WORKING PAPERS IN ECONOMICS AND STATISTICS

Report No 08 | 2024

Norm Violations and Behavioral Spillovers: Evidence From the Lab and The Field

Sebastian J. Goerg, Oliver Himmler and Tobias König

Linnæus University

Norm Violations and Behavioral Spillovers: Evidence From the Lab and The Field

Sebastian J. Goerg, Oliver Himmler and Tobias König

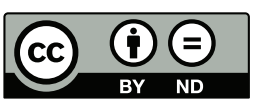
Norm Violations and Behavioral Spillovers: Evidence From the Lab and The Field

Sebastian J. Goerg, Oliver Himmler and Tobias König

Series editor: Thomas Giebe

ISBN: 978-91-8082-171-1 (pdf)
DOI: https://doi.org/10.15626/ns.wp.2024.08

Report No 08, Department of Economics and Statistics, Linnaeus University, Växjö, 2024



NORM VIOLATIONS AND BEHAVIORAL SPILLOVERS
– EVIDENCE FROM THE LAB AND THE FIELD*

Sebastian J. Goerg[1,2], Oliver Himmler[3], Tobias König[4,5]

April 2024

Abstract

This paper explores the contagion effects of norm-violating behavior across decision situations. Through a series of laboratory and field experiments, we empirically establish the conditions under which norm-breaking behavior in one decision situation leads individuals who observe this to violate norms in other, distinct decision situations. Our laboratory findings show that these spillover effects are more pronounced when the norms underlying the decision situations are perceived to be similar. However, spillovers can also affect decision situations governed by relatively dissimilar norms if the observers of norm violations have had the opportunity to first violate the same norm as the observed violator themselves. In an accompanying field experiment, we underscore the economic importance of norm similarity for spillover effects. When workers are exposed to information about celebrities evading taxes, they exhibit significantly higher rates of workplace theft than those in the control group, yet this exposure does not negatively affect work morale.

JEL classification: A12, C93, D01, D03

Keywords: Norms, Cheating, Peer Effects, Tax Evasion, Workplace Theft, Work Effort, Conditional Compliance, Unethical Behavior.

---

* We would like to thank the participants of the 2014 Southern Economic Meeting in Atlanta, the 2015 Social Norms and Institutions conference in Ascona, the 2016 ESA North America Meeting in Tuscon, the 2018 BBE Seminar in Berlin, and the 2020 CNEE Workshop in Copenhagen for their valuable comments. Financial support from the WZB Social Science Center Berlin, Hannover University (Ways-into-Science grant), and Florida State University is greatly acknowledged. We used AI-based technologies for proofreading.

[1] TUMCS for Biotechnolgy and Sustainability, Technical University Munich, Straubing, Germany, s.goerg@tum.de

[2] TUM School of Management, Technical University Munich, Munich, Germany

[3] University of Erfurt, Germany, oliver.himmler@uni-erfurt.de

[4] Linnaeus University, Sweden, tobias.koenig@lnu.se

[5] WZB Social Science Center Berlin, Germany

Introduction

Norms are a fundamental aspect of society, serving as guidelines for what constitutes appropriate behavior in a given situation. However, individuals frequently also violate these norms. In recent years, economists have extensively studied the effects of norm-breaking or rule violations on the behavior of others. When individuals break a norm, it erodes norm compliance of others facing the same decision situation. In fact, there is a body of evidence, both from the lab and the field, which suggests that observing norm violations can result in such negative peer effects in diverse situations involving, e.g., fair sharing, donations to charity, cheating on taxes, or exerting effort.[1]

There is evidence that contagion effects of norm violations may not be confined to a particular decision-making situation but can also unfold across decision situations and contexts. In a seminal paper, Keizer, Lindenberg and Steg (2008) demonstrated that environmental disorder, such as graffiti, increased the likelihood of littering and theft. Recently, Gulino and Masera (2023) found that local corruption scandals increase the propensity of supermarket customers to steal. Even though such norm spillovers have attracted the interest of researchers from multiple disciplines, there is no evidence regarding the conditions under which they manifest. It remains unclear when norm-breaking behavior in one decision situation prompts other individuals to violate norms in a different decision situation, and which factors may allow norm violations to spread far to relatively dissimilar decision situations.

This paper presents first evidence on the conditions that increase the likelihood of norm spillovers, derived from a series of laboratory experiments and an important real-world application: the workplace. We define norm spillovers as instances where individuals, by breaking an injunctive norm that governs appropriate behavior in one decision situation, influence other individuals who observe this to subsequently become less compliant with an injunctive norm in a different decision situation.[2] While the

[1] The lab evidence comes from a wide range of settings, including dictator games (e.g., Cason and Mui, 1998; Bicchieri and Xiao, 2009; Krupka and Weber, 2009), gift-exchange games (Gächter and Thöni, 2010; Thöni and Gächter, 2015; Bicchieri, Dimant, Gächter and Nosenzo, 2022), lying and cheating games (Rauhut, 2013; Diekmann, Przepiorka and Rauhut, 2015; Robert and Arnab, 2013), investment games (Mittone and Ploner, 2011), take-or-give donation games (Dimant, 2019; Bicchieri, Dimant, Gächter and Nosenzo, 2022), and public goods games where many individuals exhibit "conditional cooperation" (see, e.g., Fischbacher, Gächter and Fehr, 2001; Fischbacher and Gächter, 2010; Gächter and Thoeni, 2011). The importance of conditional cooperation has also been shown in a number of field experiments where individuals are confronted with what others have done in the same decision situation. Specifically, peer donations increase own donations (Frey and Meier, 2004; Shang and Croson, 2009; Chen, Harper, Konstan and Li, 2010; Smith, Windmeijer and Wright, 2015), tax evasion breeds more tax evasion (e.g., Alm, Bloomquist and McKee, 2017; Lefebvre, Pestieau, Riedl and Villeval, 2015), and lower work effort reduces the work effort of social peers (Ichino and Maggi, 2000; Falk and Ichino, 2006; Mas and Moretti, 2009).

[2] Keizer, Lindenberg and Steg (2008) define norm spillovers as the spread of disorder across norms (rather than situations). We employ an approach consistent with Krupka and Weber (2013), where the decision situations are characterized by certain

literature on norm spillovers so far leaves the interpretation of what constitutes different decision situations or contexts to the researcher, we employ an empirical criterion: using incentive-compatible belief elicitation methods, we classify decision situations as different, based on whether and to what extent the norms in those situations are perceived as different. Our experimental design then randomly exposes individuals (observers) to norm violations of others, varying the perceived similarity between the norms of the others and the norm in the observer's own decision situation. Our setup allows us to establish two crucial conditions that determine whether norm spillovers occur: one pertaining to the perceived similarity of the decision situations, and the other related to whether the observers of norm violations have the opportunity to themselves violate the norm of the initial norm breaker before facing a different decision situation.

In the first study of our paper, we explore spillovers in a controlled laboratory setting. Our experimental setup consists of two stages: first-mover sessions, where subjects are assigned a compliance task with a clearly defined task rule; and second-mover sessions, where other subjects (observers) are shown the behavior of a first-mover and are then required to complete a compliance task afterwards. A key experimental variation lies in whether an observer is matched with a compliant or non-compliant first-mover. This arguably influences the perceptions of how others actually behave, and thus the descriptive norms.[3] We then test whether non-compliance of the first-mover influences the observer's willingness to follow their own task rule (injunctive norm).

The tasks we employ are deliberately simple. First-movers are always given a cheating task inspired by Fischbacher and Föllmi-Heusi (2013), where they have to truthfully report a randomly generated dice number but are paid based on the reported number. In Study 1, second-movers are assigned either this same dice task or a repeated perceptual dots task (Gino, Norton and Ariely 2010), where the rule is to identify whether the left or right side of a computerized graphic contains more dots, and where in some instances following the rule conflicts with the financial incentives. Finally, second-movers may also be given the lights task from Kimbrough and Vostroknutov (2016a), where subjects can cross virtual red lights on a screen to increase their payoffs, but the rule is to wait until the lights turn green. We provide instructions for the games to subjects not involved in the decision-making experiments and use

economic features (monetary incentives, action sets, etc.), and norms are then defined in terms of the appropriate behavior for the particular situations. Ultimately, these approaches overlap.

[3]Descriptive norms refer to beliefs about what others typically do or how an average person behaves. Krupka and Weber (2009) show the subjects, prior to playing their game, the choices made by four previous participants. In our laboratory experiments, we have chosen a relatively weak treatment by showing the behavior of a single subject. This is sufficient to induce peer effects and spillovers. In the field experiment we show behavior of several others.

Krupka-Weber type norm-elicitation methods to elicit social appropriateness ratings for the behaviors in the games. We demonstrate that our task rules align with commonly shared perceptions about what is socially appropriate in these games, that is, they reflect social norms.

Our first set of results is as follows. In the treatment where both first and second-movers receive the dice task, we find that observing a rule-breaker causes a large effect on second-mover rule-breaking of 20.6 percentage points (pp), an effect size of 0.42 standard deviations (SD). We label this the 'standard peer effect', because it occurs within the same decision situation. When second-movers receive the dots decision situation, we find a substantial spillover to this different decision situation: randomly observing a norm-breaker in the dice-task causes a 19.7 pp (effect size 0.39 SD) increase in rule-breaking in the dots-task. This spillover effect is very similar in magnitude to the peer effect. Further decreasing the perceived similarity of the second-mover and the first-mover decision situation by assigning second-movers the lights task leads to a large decrease in the estimated spillover effect to 6 pp. In this case, the spillover effect is no longer statistically significant (p-value 0.18).

To shed light on the mechanisms behind the observed spillover effects, we calibrate a simple utility function that depends on both the appropriateness ratings from belief elicitation as well as the monetary payoffs of the actions available to observers. We show that second-movers give less weight to social appropriateness ratings when matched with a non-compliant first-mover, and that this decline occurs within and across games. Moreover, the decline is less pronounced when the decision situations become more dissimilar, with no significant decline in norm weights when second-movers are in the most distant decision situation. We interpret this as evidence that conditional norm compliance is a major driver behind the results (without asserting that this is the only channel).[4]

In Study 2, we test how the power of spillovers can be enhanced so that norm violations spread even to situations perceived as relatively distant. The idea of "slippery slope" suggests that individuals may adapt or become more accustomed to norm violations over time. Once they break a norm, the psychological barrier weakens, making future transgressions more likely. Based on this, if observers can be influenced by a dice cheater to become dice cheaters themselves, they may become more likely to break norms in subsequent situations, even distant ones.[5] To test this idea, as in Study 1, we show observers

[4] We discuss in Section A.3 that social learning cannot be entirely ruled out, though it appears less likely in our setup.

[5] The slippery slope concept highlights the potential for a cascade effect where a small deviation from accepted norms can lead to a series of increasingly significant violations. This idea is often discussed in the fields of psychology, sociology, and ethics when exploring how individuals navigate moral boundaries and social norms (see, e.g., Welsh, Ordóñez, Snyder and Christian, 2015). For evidence from neuroscience, refer to Garrett, Lazzaro, Ariely and Sharot (2016), who observed that the amygdala (a

the behavior of a first-mover in the dice task and then assign them the task perceived as most different (traffic lights). However, before performing the traffic light task, observers must first complete the dice task themselves. Our results support "slippery slope" arguments: Observers who are paired with a norm-breaking first-mover are more likely to break the same norm in the dice game, but they subsequently now also are more likely to break the norm in the light task.[6]

Examining spillovers in the lab allows us to identify the conditions under which they arise in a controlled environment, but questions about their relevance in the real world may remain. In Study 3, we test our first condition – that perceived norm similarity influences the occurrence of spillover – in a field experiment. In particular, we investigate spillovers in the workplace, which is often referred to as the 'hotbed of norm-guided action' (Elster, 1989). We recruit workers for a university library job. Workers need to catalog books, which involves collecting bibliographic data on a computer (e.g., ISBN number, author, publication year). Prior to this task, a random subset of workers are given newspaper articles about celebrities' tax evasion behavior. Among other information, they need to collect from the articles the monetary amounts the celebrities are reported to have evaded. This arguably manipulates the descriptive norm regarding others' tax evasion behavior. The observed outcomes are work effort and theft. The former is measured by the number of cataloged books; the latter by workers' propensity to steal pens from the work desk. Thus, we can trace how exposing workers to norm violations outside the workplace (tax evasion) affects their compliance with work-related social norms, namely, working hard and refraining from stealing office materials.[7] We elicited beliefs about the similarity of the social norms "do not evade taxes", "do not steal at the workplace", and "provide work effort." They were clearly rated as distinct from each other, with the stealing norm being closer to the tax evasion norm than the work effort norm.

