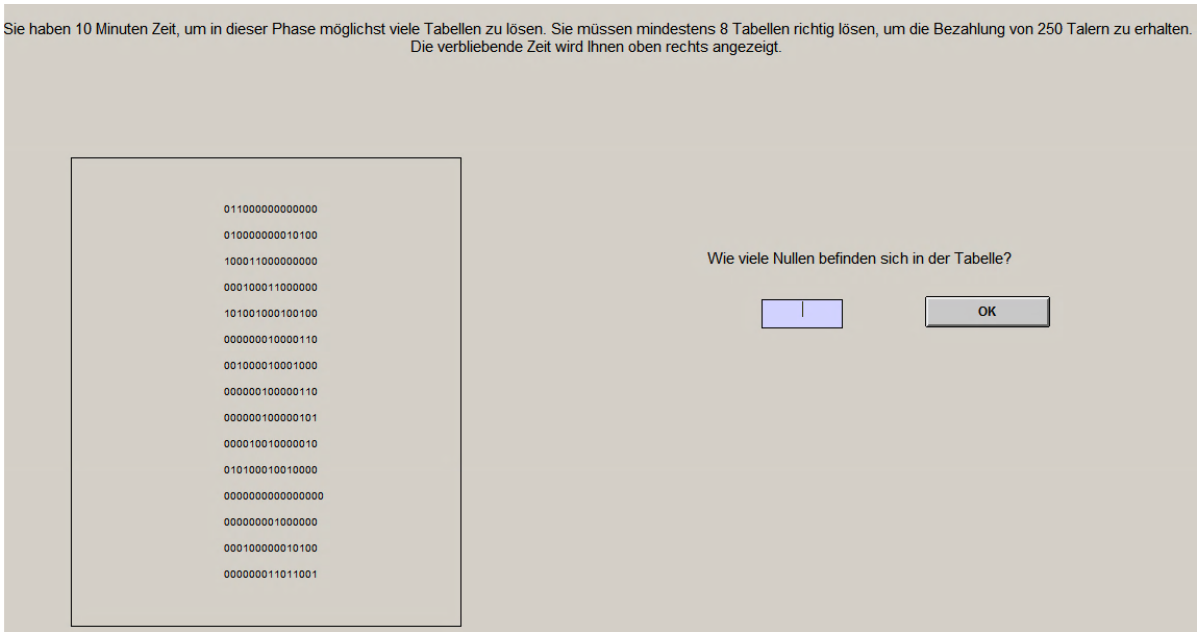
Table 6: Survey Respondents' Characteristics by Treatment

	Treatment	Control	*p*-value
***Colleague* Setting**			
Age	24.81	24.64	0.970
Male	0.524	0.436	0.429
Pro-Social Type	0.405	0.513	0.329
Reciprocal inclination (neg.)	2.690	2.906	0.304
Reciprocal inclination (pos.)	5.984	5.966	0.909
N	42	39	
***Unknown Coworker* Setting**			
Age	24.76	23.89	0.436
Male	0.540	0.492	0.594
Pro-Social Type	0.281	0.393	0.184
Reciprocal inclination (neg.)	2.786	2.644	0.619
Reciprocal inclination (pos.)	6.104	6.060	0.580
N	64	61	

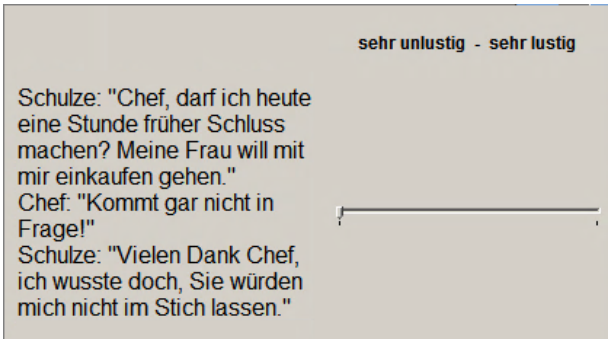
Note: p-values are obtained from Pearson's χ^2 or two-sided Wilcoxon rank-sum tests, respectively.

A.2 Screenshots of the Real Effort Tasks

The Counting Task



The Rating Task



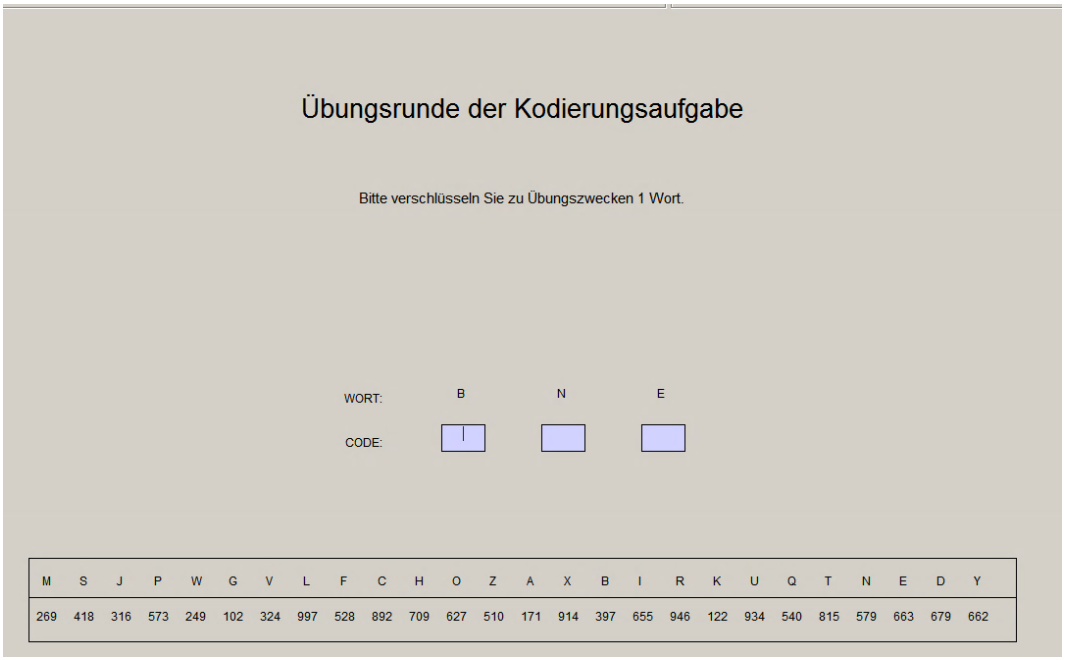
English Translation (analogously):
Schulze: "Boss, may I leave work early today? My wife wants to go shopping with me.

Boss: No, of course not. This is out of question!

Schulze: Thank you, Boss! I knew you would not let me down!

Note: The actual rating task in the experiment displayed three jokes at a time.

The Encoding Task



A.3 Instructions Laboratory Experiment

You are taking part in a scientific study in which you can earn additional money based on your decision-making, which will be added to your already assured 5€ payment. Within the experiment earnings will be realized in "taler". At the end of the experiment your taler will be converted into Euro and paid out to you in private. 100 taler are equivalent to one Euro (100 taler = 1 €). The entire experiment consists of two main stages and a closing questionnaire. All three sections will be carried out one after another.

The computer will assign you to a group and every group will incorporate three individuals. Within each group, the computer randomly assigns the role of the group leader to one participant (player 1), whereas the remaining two will act as regular group members (player 2 and player 3). As a group you will be asked to work on different tasks. Please note that the group constellation will be kept over the course of the experiment if not indicated otherwise.

You were assigned with the role of being the group leader (player 1) *[or a regular group member (player 2 or 3)]*.

Within the first stage of the experiment there will be to different types of tasks every group is asked to work on. The specific task allocation within each group is realized automatically by the computer. Every individual on its own will work on one out of the two task types.

Stage 1

Task explanations

The first task is a rating task. Here, you will be presented with a selection of jokes, which you need to rate. A concrete example will be shown to you on the next page.

Please read each joke one by one and submit your rating with the help of the provided slider. Each slider can be positioned within an interval reaching from "very unfunny" (left-hand side) to "very funny" (right-hand side). Please position (drag) the slider for each joke separately in order to meet your personal taste. The interval reaches from 0 to 100; where 0 represents "very unfunny" and 100 "very funny". Please rate as many jokes as possible but take your time in order to assess each joke's quality. For this task you will have 10 minutes and receive a fixed payment of 250 taler.

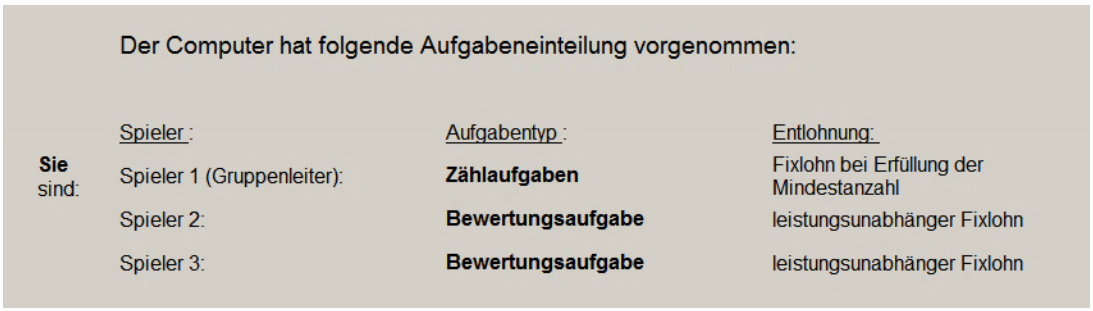
You will now have the possibility to get to know and test the task.

The second task is a counting task. Here, you will be presented with a table filled with a sequence of 0s and 1s. A concrete example will be shown to you on the next page. The task is to count the number of displayed 0s and indicate the result within the provided input mask. Afterwards, please click on the "OK"-button to confirm your answer. If your answer is correct, you will be presented with a newly generated table to be solved. If your answer was incorrect, you will have two more attempts to submit the correct number, before the program generates a new table for you. For every correctly solved table you will receive one point. Again, you have 10 minutes to count as many tables as possible. For your work on this task you receive a fixed payment of 250 taler. Please notice that you have to earn at least 8 points or solve 8 tables respectively in order to receive the fixed payment of 250 taler. If you do not manage to earn 8 points within 10 minutes, you will receive no compensation.

You will now have the possibility to get to know and test the task.

Task allocation and the group leader's decision to keep or exchange tasks

The computer realized the following task allocation within the group:



Der Computer hat folgende Aufgabeneinteilung vorgenommen:

	Spieler:	Aufgabentyp:	Entlohnung:
Sie sind:	Spieler 1 (Gruppenleiter):	**Zählaufgaben**	Fixlohn bei Erfüllung der Mindestanzahl
	Spieler 2:	**Bewertungsaufgabe**	leistungsunabhänger Fixlohn
	Spieler 3:	**Bewertungsaufgabe**	leistungsunabhänger Fixlohn

For group leaders only:

Please note: The realized task allocation here will determine your workload for both stages of the experiment. Accordingly, you will therefore work on the rating *[or counting]* task within the first and the second stage of the experiment.

Being the group leader, you have the possibility to exchange your allocated tasks with the group members. Do you want to exchange your tasks?

Please decide:

-Yes, I want to exchange.

-No, I do not want to exchange.

Please note, that irrespectively of your decision here, there will be no further interactions between you and your group members. Player 2 and 3 will work separately and completely independent from you in the second stage.

Information for group members if exchange was chosen:

Player 1 (group leader) has interfered with the task allocations realized by the computer program and exchanged his allocated tasks with you. Consequently, you and player 2 *[or 3]* will be taking over the rating *[or counting]* task while player 1 (group leader) will face the counting *[or rating]* task.

PANAS after real effort task

Now we would like to know how you feel. The following words describe different feelings and perceptions. Read every word and mark the intensity on the scale. You have the choice between five gradations. Please indicate how you feel now:

→ *Extended PANAS Questionnaire* (Watson et al. 1988)

Stage 2

For group leaders only:

This is the beginning of the second experimental stage for you. Based on your allocation you will be working now on the rating *[counting]* task. The available time will be set to 5 minutes. You will receive a fixed payment of 250 taler. Please click on „continue" and wait until your task begins. Possibly, you will have to wait for a short period. We kindly ask for your patience

Task explanation *(until the questionnaire, for group members only)*

This is the beginning of the second stage. You will be presented with a new working task. The task is to encode different words into numbers. The computer will present to you a three-letter word, which you need to translate with the help of an encryption table. You will see a concrete example on the next page.

For each word, one letter of the alphabet represents a specific three-digit number. You need to translate the entire word and submit your solution within the provided input mask. By clicking on "ok", the program checks your submitted answer. If your answer was correct, you receive a new word immediately. If your answer was incorrect, please reconsider your submitted answer until the correct solution is found. For every correctly solved word, you receive one point.

You have 5 minutes to encode as many words as possible. You will receive a fixed payment of 250 taler. The third player 2 *[or 3]* from stage 1 will work on the same task as you. By the end of the working period, an extra payout bonus of 250 taler will be given to the player with the higher point score. In case of a tie the computer draws a winner randomly.

The group leader plays independently from you a separate second stage. You will not be in any direct or indirect contact and cannot influence each other.

Tournament explanation

Before the actual task begins, you have the chance to influence the gameplay. You can modify your point account or the point account of your counterpart by a preferred amount. This way you can gain an advantage over your counterpart regarding the final determination of the total point score and increase your chances of winning the proclaimed bonus of 250 taler.

Modifying the game is costly. Every point modification costs you 10 taler. This applies for increasing your own point account as well as for reducing the point account of your counterpart. The resulting costs are subtracted from your overall payment at the end of the experiment. Please note that you can only choose between one of the two possibilities. You cannot change your own point account and your counterpart's account at the same time.

Subsequently you will see a hypothetical example for clarification:

Player A and player B both encode a total number of 5 words in the given working period. Player A decides to add 1 additional point to her own account. Player B decides to subtract 4 points from Player A. The final point scores are:

Player A: 5 + 1 – 4 = 2 points

Player B: 5 points

Player B wins the bonus of 250 taler.

The resulting costs of 10 taler for Player A and 40 taler for Player B, are subtracted from their final payoffs at the end of the experiment.

Your final point account may become negative. Consequently, the player with the highest score wins. There are no costs associated with negative point accounts. Please click on "continue" in order to choose a possible modification option. If you decide to modify the game, you can subsequently decide on the amount of points. You may choose within a range of 1 to 20 points.

Tournament decision

Please choose now from the enlisted modification possibilities:

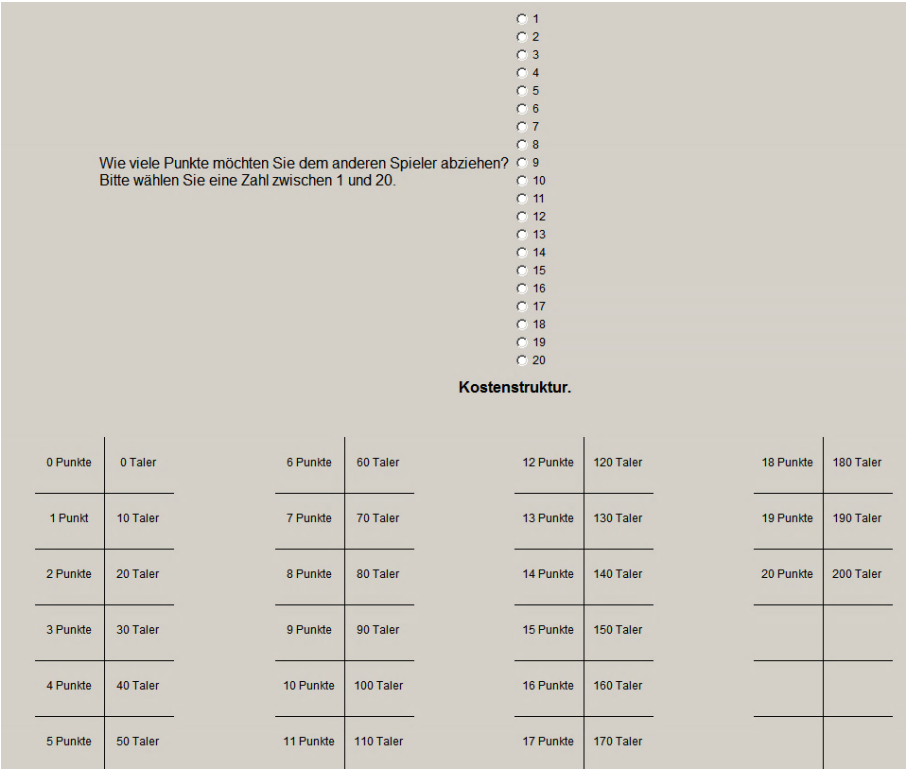
-Conduct no modification

-Modify own point account (increase)

-Modify counterparts point account (decrease)

Only if a modification was chosen:

Please indicate how many points you want to add to your point account *[or subtract from your counterpart's point account?]* Please choose a number between 1 and 20.



Tournament Belief Elicitation

Before learning about the outcomes of the second stage, we asked you to state your beliefs regarding your counterpart's behavior. Please guess whether your counterpart chose a game modification or not. If you guess correctly here, you receive an additional payoff of 150 taler. Of course, you will not lose any taler if your guess is incorrect.

-I think my counterpart chose not to modify the game.

-I think my counterpart modified his own point account.

-I think my counterpart modified my point account.

Only if modification was assumed:

Furthermore, we ask you to guess how many points your counterpart added to his point account *[or your counterpart subtracted from your point account]*. If you guess the correct amount of chosen points or within a tolerance of one you will receive an additional payoff of 500 taler. Stated values close to the correct value will be compensated as well. If your guess is further away than one point, you receive a payout of 250 taler divided by the absolute distance to the true value. The closer you are to the true value, the larger your additional payoff.

Please guess now, how many points did your counterpart add to his own point account *[or subtract from your point account]*.

Questionnaire

Socio-Demographics

How old are you?

Please indicate your gender:

Are you in your master's or bachelor's studies?

What is your field of study:

How many people do you know that took part in today's experiment?

Manipulation Checks

Please remember the first stage of the experiment:

How entertaining was the rating task in your opinion?

How entertaining was the counting task in your opinion?

For group members only:

How fair do you think was the task allocation realized by the computer?

For group members only if reallocation was chosen:

How fair do you think was the final task allocation?

For group leaders only:

Please put yourself into the position of the regular group member player 2 and 3. Do you think your intervention into the task allocation was perceived to be fair or unfair?

Please guess how many group leaders would intervene and exchange the counting for the rating tasks *[or the rating for the counting tasks]*. Please indicate your answer in percentage.

Risk Aversion and Reciprocity

Are you generally a person who is willing to take risks or do you try to avoid taking risks?

To what degree do the following statements apply to you personally? Please answer with the help of the following scale.

Social Value Orientation (Murphy et al., 2011*)* & **Big 5 Personality Test** (Costa & McCrae, 1989*)*

A.4 Instructions Survey Experiment

General Instructions:

Thank you for your participation in this study. Please answer all questions completely and faithfully. All of your answers will be used in an anonymized form in order to guarantee your privacy. Every person is allowed to participate only once. Based on your answers you can win 5€. The prerequisite to receive this compensation is a completion of all provided questions and the submission of your university email address. The email address is only used to exclude multiple participations and to inform you about your winning prize.

In the following you will be presented with a situation in which a fictional person A has to make a decision between multiple options. We will provide you with a precise description of the scenario and a detailed list of all possible decision options. Please read the description of the scenario carefully. You will be asked to evaluate the different possible choices in regard to their "moral appropriateness". By morally appropriate, we mean behavior that most people agree is the "correct" or "ethical" thing to do. The evolution can be conducted on a scale reaching from "very morally inappropriate" to "very morally appropriate". You have to possibility to vary your evaluation between these two extremes.

Based on your answers and on the answers provided by all other survey participants we will determine your payoff. At the end of the experiment we will select one of decision choices randomly. For this selected choice we will calculated most selected answer (mode). If your answer corresponds to the calculated mode you will receive a payment of 5€. On the following page of will now find the situation description and the decision choices you will need to evaluate. Important: You are asked to assess how the majority of the people evaluates the different decision choices. It is not about how you would decide to engage in this specific scenario.

The workplace scenarios

Imagine a workplace scenario in a small fast-food chain, which normally consists of one chain manager and two coequal employees.

[Additionally for treatment: Both employees had to suffer severely under the leadership of their manager, as they were often made responsible for his own mistakes. Additionally, he assigned additional work tasks to his employees which were clearly within his responsibilities. This resulted in unpaid extra hours and additional stress.*]*

The chain manager has left the fast-food company due to a better job opportunity. After his leave, the open manager position will be replaced soon. Until then, the chain is being managed externally by the headquarter of the fast food company.

[In the colleague setting: Both employees A and B extend their working contracts and stay in the particular chain.*]*

[In the unknown coworker setting: Only one of the two employees extends his contract and remains in the fast-food chain (employee A). In order to keep the business running, an additional worker is employed on an interim basis (employee B).*]*

Costumers of the fast food chain have the possibility to rate the employees' service after making their purchases with a help of provided simple rating machine (Good service? Yes/No). As usual, the employee with the highest number of positive customer ratings receives a small monetary bonus at the end of the month. Employee A and B always work on different days.

This way employee A has the chance, to submit ratings for customers who forget to do so. The rating machine is placed outside the view range of the surveillance cameras, so the risk of being caught submitting additional ratings is very low.

The table below presents a list of the possible choices available to employee A. Please indicate for each option whether the choice is being assessed as very morally inappropriate, somewhat morally inappropriate, somewhat morally appropriate, or very morally appropriate by the majority.

Employee A submits an additional positive rating for every tenth possible costumer.

Employee A submits an additional positive rating for every fifth possible costumer.

Employee A submits an additional positive rating for every third possible costumer.

Employee A submits an additional positive rating for every other possible costumer.

Employee A submits an additional positive rating for every possible costumer.

Employee A submits no additional positive ratings.

Follow-up Questionnaire

Now, please answer the following questions:

Gender:

Please indicate your age:

You study…

What degree are you striving for?

Do you exercise any kind of voluntary work in your spare time?

Reciprocity (Perugini et al., 2003)

3. The Impact of a Discriminatory Rule Setting on Individual Rule Compliance

The willingness to comply with rules is a crucial prerequisite for the functioning of societies. In this paper, I investigate whether a discriminatory rule imposition across otherwise equal group members affects the willingness to comply with a given rule. In the experiment, participants are requested to reveal costly information they gain from privately tossing a coin while some participants enjoy explicit and exclusive rule privileges. The results reveal misreporting but no differences between groups facing discriminatory and non-discriminatory rule regimes. Priorly induced group identities aiming to increase the perception of equivalence among peers do not change the results. Overall, my findings suggest that discriminatory rules per se do not necessarily erode compliance across discriminated individuals.

3.1 Introduction

The importance of rule compliance in our economy is indisputable. However, complying with rules may often contradict individuals' immediate self-interest, as acting in accordance to prescribed rules can be somewhat costly. One may think about the government asking citizens to submit their yearly income or an employer, requesting her employees to report their working hours truthfully. These examples illustrate nicely how rules require people to reveal private information while offering the chance to disobey and misreport for their own benefit. Clearly, requesting the disclosure of private information is just one way to think about rule following and may easily extend to other less surreptitious activities such as the willingness to stop at a traffic light, recycle trash, or park in designated parking areas. The reason I examine rule compliance in the context of truth telling is twofold. First, considering the potential harm dishonesty can cause across almost every economic domain makes it a particularly interesting facet of human behavior to be studied. Second, the extensive literature on dishonest behavior allows me to standardize and compare my findings to the ones gathered in the literature (see the recent meta studies by Abeler et al., 2019 and Gerlach et al., 2020). Irrespective of the way rule following is conceptualized, the crucial question remains: why do people comply with prescribed rules (or not)? From a traditional economic perspective, the answer is clear and intuitive. People break rules if they derive some benefit from it. Decision-makers simply balance prospective costs and benefits from breaking with particular rule and act accordingly (Becker, 1968). Indeed, empirical evidence derived from the lab and field indicates that impending sanctions and hence an increase in expected costs can substantially decrease misreporting (Malézieux, 2018; Holz et al., 2020). However, in contradiction to the standard economic model, studies also show that people often hesitate to disobey the request to misreport private information even under full anonymity, indicating that normative considerations, described as social costs, play an important role in decision-making (Fischbacher & Föllmi-Heusi, 2013; Engel et al. 2020). Today we understand that the decision to comply with a costly rule (or not) is the outcome of an individual trade-off between opposing motives, shaped by our social preferences across different decision-environments (Irlenbusch & Villeval, 2015). In fact, people often seek to exploit self-severing excuses to act in an unethical or egoistic way (Gino et al. 2016). A prominent example is the experimental paper by Houser et al. (2012), showing that people have a higher probability to ignore the clear request to truthfully share private information after falling victim to unfair treatment in an unrelated dictator game. Other studies confirm this general interplay between the unfair distribution of income or wages and deviant behavior – finding a higher individual probability to cheat (Birkelund & Cherry, 2020;

Dezső et al., 2022), steal (Greenberg, 1993) or destroy the property of others (Grosch & Rau, 2020). Overall, the empirical evidence shows that experienced unfairness can serve as a powerful personal justification to commit own transgressions, aiming to restore or overcome monetary inequalities (Fehr & Schmidt, 1999). From a policy perspective, this opens up important insights for the functioning of a society, as (presumably) unfair conditions may lay out the perfect breeding ground for adverse behavior within a population. I contribute to this line of literature studying the link between discrimination and rules in its most intuitive way, namely when formulated rules are unequally imposed across an otherwise equivalent population. In contrast to previous studies, unequal treatment in my paper only creates negligible monetary disparities in expectation and focuses on the psychological effect, public and unjustified rule discrimination unfolds on potential rule followers. I introduce a costly rule and request compliance while publicly releasing others from it. This allows for a precise test of whether violating the convention of rule equality – a central pillar of institutional justice in modern societies (Levi et al., 2012) – can in fact undermine compliance across populations. I want to stress an important conceptual difference that separates my paper from other studies looking at rule following across group constellations. In contrast to recent experiments, investigating how observed rule violations of others affect subsequent compliance (Dimant, 2019; Engel, 2021), I examine a setting in which singular peers are simply not subject to an otherwise binding rule.

An example, illustrating how sensitive people may react to public rule inequality can be seen in the press conference held by the prime minister of the German state of Mecklenburg-Vorpommern during the Covid-19 pandemic in March 2021. In response to a filed complaint indicting the excessive number of attendees during the public statement, local authorities disregarded any legal steps due to the apparent non-applicability of the statewide Covid-19 regulations for politicians, causing a nationwide public outcry.[19] In this particular example, the question whether unequal rules can impair subsequent compliance (here: adherence to protective health measures) becomes pressing, as it may determine the course of a pandemic along with all its far-reaching economic and societal consequences. As an additional contribution to the literature, this paper further investigates how the composition of a group or population, which sees itself confronted with unequal rules, affects subsequent behavior. Building on theoretical (Akerlof & Yellen, 1990) and empirical (e.g., Cohn et al., 2014) work, we know that social comparison between individuals heavily drives fairness perceptions.

[19] https://www.rnd.de/politik/keine-extra-corona-regeln-fur-politiker-ordnungsamt-weist-anzeige-von-piraten-vorsitzendem-kluver-zuruck-QCFZBI5IVZGLDIIMOGXJSFONJA.html

Looking at the effects of wage inequalities in the workplace, studies suggest that both occupational and social similarities can intensify the behavioral responses to unfair treatment (Obloj & Zengler, 2017; Georgellis et al., 2019; Cobb et al., 2021). In this paper, I use different intensity level of group identity among subjects and test whether the underlying group structure affects compliance in light of rule discrimination. Arguably, a stronger group cohesion and the resulting pronunciation of group member equivalence, leads to a lower acceptance for discrimination and to a stronger decline in subsequent compliance. On the other hand, Ku and Salmin (2013) suggest a reversed notion, finding a higher tolerance for inequalities among peers who share the same group identity, possibly due to the emergence of interpersonal sympathy.

To study the relationship between unequal rule imposition and subsequent rule compliance, I conducted an online experiment using a standard student sample. In the beginning of the first stage, subjects were allocated into groups, either through a random matching protocol or based on expressed consumption preferences, aiming to create a weak and a strong group identity condition. A subsequent real effort task was implemented to reinforce perceived group cohesion (or not). In the second stage, a clearly formulated rule requested subjects to report the observed outcome of a private triple coin toss (Bucciol & Piovesan, 2011). Subjects who reported three "heads" gained a monetary bonus, while all other combinations resulted in zero payoff. I investigate the effect of discriminatory rules on compliance by publicly releasing a randomly chosen group member from the original rule (favored individuals were granted four coin tosses), while requesting compliance from everyone else. A control condition in which groups experience equal treatment serves as a benchmark. My results reveal very similar compliance level across all treatment conditions. Overall, around 60% of the subjects reported payoff relevant coin tosses, clearly exceeding the expected share of 12.5% under full compliance. In contrast to my predictions, however, the discriminatory rule regime had no effect on rule compliance. Furthermore, the distinction between both group identity conditions did not change the results.

Contributing to the literature investigating the relationship between (perceived) unfairness and subsequent adverse behavior (e.g. Houser et al., 2012; Gill et al., 2013; Jeworrek & Waibel, 2021a), I focus on the psychological effect unjustified rule inequality unfolds on subsequent rule compliance. Most closely related to my research question is the paper by Birkelund and Cherry (2020), who examine how "institutional inequality" affects honesty in a real effort task. The authors induced institutional inequality by providing some subjects with all necessary task solutions, while others received no help. In contrast to the underlying analysis, the authors

confirm the general narrative described in the literature, finding more cheating among disadvantaged subjects. It is however unclear whether the decision to behave dishonest was mainly driven by the unequal "rules of the game" or by the inevitable income differences unequal treatment created. My paper contributes to that, suggesting that unfair rules alone do not necessarily lead to adverse behavior among discriminated individuals.

3.2 Experimental Design

The experiment consisted of two main parts. In the spirit of Gioia (2017), the first part was designed to induce different intensity level of group identity across subjects, building on the minimal group paradigm (Tajfel, 1970). In a short questionnaire, I gathered sociodemographic information and elicited subjects' consumption preferences for one out of two smartphones within a hypothetical lottery.[20] As in Casoria et al. (2020), I utilized their revealed phone preferences as a meaningful attribute over which I formed and induced social groups. In the strong group identity treatment (**GI**), I assorted subjects based on their favored devices into an "Iphone" or "Samsung" group. In contrast to the control condition (**NoGI**), in which subjects were randomly allocated into a "triangle" or "square" group, aiming to create a less pronounced collective identity. The corresponding group affiliation (group symbol) and with subject's unique participant ID were constantly displayed at the top of the screen to make the group membership salient (see Appendix B.3. for two exemplary screenshots). Subsequently, participants were introduced to a simple real effort task[21] (Gill & Prowse, 2012). During the task, subjects received 200 unique sliders and were instructed to drag as many sliders as possible on to a randomly determined numeric position on an interval between 0 and 200. The instructions made clear that a minimum of 30 sliders had to be correctly positioned within a fixed time span of 5 minutes to pass the task and to receive the proclaimed bonus of 3.00 EUR. Note that the threshold was set to be moderately low, ensuring that a large majority of subjects would be able to succeed. Upon successful completion, an unannounced non-monetary price was additionally awarded, aiming to reinforce the feeling of group cohesion through the accomplishment of a collective achievement. This relates to Chen and Li (2009) and Güth et al. (2009), suggesting that joint activities or shared reward communalities can fortify the shared social identity as a group. Conditioned on the successful task completion of all fellow group members, subjects in the **GI** treatment read: "*In addition, you will receive the gold cup as an*

[20] The presented devices (Samsung Galaxy Note 20 Ultra and Apple iPhone 12) were similar in price and features at the time of the experiment, which was made clear to all subjects.
[21] All instructions are provided in the Appendix B.2. Furthermore, I thank Austėja Kažemekaitytė for making her code (slider task) publicly available.

award for your group's performance, given that ALL group members successfully completed the task." Furthermore, the picture of a golden cup was displayed in the upper corner of the screen to make the collective accomplishment salient for the remaining time of the session. In order to keep both **GI** and **NoGI** settings comparable, successful participants in the control condition read a very similar message, appreciating the individual accomplishment: "*In addition, you will receive the gold cup as an award for your performance, given that you have successfully completed the task*". At the end of stage 1, I used the "Inclusion of the Other in the Self" (hereafter: IOS) scale (Aron et al., 1992; Gächter et al., 2015) to assess the feeling of perceived closeness among group members. The scale consists of seven overlapping circle pairs (see the screenshot in the Appendix B.2), from which subjects chose a fitting pair, best representing their relationship with a fellow group member.