We find that exposure to information about celebrities' tax evasion does not have a significant effect on work effort, but it increases pen theft by 21 to 32 percentage points. In two robustness treatments, we control for the impact of the added extra task (collecting information about tax evaders) and for the potential effects of reading newspaper articles, regardless of tax evasion content. Participants in one

brain region usually associated with emotion) shows fewer responses to repeated dishonesty over time.

[6] In a further treatment, the correct dice number is automatically reported for the second movers, so contagion effects on the initial domain are prevented by design. In this treatment, spillovers to the light game do not occur, reinforcing the idea that allowing for a slippery slope in the initial domain indeed facilitates the spread of norm violations to distant tasks.

[7] Unlike the lab experiments, we do not present workers with explicit task rules or norms here to avoid priming. We assume that tax evasion is not socially accepted behavior for most of the subjects, and that they would also agree that when being invited for a library job, one should provide effort and not engage in theft.

treatment arm collect celebrities' tax evasion amounts from a public database, while the other treatment group collects biographic information from the same database, where we have removed the tax evasion-related details. Once again, subsequent work effort remains unaffected but we observe a large and significant treatment effect on pen stealing. Our field experiment thus demonstrates the predictive power and applicability of perceived norm similarity for real-world settings, with spillovers occurring to the more similar norm and no spillover to the more distant norm.

Our paper advances the understanding of the spread of norm compliance in several important aspects. First, according to the broken windows theory (Wilson and Kelling 1982, Lanfear, Matsueda and Beach 2020), spillovers may amplify the detrimental effects of initial norm violation. Even small signs of disorder and neglect (e.g., graffiti) can contribute to an environment that fosters larger norm violations, resulting, for instance, in significant crimes. A major contribution of our study is that we present evidence, both from the lab and the field, that this is not necessarily the case: in various of our settings, spillovers do not occur to the distant situation. This suggests that spillovers may not be as far-reaching as suggested by the broken windows literature; they can be self-limiting.

Secondly, our field experiment, by providing first causal evidence of norm spillovers into the workplace, further underscores the relevance of spillovers in real-world economic contexts. It highlights the potential costs of spillover effects and perceived norm violations, even for uninvolved third parties, such as employers. We thus provide direct evidence that there can be hidden costs of non-compliance, such as with tax laws, which are overlooked when focusing solely on peer effects in the initial domain of non-compliance (e.g., tax evasion may not only lead to tax evasion, but may also cause norm-breaking in other domains).[8] Combined with the finding that spillovers are not unlimited, our paper can be seen as the first pointer that the actual costs of perceived non-compliance lie between the bounds suggested by the peer effects literature and the broken windows theory.[9]

Thirdly, however, we show that spillovers are not always self-limiting. An important contribution of our paper is demonstrating conditions under which concerns about far-reaching spillovers (as suggested by the broken windows theory) may be warranted. Specifically, this is the case when observers

[8]Studies on tax evasion often highlight its negative effect on others' tax compliance (see, e.g., Lefebvre, Pestieau, Riedl and Villeval, 2015). Kleven, Knudsen, Kreiner, Pedersen and Saez (2011) point out that many individuals have very limited evasion opportunities due to widespread third-party income reporting. Our paper suggests those unable to evade may adapt through costly alternatives, like workplace behavior.

[9]That there might be hidden economic costs to violations of laws can also be inferred, for instance, from Gulino and Masera (2023). The unique contribution of our paper is in bounding spillovers, demonstrating that they do not occur in distantly perceived domains.

are led on a slippery slope by being given the opportunity to cheat in the same situation as the initial violator (Study 2). That slippery slope effects may exist has been demonstrated in experimental papers in social psychology (see, e.g.,Welsh, Ordóñez, Snyder and Christian, 2015; Garrett, Lazzaro, Ariely and Sharot, 2016). Our contribution is to demonstrate how such effects play out across decision situations, by allowing people to violate different norms over time. Thus, our paper also provides insights into how spillovers into distant domains can be mitigated, specifically by reducing the opportunities for initial social contagion in the same decision situation.

The paper is closely related to recent experiments highlighting the relevance of social norms in various decision-making settings (Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2016a; Krupka, Leider and Jiang, 2017; Barr, Lane and Nosenzo, 2018). We complement these studies by demonstrating that norm-related features beyond the immediate decision-making context determine behavior, suggesting an even broader role for norms in economic behavior.[10]

Our manuscript is also related to evidence on spillovers, stemming from field studies on the impact of leaders and the influence of negative examples (such as corruption, tax evasion, or political fraud) on antisocial behavior (see, e.g., Fisman and Miguel, 2007; Gächter and Schulz, 2016; Ajzenman, 2021; Gulino and Masera, 2023). We contribute to this literature by demonstrating that instances of unethical behavior in society can affect individual norm-related behavior at work, a phenomenon not addressed by any of these studies.[11] The field experiment also contributes to the literature on fraudulent behavior and theft in the workplace (see, e.g., Greenberg, 1990; Cohn, Fehr, Herrmann and Schneider, 2014; Belot and Schröder, 2016; List and Momeni, 2020a, 2020b). What sets our study apart from existing ones is that we do not focus on how internal firm incentives (e.g., wage cuts and monitoring) influence worker cheating and theft. Rather, we demonstrate that external factors outside the workplace, such as celebrities' unethical behavior, may shape workplace theft, and can therefore impact employer-employee relationships.

[10] Gächter, Nosenzo and Sefton (2013) and Gächter, Gerhards and Nosenzo (2017) demonstrate limited evidence of norms influencing behavior in peer effects using Krupka-Weber norm-elicitation methods. Additionally, norm-nudge interventions had limited effects in a lying experiment (Dimant, Van Kleef and Shalvi, 2020). In our analysis, we estimate how the behavior of others influences the weight individuals put on norms, assuming constant norms (Lindenberg and Steg, 2013). Our findings support norm-driven adjustments to the compliance behavior of other individuals, enhancing our understanding of peer and spillover effects channels.

[11] For the effect of leaders on market participants in a laboratory environment, see d'Adda, Darai, Pavanini and Weber (2017).

STUDY 1: RULE VIOLATIONS AND BEHAVIORAL SPILLOVERS

A EXPERIMENTAL DESIGN AND PROCEDURES

A.1 FIRST AND SECOND MOVERS' TASKS

In this study, we experimentally manipulate whether individuals are exposed to rule-violating versus rule-complying behavior of others, and we analyze whether this exposure affects their own compliance behavior. To implement this, we use two types of sessions: first-mover and second-mover sessions.

The first movers receive a computerized cheating task in the spirit of the dice-rolling task by Fischbacher and Föllmi-Heusi (2013).[12] First movers see a screen for 5 seconds with six dice, ordered according to the numbers 1 to 6, and one of the dice is marked with a red cross. After ten seconds, the dice disappear, and the first movers are asked to report the number of eyes on the marked dice. Figure 1 gives an example screen with the cross and the decision button. In the instructions, it was explicitly communicated to subjects that the rule of the task is that they *are supposed to report the number of the marked dice by clicking the corresponding button*, and that they are paid the reported number in euros. Furthermore, we notified participants that their decisions might be used in future sessions and shown to random subjects who would only know that this decision was made by an unknown previous subject of the experiment. In total, we conducted two first-mover sessions involving 48 participants.

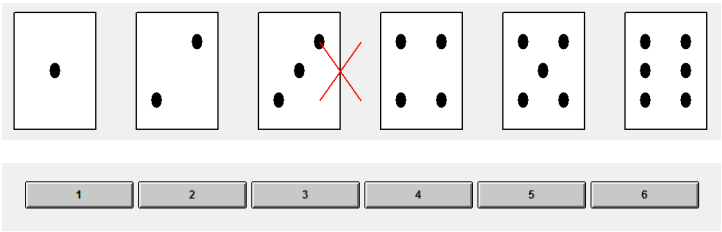


Figure 1: SCREENSHOT OF DISPLAYED DICE AND DECISION BUTTONS, FIRST-MOVER

The second-movers (observers) are matched with a subject who previously occupied the same seat in the lab during one of the earlier first-mover sessions. They are asked to read the original instructions of the first-movers and then see on the screen which dice was marked for the first mover and which button

[12] See Appendix *A*.1 for first-mover instructions.The computerized implementation is similar to Pittarello, Leib, Gordon-Hecker and Shalvi (2015). This and all other experimental sessions were conducted at the laboratory of the TU Berlin, and subjects were invited via the local ORSEE-system (Greiner, 2015). The experiments were programmed in z-Tree (Fischbacher, 2007).

they clicked. Thus, they are either exposed to rule-compliant behavior or rule-violating behavior.[13] To incentivize second-movers to accurately absorb this information, they are paid €2 when they correctly report the first-mover's marked and reported die. Thereafter, subjects receive instructions for their own decision situation.

In the treatment DICE, observers decide in exactly the same situation and perform the same task as the matched first-mover. They see the dice with the cross on the same dice, and they are also instructed to click the button displaying the number that matches the marked dice. Additionally, the decision screen reminds them of the first-mover's situation and decision (see Figure 2, Panel (a), for a screenshot).

In DOT, we implement a different task based on the perceptual dots task by Gino et al. (2010). The computer screen is divided in the middle by a vertical line. To the left and right of this line, dots are scattered, and subjects are informed in the instructions that the rule of this task is to *indicate which side of the centerline has more dots* by clicking the corresponding button (see Panel (b) in Figure 2). This task is repeated 12 times, and in 8 of these repetitions, incentives run counter to truthfully reporting, i.e., a there is a higher payoff from clicking the wrong button.[14]

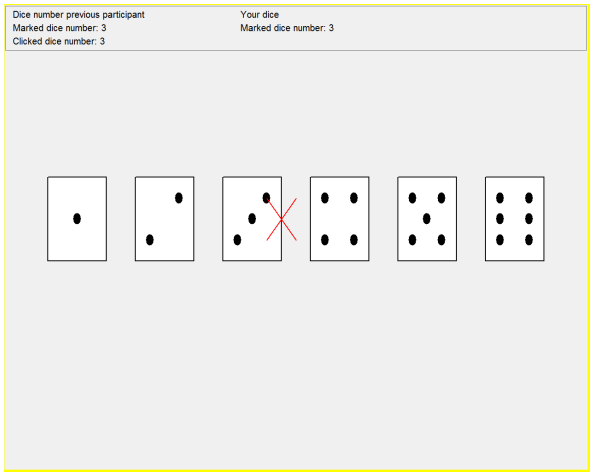
In LIGHT, we implement the rule-following task from Kimbrough and Vostroknutov (2016b). In this task, the second-movers navigate a figure across the screen. The figure will automatically stop at 5 red lights. Each light turns green six seconds after the figure stops (see Panel (c) in Figure 2). At any time subjects can press a button to make the figure start walking again. They can thus at each light induce the figure to walk across the red light and continue until it reaches the next red light. In the instructions subjects are informed that they *are supposed to wait at each red light until it turns green.* Subjects receive an endowment of €8, and they are told that for each second it takes their figure to walk or wait they will lose 10€-Cent, providing incentives to cross the screen as fast as possible.[15]

The parameters of the second-mover tasks were selected to ensure that full compliance and non-compliance would result in similar payoffs across games. All payouts were conducted double-blind,

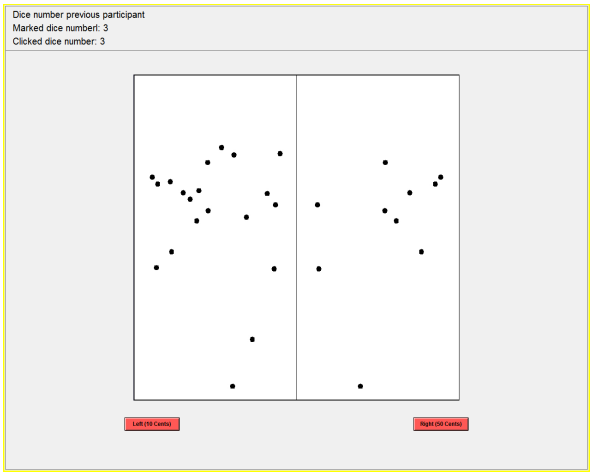
[13] In the case where a six is drawn for a first-mover, and they click a six, it is not possible to infer whether a first mover is a dice cheater. Therefore, we only use first-mover observations for dice numbers between 1 and 5 to show to second-movers. Details on how we select the first-mover observations shown to second-movers are explained in Appendix B.