In the second stage, I examined rule compliance in a modified version of the popular coin toss game (Bucciol & Piovesan, 2011). Participants were asked to take a coin[22] and to toss it three times in a row. An unambiguous rule, formulated in the instructions using bold letters, requested the reporting of all observed outcomes. It was common knowledge that only three "head" reports would result in a bonus of 2.00 EUR, while all other combinations promised no payoff. Three comprehension questions ensured a thorough understanding of the procedure and the payoff structure.

$$Payoff = \begin{cases} 2\text{ EUR} & if\ H, H, H\ is\ reported \\ 0\text{ EUR} & otherwise \end{cases}$$

Given that the coin tosses were conducted in private, subjects had the chance to break the rule by misreporting private information in order to gain the monetary price. I refrained from using words such as "lying", "honesty" or "cheating" to keep the decision as neutral as possible. The second stage consisted of two treatment conditions, which differed in the way the rule was imposed among group members. In the rule discrimination treatment (**Dis**), subjects received the following notification before tossing their coins: "*Participant XX from your group is released from the introduced rule. Participant XX is allowed to toss the coin four times and to report the three most favored outcomes*." In the corresponding control condition (**NoDis**), all group members remained subject to the original rule and hence no additional notification was shared.

[22] After signing up for an experimental session, participants received a confirmation e-mail including the request to bring any Euro coin. After entering the virtual session, all subjects had to confirm the terms and conditions for participation – including being endowed with a coin of their choice.

If public and unjustified rule exemption for some evokes the feeling of injustice, disadvantaged individuals may challenge the rule itself and justify own rule violations. Hence, I predict to find higher shares of price claims under the discrimination (**Dis**) in comparison to the non-discrimination (**NoDis**) setting. Further, assuming that a stronger group identity increases the perception of equivalence among subjects, unequal treatment may lead to a stronger sanction and a higher probability of disobeying the imposed rule. Hence, I predict to find more frequent rule violations among groups characterized by a strong group identity (**Dis**+**GI**) compared to groups with a weak group identity (**Dis**+**NoGI**) in the aftermath of rule discrimination. Before completing the experiment, I asked participants to guess (in percentage) how many of their fellow group members actually reported payoff relevant coin tosses and elicited subjects' reciprocal inclination (Perugini et al., 2003). In sum, the experiment consisted of a 2x2 between-subjects treatment design, varying the intensity of group cohesion (**GI** vs **NoGI)** and the underlying rule regime (**Dis** vs. **NoDis**).

The experiment was implemented using the software oTree (Chen et al., 2016). Overall, 244 subjects participated in 10 sessions between March and April 2021 and 6 sessions in June 2021.[23] Participants earned an average of 7.15 EUR (including a 3.00 EUR show-up fee) while each experimental session lasted approximately 17 min.[24] The experiment took place (virtually) at the Magdeburg Experimental Laboratory of Economic Research (MaxLab) located at the University of Magdeburg, using "hroot" (Bock et al., 2014) for subject recruitment. I ensured that subjects participated in only one of the four treatments. Payments were realized via bank transfer or personal collection at the MaxLab (around 4% opted for this option). The main hypotheses and the experimental design have been preregistered.[25]

3.3 Results

Figure 5 depicts the share of payoff relevant price claims (three "head" counts) across all four experimental conditions. In line with previous studies, I find substantial misreporting in the

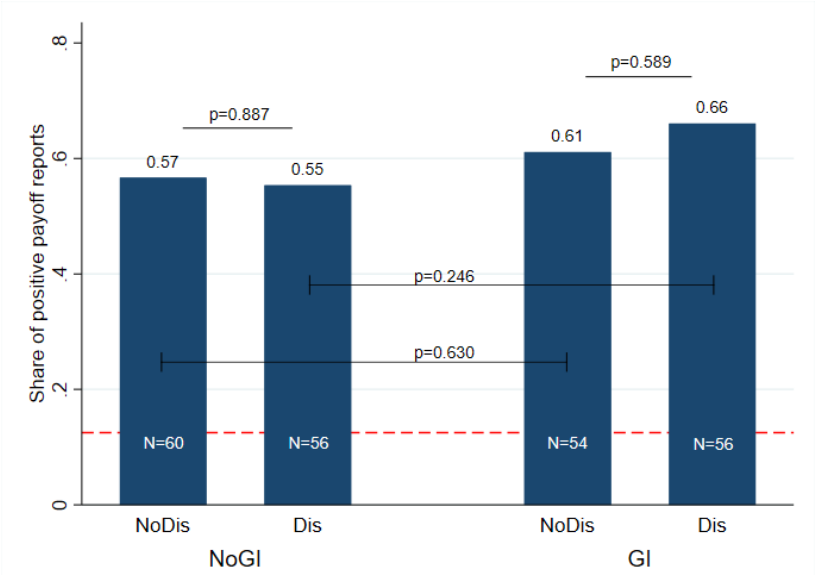
[23] Note that in the second wave, the experiment consisted of an additional (pre-announced) third stage. Here, participants were asked to answer a survey covering an unrelated topic which was part of the research project "Meta-Analyse über Praktiken des Wissenstransfers in agglomerationsfernen Räumen", commissioned by the Leibniz Institute for Regional Geography. Hence, the payments made in this study were taken both from internal funds of the author's home institution and the subcontract.

[24] In the second wave, average earnings increased to 10.11 EUR (including an additional 3.00 EUR payment in stage 3) and sessions lasted for about 23 minutes. I find no differences in coin toss reporting between both waves ($p= 0.328$, Pearson's χ^2).

[25] https://aspredicted.org/blind.php?x=8V9_M2N

data.[26] The red dashed line represents the expected share of positive reports assuming full rule compliance (12.5 %). Two-sided binomial tests indicate significant differences between actual reporting and expected reporting under full compliance (all p-values < 0.001), revealing considerable rule breaking. However, the data finds no differences in compliance between groups facing discriminatory rules and those under an equal rule regime – this holds for both the weak (0.57 vs 0.55, p = 0.887) and the strong (0.61 vs 0.66, p=0.589) group identity setting.[27] Conditioning on rule discrimination, the comparison of payoff claims between both group cohesion settings (**Dis**+**GI** vs **Dis**+**NoGI**) reveals a sizable difference of 11 percentage points (0.66 vs 0.55). Even though this difference is statistically insignificant (p=0.246), it clearly points into the predicted direction.

Figure 5: Share of Positive Payoff Reports Across Treatment Conditions



Note: The dashed horizontal line marks the expected share under complete rule following / honest reporting (12.5%).

Subsequently, I ran an additional probit regression analysis to check for other determinants that may drive the decision to rule following and to control for observable differences between treatment and control groups.[28] Again, the analysis focusses on subjects' individual coin toss report. I used a binary variable indicating whether subjects claimed the monetary price as the dependent variable. The estimation of interaction terms within non-linear models is known to be problematic as it can lead to biased results (see Ai & Norton, 2003). Hence, I additionally

[26] Please note that all rule-favored subjects (n=18) were excluded from the analysis. Overall, 15 out of 18 rule-favored subjects (83.33%) claimed the bonus, clearly exceeding the threshold of 31.25% expected under full rule compliance (p<0.001, Two-sided binomial test).
[27] If not stated otherwise, the p-values are obtained from Pearson's χ^2.
[28] Table 8 in Appendix B.1 provides summary statistics on subjects' socio-demographic information and personality traits.

report my results using a linear probability model (LPM) (see columns 3 & 5) and find very similar results across both model specifications. Columns 1 and 2 (3) in Table 7 confirm the previous findings, as both treatment dummies (*Discrimination* and *Group Identity*) and their interaction (*Discrimination* * *Group Identity*) remain insignificant in explaining rule following. Proceeding the analysis in an exploratory way, column 4 (5) introduces a rich set of controls. I find that subjects' beliefs regarding other group members' reporting behavior significantly affects reporting. Note that the elicitation of beliefs was not incentivized. Hence, I cannot exclude the possibility that subjects used this question to justify their own rule breaking. Nevertheless, the results confirm the importance of peer behavior in the decision process (Gino et al., 2009; Engel, 2021; Bicchieri et al., 2022).[29] Economics students have a higher probability of claiming the bonus, supporting the suggestion of a stronger inclination towards dishonesty within this particular sub-population (Lundquist et al., 2009). Furthermore, I included information on subjects' prior experimental experience derived from the online recruitment system "hroot". The variable indicates the number of participants' previous participations in other laboratory experiments and serves as a proxy for their attained experience in economic experiments. The regression shows that experienced participants are more likely to violate the rule and to claim the monetary price. This finding is not new and previous studies show that experienced subjects behave more self-oriented and are more willing to maximize own profits (Matthey & Regner, 2013 & Benndorf et al., 2017). Following Schmidt et al. (2019), this can be explained through a shift in social norm perceptions caused by repeated participations in economic experiments. This would mean that trained subjects judge rule violations to be less inappropriate compared to untrained ones - leading to less social and more selfish behavior. However, I find no evidence for heterogonous treatment effects regarding rule compliance when I differentiate between experienced and unexperienced subjects.[30] Subjects, who allegedly tossed a low-value coin (cent), claim the bonus with a significantly lower probability. Here, it is unclear why holding different valued coins should affect reporting behavior, given that the coins were already in possession of the decision-maker and therefore not at stake.[31] Finally, I find that exerted effort (in stage 1), the explicit cellphone preference, age, gender as well as subjects' reciprocal inclination do not predict reporting behavior in the experiment.

[29] There is no difference in stated beliefs between both Dis and NoDis conditions (both 77.9 %).

[30] See Table 9 in Appendix B.1 for a detailed subgroup analysis looking at experienced und unexperienced individuals (categorized via median split) separately. Furthermore, I present regression results looking at participants without an economic background. Again, I find no evidence for heterogonous treatment effects regarding rule compliance.

[31] Studies show that people perceive higher psychological costs from a potential money loss when the money is held in hand (cash) compared to abstract book money (e.g., Reinstein & Riener, 2012).

Table 7: Positive Payoff Reports

	(1) Probit	(2) Probit	(3) LPM	(4) Probit	(5) LPM
Discrimination	0.019	-0.013	-0.013	-0.020	-0.018
	(0.060)	(0.092)	(0.095)	(0.076)	(0.078)
Group Identity		0.044	0.044	-0.063	-0.060
		(0.086)	(0.088)	(0.103)	(0.104)
*Discrimination*Group Identity*		0.064	0.063	0.101	0.096
		(0.117)	(0.118)	(0.122)	(0.122)
Beliefs				0.005***	0.005***
				(0.001)	(0.001)
Experience				0.011**	0.011**
				(0.005)	(0.004)
Effort				-0.0001	-0.0001
				(0.003)	(0.003)
Economics student (0/1)				0.146**	0.141**
				(0.071)	(0.070)
Age				-0.005	-0.005
				(0.010)	(0.010)
Female (0/1)				-0.015	-0.018
				(0.058)	(0.061)
Cent (0/1)				-0.105**	-0.108**
				(0.053)	(0.058)
Iphone (0/1)				0.085	0.081
				(0.063)	(0.069)
High neg. reciprocal inclination (0/1)				-0.097	-0.098
				(0.068)	(0.072)
Constant			0.567***		0.292
			(0.070)		(0.314)
N	226	226	226	223	223
Pseudo R-squared	0.0003	0.0055		0.1048	
R-squared			0.0073		0.1315

Note: Columns with probit regressions show average marginal effects. The dependent variable is a dummy indicating whether the participant reported the positive payoff outcome. "High neg. reciprocal inclination" is a dummy variable created by a median split. "Experience" is the number of participation in prior lab experiments at the MaxLab. "Effort" is the number of correctly placed sliders in stage 1. Standard errors are in parentheses and clustered on the session level, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In stage 1, I aimed to induce different intensity level of group identity to derive a better understanding of how discriminatory rule enforcement may affect compliance across different group constellations. As discussed above, the results show no variation in reporting across both group conditions. I shortly summarize the results derived in the first part. As intended, the overwhelming majority of subjects managed to pass the threshold set in the slider task (97.3%). In general, subjects exerted more effort than required, with an average of 59.9 sliders. A closer look reveals a high variation in performance, indicating an average of 76.05 sliders for high performers (representing the 75th percentile) and 42.63 sliders for the less ambitious subjects (representing the 25th percentile). As mentioned in the regression analysis above, effort did not predict reporting behavior. However, a more detailed examination reveals a noteworthy pattern

at the extreme ends of the effort distribution. Top performers seem to have a higher probability to claim the bonus compared to low performers (0.67 vs 0.58, p=0.294), suggesting that subjects who wore themselves out during the task, may have felt entitled to additional compensation and justified rule violations more frequently. I find that the higher willingness to break with the rule was driven by those high performers who saw themselves confronted with the unequal rule regime (0.77 vs 0.58, p= 0.122).[32] Clearly, these explorative results do align with the formulated research hypothesis but should be taken with caution given their statistical insignificance. Furthermore, I had no prior hypothesis on whether exerted effort in stage 1 would interact with rule following behavior in stage 2.

Similar to Jacquemet et al. (2021) and Dimant (2021), who successfully utilized the IOS scale to assess the perceived closeness between in- and out-group members along natural groups, I build on the IOS score to quantify the induced intensity level of group cohesion across both experimental conditions (**GI** & **NoGI)**. My findings reveal no significant difference in IOS scores between both group environments (**GI**=2.94 vs **NoGI**=3.15, p = 0.970, two-sided Wilcoxon rank-sum test). Hence, I find no support for different group intensity level across both group types.[33] Arguably, even the random allocation protocol and the usage of geometric symbols – making group membership salient throughout the experiment – may have been sufficient to create a rather strong collective identity among participants in the control condition (Chen & Li, 2009).

3.4 Discussion & Conclusion

Contributing to the literature on rule following, I study whether the discriminatory imposition of a costly rule across equivalent peers affects the willingness to comply with the given rule. With the help of an online experiment, my results show that the public and unjustified rule waiving for singular individuals leads to no change in rule compliance among disadvantaged subjects. The reinforcement of group identities, intended to increase the feeling of equivalence among subjects, does not change behavior. Building on previous findings highlighting a generally strong interplay between perceived unfairness and adverse behavior, my results suggest that unequal treatment per se seems to be insufficient to impair compliance. The decision to react to unfair treatment by justifying own deviant behavior is likely to be driven by pecuniary motives and the desire to overcome these inequalities. In the absence of strong

[32] For low performers the effect direction is reversed but remains insignificant (0.55 vs 0.62, p=0.549).
[33] The exclusion of groups and individuals who did not receive the gold cup does not change the result (**GI**=2.84 vs **NoGI**=3.19, p=0.970, two-sided Wilcoxon rank-sum test).

immediate monetary disadvantages, discriminatory treatment does not necessarily trigger severe behavioral reactions. From a policy perspective, this observed stable adherence to formulated rules is somehow good news and may allow for some margin of error in rule setting. Nevertheless, my findings should be taken with some caution given the generally high level of rule breaking across all treatment conditions. Furthermore, one might ask whether the unequal rule imposition created a feeling of unfairness in the first place. Clearly, research shows that inequality alone may not necessarily be considered as unfair – for example when inequalities arise from individual achievements earned through merit or ability (Konow, 2000; Cappelen et al., 2020). In this paper, subjects could not attribute differential treatment to any fair (e.g. competitive) process, as no information was shared justifying the singular favoritism in rule setting. Hence, there is no reason to believe that participants rationalized their disadvantaged position in the experiment.