[14] Half of the subjects receive 50€-Cent when clicking the left button and 10€-Cent when clicking the right button.To counterbalance, the other half of subjects receive 10€-Cent when clicking the left button and 50€-Cent when clicking the right button, using the same distribution of dots but mirrored along the vertical line.

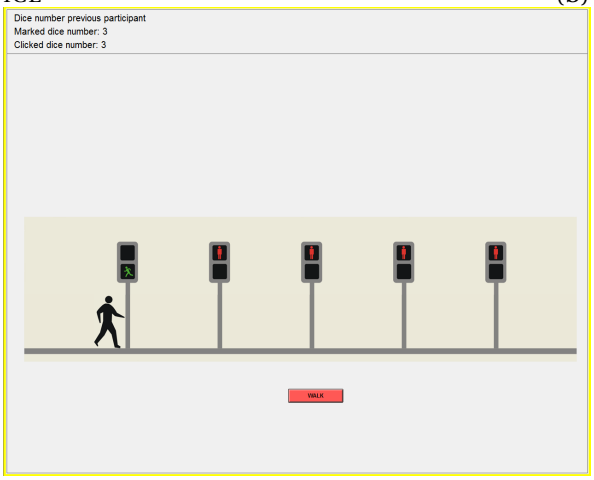
[15] It takes 24 seconds to cross the screen, excluding waiting at the traffic light. The maximum earnings are €5.60 (with no waiting) and €2.60 (if fully rule-compliant with six seconds of waiting at each light).



(a) DICE



(b) DOT



(c) LIGHT

Figure 2: SCREENSHOTS OF OBSERVER GAMES.

and the person in charge of the payments did not know the names of the subjects or the tasks they participated in. In addition to the earnings from the task, subjects received a €5 show-up fee.

Table 1 provides an overview of the treatments and subjects.[16] Overall, we collected 12 observer sessions with 4 sessions per task. Within each session, we randomly matched half of the subjects with a rule following first-mover (FIRSTMOVER:COMPLIANT) and half with a rule breaking first-mover (FIRSTMOVER: VIOLATION).[17] In total, we gathered the data of 278 subjects participating in the observer sessions.

[16]Within each game, there are no significant differences between the groups exposed to rule violations and the unexposed groups in terms of age (all $p > 0.26$), gender (all $p > 0.83$), number of siblings (all $p > 0.42$), and prior number of experiment participation (all $p > 0.11$). No significant differences exist for these variables when comparing them across games (for all variables $p > 0.25$).

[17]Different numbers of subjects are the result of no-shows. Testing the distribution of dice shown to the first movers used in

Table 1: TREATMENTS AND SUBJECTS STUDY 1

TREATMENT	SUBJECTS	OBS.	MEAN AGE	FEMALE
DICE				
FIRSTMOVER:COMPLIANT	47	47	23.1	45%
FIRSTMOVER:VIOLATION	45	45	23.9	44%
DOT				
FIRSTMOVER:COMPLIANT	47	564	23.9	47%
FIRSTMOVER:VIOLATION	47	564	24.7	51%
LIGHT				
FIRSTMOVER:COMPLIANT	46	230	25.3	52%
FIRSTMOVER:VIOLATION	46	230	24.3	48%

NOTE: One decision in DICE, multiple decisions in DOT (12×) and LIGHT (5×).

A.2 APPROPRIATENESS RATINGS

The rules in our decision experiments are designed to establish focal points for appropriate behavior. We also test whether these rules align with what the participants regard as social norms, that is, whether the actions prescribed by the rules are seen as socially appropriate or ethically right behavior in the particular decision tasks. We employ the frequently utilized Krupka and Weber (2013) elicitation method to obtain social appropriateness ratings for the available actions in each decision situation through incentivized coordination games.[18]

We conduct two sessions with individuals not participating in the choice experiments (N=49).[19] These participants receive a description of all three games (DICE, DOT, and LIGHT) used in the choice experiment. They are then prompted to assess the appropriateness of all potential actions for each game by selecting one option on a four-point ordered scale: "Very appropriate," "Somewhat appropriate," "Some-

the observer sessions reveals no significant differences between the dice of the first movers who violated the norm and who did not ($p = 0.55$ in DICE-DICE, $p = 0.48$ in DICE-DOTS, and $p = 0.41$ in DICE-LIGHTS, using the Epps-Singleton test implementation by Goerg and Kaiser, 2009). See Appendix B for more details on first-mover behavior and the used observations.

[18]This method is tied to a concept of social (injunctive) norms, defining them as collective perceptions or commonly shared beliefs about the appropriateness of the actions available in a particular decision situation. Though the coordination games involve multiple equilibria, as respondents are rewarded for accurately reporting their beliefs about how appropriate a particular behavior is commonly regarded in a group of individuals, and thus elicit second-order beliefs of appropriateness, the results of the method have been shown to be robust against using other belief-elicitation methods (see, e.g., Lane, Nosenzo and Sonderegger, 2023).

[19]Once again, the sessions were held at the laboratory of the TU Berlin, and subjects were invited through the local ORSEE-system (Greiner, 2015). The survey was programmed in z-Tree (Fischbacher, 2007). Refer to Appendix A.3 for instructions.

what inappropriate," or "Very inappropriate."[20] For the DICE task, participants are asked, for instance, about the appropriateness of clicking the number 6 when the dice shows the number 1. In the DOT task, for example, subjects are queried about the appropriateness of reporting the wrong side 3 out of 8 times. Regarding the LIGHT task, participants are, for example, asked about the appropriateness of crossing 2 out of the 5 traffic lights.[21]

Figure A.C.4 in the appendix visualizes the results of the belief-elicitation, where we assigned numerical scores to the subjects' responses, ranging from 1 ("very inappropriate") to 4 ("very appropriate"). The figure displays the mean answers on the obtained 1-4 scale. As shown, for all games, the action prescribed by the rules of the tasks receives by far the highest mean social appropriateness ratings. In addition, in each game, the mean ratings monotonically decline with the severity of the norm violation, with relatively sharp drops already for slight norm violations.[22] The profile of the appropriateness ratings thus corresponds to the rules of the games we imposed, suggesting that the rules resonate with beliefs regarding social norms of appropriate behavior within the subject pool.

A.3 DISTANCE RATINGS

Our particular focus is on whether possible spillovers across decision situations depend on the (perceived) similarity of the rules underlying the decision situations. The rules in all three games—the dice game, the dot game, and the light game—each read differently. In the dice game, the rule is to truthfully report the eyes on the dice; in the dot game, one has to indicate on which side of the screen there are more dots, and in the lights game, the rule is to stop at each traffic light. Intuitively the rules in the first two decision situations can be perceived to be more similar as they both involve truthful reporting, whereas in the light game, subjects have to follow a simple traffic rule. However, it can be argued that all the rules can be subsumed under some kind of umbrella norm, such as "follow the instructions of the experimenters", making them examples of the same norm. In this case, all three decision situations could also be interpreted as representing one underlying, broad decision situation (e.g., sitting in front of

[20] In the instructions, we explain to subjects that by "appropriate behavior," we mean behavior that is "consistent with generally accepted ethical or moral rules," and most people would consider this to be the "correct" or "ethically right" behavior.

[21] To keep the experiments concise, for the DICE task, we did not assess the appropriateness of reporting a number lower than the actual dice number (under-reporting is also not observed in the decision experiments).

[22] In the DICE-game, the mean ratings monotonically decrease as the wrongly reported number increases, with the steepest drop occurring for slight over-reporting (e.g., reporting "2" if the dice number "1" is shown). For the DOT-game and the LIGHT-game, the mean ratings decline with the number of rule violations in the games, with the steepest decline occurring for the first violation (reporting 1 instead of 0 times that the wrong side has more dots, crossing 1 instead of 0 red lights).

a computer screen in the lab and engaging in a rule-following task). We do not want to rule out that the rules we use might be aggregated further. In our analysis, however, we rely on actual elicited assessments of the similarity of the norms. We thus apply an empirical approach and metric to norm distance, using incentive-compatible coordination games, not only for measuring norms (see Section A.2) but also to assess their perceived distance.

To do so, we let the respondents from the norm-elicitation experiments, after assessing the social appropriateness of the actions, also assess the relative distance of the rules in the decision situations. Specifically, subjects are asked to rate the distance of the rules in the DOT- and the LIGHT-game compared to the rule of the DICE-game, respectively, on a scale from 0 to 10. Here, 0 means a rule is exactly the same as in the dice game, and 10 means that the two rules in question are maximally distant. Respondents are paid if their response matches the modal response of other participants in their session.[23]

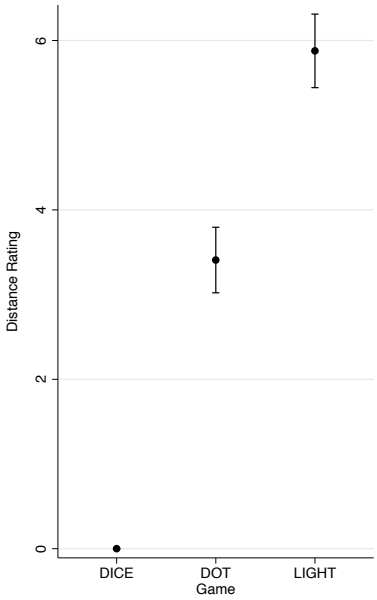


Figure 3: DISTANCE RATINGS (MEANS AND STANDARD ERRORS)
Note: Perceived norm distances in comparisons to the rule in the dice game on a scale from 0 to 10, where 0 means that two rules are identical, and 10 means that they are maximally distant. The graph shows the mean answers on the scale from a sample of students not taking part in the decision experiments ($N = 48$).

If all respondents assign a zero for the DOT-game and the LIGHT-game, then all rules would be considered the same. Figure 3, displaying the mean distance ratings between the rules in the three decision situations, shows that this is not the case. Compared to the DICE-game, for which we set the norm dis-

[23] Refer to Appendix A.3 for instructions.

tance score to be zero, distance scores are higher for the DOT-game and reach the highest values for the LIGHT-game. Two-sided Wilcoxon signed-rank tests indicate that all norm distance ratings are statistically significantly different from one another (DOT vs. DICE ratings: $z = 6.035$, $p \leq 0.001$; LIGHT vs. DICE: $z = 6.074$, $p \leq 0.001$; LIGHT vs. DOT $z = 4.091$, $p \leq 0.001$).

Overall, these results show that all the rules of the decision situations are evaluated to be different, with the norm in the DOT-situation being more similar to the norm in the DICE-situation than the norm in the LIGHT-situation.

B RESULTS FROM THE CHOICE EXPERIMENTS: EXPOSURE TO RULE VIOLATIONS AND BEHAVIORAL SPILLOVERS

Figure 4 shows the relative frequency of rule violations for each game, split by being matched with a first-mover violating the rule or with a first-mover complying with the rule.[24] When observers are assigned the dice task DICE-DICE 64% of decisions in the observer sessions violate the rule of truthful reporting if matched first movers complied with the rule, while 84% of decisions violate the rule if matched first movers violated the rule. The Pearson's χ^2 test confirms that the difference in rule violations is statistically significant ($\chi^2(1) = 5.067$, $p = 0.024$). This corresponds to an effect size of 0.42 SD. To put the treatment effect into perspective, we also report persuasion rates by relating the changes in rule violations due to the treatment to the base rates in the control group.[25] We find that about 56% of individuals who would not have violated a rule if they had observed a rule complier were persuaded to violate the rule by observing a rule violator. This indicates that observing rule violations strongly influences the behavior of those who can potentially be persuaded to do so (i.e., whose behavior can actually be changed). Taking together these results, we find economically and statistically significant peer effects within a decision situation, serving as a benchmark scenario.

In DICE-DOT 52% of decisions violate the rule if observers are matched with a rule-abiding first-mover.[26] In contrast, 72% of decisions made by observers violate the rule when matched with rule vio-

[24] First mover results are presented in Appendix B.