In line with the literature studying dishonest behavior using random devices (mostly coins or dice) to generate private information, I find substantial misreporting. However, my results indicate a rather high level of dishonesty compared to similar designed experiments (see Kajackaite & Gneezy, 2017 or Garbarino et al., 2019). To make my findings somewhat comparable to the ones reported in the literature, I use the standardization approach proposed by Garbarino et al. (2018) to calculate the mean lying rate. I find that approximately 53.95 % of my participants – considering those who actually observed zero payoff tosses (with a probability of 87.5%) – lied and claimed the monetary price. This lying rate is larger than the average rate reported in the recent meta-analysis by Garbarino et al. (2019). Based on 81 studies, the authors derive a mean lying rate of 24.56% for similar tuned experiments (here: experiments with a probability larger or equal to 75% of receiving the lowest payoff). Several reasons may help to explain the high lying rates in this paper. First, it is important to point out that the study was implemented as an online experiment using a standard student subject pool. Arguably, subjects deciding remotely from home may feel less socially controlled and observed, making misreporting less costly (Abeler et al., 2019). However, the empirical evidence is inconclusive on that matter and recent studies suggest no difference in lying across online and on-site environments (Gerlach et. al (2020); Dickinson & McEvoy, 2021).[34] Second, the experiment took place during the third Covid-19 wave in Germany (2021) with raising infection numbers in March and April and declining cases in June. There is an ongoing debate

[34] It is worth mentioning that I was among the first researchers to run a lab-like online experiment at the MaxLab after on-site sessions had been expelled due to the Covid-19 restrictions. I cannot rebut that this "pool naivety" towards online experiments may have influenced behavior and the perceived level of anonymity.

on whether the exposure to Covid-19 or better its challenging side effects (e.g. social distancing or anxiety) influenced social preferences and consequently economic behavior. Indeed, recent studies suggest that in the wake of the pandemic, people became more selfish (Branas-Garza et al., 2022; Jeworrek & Waibel, 2021b) and antisocial (Lohmann et al., 2020), which may translate directly into a lower hesitation towards rule violation and lying. Overall, it is conceivable that the seemingly low costs for lying in the experiment and the resulting excessive claim shares overlaid any potential behavioral differences triggered by the unequal rule regime. Therefore, future research should further investigate the psychological effect of unjustified rule discrimination (with and without monetary consequences) within settings that are less prone to high lying rates and across domains that look at rule following aside from dishonest behavior (e.g., Kimbrough & Vostroknutov, 2016, 2018; Engel, 2021).

References

Abeler, J., **Nosenzo**, D., & **Raymond**, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4), 1115-1153.

Ai, C., & **Norton**, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1), 123-129.

Akerlof, G. A., & **Yellen**, J. L. (1990). The fair wage-effort hypothesis and unemployment. The *Quarterly Journal of Economics*, 105(2), 255-283.

Aron, A., **Aron**, E. N., & **Smollan**, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63(4), 596–612.

Becker, G. S. (1968). Crime and punishment: An economic approach. In The economic dimensions of crime, 13-68. Palgrave Macmillan, London.

Benndorf, V., **Moellers**, C., & **Normann**, H. T. (2017). Experienced vs. inexperienced participants in the lab: Do they behave differently?. *Journal of the Economic Science Association*, 3(1), 12-25.

Bicchieri, C., **Dimant**, E., **Gächter**, S., & **Nosenzo**, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, 132, 59-72.

Birkelund, J., & **Cherry**, T. L. (2020). Institutional inequality and individual preferences for honesty and generosity. *Journal of Economic Behavior & Organization*, 170, 355-361.

Brañas-Garza, P., **Jorrat**, D., **Alfonso**, A., **Espín**, A. M., **Muñoz**, T. G., & **Kovářík**, J. (2022). Exposure to the COVID-19 pandemic environment and generosity. *Royal Society open science*, 9(1), 210919.

Bucciol, A., & **Piovesan**, M. (2011). Luck or cheating? A field experiment on honesty with children. *Journal of Economic Psychology*, 32(1), 73-78.

Cappelen, A. W., **Falch,** R., & **Tungodden,** B. (2020) Fair and Unfair Income Inequality. In: Zimmermann, K. (Ed), Handbook of Labor, Human Resources and Population Economics, Springer, Cham.

Casoria, F., **Reuben**, E., & **Rott**, C. (2020). The effect of group identity on hiring decisions with incomplete information. Hshs-03007766.

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

Chen, D. L., **Schonger**, M., & **Wickens**, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88-97.

Cobb, J. A., **Keller**, J. R., & **Nurmohamed**, S. (2021). How Do I Compare? The Effect of Work-Unit Demographics on Reactions to Pay Inequality. *ILR Review*, 00197939211001874.

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.

Dezső, L., **Alm**, J., & **Kirchler**, E. (2022). Inequitable wages and tax evasion. *Journal of Behavioral and Experimental Economics*, 96, 101811.

Dickinson, D. L., & **McEvoy**, D. M. (2021). Further from the truth: The impact of moving from in-person to online settings on dishonest behavior. *Journal of Behavioral and Experimental Economics*, 90, 101649.

Dimant, E. (2019). Contagion of pro-and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology*, 73, 66-88.

Dimant, E. (2021). Hate trumps love: The impact of political polarization on social preferences. *CESifo Working Paper* No. 9073.

Engel, C., **Mittone**, L., & **Morreale**, A. (2020). Tax morale and fairness in conflict an experiment. *Journal of Economic Psychology*, 81, 102314.

Engel, C. (2021). Crime as Conditional Rule Violation. *MPI Collective Goods Discussion Paper* (2021/20).

Fehr, E., & **Schmidt**, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3), 817-868.

Fischbacher, U., & **Föllmi-Heusi**, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3), 525-547.

Gächter, S., **Starmer**, C., & **Tufano**, F. (2015). Measuring the closeness of relationships: a comprehensive evaluation of the 'inclusion of the other in the self' scale. *PloS one*, 10(6), e0129478.

Garbarino, E., **Slonim**, R., & **Villeval**, M. C. (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. C. (2019). Loss aversion and lying behavior. *Journal of Economic Behavior & Organization*, 158, 379-393.

Georgellis, Y., **Garcia**, S. M., **Gregoriou**, A., & **Ozbilgin**, M. (2019). Pay referents and satisfaction with pay: Does occupational proximity matter?. *British Journal of Management*, 30(3), 578-592.

Gerlach, P., **Teodorescu**, K., & **Hertwig**, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, 145(1), 1.

Gill, D., & **Prowse**, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1), 469-503.

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., **Ayal**, S. and **Ariely**, D. (2009) Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science* 20(3): 393–398.

Gino, F., **Norton**, M. I., & **Weber**, R. A. (2016). Motivated Bayesians: Feeling moral while acting egoistically. *Journal of Economic Perspectives*, 30(3), 189-212.

Gioia, F. (2017). Peer effects on risk behaviour: the importance of group identity. *Experimental Economics*, 20(1), 100-129.

Greenberg, J. (1993). Stealing in the name of justice: Informational and interpersonal moderators of theft reactions to underpayment inequity. *Organizational Behavior and Human Decision Processes*, 54 (1), 81(103).

Grosch, K., & **Rau**, H. A. (2020). Procedural unfair wage differentials and their effects on unethical behavior. *Economic Inquiry*, 58(4), 1689-1706.

Güth, W., **Ploner**, M., & **Regner**, T. (2009). Determinants of in-group bias: Is group affiliation mediated by guilt-aversion?. *Journal of Economic Psychology*, 30(5), 814-827.

Holz, J. E., **List**, J. A., **Zentner**, A., **Cardoza**, M., & **Zentner**, J. (2020). The $100 million nudge: Increasing tax compliance of businesses and the self-employed using a natural field experiment (No. w27666). *National Bureau of Economic Research.*

Houser, D., **Vetter**, S., & **Winter**, J. (2012). Fairness and cheating. *European Economic Review*, 56(8), 1645-1655.

Irlenbusch, B., & **Villeval**, M. C. (2015). Behavioral ethics: how psychology influenced economics and how economics might inform psychology?. *Current Opinion in Psychology*, 6, 87-92.

Jacquemet, N., **Luchini**, S., **Rosaz**, J., & **Shogren**, J. F. (2021). Can we commit future managers to honesty?. *Frontiers in Psychology*, 2785.

Jeworrek, S., & **Waibel**, J. (2021a). Unethical employee behavior against coworkers following unkind management treatment: An experimental analysis. *Managerial and Decision Economics*, 42(5), 1220-1234.

Jeworrek, S., & **Waibel**, J. (2021b). Alone at home: The impact of social distancing on norm-consistent behavior. *IWH Discussion Papers* No. 08/2021.

Kajackaite, A., & **Gneezy**, U. (2017). Incentives and cheating. *Games and Economic Behavior*, 102, 433-444.

Kimbrough, E. O., & **Vostroknutov**, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3), 608-638.

Kimbrough, E. O., & **Vostroknutov**, A. (2018). A portable method of eliciting respect for social norms. *Economics Letters*, 168, 147-150.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review*, 90(4), 1072-1091.

Ku, H., & **Salmon**, T. C. (2013). Procedural fairness and the tolerance for income inequality. *European Economic Review*, 64, 111-128.

Levi, M., **Tyler**, T. R., & **Sacks**, A. (2012). The reasons for compliance with law. In R. Goodman, D. Jinks, & A. Wood (Eds), Understanding social action, promoting human rights, 70–99. New York, NY: Oxford University Press.

Lundquist, T., **Ellingsen**, T., **Gribbe**, E., & **Johannesson**, M. (2009). The aversion to lying. *Journal of Economic Behavior & Organization*, 70, 81–92.

Lohmann, P., **Gsottbauer**, E., You, J., & **Kontoleon**, A. (2020). Social preferences and economic decision-making in the wake of COVID-19: experimental evidence from China. *Available at SSRN 3705264.*

Malézieux A. (2018). A Practical Guide to Setting Up your Tax Evasion Game. *Journal of Tax Administration,* 4:12018.

Matthey, A., & **Regner**, T. (2013). On the independence of history: experience spill-overs between experiments. *Theory and Decision*, 75(3), 403-419.

Obloj, T., & **Zenger**, T. (2017). Organization design, proximity, and productivity responses to upward social comparison. *Organization Science*, 28(1), 1-18.

Perugini, M., **Galucci**, M., **Presaghi**, F. & **Ercolani**, A. P. (2003). The personal norm of reciprocity. *European Journal of Personality*, 17(4), 251-283.

Reinstein, D., & **Riener**, G. (2012). Decomposing desert and tangibility effects in a charitable giving experiment. *Experimental Economics*, 15(1), 229-240.

Schmidt, R., **Schwieren**, C., & **Sproten**, A. N. (2020). Norms in the lab: Inexperienced versus experienced participants. *Journal of Economic Behavior & Organization*, 173, 239-255.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223(5), 96–102.

Appendix B

B.1 Tables

Table 8: Descriptive Statistics

	NoDis + NoGI	Dis + NoGI	NoDis + GI	Dis + GI	*p-value*
Experienced (0/1)	0.550	0.518	0.370	0.536	0.206
Economics student (0/1)	0.283	0.196	0.333	0.143	0.082
Age	25.367	25.607	24.667	24.745	0.418
Female (0/1)	0.600	0.589	0.556	0.500	0.704
Cent (0/1)	0.483	0.418	0.264	0.286	0.043
Reciprocal inclination (neg.)	2.883	2.964	2.716	2.780	0.822
Reciprocal inclination (pos.)	6.211	6.012	6.117	6.042	0.821

Note: *p*-values obtained from Kruskal-Wallis tests and Chi² tests, respectively.

Table 9: Coin Toss Reporting Across Subgroups

Subgroup	Experienced		Unexperienced		Non-Econ	
	Probit	LPM	Probit	LPM	Probit	LPM
	(1)	(2)	(3)	(4)	(5)	(6)
Discrimination	0.031	0.028	-0.060	-0.057	-0.081	-0.077
	(0.071)	(0.086)	(0.100)	(0.103)	(0.073)	(0.078)
Group Identity	0.040	0.053	-0.190	-0.213	-0.065	-0.061
	(0.077)	(0.088)	(0.179)	(0.190)	(0.126)	(0.134)
Discrimination *	0.100	0.075	0.119	0.135	0.192	0.180
Group Identity	(0.134)	(0.126)	(0.190)	(0.202)	(0.138)	(0.143)
Beliefs	0.003***	0.003**	0.006***	0.006**	0.005***	0.005***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
Experience	0.010	0.009	-0.023	-0.021	0.011**	0.011**
	(0.007)	(0.007)	(0.026)	(0.028)	(0.005)	(0.005)
Effort	-0.003	-0.003	0.006	0.006	0.001	0.001
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Economics student	0.108	0.109	0.202**	0.198*		
(0/1)	(0.100)	(0.104)	(0.087)	(0.098)		
Age	-0.017	-0.018	0.001	0.0004	-0.003	-0.004
	(0.013)	(0.014)	(0.015)	(0.017)	(0.011)	(0.011)
Female (0/1)	-0.022	-0.026	-0.009	-0.015	0.024	0.018
	(0.098)	(0.107)	(0.058)	(0.064)	(0.081)	(0.083)
Cent (0/1)	-0.189**	-0.197*	0.019	0.017	-0.092	-0.093
	(0.084)	(0.104)	(0.088)	(0.101)	(0.076)	(0.084)
Iphone (0/1)	-0.011	-0.026	0.190***	0.198**	0.013	0.014
	(0.098)	(0.110)	(0.064)	(0.075)	(0.080)	(0.086)
High neg. reciprocal	-0.057	-0.053	-0.184**	-0.193**	-0.035	-0.036
inclination (0/1)	(0.102)	(0.109)	(0.078)	(0.087)	(0.071)	(0.074)
Constant		0.939**		-0.249		0.162
		(0.346)		(0.541)		(0.381)
N	111		112	112	169	169
Pseudo R-squared	0.0998		0.1840		0.0908	
R-squared		0.1170		0.2276		0.1166

Note: Columns with probit regressions show average marginal effects. The dependent variable is a dummy indicating whether the participant reported the positive payoff outcome. "High neg. reciprocal inclination" is a dummy variable created by a median split. "Experience" is the number of participation in prior lab experiments at the MaxLab. "Effort" is the number of correctly placed sliders in stage 1. Standard errors are in parentheses and clustered on the session level, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.2 Instructions

(translated from German)

Welcome to today's online experiment.

Before you can start, please read and confirm all terms and conditions.

- I confirm that I participate alone and without the help of a third party.
- I confirm that I am located at a quiet place without any disturbances and I conduct this experiment on a desktop-PC or laptop.
- I confirm that I am endowed with a coin that I brought to this session.

[On the next screen]

The entire experiment consists of two parts - in which you can earn money. Please read all instructions carefully. Please conduct the experiment using the "full screen mode" of your browser.

The money you earn throughout the experiment will be added to your show-up fee of 3,00 €. Please note that you have to finish this experiment to receive any payment (and your payment code). If you are inactive, the computer program may remove you from this session. All your decisions are confidential and your data will be gathered and stored anonymously.

[On the next screen]

For the entire duration of this experiment you will receive an internal identification number: XX. This identification number will be displayed throughout the session in the upper left corner of your screen.

Please answer the following questions:

- Please indicate your age.
- Please indicate your gender.
- Which university do you belong to?
- Which faculty do you belong to?
- What degree are you striving for?

[On the next screen]

Please imagine the following hypothetical scenario. You urgently need a new cellphone and you won the star prize in a raffle. You can choose between two phone models that are similar in price and features: Samsung Galaxy Note 20 Ultra and Apple iPhone 12.
Which of the devices would you choose?