[25] The persuasion rate is calculated as $\frac{y_T - y_C}{e_T - e_C} \cdot \frac{1}{1-y_0}$, where y_T and y_C are the shares in the treatment and control group exhibiting the behavior of interest; e_T and e_C are the shares in treatment and control receiving the treatment (see DellaVigna and Kaplan, 2007, and DellaVigna and Gentzkow, 2010). Y_0 denotes the share of individuals that adopt the behavior of interest absent treatment. We set $e_T = 1$ and $e_C = 0$, and as in DellaVigna and Gentzkow (2010), we set $y_0 = y_C$.

[26] In eight of the 12 decisions of the DICE-DOT setup payments were higher for reporting the wrong side and violating the rule of truthfully reporting. In the remaining 4 decisions incentives were aligned with the rule. In the analysis we drop the decisions for the situations without a conflict between incentives and rule. Keeping these observations does not change our

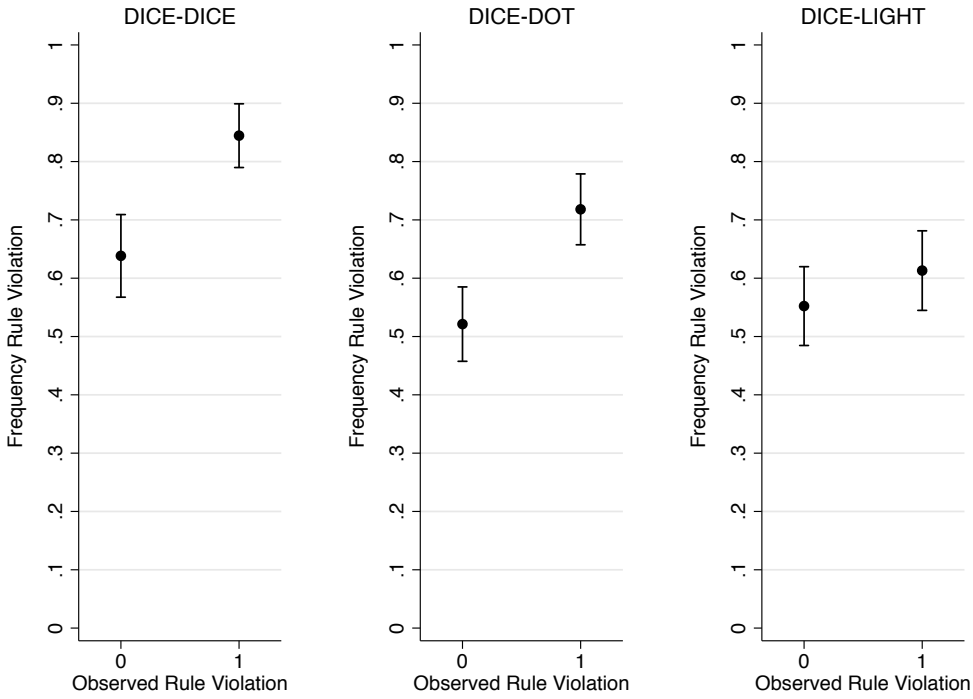


Figure 4: RELATIVE FREQUENCIES OF RULE VIOLATIONS

Note: The figure illustrates the mean (and standard errors) of second-mover rule violations for all second-mover games, conditional on whether a subject is matched with a first-mover violating the rule ("1") or with a first-mover complying with the rule ("0").

lating first movers. Comparing observers' average rule violation in the two conditions with a two-sided Mann-Whitney (rank-sum) test ($z = 2.262$, $p = 0.024$) confirms this difference to be statistically significant. The difference in norm violations corresponds to an effect size of 0.39 SD, with a persuasion rate of roughly 42%. The spillover effect is thus approximately comparable in size to the direct peer effect.

While this comparison is based on the average of multiple decisions per observer, it is also intriguing to see how witnessing rule violation affects the dynamics of decisions in the dot game. In Figure 5, we report the frequencies of rule violations by observers in DICE-DOT for each of the twelve decisions and for being matched with a rule compliant first-mover versus a non-compliant first-mover. We observe a significant treatment effect from being matched with a norm-breaking first-mover for 6 out of the 8 decisions where financial incentives are misaligned with norm compliance (decisions 1-2, 4-5, 7-8, and

results but artificially reduces the incidences of rule violations in both treatments. Less than 1% of responses were incorrect when incentives and rules were aligned.

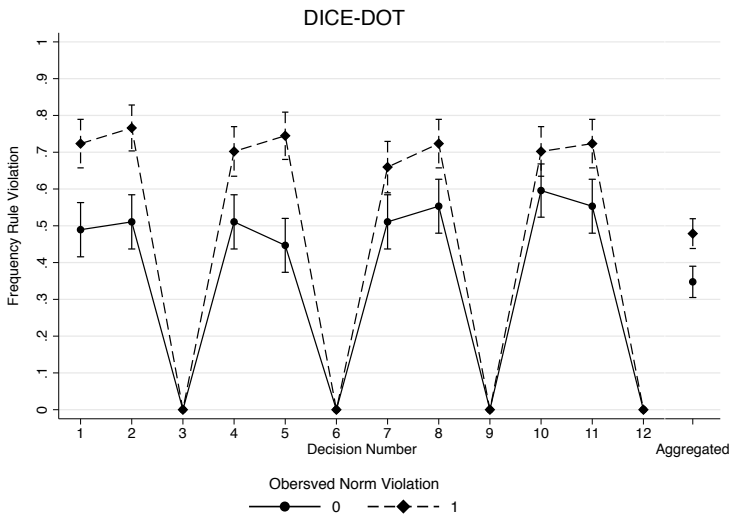


Figure 5: FREQUENCIES OF RULE VIOLATIONS IN THE DOT GAME, FOR EACH DECISION SECOND-MOVER ACTION
Note: The figure plots the relative frequencies of second-movers' rule violations for each of the twelve decisions in the dot game, based on whether they are matched with a rule-compliant first-mover ("0") or a rule-violating first-mover ("1"). Decision number "1" refers to the decision on the first screen with dots shown, "2" to the second dot screen, etc. Rule violation is defined as not reporting the right side to have more dots. The aggregate includes all 12 decisions, including those where incentives are aligned with following the dot norm (decisions "3", "6", "9", and "12").

10-11). This treatment effect is already present in the first decision (49% vs. 72%) and statistically significant when applying Pearson's χ^2 test ($\chi^2(1) = 5.393$, $p = 0.020$). Moreover, the magnitude of the upward shift in the dashed line, representing the treatment effect, is comparable for all of these decisions.[27] This demonstrates that individuals systematically violate the norm more often in the DOT-game when they observe norm violations in a different game, the DICE-game.

In DICE-LIGHT we also observe a higher frequency of rule violations (share of red lights crossed) by observers who have been matched with non-compliant first-movers (61%) compared to observers who have been matched with compliant first-movers (55%). Yet, this effect is small in size and not statistically significant on the aggregate level ($z = 0.657$ with $p = 0.51$, Mann-Whitney (rank-sum) test); the implied persuasion rate is 13%. In the appendix, Figure A.C.5, we depict the frequency of rule violations in DICE-LIGHT for all individual lights. It shows that there are also no significant treatment effects for any of the lights.

Taken together with the results from the elicited distance rating experiment, we conclude that there

[27] The figure also shows that norm violations only occur when they are financially advantageous: If incentives are aligned with the norm of the task (in decisions 3, 6, 9, and 12), almost all subjects choose to report the side on which they see more dots.

is a large and significant spillover effect on the task where the rule is judged to be similar. However, there is only a small and statistically insignificant spillover effect into the task considered to be most different from the rule of the first mover's task.

C Discussion of Channels: Conditional Norm Compliance versus Social learning

What explains the observed spillover effects? We discuss two channels: conditional norm compliance and social learning.

We conceive of conditional norm compliance as a model where, (i) individuals not only value material payoffs but also explicitly assign value to actions if they align with social norms – that is, people have norm-based preferences (see, e.g., Krupka and Weber, 2013); (ii) these preferences are conditional: the relative weight given to social norms depends on the social context, that is, on whether or not others violate norms. In particular, individuals are less willing to follow norms themselves when exposed to norm violations, so the preference weight given to norms should decrease.[28] If conditional norm compliance is behind the result that spillover effects are smaller for more dissimilar decisions, we would expect (iii) the drop in the weight of norms caused by seeing a norm violator to be smaller when the norms of the decision situations are perceived to be more dissimilar.

Combining the choice with the belief elicitation data allows us to explicitly test these ideas. We assume that the second mover's preferences are described by the following utility function

$$U(x_i) = \beta\pi(x_i) + \varphi N(x_i), \tag{1}$$

where $\pi(x_i)$ is the payoff associated with action x_i, and $N(x_i)$ is the social appropriateness rating of that action, as derived from the norm-belief elicitation experiment.[29] The parameter ratio φ/β represents the relative utility weight individuals attach to the social appropriateness ratings compared to monetary payoff.[30] We use a conditional logistic choice model to estimate the preference parameters β and φ. We conduct separate regressions for subjects exposed to rule-abiding and rule-violating behavior

[28] Ideas of such conditional preferences are often found in the literature on peer effects in norm-breaking (see, e.g., Bicchieri and Xiao, 2009).

[29] The norm ratings $N(x_i)$ take on the values 1-4, depending on whether an action is classified as 'very inappropriate' (1) to 'very appropriate' (4).

[30] The ratio φ/β represents the willingness to pay for a marginal increase in the social appropriateness rating of a certain action.

in each game, so that we can, for each game, calculate the change in the relative norm weights when being matched with a cheater. This change we denote by $\Delta := \Delta(\varphi/\beta)$.[31]

The regression results are reported in Table 2.[32] As can be seen, in all models, the estimates for both weight parameters are statistically significant, indicating that norm-based preferences can successfully explain behavior in our experiment. Moreover, in both DICE-DICE and DICE-DOT, the relative norm weight significantly drops for subjects matched with a rule-violating first-mover. By contrast, in DICE-LIGHT, for which we observed no behavioral spillover, the decrease in the relative norm weight is relatively small and insignificant. In fact, the magnitude of change in the relative norm weight, highest in DICE-DICE ($\Delta(\varphi/\beta) = -46\%$), intermediate in DICE-DOT ($\Delta(\varphi/\beta) = -12.9\%$), and lowest in DICE-LIGHT ($\Delta(\varphi/\beta) = -3\%$), is (inversely) related to the relative norm distance ratings from Figure 3. This qualitatively also aligns with the size of the observed behavioral spillover effects from Section B.

Table 2: CONDITIONAL LOGISTIC ESTIMATION USING AVERAGE APPROPRIATENESS RATINGS

	DICE-DICE		DICE-DOT		DICE-LIGHT	
	No Exposure	Exposure	No Exposure	Exposure	No Exposure	Exposure
Payoff (β)	1.143***	1.663***	1.858***	3.195***	2.025***	2.444***
	(0.287)	(0.452)	(0.485)	(0.717)	(0.350)	(0.414)
Rating (φ)	1.089***	0.849**	2.595***	3.911***	2.156***	2.525***
	(0.306)	(0.372)	(0.594)	(0.906)	(0.387)	(0.465)
φ/β	.95	.51	1.40	1.22	1.06	1.03
$\Delta(\varphi/\beta)$	-46%		-12.9%		-3%	
$\chi^2(1)$	3.36		2.88		0.16	
Prob > χ^2	0.067		0.090		0.692	
Observations	181	177	423	423	276	276
Clusters	47	45	47	47	46	46

Note: Results for the significance of ratios and differences in ratios are obtained using nonlinear Wald-type tests. Standard errors are adjusted for clustering at the individual level, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

An alternative interpretation of the results involves social learning. By observing others breaking

[31] Note that we are not comparing the ratios between games as such, as they can vary depending on the situation. In some contexts, the social appropriateness of following a rule might be more important than in others. Instead, we can compare the changes in these ratios, that is, the treatment effects on the relative normative weights, across games.