[On the next screen]

One condition for participation was to bring any Euro coin.
Please, answer the following questions regarding your coin.

What kind of Euro coin did you bring (tails)?

What kind of motive can be found on your coin (heads)?

[On the next screen]

For the rest of the experiment you have been allocated into one of two groups. The group allocation was based on your stated phone preference [*was random*]. You are member of the "IPhone" / "Samsung" group, which only consists of participants who also preferred the Iphone / Samsung. [*You are member of the "triangle" / "square" group.*]

You will remain in your group until the end of the experiment. Your group symbol will be displayed in the upper right corner of the screen.

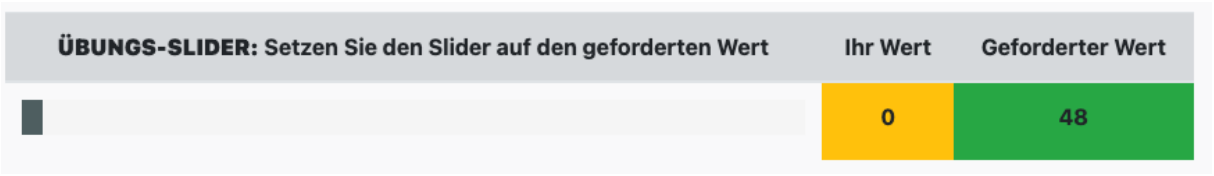
[On the next screen]

Part 1: Slider Task

In the first part of the experiment, you and your group members are working independently on the subsequent slider task.

In this task, you will have to position as many sliders as possible onto a particular position within 300 seconds. You can freely move each slider on an interval ranging from 0 to 200 – please use your mouse (recommended) or your touchpad. Note that the requested position for each slider is going to be different (see "requested value").

In this example, you are asked to drag the slider onto the value of "48". The task is correctly solved if both numbers match.



You have 300 seconds (5 minutes) to place as many sliders as possible onto their correct position. If you manage to place at least 30 sliders correctly, you will receive a payoff of 3€. If you don't, there will be no payoff. After 300 seconds, the task will stop automatically and you will receive a short summary of your work.

[On the next screen]

Slider task

[On the next screen]

You successfully completed the slider task and earned a payment of 3.00 €.

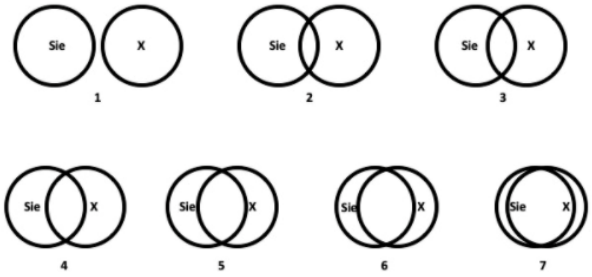
In addition, you will receive the gold cup as an award for your group's performance, given that ALL group members successfully completed the task.

[In addition, you will receive the gold cup as an award for your performance, given that you have successfully completed the task. XX % of your group members managed to pass the task successfully.]

[On the next screen]

Look at the circles closely. Please decide which circle pair represents your (here: YOU) relationship with a randomly chosen group member (here: X) from your Iphone-group *[square-group]* best.

○1 ○2 ○3 ○4 ○5 ○6 ○7



[On the next screen]

Part 2:

You are now starting with the second part of the experiment. Here, you have the possibility to earn an additional payoff of 2.00 €. Here, you will need the coin you have brought.

Rule: Please throw your coin three times and write the observed outcomes into the entry mask.

- For an outcome of "head", "head", "head" you will receive 2.00€
- For all other outcomes you will receive 0.00€

Please answer the following comprehension questions in order to proceed.

Assume your triple coin toss produces "tail", "tail", "head". What would be your payment here?

Assume your triple coin toss produces "head", "tail", "head". What would be your payment here?

Assume your triple coin toss produces "head", "head", "head". What would be your payment here?

[On the next screen]

[Participant XX from your group is released from the introduced rule. Participant XX is allowed to toss the coin four times and to report the three most favored outcomes.]

[On the next screen]

Now it is your turn. Please throw your coin three times and report the observed outcomes with the help of the entry mask.

[On the next screen]

Please answer the following questions to finish the experiment.

Please guess: how many of your fellow group members reported “head”, “head”, “head” and received 2.00 €?

[On the next screen]

Reciprocal inclination (*6 items, following Perugini et al., 2003*)

B.3 Screenshot Examples for both Group Conditions

Iphone Group and Gold Cup

Square Group and Gold Cup

4. Alone at Home: The Impact of Social Distancing on Norm-consistent Behavior

The Covid-19 pandemic has turned daily life upside down, with social distancing being the most effective method to contain the virus in the absence of herd immunity. By conducting two experiments and manipulating social isolation recollections, we study whether social distancing has affected norms of prosociality and norm compliance —a question that is also relevant beyond the pandemic in a world that becomes more digitalized every day, leading to a decline in in-person interactions. The normative expectations of what behaviors others would (dis)approve in our experimental setting did not change. Looking at actual behavior, however, we find a decline in prosociality even after the relaxation of social distancing rules. Our results also contain some good news since subjects still seem to care for norms and become more prosocial once again after we draw their attention to the empirical norm of how others have previously behaved in a similar situation.

This chapter is joint work with Sabrina Jeworrek and is based on a revised version of the same-titled IWH Discussion Paper No. 08/2021.

4.1 Introduction

Around the globe, the Covid-19 pandemic has turned daily life upside down. Besides comprehensive hygiene rules and widespread testing, social distancing is probably the most effective method to contain a pandemic as long as there is no herd immunity. Given that a holistic monitoring and an aggressive enforcement of far-reaching social distancing rules is hardly feasible, voluntary compliance is crucial. Voluntary compliance with social distancing rules resembles a classical collective action problem for which the development of social norms increase the probability of individuals solving these problems (Ostrom 2000). Using a sample of almost 90k individuals from 39 countries, Ludeke et al. (2021) show that local social norms (i.e. the perceived societal consensus on the importance of the social distancing rules) are indeed an important determinant of social distancing behaviors, even trumping personality-driven compliance behavior. In a similar spirit, Bian et al. (2022) exploit the US-county specific cultural norm of individualism and find lower engagement in social distancing behavior across counties that are characterized by a high level of individualism. For a college aged U.S. sample, Eckel et al. (2021) find that the correlation between precautionary behaviors and the beliefs on what one ought to do (injunctive norms) was weak at the beginning of the pandemic but became stronger in a later phase when individuals got tired of the restrictions. Casoria et al. (2021) confirm that the perceptions of social distancing norms (and therewith also behavior) shifted quickly during the pandemic. However, they highlight the role of the law in this process, as government's recommendations alone were not sufficient to alter the perceived social norm regarding meeting friends within their observed sample of French students.

The question whether social distancing might in turn affect norms and norm compliance has not been addressed yet.[35] Especially other-regarding behavior —which is one fundamental feature of well-functioning societies— is often conditional. In that case, one's own behavior is either influenced by expectations about how others act in similar situations (empirical expectations) or by expectations about what behaviors others would approve or disapprove (normative expectations). Whereas new norms can emerge quite quickly, existing norms are rather stable but can adjust within individuals' reference networks (i.e. the people whom we care about when making particular decisions). Even though these networks need not to be

[35] Without discussing the role of social norms but somehow related, Bland et al. (2022) observe that social contact with friends and family is related to the offers proposed in an ultimatum game and Shachat et al. (2021) find lower levels of trust and cooperation for experimental subjects quarantined in Wuhan during the lockdown.

physically present (Bicchieri 2016), they may still change due to extensive social distancing rules and thus also the underlying normative expectations.

Even if the norm of prosociality is not affected itself, norm compliance might be. It has been shown that individuals seek social cues that justify own non-compliance. As such, there is an asymmetric response to information about others' behavior, as observed norm violations have much larger effects on individuals' willingness to comply than observed norm compliance (Bicchieri et al. 2022, Dimant 2019), especially in the absence of a clear identification with a group. Similarly, Christensen et al. (2004) argue that positive emotions underlie conformity with social norms. The greater the identification with a group is the more positive emotions are associated with norm compliance. If persistent social distancing weakens social ties and, therewith, individuals' perceived group identification, norm compliance is likely to decrease not only because of less positive emotions derived from norm compliance but also because it might be easier to find a moral excuse for non-compliance. In the worst case, social isolation might lead to a situation in which people simply do not think about the norms of society anymore since they feel left behind or have other issues to handle. Brodeur et al. (2021), for example, suggest that people's mental health may have been severely affected by the pandemic and lockdown. At the same time, however, Hensel et al. (2022) show that lockdowns in the early phase of the pandemic had a positive effect on mental health, especially for those who personally adhered to the normative expectations of COVID-19 containment behavior in society. Hence, whether and how the social isolation experiences affect perceived social norms and norm compliance is an open empirical question. Still, putting the different arguments together, one would most likely expect a negative effect on norm compliance.

In our paper, we present the results from two experiments — conducted with different participants from the same standard student subject pool at a German university — which are based on a Take-or-Give (ToG) game similar to Bicchieri et al. (2022). In this game, subjects are matched with a charitable organization and both the charity and the subject receive an endowment of 5 EUR. Subjects act as decision-makers and can either retain the initial allocation, take money from or give money to the charity. Whereas in the first experiment we elicit the normative expectations how one should behave in this game (Krupka and Weber 2013), we observe actual behavior in the second experiment. To test the causal impact of social distancing experiences on the norm of pro-social behavior and norm compliance, we use the priming method to manipulate the saliency of these social isolation recollections. Both experiments consisted of three parts. Primed subjects started with answering questions on their

social distancing behavior and their feelings during the last six months of lockdown (November 2020 to April 2021). In the second part, subjects had to judge the social appropriateness of the different behaviors in the ToG game (experiment 1) or played the ToG game (experiment 2). Afterwards, subjects had to answer questions on socio-demographics and personality traits. The latter and the priming part were designed to last about the same time so that we could simply swap the two parts to create a control and a priming group.

Our results show that the normative expectations of appropriate behaviors did not change after recalling social isolation experiences. When looking at norm compliance (i.e. actual behavior), we do not find a statistically significant effect for the full sample. About one fourth of our participants, however, reported to have hardly ever felt socially isolated during the lockdown. When excluding these subjects, we do find a statistically significant negative impact on donations. Subjects primed on their isolation experiences took on average 1.81 EUR from the charity whereas non-primed subjects took only 0.31 EUR. To explore whether subjects still care for norms, we additionally conducted two information treatments — combined with priming — in which we draw their attention to either the empirical or the normative expectation based on the findings by Bicchieri et al. (2022). The average amount taken declined to 0.71 EUR for the normative and 0.66 EUR for the empirical expectations but only for the latter treatment, the effect is statistically significant.

We conducted our experiment at the end of May in an area in which the "nationwide emergency brake" (Bundesnotbremse)[36] was already suspended with rapidly declining incidence rates and reopening of shops, restaurants and cultural activities. Hence, our findings show that persistent social distancing indeed causes a decline in prosociality even after the relaxation of social distancing rules and in times of optimism. At the same time, our results also contain good news since subjects still seem to care for norms and become more prosocial once again after we draw their attention especially to the empirical norm.

[36] For details, see: https://www.bundesregierung.de/breg-en/news/nationwide-emergency-brake-1889136

4.2 Experimental Design

4.2.1 Experiment 1 – Norm Elicitation

In the first experiment, we asked our participants to read the instructions of a hypothetical Take-or-Give (ToG) donation game similar to Dimant (2019) and Bicchieri et al. (2022). Both an active player (person A) and a passive charitable organization[37] receive an endowment of 5 EUR by the experimenter. As the active decision-maker, person A can overrule the equal split and take any desired amount from or give any desired amount to the charity, while restricting the choice set to integer values between -5 and +5. In order to elicit the normative expectations regarding the social appropriateness of taking from or giving money to the charity, we followed the experimental procedure introduced by Krupka and Weber (2013). After being familiarized with the ToG game, participants had to rate all 11 possible allocation options available to person A, using a four-point scale (very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, very socially appropriate). We clarified that "social appropriateness" should be understood as a behavior that most people would agree to be the "correct" thing to do in a given situation.

We designed this first experiment as an online survey experiment so that participation was possible anytime during the 12 days when the survey was live. Potential participants received their invitation via a subject unique survey link. The 275 participants spent on average slightly less than nine minutes on answering the whole survey. For their participation itself, they did not receive any payment. To reveal participants' true normative expectations, however, we incentivized their responses in the norm elicitation part by randomly choosing one allocation option as being payoff relevant after the experiment was completed. From all participants whose assessment corresponded to the modal response for the selected allocation option, we randomly picked 10 participants who received 50 EUR. The payoff procedure was thoroughly explained before ratings took place.[38]

In order to causally assess whether experiencing persistent social isolation affects normative expectations, we used the priming technique (for an overview about priming in economics, see Cohn and Maréchal 2016) to recollect subjects' experiences gained during the second and third

[37] We opted for a popular German charity named „Brot für die Welt" ("Bread for the world"), which indeed aims at ensuring food security but not exclusively. Hence, we added the following information for our participants: *"Brot für die Welt" supports more than 1500 projects in Africa, Asia, Latin America and Eastern Europe. The charity's goal is to ensure food security, the promotion of education and health, the strengthening of democracy, the respect for human rights, and the promotion of equality between men and women.*

[38] Descriptive statistics and all experimental instructions (translated from German) are provided in the Appendix C.

wave of the Covid-19 pandemic in Germany.[39] First, we implemented five questions to derive an index capturing individuals' willingness to adhere to social distancing rules as proposed by Pedersen and Favero (2020) — plus two additional questions on social contacts. On the next screen, we pointed participants to a recently published study by Clair et al. (2021), showing that especially young adults suffered from social isolation during the pandemic. Subsequently, we asked them to state how often they felt socially isolated (hardly ever, sometimes, often). In order to make own isolation experiences as salient as possible, we further posed two open questions in which subjects had the chance to describe a typical day or situation in which they felt particularly isolated (similar to Callen et al. 2014) and to name the social activity they missed the most.

In our *Prime* condition, participants answered the just described questions first, followed by the norm elicitation task. In a third part, we then posed questions on socio-demographic information, inclination towards reciprocal behavior (Perugini et al. 2003) and personality traits using the reduced form of the Big Five personality framework (Costa & McCrae 1989). In contrast, participants in the *NoPrime* condition had to answer the socio-demographic and personality traits questions first and the social distancing "priming questions" last. Even though we were not particularly interested in the personality traits, for example, we added these questions to ensure that the two parts, which we exchanged in order to create a control and a treatment group, lasted about the same time (i.e. about three minutes each). Random treatment assignment was ensured by the software "SoSci Survey" (Leiner 2019) at the individual level.

4.2.2 Experiment 2 – Norm Compliance

In the second experiment, participants actually played the ToG donation game and made a payoff-relevant allocation choice between themselves and the passive charity. Contrary to the first experiment, the second one was not designed as a survey experiment. The newly recruited 254 participants registered for one of the 19 online sessions that took place on eight days during the last two weeks of May 2021. We needed the session structure since, similar to Bicchieri et al. (2022) and Dimant (2019), we opted for a "pay-one" approach (Charness et al. 2016) to reduce concerns about peer interdependencies in the decision-making process. It was clearly stated that at the end of each experimental session, the allocation decision of only one randomly chosen subject would be executed, while all other subjects would instead receive a fixed payment of 5 EUR. All participants were informed about the possibility to receive a copy of the

[39] Similarly, Cappelen et al. (2021) have used a reminder about the crisis to investigate the pandemic's causal effect on people's views on solidarity and fairness.

donation recipe. Four comprehension questions ensured a thorough understanding of both the ToG allocation and payment procedure.