[32] The estimation is based on a logistic choice model, which bases the likelihood of an action x on the relative utility resulting from that action compared to the utility resulting from all other actions: $P(x_i = x^j) = \exp(U(x^j))/\sum_k \exp(U(x^k))$.

norms, individuals may learn that cheating is possible and then simply mimic the cheating behavior of someone else, without necessarily considering social norms or ethical standards. This learning effect might be more relevant when the decision situations are sufficiently similar. Note that our instructions include control questions that explicitly check whether second movers understand the monetary consequences of following or deviating from the norm (both for their own and the first-mover decision situations) and the experiment proceeds only if all subjects have answered the control questions correctly. We thus argue that observers are fully aware of their choice options and the specific consequences they entail. Moreover, in our setting, there is also no punishment or fine for norm violations, so detection probability and beliefs thereof should not play a role for behavior. We thus consider it less plausible that social learning about features of the decision situation from being matched with a norm violator is a major driving factor of the spillover results. Yet, even if social learning were driving the spillover effects, the findings would still be relevant, as they provide insights into when behavioral adjustments across games are more likely to occur, a topic about which little is known.

Overall, combining the choice data with belief-elicitation methods shows that the results are in line with ideas of conditional cooperation, where the decrease in the relative preference for norm compliance caused by observing others' norm violation decreases with the perceived distance of the norms. We therefore argue that conditional norm compliance is one important, though not necessarily the only, channel driving the results.

STUDY 2: BOOSTING THE POWER OF SPILLOVERS

A Experimental Design and Procedures

The experimental design of Study 2 is similar to that of the DICE-LIGHT treatment in Study 1, with the addition of a slippery slope stage before the light task. Observers are exposed to first movers who either comply (FIRST-MOVER: COMPLIANT) or violate the norm in the dice task (FIRST-MOVER: VIOLATION). Thereafter, observers complete the dice task themselves before proceeding to the light task (we label this CHOICE-LIGHT because observers can choose what to report in the dice task).[33]

In an additional treatment, again observers are first exposed to the dice task and the decision of the

[33] Again, the experiments were conducted at the laboratory of TU Berlin. The instructions were the same as for Study 1, with the exception that, after subjects saw the decision of the first-mover, they needed to play the dice game themselves.

matched first-mover before they complete the light task. However, after seeing the first-movers, these second-movers do not have a choice in the dice task. Instead, the computer indicates the correct answer, and observers have only one option: to confirm the correct number (as second-movers have no choice, we label this treatment NOCHOICE-LIGHT). This forces compliance and thus prevents the possibility of being led on a slippery slope via contagion in the initial decision situation the dice task.[34] We thus can investigate whether any potential spillovers into the light task are actually the result of observers' own norm violations in the dice task, as opposed to being generated by adding a further stage to the experiment.

As in Study 1, we conducted two first-mover sessions to collect the data shown to the observers. We conducted 7 observer sessions: 4 CHOICE-LIGHT (N = 95), with 52 second movers matched with compliant and 43 with non-compliant first movers, and 3 NOCHOICE-LIGHT (N = 67), with 37 second movers matched with compliant and 30 with non-compliant first movers, for a total of 162 subjects (for an overview of treatments, see Table 3).[35]

Table 3: TREATMENTS AND SUBJECTS STUDY 2

		FREQ. NORM VIOLATION			
		DICE		LIGHT	
TREATMENT	N	MEAN	SD	MEAN	SD
CHOICE-LIGHT					
FIRST-MOVER: CORRECT	52	0.40	0.44	0.61	0.45
FIRST-MOVER: NORM VIOLATION	43	0.67	0.47	0.79	0.37
NOCHOICE-LIGHT					
FIRST-MOVER: CORRECT	37	0	0	0.71	0.43
FIRST-MOVER: NORM VIOLATION	30	0	0	0.71	0.41

Note: Frequency of norm violation DICE: indicator for an observer breaking the norm in the dice task. Frequency of norm violation LIGHT: share of the five lights where an individual violated the norm and crossed a red light.

[34] In the instructions, we describe this treatment as follows "We have already pre-set the dice number to be clicked on the button bar for you. The experiment will only continue once you confirm this number by clicking on it. You cannot select a different number." Everything else is the same as in the instructions for CHOICE-LIGHT.

[35] The reason for the imbalance in observations between FIRST-MOVER: COMPLIANT and FIRST-MOVER: VIOLATION is due to no-shows and the pre-assignment of treatments to cubicles. The distribution of dice numbers shown to first movers in NOCHOICE-LIGHT and CHOICE-LIGHT do not differ significantly (p = 0.93, Epps-Singleton test). There are also no significant differences between the distributions of dice numbers shown to first-mover who violated the norm and first movers who followed the norm in CHOICE-LIGHT (p = 0.81, Epps-Singleton test).

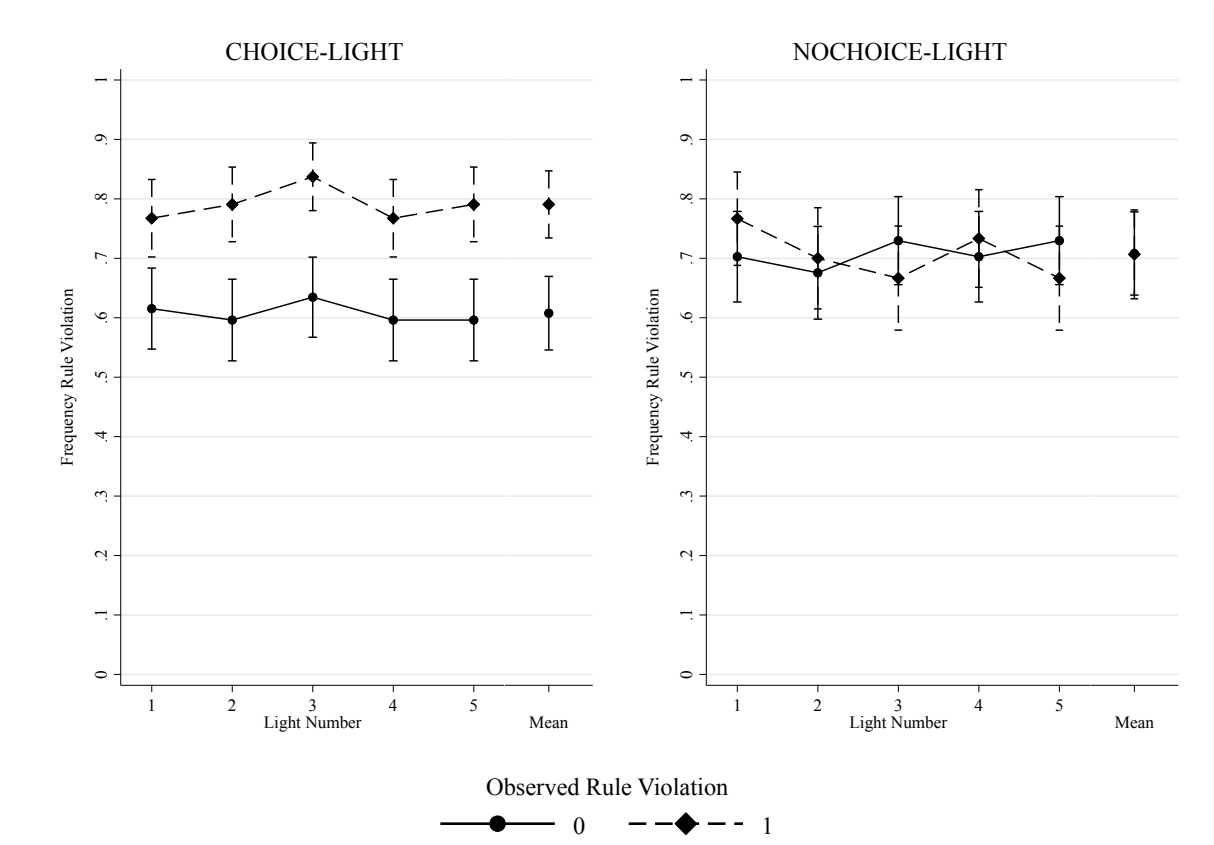


Figure 6: FREQUENCIES OF RULE VIOLATIONS (MEANS AND STANDARD ERRORS)
Note: Rule violation defined as walking before light changes to green.

B Results

We start by reporting the effects of being matched with a compliant versus a non-compliant first mover on the second mover's behavior in the dice task (CHOICE-LIGHT). Observers matched with a dice cheater are more inclined to report a wrong dice number compared to observers matched with compliant first movers (67% vs. 40%; $\chi^2(1) = 6.911$, $p = 0.009$, Pearson's χ^2). Thus, there is a peer effect of norm violations, which is broadly consistent with what we observed previously in Study 1 (27pp in CHOICE-LIGHT vs. 20pp in DICE-DICE; effect size 0.61SD vs. 0.41SD; persuasion rate 0.45 vs. 0.55).

The results for the traffic light task, which is our primary focus here, are visualized in Figure 6. In stark contrast to Study 1, in textscChoice-Light we now observe a spillover effect to the LIGHT game. Observers matched with norm-breaking first-movers are 18pp more likely to break the norm in the light task than observers matched with complying first movers ($z = 2.55$, $p = 0.011$, Mann-Whitney test; effect size 0.4

SD; persuasion rate 46%). This difference can be observed at each of the five lights and it turns out to be significant for most of them (Light 1: $\chi^2 = 2.52$ with $p = 0.11$; Light 2: $\chi^2 = 4.12$ with $p = 0.042$; Light 3: $\chi^2 = 4.86$ with $p = 0.027$; Light 4: $\chi^2 = 3.14$ with $p = 0.076$; Light 5: $\chi^2 = 4.12$ with $p = 0.042$, all Pearson's χ^2 test). This demonstrates that the occurrence of spillovers in CHOICE-LIGHT is systematic. By contrast, in NOCHOICE-LIGHT, where observers are forced to accept the rule-abiding input in the dice task, there are no spillovers to the light task (see the right panel of Figure 6); spillovers are effectively switched off again.

A plausible mechanism that can drive the extension of spillover effects to the lights task in CHOICE-LIGHT is that observers are led down a slippery slope of norm violations. Observing a norm violator in the dice task can lead to own norm-breaking in the dice task, and this may subsequently lead to norm violations in the traffic light task. Our experimental design allows us to employ an instrumental variable approach to causally estimate the magnitude of this slippery slope effect, i.e. the effect of being turned into a norm violator in the dice task on norm violation in the lights task. We instrument an observer's norm violation in the dice task with random exposure to a norm violation by a first-mover, and report the 1st and 2nd stages in Table 4. The 2nd stage dependent variable is an indicator for crossing a red light, where each observation represents one light (five per individual). About 27% of individuals are 'compliers' and will be moved to break the dice norm when they observe a norm-breaking first mover (Column 1). The estimated effect for an observer of breaking the dice-norm themselves on breaking the lights norm is then 0.6764 ($p = 0.007$; Column 3). Being turned into a norm-breaker in the dice game by being matched with a norm-breaking first mover therefore increases the probability of crossing a red light by 67 pp.[36]

STUDY 3: SPILLOVERS IN THE FIELD

The lab experiments provide clear evidence of spillover effects and show how norm similarity influences their occurrence in a highly controlled, yet stylized setting. In the following sections, we will analyze the importance of norm similarity for spillover effects in a natural setting: the workplace.

[36] This is the effect of own dice cheating on lights crossing for those who are turned into a dice cheater by observing a first mover dice cheater, i.e. the complier average causal effect (or local average treatment effect). Backing out from our data the share of norm-breaking compliers in the control group at roughly 21% implies that this share will grow to 88% when individuals are put on a slippery slope.

Table 4: INSTRUMENTAL VARIABLE REGRESSION RESULTS

	1st Stage (dep. var.: own dice cheating)		2nd Stage (dep. var: lights cheating)	
	Coefficient	Std. Err.	Coefficient	Std. Err.
Own dice cheating			0.6764**	(0.2911)
First mover dice cheater	0.2706***	(0.0993)		
Constant	0.4038***	(0.0685)	0.3345**	(0.1587)
Observations	475		475	
R-squared	0.0728		0.1375	

Note: The table presents results from an instrumental variable regression where own dice cheating is instrumented with random variation in observing a first-mover dice cheater. The dependent variable in the 2nd stage is an indicator for crossing a red light, where each observation represents one light (five per individual). Standard errors are adjusted for clustering at the individual level, *** p<0.01, ** p<0.05, * p<0.1.

A Experimental Design

In our field experiment, workers are exogenously exposed to information about (alleged) tax evasion by celebrities – they thus observe violations of the norm of honestly paying taxes. We observe two outcomes: work output and stealing office material. Increased work effort leads to higher output. Since Workers receive a fixed payment, work output can serve as a proxy for the worker's intrinsic willingness to comply with the norm of exerting effort at work. Taking home office material violates the norm of not stealing. Similar to the lab experiments, we conduct an incentivized survey experiment with participants who are not part of the choice experiment, in order to elicit their beliefs regarding the distances between the three norms "do not evade taxes", "provide work effort" and "do not steal at the workplace." It turns out that the tax evasion and theft norms are more similar than tax evasion and the work effort norm (see Figure B.A.1).[37] Seeing stronger spillovers for theft would therefore be consistent with the lab results.

A.1 Procedures

Workers were recruited via job postings on a university bulletin board requesting assistance for the library in a data entry task. They were not aware that this was an experiment. For a more detailed description of the procedures and the protocol used, see Appendix C.

Upon arrival, workers received a fixed payment of €30 in an envelope and signed its receipt. Then, a

[37] Appendix A provides survey details.

research assistant explained the main data entry task to the worker, which was the same in all four treatments. As in Kube et al. (2012) workers were seated in front of a computer and had to enter bibliographic information from a selection of books into a database; Figure 7 shows pictures of the workplace.[38] The information to be entered was the book's authors, title, year of publication, and ISBN number. Workers were informed that they would perform this task for 2.5 hours but would initially start with another task for 25 minutes. The first task is where the experimental manipulation occurs: all treatments differ solely in the type of data entered into the computer during these initial 25 minutes.

(a) ROOM (b) DESKTOP AND PENS

Figure 7: PICTURES OF THE WORKPLACE.

In the first treatment (NEWS-NEUTRAL), the 25 minute task subjects performed was the same data entry task as in the main task.[39] In the second treatment (NEWS-TAX EVASION), subjects were exposed to tax evasion and had to enter information from newspaper articles on celebrities who either evaded taxes or were accused of evading taxes.[40] Subjects had to enter the name of the celebrity the article was mostly about, the year the article was published and the (purportedly) withheld amount of taxes. To control whether the overall framing of the newspaper articles, the celebrities, or the exposure to tax evasion drives possible effects we added two additional treatments. In these treatments subjects had to collect information from neutral biographies of the same celebrities used in the newspaper treatment. These biographies were taken from a professional provider of contemporary information services (Munzinger Online). In the treatment BIO-NEUTRAL, all references to tax evasion were deleted from the biographies

[38] We decided to use data entry tasks as they provide an easy way to measure individual output and work quality. Moreover, they are frequently used in other field experiments, e.g., Gneezy and List (2006), Kosfeld and Neckermann (2011), Hennig-Schmidt, Sadrieh and Rockenbach (2010), and Kube, Maréchal and Puppe (2013).

[39] In this treatment, participants were told that they have to work on the book task for 175 minutes.

[40] We did not distinguish between these two cases. If anything, we would assume this lack of distinction to downward bias our estimates.

and subjects had to collect the name of the celebrity, year the biography was updated, year of marriage and number of children. In the treatment BIO-TAXEVASION, the information on tax evasion was present in the biographies and subjects had to collect the same details as in NEWS-TAX EVASION. Example texts used in the treatments are available in Appendix D (specifically for the former tennis professional Boris Becker).[41]

After explaining the tasks, the research assistant left the room, and the worker began the first task for 25 minutes. After this period, the research assistant returned to instruct the workers to start on the second task. In all treatments it was common knowledge that the research assistant would return after 25 minutes. After the task switch and addressing any remaining questions, the research assistant left again. For the next 2.5 hours, all subjects worked on the main task. During this time, workers were unmonitored; they had the key to the room and were free to take breaks as needed. After 2.5 hours, the research assistant would return, and the subjects could then directly leave.

A.2 Dependent Variables

We measure work output by the number of books entered during the final 2.5-hour period. Our setting enables us to unobtrusively assess work output and its development over time, as the database used for the bibliographic task records the exact time of each data entry. To measure theft incidences, we placed 20 pens in a small bin on the table (see Panel (b) of Figure 7); these pens were not required for the task. The number of pens missing after the worker left the room serves as the measure of theft.[42]

B Results

Table 5 reports the summary statistics for the four treatments, each aiming to collect 25 independent observations.[43] Subjects in all treatments produce, on average, outputs ranging from 103 to 111 book entries. Neither in NEWS ($z = 0.301$ with $p = 0.76$, Mann-Whitney rank-sum test) nor in BIO ($z = -0.95$ with $p = 0.34$) do the outputs between NEUTRAL and TAXEVASION differ significantly. A comparison of the outputs across all four treatments again demonstrates no significant difference ($\chi^2(1) = 0.717$ with

[41] In very few cases did the biographies not include the information on tax evasion; we then added a paragraph with the necessary information.

[42] Otherwise, the rooms were kept clean with only a limited number of work items (such as two blocks of sticky notes) where theft would be immediately visible. In contrast, there were 20 pens in the box, making theft from there more anonymous.

[43] One worker did not show up in BIO-TAXEVASION and we ended up collecting 24 observations for this treatment.

Table 5: TREATMENTS AND SUBJECTS FIELD EXPERIMENT

		OUTPUT		FREQ. THEFT	
TREATMENT	N	MEAN	SD	MEAN	SD
NEWS:					
NEUTRAL	25	102.60	29.18	0.04	0.20
TAXEVASION	25	104.88	26.98	0.32	0.48
BIO:					
NEUTRAL	25	103.24	25.44	0.00	0.00
TAXEVASION	24	110.67	28.56	0.25	0.44

p = 0.397, Kruskal-Wallis test). OLS regressions presented in Table 6 demonstrate that the inclusion of additional controls, such as ability, does not alter these comparisons between treatments.

Table 6: ROBUSTNESS CHECK: EXPOSURE TO TAX EVASION AND OUTPUT

Dependent variable	Output			
	(1)	(2)	(3)	(4)
NEWS -NEUTRAL	6.440	-6.609	-5.127	0.270
	(7.738)	(7.487)	(7.656)	(10.25)
NEWS - TAXEVASION	1.640	-3.814	-3.547	-3.996
	(7.417)	(6.555)	(6.545)	(6.553)
BIO - TAX EVASION	7.427	0.546	1.122	2.439
	(7.735)	(6.693)	(6.608)	(7.101)
Ability		-0.637***	-0.620***	-0.597***
		(0.103)	(0.108)	(0.123)
Constant	103.2***	166.2***	189.7***	191.4***
	(5.088)	(11.62)	(23.11)	(24.75)
Additional Controls				
Gender and Age	No	No	Yes	Yes
Week, Day of Week, Time	No	No	No	Yes
Observations	99	99	99	99
R-squared	0.013	0.292	0.304	0.380

Note: Table reports the results of OLS regressions with robust standard errors in parentheses. The dependent variable in columns 1 to 4 is the number of books entered into the database during 2.5 hours. In treatments with TAX EVASION subjects were exposed to tax evasion in the 25 minutes of the first stage. Coefficients for the treatments are with respect to treatment NEUTRAL - BIO. The ability measure is based on the average time a subject needed to enter the test books. Ability was measured in a training period before the start of the treatments. Additional controls include gender, age and dummies for the running week number of the experiment (1-7), the day of the work week (1-5), and the time slot on a particular day (1-3). *** p<0.01, ** p<0.05, * p<0.1

Turning to spillovers into the more similar domain, Table 5 and Figure 8 provide the frequencies of

theft for all four treatments. In the NEWS treatments 4% of the workers steal at least one pen in NEUTRAL, while 32% do so in TAXEVASION ($\chi^2(1) = 6.64$ and $p = 0.01$, Pearson's χ^2 test). In the BIO treatments, we observe the same pattern: no worker stealing a pen in NEUTRAL, but 25% stealing in TAXEVASION ($\chi^2(1) = 7.12$ and $p = 0.008$, Pearson's χ^2 test). Finding consistently the same results in the NEWS and BIO treatments shows that our spillover results from the exposure to tax evasion, i.e., norm violation, rather than the manner in which the norm violation is presented. Pooling our data from all four treatments shows that significantly more workers steal a pen after being exposed to tax evasion ($\chi^2(1) = 13.59$ and $p < 0.001$, Pearson's χ^2 test).

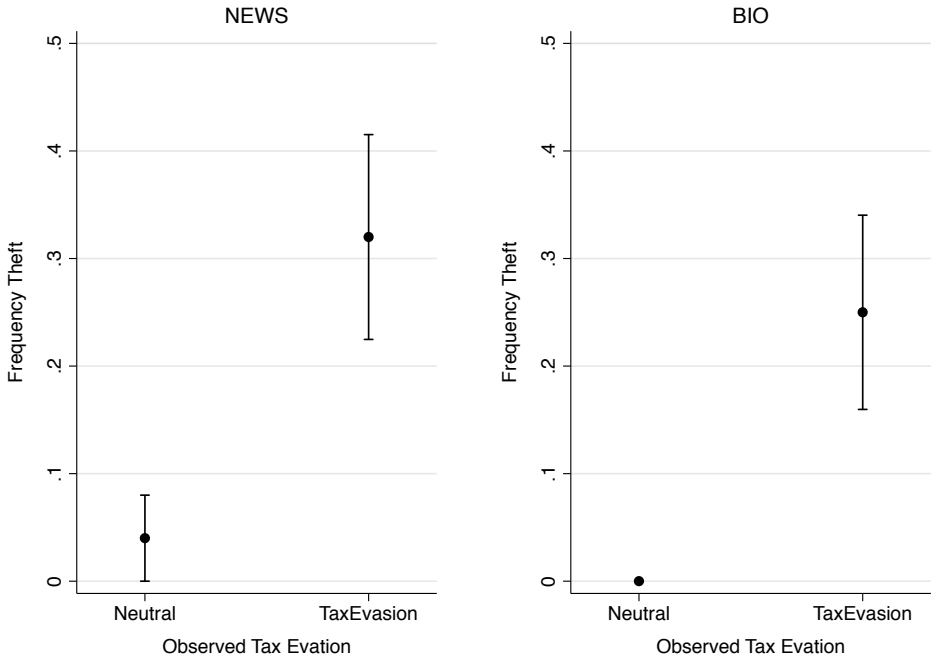


Figure 8: FREQUENCIES OF THEFT (MEANS AND STANDARD ERRORS).

More theft incidences in the TAXEVASION treatments result in significantly more stolen pens ($z = 3.6$ with $p < 0.001$, Mann-Whitney (rank-sum) test). Workers who steal take one (40%), two (20%), or three (40%) pens. None of the workers take more than three of the 20 pens.

Finally, Table 7 demonstrates that the effects of norm violations on theft are robust when controlling for additional influences. Overall, these results show that being exposed to tax evasion can affect behavior at the workplace and that these spillovers occur in a similar domain (stealing) but not in a more dissimilar one (work effort).

Table 7: ROBUSTNESS CHECK: EXPOSURE TO TAX EVASION AND THEFT

Dependent variable	Probability of Theft		Number Pens Stolen	
	(1)	(2)	(3)	(4)
NEWS -NEUTRAL	0.0436	-0.00845	0.130	0.0538
	(0.0395)	(0.0917)	(0.114)	(0.200)
NEWS - TAXEVASION	0.320***	0.322***	0.593***	0.601***
	(0.096)	(0.093)	(0.205)	(0.203)
BIO - TAX EVASION	0.251***	0.245***	0.502**	0.468**
	(0.092)	(0.091)	(0.205)	(0.226)
Constant	0.0371	0.114	-0.163	0.244
	(0.248)	(0.274)	(0.563)	(0.627)
Additional Controls				
Gender and Age	Yes	Yes	Yes	Yes
Week, Day of Week, Time	No	Yes	No	Yes
Observations	99	99	99	99
R-squared	0.145	0.320	0.111	0.259

Note: Columns 1 and 2 present results of linear probability models. The dependent variable is a binary indicator which is one if at least one pen was stolen and zero otherwise. Columns 3 and 4 give OLS estimates for the number of pens missing at the end of the session as the dependent variable. In treatments with TAX EVASION subjects were exposed to tax evasion in the 25 minutes of the first stage. Coefficients for the treatments are with respect to treatment NEUTRAL - BIO. Additional controls include gender, age and dummies for the running week number of the experiment (1-7), the day of the work week (1-5), and the time slot on a particular day (1-3).e Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

CONCLUDING DISCUSSION

In this paper, within a unified experimental framework we identify conditions for spillovers and determine when they are more or less likely to occur. Lab results show that spillovers are more likely when the norms of the decision situations are perceived as similar. However, spillovers can also affect dissimilar decision situations if there is an opportunity for self-contagion in the initial decision domain. The field experiment demonstrates the relevance of the similarity condition for effects on workplace behavior, thus showing the replicability of norm spillovers in a relevant economic setting.