The construction of both the treatment and control group was identical to the one described in experiment 1 —using the same social distancing questions to prime subjects to recollect experiences of social isolation. In addition to the *Prime* and *NoPrime* condition, we introduced two additional information treatments to investigate the effect of norm reminders, explicitly manipulating either the empirical or the normative expectations, on subsequent norm compliance under priming. We utilize the findings derived by Bicchieri et al. (2022) to provide participants with explicit information on how other student subjects behaved or judged behavior in a very similar ToG donation game setting.[40] The given information before making the final decision reads as follows:

PrimeNormative condition: *"Participants in a recently published study stated in a very similar situation that it would be socially appropriate to keep the initial allocation or to share parts or the entire personal endowment with the charity organization."*

PrimeEmpirical condition: "*The majority of participants in a recently published study - facing a very similar situation - decided to keep the initial allocation or to share parts or the entire personal endowment with the charity organization."*

We implemented both the *PrimeNormative* and the *PrimeEmpirical* conditions since we had no clear prediction on which of the two is more likely to affect participants' behavior given the ambiguity in previous findings. Even though most studies pronounce the effectiveness of empirical expectations, some studies find no effect for empirical but for normative expectations only (see, e.g. Raihani and McAuliffe 2014) or that both are considered to be a strong mechanism of social influence (e.g. Minguez and Sese, 2021).[41]

At the beginning of the experiment, we reminded all participants that the experiment would last about 30 minutes and that they had to finalize the entire experiment to receive their final payoff. Participants earned on average 11.30 EUR (including a 3 EUR show-up fee) and each

[40] In the beginning of the experiment, we declared that all information shared within the study is true and could be verified by each participant upon request.

[41] We have pre-registered our study on aspredicted.org, see https://aspredicted.org/blind.php?x=gw6sx8. We have originally planned to implement a full factorial design but when observing the registrations for our experiment, we have quickly discovered that it would be rather impossible to obtain the necessary sample size in the time period we have set so that we decided to focus on the treatment conditions of main interest (i.e. the priming treatments and the no priming control condition) to ensure a reasonable power of the statistical tests. Based on the experimental design described in our pre-registration, we have also obtained the Institutional Review Board Certificate of the German Association for Experimental Economic Research e.V. (GfeW), No. 32c4CG7J.

experimental session lasted approximately 23 min.[42] Both experiments were conducted in May 2021 with subjects from the Magdeburg Experimental Laboratory of Economic Research (MaxLab) located at the University of Magdeburg, using "hroot" (Bock et al. 2014) for subject recruitment. The experiment was implemented using the computer program Otree (Chen et al. 2016). We ensured that subjects participated in only one of the two experiments. Payments were realized either by bank transfer or personal collection at the MaxLab (less than 5% opted for this option).

4.2.3 The Priming of Social Isolation

One common concern of priming studies is that the priming did not work as it was intended to. The open questions that we asked (i.e. to describe a typical day or situation in which they felt particularly isolated) allow us to check whether participants indeed report about social isolation or more generally about the pandemic, e.g. about economic anxiety (Binder 2020, Fetzer et al. 2021). The clear majority of participants describes to suffer from sitting alone at home, especially at the weekends. They report to spent most of their time sitting in front of the computer and to have contact with others only via video calls or instant messaging. Moreover, close social contacts with whom one might discuss problems or fears seem to be reduced to only one or a very few persons, contrary to before the pandemic. Finally, there is a significant number of individuals who report to miss especially visiting their families and having joint birthday parties, for example. The ones who stated to have hardly ever felt socially isolated are mostly those who are living in a shared apartment with other students. Hence, we cannot fully exclude that participants also have other thoughts after being asked about the lockdown, but the statements clearly show that they did recollect their own, personal social isolation experiences, if existent.

[42] After completing the above-described choice experiment, participants were asked to answer a survey covering an unrelated topic which was part of the research project "Meta-Analyse über Praktiken des Wissenstransfers in agglomerationsfernen Räumen", commissioned by the Leibniz Institute for Regional Geography. Hence, the payments made in this study were taken both from internal funds of the authors' home institution and the subcontract.

4.3 Results

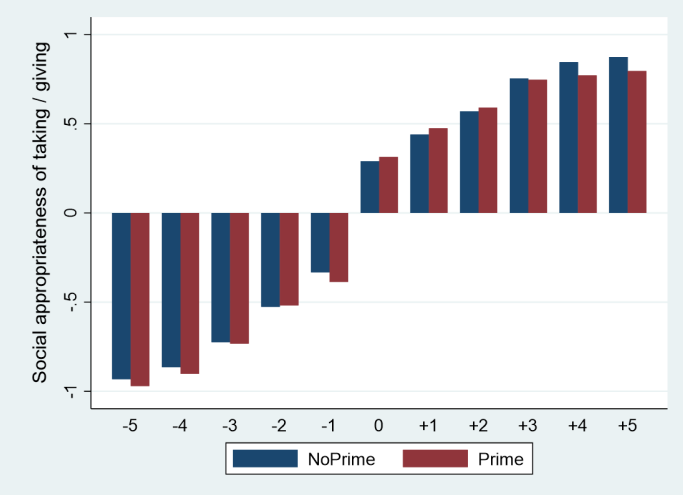
4.3.1 Norm Elicitation

Following the previous studies that use the Krupka-Weber method, we assign evenly-spaced numeric values from -1 to +1 to the four appropriateness ratings with +1 as "very socially appropriate". Separated by treatment group, Figure 6 shows the average value for each possible action in the ToG game. Taking money from the charity is seen as socially inappropriate (negative ratings) but less inappropriate for smaller amounts taken. Keeping the equal split and giving money to the charity are seen as socially appropriate (positive ratings) with giving all the allocated money to the charity as the most socially appropriate behavior. These results are in line with findings from previous studies such as Bicchieri et al. (2022).

When comparing the average ratings of the *NoPrime* (N = 138) with the *Prime* (N = 137) treatment, we only observe minor differences. Taking multiple hypothesis testing into account and adjusting p-values as proposed by List et al. (2019), we do not find any statistically significant differences (all p-values > 0.6).[43] The same is true for the reduced sample without subjects who are rather unlikely to respond to our priming intervention, i.e. the ones who felt hardly ever isolated during the lockdown —roughly one fifth of our sample. We also controlled for the observable characteristics that we collected during the experiment in an ordered probit regression model but again, there is no significant effect of the *Prime* treatment when correcting for multiple hypothesis testing. To summarize, we do not observe a causal impact of salient social distancing recollections on normative expectations about what behaviors others would approve or disapprove in this experimental setting.

[43] Calculating p-values with the Wilcoxon rank-sum test and applying e.g. the Bonferroni or the Benjamini-Hochberg procedure results in the same conclusion.

Figure 6: The Social Appropriateness of Possible Behaviors in the ToG Game



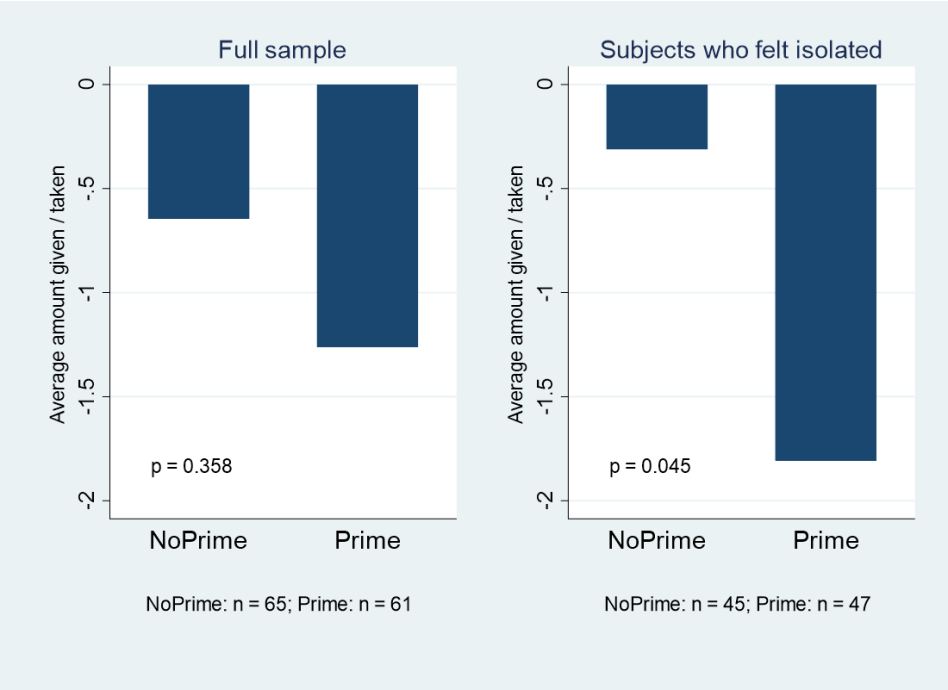
4.3.2 Norm Compliance

After showing that normative expectations have not been affected by social distancing recollections, we now turn to subjects' actual behavior in the ToG game. Even though the norm elicitation experiment clearly showed that taking money from the charity is seen as socially inappropriate, 47.64% of our subjects did so with an average amount taken of 3.81 EUR. Hence, almost half of our subjects actively engaged in violating the norm.

Figure 7 shows whether social distancing affected norm compliance by depicting the average donations for our main treatment groups, *NoPrime* and *Prime*. For the full sample in the left panel, we observe no statistically significant effect but the averages already indicate that norm compliance might diminish due to salient social distancing memories. Given that our priming intervention can reasonably recall social isolation experiences only for those who indeed felt isolated —which is not true for about one fourth of our participants— we show the results for this reduced sample in the right panel. The average amount taken increases from 0.31 EUR in the *NoPrime* to 1.81 EUR in the *Prime* treatment group. This difference becomes statistically significant and is also a medium effect size in terms of Cohen's d ($d = 0.448$).[44] The subsample analysis, however, was not stated in our pre-registration since we did not expect such a relatively high share of individuals claiming to have hardly ever felt isolated during such a long period of lockdown.

[44] As stated in our pre-registration, we have checked whether participants took the question seriously in which situation they have felt particularly isolated. Among those who have stated to have felt isolated during the lockdown, we have identified five individuals who were not able to mention any day or situation in which they indeed felt isolated. When excluding them from our sample, our results stay quantitatively and qualitatively the same, e.g. the effect size of the *Priming* treatment increases slightly to $d = 0.487$ with $p = 0.029$.

Figure 7: The Impact of Social Distancing on Norm Compliance



Note: *p*-values are obtained from Wilcoxon rank-sum tests.

Next, we check the robustness of our reduced sample results using regression analyses that are shown in Table 10. These clearly reveal the robust nature of our treatment effect across all OLS specifications, as well as the Tobit model to control for the censored choice set of decision-makers, as suggested by Engel (2011).[45]

[45] Given that the variable *having a job* is not perfectly balanced between the two groups, we also included an interaction term between having a job and being primed. As a result, the estimated treatment effect based on specification (3) from Table 10, for example, becomes significant even at the 1% level, without the interaction term being significant.

Table 10: Main Regression Results

	(1) OLS	(2) OLS	(3) OLS	(4) Tobit
Prime	-1.497**	-1.600**	-1.393**	-1.394***
	(0.536)	(0.604)	(0.598)	(0.482)
Female		1.823**	2.093***	2.293***
		(0.582)	(0.472)	(0.499)
Age		-0.030	-0.060	-0.044
		(0.089)	(0.069)	(0.064)
Adherence to social distancing rules (0/1)		1.593**	1.526**	1.317**
		(0.607)	(0.599)	(0.560)
More satisfied with life (0/1)		0.436	1.070	1.220
		(0.902)	(0.814)	(0.760)
Having a job		-0.993*	-1.240	-1.100
		(0.492)	(0.759)	(0.773)
Economics student		-0.914	-0.346	-0.334
		(0.687)	(0.545)	(0.550)
Constant	-0.311	-0.525	-2.590	
	(0.254)	(2.589)	(3.789)	
Big Five	No	No	Yes	Yes
N	92	92	92	92
Adjusted R^2	0.038	0.199	0.259	

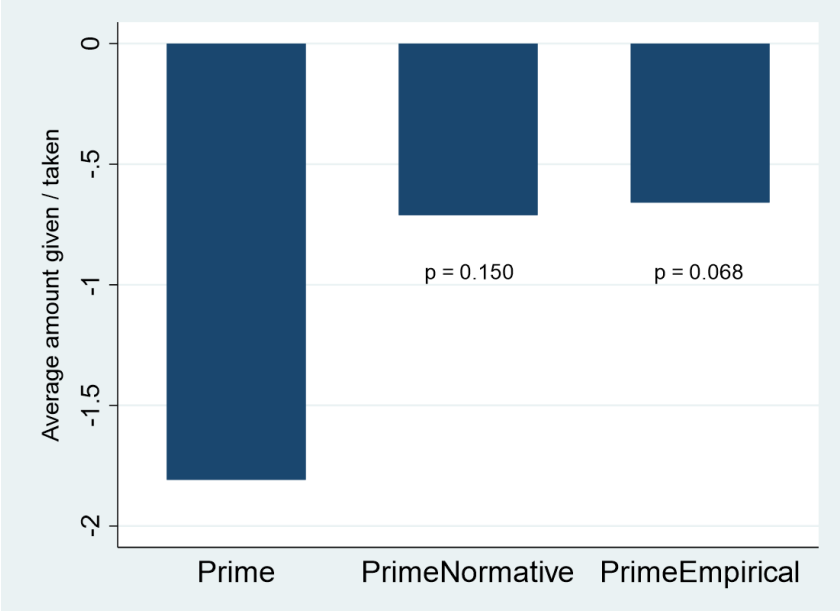
Note: Individuals who felt hardly ever isolated due to the social distancing rules have been excluded. "Adherence to social distancing rules" and "More satisfied with life" are dummy variables created based on a median split. Average marginal effects are shown for the Tobit model. Robust standard errors clustered on the session level in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Besides the robustness of our main treatment effect, our regression analyses reveal a rather large gender effect (the average donation of women is about 2 EUR higher) and a higher average donation of participants who declared to have followed the social distancing rules more strictly —measured by a 5-item index as proposed by Pedersen and Favero (2020).[46] Following Kimbrough and Vostroknutov (2016), individuals who care much about the norm in one setting (i.e. following the social distancing rules) also seem to care more about the norm in another setting (i.e. not taking money from the charity), even without norm reminders. This also fits to recent findings by Müller and Rau (2021) showing that pre-crisis social responsible behavior is positively related with compliance to social distancing. Given the findings by Ludeke et al. (2021) that especially the personality trait agreeableness is associated with sensitivity and obedience to norms, we also control for the Big Five in specifications (3) and (4) without any impact on our result. Moreover, when checking whether one of the five traits might moderate the treatment effect (results not shown here because of the explorative nature of this analysis and small sample sizes), we do not find any striking patterns.

[46] Due to the limited sample size, we applied a median-split for this analysis. This median-split results in an average social distancing index of 3.927 and 7.290 (out of a maximum of 10) for individuals who adhered less or more strictly to the social distancing rules during the lockdown.

The impact of norm reminders for subjects who felt isolated during the lockdown is shown in Figure 8 —the results for the full sample are, as before, statistically insignificant. The average amount taken from the charity declines to 0.71 EUR for the normative and to 0.66 EUR for the empirical expectations manipulation. Whereas these values are very close to each other, there is more variation in individuals' choices in the *PrimeNormative* condition (std = 3.643 vs. 2.987 in *PrimeEmpirical*) so that we do not observe a statistically significant treatment effect here, neither using the Wilcoxon rank-sum test nor by conducting regression analyses that are presented in Table 11. Hence, subjects seem on average to care a little more for empirical than normative expectations and highlighting the fact that the majority of individuals comply with the norm of not taking money from the charity makes primed subjects to behave similar to subjects who have not been primed on their social distancing experiences. The regression analyses (see Table 11) confirm the robustness of this finding. Taking into account the censored choice set of decision-makers by using the Tobit mode, the treatment effect of the *PrimeEmpirical* condition is even significant at the 1% level.

Figure 8: Sensitivity to Social Expectations



Note: Reduced sample of subjects who felt isolated. Prime: n = 47; PrimeNormative: n = 52; PrimeEmpirical: n = 47.

As regards the remaining controls, we again find a robust and economically significant gender effect, but the positive effect of participants who declared to have followed the social distancing rules more strictly has vanished. This finding, however, confirms that even if individuals are less sensitive to a norm (i.e. do not care much about the norm itself), salient social expectations make them to comply with the norm (Bicchieri 2016). As a result, the observed behavioral difference between individuals who are more or less sensitive to social norms disappears.

Finally, we find that economics students behave more selfishly, a rather common finding in experiments that involve monetary allocation decisions. Gerlach (2017) suggests that economists make lower offers because they expect others not to comply with the shared fairness norm and our data confirms this suggestion. Whereas the average amount taken by economists is 2.44 EUR and 2.10 EUR in the *Prime* and the *PrimeNormative* condition, respectively, it declines to 0.50 EUR in the *PrimeEmpirical* condition. When including interaction terms to our regression analysis, as shown in specification (5) in Table 11, it confirms that economics students do not care much for normative expectations —the point estimate for *PrimeNormative* is equal in size as the interaction term of being an economics student in the *PrimeNormative* condition but with reversed signs. However, empirical expectations seem to be even more important for them compared to the remaining sample (i.e. positive but insignificant point estimate of the interaction term). Admittedly, our sample size is not sufficiently big for such subsample analyses and, hence, the results have to be taken with care. However, this latter analysis indicates that not only the empirical expectations ($p = 0.064$) but also the normative expectations ($p = 0.087$) can help to overcome the negative effects of social distancing, at least for certain individuals.

Table 11: Sensitivity to Social Expectations — Robustness Checks

	(1) OLS	(2) OLS	(3) OLS	(4) Tobit	(5) Tobit
PrimeNormative	1.097	1.089	1.056	1.200	1.105*
	(0.672)	(0.634)	(0.744)	(0.737)	(0.645)
PrimeEmpirical	1.149*	1.003*	1.033**	1.146***	0.785*
	(0.558)	(0.560)	(0.456)	(0.390)	(0.423)
Female		2.389***	3.050***	3.155***	2.581***
		(0.769)	(0.602)	(0.602)	(0.480)
Age		0.016	0.019	0.049	0.167
		(0.050)	(0.056)	(0.054)	(0.046)
Adherence to social distancing rules (0/1)		0.575	0.408	0.244	0.373
		(0.495)	(0.517)	(0.497)	(0.423)
More satisfied with life (0/1)		0.0903	-0.184	-0.396	-0.175
		(0.522)	(0.524)	(0.494)	(0.419)
Having a job		0.0572	-0.129	-0.130	-0.080
		(0.592)	(0.555)	(0.516)	(0.447)
Economics student		-1.118**	-1.161**	-1.061**	-0.759
		(0.469)	(0.539)	(0.516)	(0.555)
Econ x *PrimeNormative*					-1.036
					(0.721)
Econ x *PrimeEmpirical*					0.313
					(0.649)
Constant	-1.809***	-3.650**	-0.608		
	(0.401)	(1.517)	(3.367)		
Big Five	No	No	Yes	Yes	Yes
N	146	146	146	146	146
Adjusted R^2	0.012	0.115	0.136		

Note: Individuals who felt hardly ever isolated due to the social distancing rules have been excluded. "Adherence to social distancing rules" and "More satisfied with life" are dummy variables created based on a median split. Average marginal effects are shown for the Tobit model. Robust standard errors clustered on the session level in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.4 Conclusion

Social norms are an important driver of individual behavior in many areas, whether in a variety of labor market contexts (Görges and Nosenzo 2020), in solving collective action problems such as fighting climate change (Doherty and Webler 2016) or in the question why people vote (Gerber et al. 2016). From a broader perspective, norms also play an important role for tax compliance and therewith affect the economy's development path (Varvarigos 2017). Guiso et al. (2015) even claim that informal institutions such as social norms are at least as important as formal institutions for national prosperity. The question therefore arises whether the massive changes in our everyday lives, caused by the Covid-19 pandemic and its imposed social distancing rules, had an impact on how we collectively perceive and follow shared norms. By conducting two experiments we investigate the effects of persistent social isolation on

normative expectations and norm compliance in the context of other-regarding behavior. Whereas the normative assessments of the behavioral options in a Take-or-Give donation game are basically unaffected, we find that norm compliance decreases substantially among those subjects who stated to have felt isolated after being primed to reflect on these particular memories.

Our paper contributes to the ongoing discussion about the societal and economic damages caused by the Covid-19 pandemic. On the one hand, our findings are alarming as they show a clear negative shift in the willingness to adhere to the norms we share as a society. Even though our study provides causal evidence for only one specific norm, Kimbrough and Vostroknutov (2016) observed that the preference for following norms carries over from one context to another unrelated context. Our data supports this narrative, as participants who had indicated that they mostly followed the social distancing rules during the lockdown took away about 1.30 EUR less from the charity than those for whom the distancing rules seemed less worthy to comply with. Therefore, the transferability of our results to other settings and norms is quite conceivable. We also cannot analyze any long-run behavioral effects with our experiment, it is, however, unlikely that everything will be forgotten in the moment the pandemic will be declared over. The reason is that we conducted our experiment at a time and in an area in which the "nationwide emergency brake" (Bundesnotbremse) was already suspended —leading to the abolishment of the night-time curfew, the permit for various outdoor activities (e.g. sports and cultural activities up to 25 or 100 participants, respectively) and the re-opening of restaurants and shops. Hence, we observed a decline in prosociality even after the relaxation of social distancing rules and in times of optimism, suggesting rather long-lasting behavioral distortions within affected groups. In case of future waves of infection, these findings should be taken into account —especially since models show that a strategic (i.e. with repeated contacts) reduction of interaction by only 50% decreases the number of infections sizably (Block et al. 2020).

On the other hand, our analysis can also be read in a more optimistic way. First, even severe social isolation of about six months seems to be incapable of substantially changing the basic norms we uphold. Second, we reveal that simple (especially empirical) norm reminders can in turn increase prosocial behavior once again. We therewith not only show that people still care for the norms of their society but also offer a potential path to overcome some of the detrimental effects caused by the Covid-19 pandemic. Hence, our study may help to underline the value of putting exemplary behavior (e.g. voluntary work) into the societal spotlight, as it can serve as a powerful instrument to buffer the less obvious behavioral damages caused by social distancing

in times of crisis. Such a measure, however, might only work as long as the erosion of norm compliance is not yet broadly visible since observed norm violations seem to have much larger effects on individuals' willingness to comply with a norm than observed norm compliance (Bicchieri et al. 2022, Dimant 2019).

Even though our paper primarily helps to understand changes in human behavior caused by the pandemic, our results call for further research in a world that becomes more digitalized every day. Our student subject pool was certainly able to stay in touch with their family and friends via video telephony, social media and alike. Still, they perceived to be socially isolated. Even though the lockdown was an extreme situation, the use of social media has surely replaced one or the other in-person interaction already before the pandemic, as daily time spent on social networking has increased from 90 minutes in 2012 to 145 minutes in 2019.[47] If working from home is likely to stay after the pandemic —the software company SAP even announced to give its employees complete freedom in the choice where to work[48]— in-person interactions will decline even further. Hence, our findings point to an important aspect to consider e.g. when discussing the pros and cons of remote working. Here, in addition to potential productivity losses or the disruption of workers' work-life balance, future research should be interested in asking how much in-person interaction might be necessary to prevent the impending behavioral damages caused by (perceived) social isolation. Already small negative effects on norm compliance might add up to substantial damages from both a societal and economical perspective if not only the norm of prosociality is affected but norm compliance in more general.

[47] https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/
[48] https://www.dw.com/en/working-from-home-a-new-status-symbol/a-57797924

References

Bian, B., **Li**, J., **Xu**, T., & **Foutz**, N. Z. (2022). Individualism during crises. *Review of Economics and Statistics*, 104(2), 368-385.

Bicchieri, C. (2016). Norms in the wild: How to diagnose, measure, and change social norms. Oxford University Press.

Bicchieri, C., **Dimant**, E., **Gächter**, S. & **Nosenzo**, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior,* 132, 59-72.

Binder, C. (2020). Coronavirus fears and macroeconomic expectations. *Review of Economics and Statistics*, 102(4), 721-730.

Bland, A. R., **Roiser**, J. P., Mehta, M. A., **Sahakian**, B. J., **Robbins**, T. W., & **Elliott**, R. (2022). The impact of COVID-19 social isolation on aspects of emotional and social cognition. *Cognition and Emotion*, 36(1), 49-58.

Block, P., **Hoffman**, M., **Raabe**, I.J., **Dowd**, J.B., **Rahal**, C., **Kashyap**, R. & **Mills**. M.C. (2020). Social network-based distancing strategies to flatten the Covid-19 curve in a post-lockdown world. *Nature Human Behavior*, 4, 588-596.

Bock, O., **Nicklisch**, A. & **Baetge**, I. (2014). Hroot: Hamburg registration and organization online tool. *European Economic Review*, 71, 117–120.

Brodeur, A., **Clark**, A.E., **Fleche**, S.F. & **Powdthavee**, N. (2021). Covid-19, lockdowns and well-being: Evidence from Google trends. *Journal of Public Economics*, 193, 104346.

Callen, M., **Isaqzadeh**, M., **Long**, J. & **Sprenger**, C. (2014). Violence and risk preference: Experimental evidence from Afghanistan. *The American Economic Review*, 104(1), 123-148.

Cappelen, A. W., **Falch**, R., **Sørensen**, E. Ø., & **Tungodden**, B. (2021). Solidarity and fairness in times of crisis. *Journal of Economic Behavior & Organization*, 186, 1-11.

Casoria, F., **Galeotti**, F. & **Villeval**, M. (2021). Perceived social norm and behavior quickly adjusted to legal changes during the COVID-19 pandemic. *Journal of Economic Behavior & Organization,* 190, 54-65.

Charness, G., **Gneezy**, U. & **Halladay**, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, 131:141–150.

Chen, D.L., **Schonger**, M. & **Wickens**, C. (2016). oTree - An open-source platform for laboratory, online and field experiments. *Journal of Behavioral and Experimental Finance*, vol 9: 88-97.

Christensen, P. N., **Rothgerber**, H., **Wood**, W. & **Matz**, D. C. (2004). Social norms and identity relevance: A motivational approach to normative behavior. *Personality and Social Psychology Bulletin*, 30(10), 1295-1309.

Clair, R., **Gordon**, M., **Kroon**, M. & **Reilly**, C. (2021). The effects of social isolation on well-being and life satisfaction during pandemic. *Humanities & Social Sciences Communications*, 8(1), 1-6.

Cohn, A. & **Maréchal**, M. A. (2016). Priming in economics. *Current Opinion in Psychology*, 2016, 12, 17–21.

Costa, P. T. & **McCrae**, R. R. (1989). The NEO-PI/NEO-FFI manual supplement. Odessa, FL.: Psychological Assessment Resources.

Dimant, E. (2019). Contagion of pro- and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology*, 73, 66-88.

Doherty, K.L. & **Webler**, T.N. (2016). Social norms and efficacy beliefs drive the alarmed segment's public-sphere climate actions. *Nature Climate Change*, 6, 879-884.

Eckel, C., **Hoover**, H., **Krupka**, E., **Ojumu**, O., **Rosenblatt**, T., **Sinha**, N. & **Wilson**, R.K. (2021). Norm and behavior change among college aged republicans and democrats in response to Covid-19. Manuscript.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14, 583– 610.

Fetzer, T., **Hensel**, L., **Hermle**, J. & **Roth**, C. (2021). Coronavirus perceptions and economic anxiety. *The Review of Economics and Statistics*, 130(5), 968-978.

Gerber, A.S., **Huber**, G.A., **Doherty**, D. & **Dowling**, C.M. (2016). Why people vote: Estimating the social returns to voting. *British Journal of Political Science*, 46(2), 241-264.

Gerlach, P. (2017). The games economists play: Why economics students behave more selfishly than other students. *PLoS ONE*, 12(9), e0183814.

Görges, L. & **Nosenzo**, D. (2020). Social norms and the labor market. In Zimmermann, K. (Ed), Handbook of Labor, Human Resources and Population Economics, Springer, Cham.

Guiso, Luigi, **Sapienza**, P. & **Zingales**, L. (2015). Corporate culture, societal culture, and institutions. *American Economic Review: Papers & Proceedings*, 105(5), 336-339.

Hensel, L., **Witte**, M., **Caria**, A.S., **Fetzer**, T., **Fiorin**, S., **Götz**, F.M., **Gomez**, M., **Haushofer**, J., **Ivchenko**, A., **Kraft-Todd**, G., **Reutskaja**, E., **Roth**, C., **Yoeli**, E. & **Jachimowicz**, J.M. (2022). Global behaviors, perceptions, and the emergence of social norms at the onset of the COVID-19 pandemic. *Journal of Economic Behavior & Organizations*, 193, 473-496.

Kimbrough, E. O. & **Vostroknutov**, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14, 608–638.

Krupka, E. L. & **Weber**, R. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3), 495-525.

Leiner, D. J. (2019). SoSci Survey (Version 3.1.06) [Computer software]. Available at https://www.soscisurvey.de.

List, J.A., **Shaikh**, A.M. & **Yang**, X. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22, 773-793.

Ludeke, S. G., **Vitriol**, J. A., **Larsen**, E. G. & **Gensowski**, M. (2021). Personality in a pandemic: Social norms moderate associations between personality and social distancing behaviors. *Personality and Individual Differences*, 177, 110828.

Minguez, A. & **Sese**, F.J., (2021). Understanding the effectiveness of social influence appeals in charitable giving: The roles of affinity with the cause, and past giving behavior. *Journal of Marketing Theory and Practice, 29(3),375-386*.

Müller, S. & **Rau,** H. (2021). Economic preferences and compliance in the social stress test of the Covid-19 crisis. *Journal of Public Economics*, 194, 103422.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, 14(3), 137-158.

Pedersen, M. J. & **Favero**, N. (2020). Social distancing during the Covid-19 pandemic: Who are the present and future noncompliers? *Public Administration Review*, 80(5), 805-814.

Perugini, M., **Galucci**, M., **Presaghi**, F. & **Ercolani**, A. P. (2003). The personal norm of reciprocity. *European Journal of Personality*, 17(4), 251-283.

Raihani, N.J. & **McAuliffe** K. (2014). Dictator game giving: the importance of descriptive versus injunctive norms. *PLoS ONE*, 9(12), e113826.

Shachat, J., **Walker**, M.J. & **Wie**, L. (2021). The impact of the Covid-19 pandemic on economic behaviours and preferences: Experimental evidence from Wuhan. *AEA Papers and Proceedings,*111,302-306.

Varvarigos, D. (2017). Cultural norms, the persistence of tax evasion, and economic growth. *Economic Theory*, 63, 961-995.

Appendix C

C.1 Descriptive Statistics

Table 12: Norm Elicitation

	NoPrime	Prime	p-value
Female	0.500	0.504	0.952
Age	25.297	25.182	0.748
Adherence to social distancing rules	5.629	5.826	0.412
Feeling isolated due to distancing	0.797	0.832	0.455
Current life satisfaction	7.130	6.708	0.112
Having a job	0.587	0.701	0.049
Economics student	0.217	0.095	0.005
Conscientiousness	5.031	5.083	0.323
Agreeableness	5.331	5.411	0.234
Openness	5.000	4.842	0.394
Extraversion	4.529	4.805	0.138
Neuroticism	4.290	4.304	0.685
N	138	137	

Note: *p*-values obtained from Kruskal-Wallis equality-of-populations rank tests and Chi^2 tests, respectively.