There are several aspects of our study that warrant further discussion. First, we show that norm similarity consistently plays a role in the lab as well as in an important economic field setting. Assessing the applicability of the similarity condition for real-world domains beyond the workplace remains a task for future research. In this regard, our analysis may guide policymakers and future research by suggesting

where to look for spillovers: specifically in domains with norms that are rated to be similar by means of economic belief elicitation methods.

Second, it is also worth noting that our setups abstract away from complex social interactions. For instance, in the lab, we utilized simple cheating and rule-following games that focus on social contagion in completely anonymous decision settings. While this approach allows us to cleanly isolate the spillover effect and control for confounding effects from other social concerns, such as reputation, it would be interesting to explore whether social interactions intensify spillovers across games and decision situations.

Third, the results suggest that there can be welfare costs associated with spillover effects which would be overlooked when merely considering the behavioral adjustment to perceived norm violation within the same behavioral outcome dimension (e.g. tax evasion only leading to more tax evasion). This observation might be interpreted to imply that spillover effects (and the hidden costs they entail) necessitate increased policy intervention to prevent social contagion. However, it is important to note that we have adopted a positive rather than a normative approach in our framework. For instance, measures aimed at suppressing spillover effects, such as enforcing norm-compliant behavior, may come with other costs (monetary and non-monetary) that would need to be taken into account.

Finally, our approach defines norms as referring to perceptions of appropriate behavior for a particular situation. This aligns with the method of measuring social norms in recent economic experiments, which assigns social appropriateness ratings for the set of actions available in a specific decision situation or laboratory game (such as in Krupka and Weber, 2013). However, it might be argued that such a concept of norms is narrow and that meta- or umbrella norms that apply across different contexts can also exist. For example, in our experimental setting, the dot and dice decision both pertain to lying, so they might be interpreted as instances of a meta-norm, capturing various aspects of truth-telling. Nonetheless, individuals in our study de facto perceive the norms to be different when the situations are explained to them, and this forms the basis of our categorizing of behavioral adjustments across decision situations as spillovers. This approach is purely empirical and descriptive, yet it enables us to judge where spillovers may predominantly occur in our setting, namely across norms perceived as similar. It would be intriguing to develop theoretical foundations for how norms can be aggregated further, and then explore which meta-norms are more successful in explaining economic behavior than others, as well as looking into spillovers across different types of meta-norms or decision situations. We leave this avenue open for future research, as well as a deeper exploration of the dynamic aspects of how meta-norms form and evolve, and their interplay with compliance behavior from a long-term perspective.

References

Ajzenman, Nicolás, "The power of example: Corruption spurs corruption," *American Economic Journal: Applied Economics*, 2021, *13* (2), 230–257.

Alm, James, Kim M Bloomquist, and Michael McKee, "When you know your neighbour pays taxes: Information, peer effects and tax compliance," *Fiscal Studies*, 2017, *38* (4), 587–613.

Barr, Abigail, Tom Lane, and Daniele Nosenzo, "On the social inappropriateness of discrimination," *Journal of Public Economics*, 2018, *164*, 153–164.

Belot, Michèle and Marina Schröder, "The Spillover Effects of Monitoring: A Field Experiment," *Management Science*, 2016, *62* (1), 37–45.

Bicchieri, Cristina and Erte Xiao, "Do the right thing: but only if others do so," *Journal of Behavioral Decision Making*, 2009, *22* (2), 191–208.

____, Eugen Dimant, Simon Gächter, and Daniele Nosenzo, "Social proximity and the erosion of norm compliance," *Games and Economic Behavior*, 2022, *132*, 59–72.

Cason, Timothy N and Vai-Lam Mui, "Social influence in the sequential dictator game," *Journal of mathematical psychology*, 1998, *42* (2-3), 248–265.

Chen, Yan, F Maxwell Harper, Joseph Konstan, and Sherry Xin Li, "Social comparisons and contributions to online communities: A field experiment on movielens," *American Economic Review*, 2010, *100* (4), 1358–1398.

Cohn, Alain, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider, "Social Comparison and Effort Provision: Evidence From a Field Experiment," *Journal of the European Economic Association*, 2014, *12* (4), 877–898.

DellaVigna, Stefano and Ethan Kaplan, "The Fox News effect: Media bias and voting," *The Quarterly Journal of Economics*, 2007, *122* (3), 1187–1234.

____ and Matthew Gentzkow, "Persuasion: empirical evidence," *Annual Review of Economics*, 2010, *2* (1), 643–669.

Diekmann, Andreas, Wojtek Przepiorka, and Heiko Rauhut, "Lifting the veil of ignorance: An experiment on the contagiousness of norm violations," *Rationality and Society*, 2015, *27* (3), 309–333.

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

____ **, Gerben A Van Kleef, and Shaul Shalvi**, "Requiem for a nudge: Framing effects in nudging honesty," *Journal of Economic Behavior & Organization*, 2020, *172*, 247–266.

d'Adda, Giovanna, Donja Darai, Nicola Pavanini, and Roberto A Weber, "Do leaders affect ethical conduct?," *Journal of the European Economic Association*, 2017, *15* (6), 1177–1213.

Falk, Armin and Andrea Ichino, "Clean evidence on peer effects," *Journal of labor economics*, 2006, *24* (1), 39–57.

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

____ **and Franziska Föllmi-Heusi**, "Lies in disguise—an experimental study on cheating," *Journal of the European Economic Association*, 2013, *11* (3), 525–547.

____ **and Simon Gächter**, "Social preferences, beliefs, and the dynamics of free riding in public goods experiments," *American economic review*, 2010, *100* (1), 541–556.

____ **,** ____ **, and Ernst Fehr**, "Are people conditionally cooperative? Evidence from a public goods experiment," *Economics letters*, 2001, *71* (3), 397–404.

Fisman, Raymond and Edward Miguel, "Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets," *Journal of Political economy*, 2007, *115* (6), 1020–1048.

Frey, Bruno S and Stephan Meier, "Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment," *American economic review*, 2004, *94* (5), 1717–1722.

Gächter, Simon and Christian Thoeni, "Micromotives, microstructure, and macrobehavior: The case of voluntary cooperation," *The Journal of Mathematical Sociology*, 2011, *35* (1-3), 26–65.

____ **and Christian Thöni**, "Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis," *Journal of Economic Behavior & Organization*, 2010, *76* (3), 531–543.

____ **and Jonathan F Schulz**, "Intrinsic honesty and the prevalence of rule violations across societies," *Nature*, 2016, *531* (7595), 496.

____ **, Daniele Nosenzo, and Martin Sefton**, "Peer effects in pro-social behavior: Social norms or social preferences?," *Journal of the European Economic Association*, 2013, *11* (3), 548–573.

____ **, Leonie Gerhards, and Daniele Nosenzo**, "The importance of peers for compliance with norms of fair sharing," *European Economic Review*, 2017, *97*, 72–86.

Garrett, Neil, Stephanie C Lazzaro, Dan Ariely, and Tali Sharot, "The brain adapts to dishonesty," *Nature neuroscience*, 2016, *19* (12), 1727–1732.

Gino, Francesca, Michael I Norton, and Dan Ariely, "The counterfeit self: The deceptive costs of faking it," *Psychological science*, 2010, *21* (5), 712–720.

Gneezy, Uri, Agne Kajackaite, and Joel Sobel, "Lying Aversion and the Size of the Lie," *American Economic Review*, 2018, *108* (2), 419–53.

____ **and John A List**, "Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments," *Econometrica*, 2006, *74* (5), 1365–1384.

Goerg, Sebastian J and Johannes Kaiser, "Nonparametric testing of distributions—the Epps–Singleton two-sample test using the empirical characteristic function," *The Stata Journal*, 2009, *9* (3), 454–465.

Greenberg, J., "Employee theft as a reaction to underpayment inequity The hidden cost of pay cuts," *Journal of Applied Psychology*, 1990, *75* (5), 561–568.

Greiner, Ben, "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 2015, *1* (1), 114–125.

Gulino, Giorgio and Federico Masera, "Contagious dishonesty: Corruption scandals and supermarket theft," *American Economic Journal: Applied Economics (Forthcoming)*, 2023.

Hennig-Schmidt, Heike, Abdolkarim Sadrieh, and Bettina Rockenbach, "In search of workers' real effort reciprocity—A field and a laboratory experiment," *Journal of the European Economic Association*, 2010, *8* (4), 817–837.

Ichino, Andrea and Giovanni Maggi, "Work environment and individual background: Explaining regional shirking differentials in a large Italian firm," *The Quarterly Journal of Economics*, 2000, *115* (3), 1057–1090.

Kajackaite, Agne and Uri Gneezy, "Lying costs and incentives," *UC San Diego Discussion Paper*, 2015.

Keizer, Kees, Siegwart Lindenberg, and Linda Steg, "The spreading of disorder," *Science*, 2008, *322* (5908), 1681–1685.

Kimbrough, Erik O and Alexander Vostroknutov, "Norms make preferences social," *Journal of the European Economic Association*, 2016, *14* (3), 608–638.

____ and ____ , "Norms make preferences social," *Journal of the European Economic Association*, 2016, *14* (3), 608–638.

Kleven, Henrik Jacobsen, Martin B Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez, "Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark," *Econometrica*, 2011, *79* (3), 651–692.

Kosfeld, Michael and Susanne Neckermann, "Getting more work for nothing? Symbolic awards and worker performance," *American Economic Journal: Microeconomics*, 2011, *3* (3), 86–99.

Krupka, Erin and Roberto A Weber, "The focusing and informational effects of norms on pro-social behavior," *Journal of Economic psychology*, 2009, *30* (3), 307–320.

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

____ , Stephen Leider, and Ming Jiang, "A meeting of the minds: informal agreements and social norms," *Management Science*, 2017, *63* (6), 1708–1729.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe, "Do wage cuts damage work morale? Evidence from a natural field experiment," *Journal of the European Economic Association*, 2013, *11*, 853–870.

Lane, Tom, Daniele Nosenzo, and Silvia Sonderegger, "Law and Norms: Empirical Evidence," *American Economic Review*, May 2023, *113* (5), 1255–93.

Lanfear, Charles C, Ross L Matsueda, and Lindsey R Beach, "Broken windows, informal social control, and crime: assessing causality in empirical studies," *Annual review of criminology*, 2020, *3*, 97–120.

Lefebvre, Mathieu, Pierre Pestieau, Arno Riedl, and Marie Claire Villeval, "Tax evasion and social information: an experiment in Belgium, France, and the Netherlands," *International Tax and Public Finance*, 2015, *22*, 401–425.

Lindenberg, Siegwart and Linda Steg, "Goal-framing theory and norm-guided environmental behavior," in "Encouraging sustainable behavior," Psychology Press, 2013, pp. 37–54.

List, John A. and Fatemeh Momeni, "Leveraging upfront payments to curb employee misbehavior: Evidence from a natural field experiment," *European Economic Review*, 2020, *130*, 103601.

List, John and Fatemeh Momeni, "When Corporate Social Responsibility Backfires: Evidence from a Natural Field Experiment," *Management Science*, 06 2020, *67*.

Mas, Alexandre and Enrico Moretti, "Peers at work," *American Economic Review*, 2009, *99* (1), 112–145.

Mittone, Luigi and Matteo Ploner, "Peer pressure, social spillovers, and reciprocity: An experimental analysis," *Experimental Economics*, 2011, *14*, 203–222.

Pittarello, Andrea, Margarita Leib, Tom Gordon-Hecker, and Shaul Shalvi, "Justifications shape ethical blind spots," *Psychological Science*, 2015, *26* (6), 794–804.

Rauhut, Heiko, "Beliefs about lying and spreading of dishonesty: Undetected lies and their constructive and destructive social dynamics in dice experiments," *PloS one*, 2013, *8* (11), e77878.

Robert, Innes and Mitra Arnab, "Is dishonesty contagious?," *Economic Inquiry*, 2013, *51* (1), 722–734.

Shang, Jen and Rachel Croson, "A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods," *The economic journal*, 2009, *119* (540), 1422–1439.