Table 13: Take-or-Give Game

	NoPrime	Prime	Prime Normative	Prime Empirical	p-value
Female	0.508	0.508	0.578	0.516	0.828
Age	25.323	25.590	25.516	25.313	0.603
Adherence to social distancing rules	5.357	5.495	5.325	5.991	0.169
Feeling isolated due to distancing	0.692	0.770	0.813	0.734	0.437
Current life satisfaction	6.877	5.705	5.984	6.266	0.012
Having a job	0.508	0.672	0.672	0.516	0.080
Economics student	0.277	0.230	0.188	0.234	0.694
Conscientiousness	5.256	5.448	5.078	4.953	0.086
Agreeableness	5.195	5.322	5.203	5.411	0.446
Openness	5.072	4.913	4.865	4.594	0.170
Extraversion	4.836	4.672	4.766	4.469	0.556
Neuroticism	4.164	4.322	4.464	4.453	0.605
N	65	61	64	64	

Note: p-values obtained from Kruskal-Wallis equality-of-populations rank tests and Chi² tests, respectively.

Table 14: Take-or-Give Game —Subgroup of Individuals Who Have Felt Isolated

	NoPrime	Prime	Prime Normative	Prime Empirical	p-value
Female	0.578	0.553	0.577	0.617	0.939
Age	24.889	25.319	25.712	24.426	0.151
Adherence to social distancing rules	5.311	5.485	5.404	6.102	0.151
Current life satisfaction	6.489	5.553	5.558	5.681	0.097
Having a job	0.511	0.660	0.673	0.532	0.238
Economics student	0.311	0.191	0.192	0.255	0.462
Conscientiousness	5.304	5.539	5.026	4.943	0.034
Agreeableness	5.222	5.355	5.147	5.390	0.530
Openness	4.970	5.028	4.647	4.532	0.152
Extraversion	4.889	4.837	4.660	4.426	0.338
Neuroticism	4.370	4.397	4.596	4.709	0.616
N	45	47	52	47	

Note: p-values obtained from Kruskal-Wallis equality-of-populations rank tests and Chi^2 tests, respectively.

C.2 Instructions

(*Instructions translated from German*)

General Instructions (i.e. Welcome Page)

Experiment 1 - Norm elicitation

Thank you for participating in this study. You can only answer this survey once. You will need approximately 10 minutes to complete.

This study consists of three parts. In the second part, you will be asked to think yourself into a particular situation —the more successful you are, the higher your chances of winning 50 €. Overall, we pay out 10x50 € prices via bank transfer. You will receive all necessary information in the second part.

The prerequisite to receive any compensation is the completion of all survey questions. Please answer all questions faithfully. Your answers will be used in an anonymized form in order to guarantee your privacy.

Experiment 2 -Take-or-Give Game

Welcome to the experiment. Please note that you have to finalize the entire experiment to receive your final payoff (3 € show-up fee + subsequent earnings). As usual in the MaxLab, all shared information in this experiment are true and can be verified by you upon request.

The entire experiment consist of four short parts (A, B, C, D). You will need around 30 minutes to complete the entire experiment. All necessary information will be shared on your screen.

(Part D consisted of an unrelated survey study not discussed in the paper)

Priming & General Questionnaire *(identical for both experiments)*

Priming (Part I in *Prime* condition or Part III in *NoPrime*, respectively)

We want to ask you some questions regarding a topic that currently concerns all of us:
the Covid-19 pandemic and its imposed social distancing measures.

Please think about the last five months starting with the beginning of the current lockdown and indicate with the help of the provided scale ranging from 0 to 10 (0 = "strongly disagree" to 10 = "strongly agree") to what extent you agree with the following statements (f*ollowing Pedersen & Favero, 2020)*:

- I met with friends or relatives who live outside my own household.
- I made the fewest possible trips to the grocery store.
- I was at places where other people were as well (café, restaurant, specialty shops, church, etc.).
- I avoided all social gatherings and adhered to the 'social distancing' rules.
- I strongly encouraged others to avoid all social contact and to adhere to the 'social distancing' rules.

How is your current living situation?
(*alone; shared apartment; with family / partner; different living situation*)

Do you have a part-time job with social contacts?
(no part-time job; remote part-time job; part-time job with direct social contact)

[On the next screen]

A recently published study showed that due to social distancing measures especially young adults felt isolated and experienced negative effects on their psychological well-being. *(Study: Clair, Gordon, Kroon und Reilly in Humanities and Social Sciences Communications, 2021)*

Hence, we now would like to know how *you* feel due to the imposed social distancing measures.

- How often have you felt socially isolated? (*hardly ever; sometimes; often*)

Please use the provided scale ranging from 0 to 10 (0 = "strongly disagree" to 10 = "strongly agree") to indicate to what extent you agree with the following statement:

- Due to growing feeling of being socially isolated, I started to adhere less to social distancing measurements than in the beginning of the lockdown.

Please use the provided scale ranging from 0 to 10 (0 = " not satisfied at all" to 10 = "completely satisfied") to answer the following question:

- How satisfied are you at present, all in all, with your life?

Please describe a situation (or a typical day) in which you felt (or feel) particularly socially isolated: ____

Which social activity do you miss the most: ____

General Questionnaire (Part III in *Prime* condition or Part I in *NoPrime*, respectively)

Please answer the following questions.

- Please indicate your age.
- Please indicate your gender.
- Which university do you belong to?
- Which faculty do you belong to?
- What degree are you striving for?

Reciprocal inclination *(6 items, following Perugini et al., 2003)*

Big five personality traits *(15 items, following Costa & McCrae, 1989)*

Norm Elicitation *(Part II in experiment 1)*

In the following, you will receive the instruction to an experiment, which has been run multiple times in a similar way. Please read the description of the experiment carefully. You will be asked to evaluate all possible choices in regard to their "social appropriateness". By social appropriate we mean behavior that most people agree is the "correct" or "ethical" thing to do.

The evaluation can be conducted on a scale ranging from "very socially inappropriate" to "very socially appropriate". You have the possibility to vary your evaluation between these two extremes.

Based on your answers and on the answers provided by all other survey participants, your payoff will be determined. In the end of the experiment, we will select one of the choice options randomly. For this selected choice, we will calculate the most selected answer (mode). If your answer corresponds to the calculated mode value, you will qualify for the 50 € payment. Among all qualified participants, we will randomly draw 10 individuals who receive 50 €.

On the following page of will find the description of the experiment and all the possible options you need to evaluate.

Important: You are asked to assess how the majority of people evaluates the different choices. The better your estimates are, the higher are your chances of winning 50 €.

[On the next screen]

In the experiment, a participant ("person A") is being matched with the charity organization "Brot für die Welt". Person A and "Brot für die Welt" both receive 5 € from the experimenter.

"Brot für die Welt" supports more than 1500 projects in Africa, Asia, Latin America and Eastern Europe. The charity's goal is to ensure food security, the promotion of education and health, the strengthening of democracy, the respect for human rights, and the promotion of equality between men and women."

The decision regarding the final allocation of the money (10 €) falls to person A. Therefore, person A determines her own personal payoff and the donation size "Brot für die Welt" receives:

- Person A can take parts or the entire endowment (5 €) allocated to "Brot für die Welt".
- Person can keep the equal allocation between both parties.
- Person A can give parts or her entire endowment (5 €) to "Brot für die Welt".

Please assess all possible options person A is facing.

Please indicate for each option whether the choice is being assessed as "very socially inappropriate", "somewhat socially inappropriate", "somewhat socially appropriate", or "very socially appropriate" by the majority of people.

	sehr sozial unangemessen	eher sozial unangemessen	eher sozial angemessen	sehr sozial angemessen
Person A nimmt 5 € von „Brot für die Welt“ (Person A erhält 10€ / „Brot für die Welt“ erhält 0€)	○	○	○	○
Person A nimmt 4 € von „Brot für die Welt“ (Person A erhält 9€ / „Brot für die Welt“ erhält 1€)	○	○	○	○
Person A nimmt 3 € von „Brot für die Welt“ (Person A erhält 8€ / „Brot für die Welt“ erhält 2€)	○	○	○	○
Person A nimmt 2 € von „Brot für die Welt“ (Person A erhält 7€ / „Brot für die Welt“ erhält 3€)	○	○	○	○
Person A nimmt 1 € von „Brot für die Welt“ (Person A erhält 6€ / „Brot für die Welt“ erhält 4€)	○	○	○	○
Person A nimmt / gibt 0 € von / an „Brot für die Welt“ (Person A erhält 5€ / „Brot für die Welt“ erhält 5€)	○	○	○	○
Person A gibt 1 € an „Brot für die Welt“ (Person A erhält 4€ / „Brot für die Welt“ erhält 6€)	○	○	○	○
Person A gibt 2 € an „Brot für die Welt“ (Person A erhält 3€ / „Brot für die Welt“ erhält 7€)	○	○	○	○
Person A gibt 3 € an „Brot für die Welt“ (Person A erhält 2€ / „Brot für die Welt“ erhält 8€)	○	○	○	○
Person A gibt 4 € an „Brot für die Welt“ (Person A erhält 1€ / „Brot für die Welt“ erhält 9€)	○	○	○	○
Person A gibt 5 € an „Brot für die Welt“ (Person A erhält 0€ / „Brot für die Welt“ erhält 10€)	○	○	○	○

Take-or-Give Game *(Part II in experiment 2)*

In part B of the experiment, you are asked to make an active decision which determines your final payoff.

In this part, you are the active decision-maker and you have been matched with the passive receiver —the charity organization “Brot für die Welt”.

“Brot für die Welt” supports more than 1500 projects in Africa, Asia, Latin America and Eastern Europe. The charity’s goal is to ensure food security, the promotion of education and health, the strengthening of democracy, the respect for human rights, and the promotion of equality between men and women.”

You and “Brot für die Welt” both receive 5 €.

This means that 5 € will be added to your final payoff while "Brot für die Welt" receives a 5 € donation made by the MaxLab after the finalization of the experiment.

(The donation receipt will be send to you upon request. Please turn to the MaxLab if you want to receive the final receipt!)

You have the chance to change the initial allocation of the sum of 10 € to your desire.

[On the next screen]

Important payoff information:

We will randomly select one participant from today's session, whose allocation decision will be executed! All other participants receive a fixed endowment of 5 € instead. For the non-selected individuals, no donations will be made.

- You can take parts or the entire endowment (5 €) allocated to "Brot für die Welt".
- You can keep the equal allocation of money.
- You can give parts or the entire endowment (5 €) allocated to you to "Brot für die Welt".

Before making your final allocation decision, please answer the following comprehension questions:

- Imagine you decide to give 4 € to "Brot für die Welt". What would be your final payoff and the donation made to "Brot für die Welt", if the computer picks you at the end of the experiment?
- Imagine you decide to take 4 € from "Brot für die Welt". What would be your final payoff and the donation made to "Brot für die Welt", if the computer picks you at the end of the experiment?
- Imagine you decide to give 2 € to "Brot für die Welt". What would be your final payoff and the donation made to "Brot für die Welt", if the computer doesn't pick you at the end of the experiment?
- Imagine you decide to keep the equal allocation between you and "Brot für die Welt". What would be your final payoff and the donation made to "Brot für die Welt", if the computer picks you at the end of the experiment?

[On the next screen]

Now it is your turn. Please decide how you want to allocate the money (10 €) between you and "Brot für die Welt".

[Additionally for PrimeNormative: Participants in a recently published study stated in a very similar situation that it would be socially appropriate to keep the initial allocation or to share parts or the entire personal endowment with the charity organization.*]*

[Additionally for PrimeEmpirical: The majority of participants in a recently published study, facing a very similar situation, decided to keep the initial allocation or to share parts or the entire personal endowment with the charity organization.*]*

Please make your choice now:

Handlungsoption		Auszahlung
Ich nehme 5 € von Brot für die Welt	○	Sie erhalten 10 € Brot für die Welt erhält 0 €
Ich nehme 4 € von Brot für die Welt	○	Sie erhalten 9 € Brot für die Welt erhält 1 €
Ich nehme 3 € von Brot für die Welt	○	Sie erhalten 8 € Brot für die Welt erhält 2 €
Ich nehme 2 € von Brot für die Welt	○	Sie erhalten 7 € Brot für die Welt erhält 3 €
Ich nehme 1 € von Brot für die Welt	○	Sie erhalten 6 € Brot für die Welt erhält 4 €
Ich nehme / gebe 0 € von / an Brot für die Welt	◉	Sie erhalten 5 € Brot für die Welt erhält 5 €
Ich gebe 1 € an Brot für die Welt	○	Sie erhalten 4 € Brot für die Welt erhält 6 €
Ich gebe 2 € an Brot für die Welt	○	Sie erhalten 3 € Brot für die Welt erhält 7 €
Ich gebe 3 € an Brot für die Welt	○	Sie erhalten 2 € Brot für die Welt erhält 8 €
Ich gebe 4 € an Brot für die Welt	○	Sie erhalten 1 € Brot für die Welt erhält 9 €
Ich gebe 5 € an Brot für die Welt	○	Sie erhalten 0 € Brot für die Welt erhält 10 €

You will be informed by the end of the experiment, whether your allocation decision will be executed.

5. Concluding Remarks

Unethical behavior may come in different forms and shapes and has the potential to cause severe damage to the economy. The economic costs make it essential to derive a clear picture of why individuals choose to act unethical, allowing policy makers to build upon when drafting measures to contain these detrimental effects. The aim of this thesis is to add further pieces to the literature, paving the long road to a more conclusive understanding of this particular aspect of human behavior. All three articles presented in this thesis build on the idea that individual economic decision-making is a dynamic and rather malleable process. Especially across settings in which opposing motives are colliding, experiences or specific circumstances may allow decision-makers to justify otherwise questionable behavior. Chapter 2 and 3 causally test this narrative and investigate whether individuals utilize the experience of prior mistreatment to justify unethical behavior. In contrast to Chapter 4, which examines how profound societal changes influence the normative evaluation of behavior and the willingness to adhere to these norms.

The article presented in Chapter 2, is the first to causally address the relevance of generalized negative reciprocity in a competitive workplace setting. The analysis reveals the determining role of the target's identity on whether or not people justify the displacement of unkindness towards innocent others. Our findings show that under certain circumstances, unkind supervisor treatment can set off a subsequent chain of unkindness, potentially endangering firm productivity. More importantly, the paper emphasizes the sensitivity of normative considerations across different social constellations and stresses the crucial role of personal relations in the workplace. This points to the importance for firms to strengthen the social ties between employees also as a means to reduce the risk of spreading unethical and counterproductive behavior across the workforce.

Chapter 3 presents an experimental paper, which investigates whether the discriminatory imposition of a costly rule across equivalent peers impairs the willingness to comply with the given rule. In contrast to previous studies, I focus on the psychological effect of discrimination, while keeping the monetary disparities resulting from unequal treatment to a minimum. I find that groups who are subject to a discriminatory rule regime are not more likely to violate the rule, compared those groups who enjoy equal treatment. From a societal perspective, this apparent insensitivity towards unequal rule settings offers some good news. Given that rules may often appear to be unequally valid in the public eye due to incomplete information or

miscommunication, my findings suggest a rather stable adherence level. An open question is however, whether this unpredicted indifference against public rule discrimination holds against other measures of rule following (apart from lying). If so, does it hold in the field as well?

In Chapter 4, we investigate the effect of persistent social isolation, experienced during Covid-19 pandemic, on the normative evaluation of both pro- and antisocial behavior and the willingness to comply with these norms. We find that salient recollections of social isolation did not shift the normative perception of behavior in our experiment. However, our results show a decline in norm compliance as subjects behave less prosocial after recalling isolation memories. First, the paper uncovers less apparent behavioral changes in the wake of persistent social distancing and complements our understanding of the overall economic damages caused by the Covid-19 pandemic. More importantly however, we exploit the pandemic to illustrate how far-reaching societal changes can quickly alter behavior across relevant domains. The described detrimental effects of social isolation further stress the value of actual face-to-face interactions in daily life, an insight particularly interesting in a world that becomes more digitalized every day. Researchers as well as practitioners should keep these potential risks in mind, for example, when weighing the costs and benefits of remote work arrangements.

Clearly, an ever-changing world creates an ever-changing reality. This urges us to constantly reassess findings from different perspectives and against new backgrounds. In that regard, the three research articles presented in this thesis add some interesting findings to the literature and contribute to a better understanding of how different settings and experiences may affect the willingness to act unethical.