Smith, Sarah, Frank Windmeijer, and Edmund Wright, "Peer effects in charitable giving: Evidence from the (running) field," *The Economic Journal*, 2015, *125* (585), 1053–1071.

Thöni, Christian and Simon Gächter, "Peer effects and social preferences in voluntary cooperation: A theoretical and experimental analysis," *Journal of Economic Psychology*, 2015, *48*, 72–88.

Welsh, David T, Lisa D Ordóñez, Deirdre G Snyder, and Michael S Christian, "The slippery slope: how small ethical transgressions pave the way for larger future transgressions.," *Journal of Applied Psychology*, 2015, *100* (1), 114.

Wilson, James Q and George L Kelling, "Broken windows," *Atlantic monthly*, 1982, *249* (3), 29–38.

Appendix

A Supplementary Material Study 1

A Instructions: Study 1

A.1 First-Mover Instructions

Thank you for participating in our experiment (Hand out)

Please use only the programs and functions provided for the experiment. Do not talk to other participants. If you have a question, please raise your hand. We will come to you and answer your question in silence. Please do not ask your question out loud. If the question is relevant to all participants, we will repeat it out loud and answer it.

Your payment for participating in the experiment consists of a show-up fee of €5 and a flexible amount. The flexible amount will depend on the decision you make during the experiment. These instructions describe your decision situation. All participants in the experiment have received the same instructions. It is important that you read these instructions thoroughly and accurately so that you understand the decision situation.

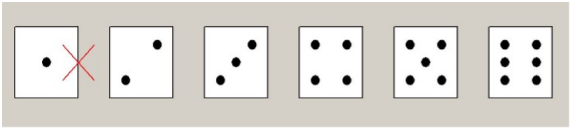
Your decision is anonymous. The payout is also completely anonymous by transmitting the income accrued to your seat number to the payout office at the end of the experiment. The amount of money collected for each seat number will be paid in cash. The person at the payout office receives a list that only includes seat numbers and the income collected on them. That means she cannot observe the decision you made during the experiment or the random components realized for you.

Your decision may be communicated to a third, randomly selected person in a later session. This person will only be informed that this is the behavior of an unknown person who participated in the experiment at an earlier time.

The Decision Situation (on-screen)

Now your decision situation will be described. You will soon see six dice on your screen. The number of dots on the dice increases from left to right—from 1 to 6. In addition, one of the dice with the numbers 1

to 6 is marked with a red cross. Which of these six dice is marked with the cross is **randomly determined for each participant**. Each of the six dice has an equal probability of being marked with the cross.



Example: The dice will be displayed for a duration of 5 seconds. Then a new screen appears.

Your task is as follows: You should indicate the number of dots on the marked dice by clicking on the corresponding button.

For the above example, you should click on "1":

You will receive the **number of € corresponding to the button you chose as your payout. If you press the "1" button, you will receive €1. If you press the "2" button, you will receive €2, and so on, up to €6 for pressing the "6" button.**

First, you can familiarize yourself with the decision situation by performing **three practice rounds**. These practice rounds are not payout relevant. The subsequent **fourth round is payout relevant**. You will be informed separately on the screen when the payout-relevant round starts.

Comprehension Questions *Example situation: Suppose the red cross marks the number "3".* Which button should you click? Enter the answer in the box: (text box) Is it possible to click the "2" button or the "5" button? (Yes, No) What payout will you receive if you click the "4" button? (text box) The experiment will not continue until everyone has answered the comprehension questions. Someone will come to you to check your comprehension answers. Please wait for the experiment to start on the computer.

A.2 Second-Mover Instructions

Thank you for participating in our experiment (Hand out)

General Information

Please only use the programs and functions provided for the experiment. Please do not talk to the other participants. If you have a question, please raise your hand. We will come to you and answer your question quietly. Please do not ask your question out loud. If the question is relevant to all participants, we will repeat it out loud and answer it.

Your payment for participating in the experiment consists of a show-up fee of €5 and a flexible amount. The flexible amount will depend on the decisions you make during the course of the experiment. These instructions describe your decision-making situations. All participants in the experiment have received the same instructions. It is important that you read these instructions thoroughly and accurately so that you understand the decision-making situations.

Your decisions are anonymous. Payment is also completely anonymous, with the income collected for your seat number being transmitted to the payment station after the end of the experiment. The money collected for your seat number will be paid out in cash. The person at the payment station will receive a list that only contains the seat numbers and the income collected for them. This means that they cannot observe the decisions you made during the experiment or the random components that were realized for you.

Experiment Setup

The experiment consists of two parts. You will now be given instructions for the first part of the experiment. Instructions for part 2 will be explained to you later on the screen.

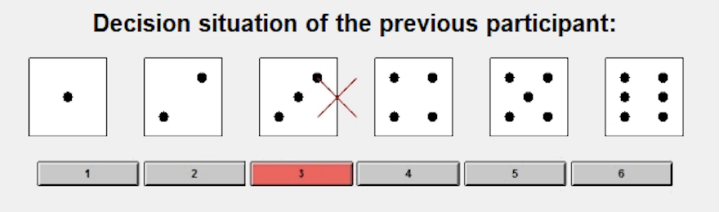
Part 1

First, we would like you to understand the decision-making situation of participants in a previous experiment. To do this, you will be presented with the **instructions from the previous experiment. Please read them carefully.** At the end of the instructions for the previous experiment, you will find comprehension questions. The experiment on the screen will not continue until all participants have answered these comprehension questions correctly. Someone will come to you to check the comprehension questions.

[On Screen]

You will soon see how a former experiment participant behaved in the just-described decision situa-

tion. We will show you the decision of a person who participated in a session that took place this Monday and who sat exactly in the same cubicle that was assigned to you today. On the next screen, you will see which number was marked for the previous participant in the payout-relevant round, and which number this person then selected on the decision screen by clicking the button. Please click on CONTINUE to see this information.



Please answer the following questions. If both answers are correct, your show-up fee will be increased by €1. What number was marked for the previous participant? (text box) What number did the previous participant click? (text box) Please wait, we will continue shortly.

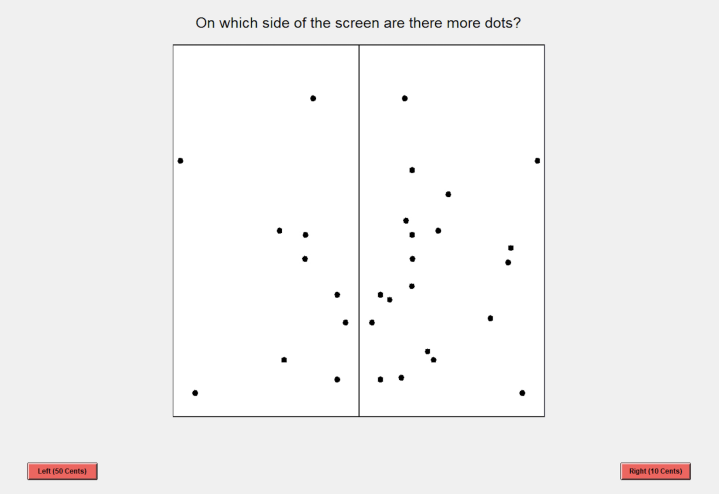
In the first part of the experiment, you earned €1. We will now start the second part of the experiment. The instructions for the second part will be displayed on your screen. Please read the instructions carefully. Click to the next screen only when you have understood everything. If something is unclear, please raise your hand. Someone will come to assist you.

[Treatment: Dice-Dice] You are now also in the decision situation of the previous participant. You will soon see six dice on your screen. One of the dice is marked with a red cross. It is the same dice that was randomly marked for the previous participant. The dice will be displayed for a period of 5 seconds. Your task is the same as the previous participant: You should indicate the number on the marked dice by clicking on the corresponding button. You will receive as payout the number of the button you have selected in € . Pressing the button "1" will give you 1€. Pressing the button "2" will give you 2€ and so on, up to 6€ by pressing the button "6". Please press START to begin your payout-relevant round. There is no practice round.

[Treatment: Dice-Dot] In this part of the experiment, you will be shown 12 different screens, each displaying a picture. Each picture is divided into two halves by a vertical centerline. Black dots are located

on the left and right sides of the centerline. In one half of the picture, there are more dots than in the other. Below the picture are two buttons—"left" and "right". The rule in this game is as follows: You should indicate for each picture which side of the centerline has more dots. To do this, select the respective "left" or "right" button. Your payout will be determined for each picture as follows: If you choose the left [treatment: right] button, you will receive 50 cents. If you choose the right [left] button, you will receive 10 cents. After you have made your decision, the next picture will be displayed. If you click OK, the comprehension questions will be displayed to you.

Comprehension Questions: Please refer to the example picture below. For the example picture, should you press the "Left" button or the "Right" button? (Left, Right) What amount will you receive if you press the "Right" button for the example picture? (10 Cents, 50 Cents) What amount will you receive if you press the "Left" button for the example image? (10 Cents, 50 Cents)

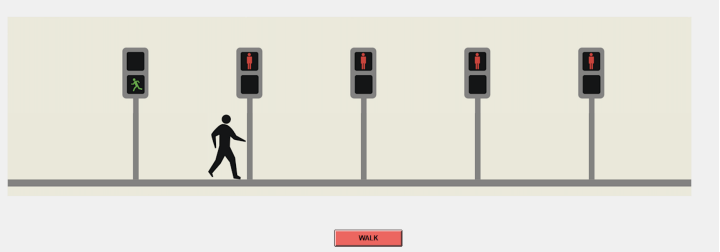


Thank you, you have answered the comprehension questions correctly. Click START to begin the game.

[Treatment: Dice-Light] You control a stickman that can move from left to right on the screen. When you click the "Start" button, the stickman starts running. Your character will run into a series of red traffic lights and stop at each traffic light to wait. Your stickman will only start running again at a traffic light when you click the "Walk" button displayed in the middle of the screen (one click is enough). The rule in this game is the following: You are supposed to wait at each red traffic light until it turns green. Your

payout is determined by how much time (in seconds) your character takes to run from the left to the right side of the screen. Your initial amount is €8. For each second that your character takes to run from left to right, this amount is reduced by 10 cents. If you click OK, the comprehension questions will be displayed.

Comprehension Questions: In the example situation, should you wait until the traffic light turns green or press "Walk"? (Wait until light turns green, Press "WALK") In the example situation, does your stickman start walking when you press "Walk"? (Yes,No) If your character waits for two seconds longer in the example situation, how many cents will your income decrease by? (0 cents, 20 cents)



Thank you, you have answered the comprehension questions correctly. Click START to begin the game.

A.3 Instructions: Norm Elicitation Experiment

Welcome to our experiment (Hand out)

General information

Please only use the programs and functions provided for the experiment. Please do not talk to the other participants. If you have a question, please raise your hand. We will then come to you and answer your question silently. Please do not ask your question out loud. If the question is relevant to all participants, we will repeat it out loud and answer it. The instructions are the same for all participants.

Your payment for participating in the experiment consists of a fee of €8 and a flexible amount. The flexible amount will depend on the answers you and the other participants give to the questions asked during the course of the experiment.

In today's study, you will read descriptions of a series of situations in which "Person A" must decide

on a course of action. The description includes several available options for Person A. After reading the descriptions, you will be asked to evaluate each of the different courses of action. You are to judge whether the actions or behavior are "appropriate and consistent with generally accepted ethical or moral rules of behavior" or "inappropriate and not consistent with generally accepted ethical and moral rules of behavior." By appropriate behavior, we mean behavior that most people would say is "correct" or "ethically right." Another way to describe this is that if Person A behaves inappropriately, someone else could be upset with Person A's behavior because they believe that Person A should have acted differently in that situation.

We ask that you answer each question as truthfully as possible and based on what you classify as appropriate or inappropriate behavior. To give you an idea of how the experiment will proceed, we will show you an example of how to answer your questions.

You will find the example on the next page. At the end of the instructions, you will find comprehension questions that you should answer after reading the instructions completely and silently. The experiment will not proceed until all participants have answered the comprehension questions correctly. If something is unclear to you, or if you have answered the comprehension questions, please raise your hand. Someone will come to you to answer your question or check the comprehension questions.

Example situation: Person A is in a café. A notices that someone has left a wallet on one of the tables. A must decide how to behave. A has four different options: take the wallet; ask other guests if the wallet belongs to them; leave the wallet where it is; give the wallet to the counter.

A can only choose one of the four options. The four possible actions are shown in the table below. You are now asked to indicate how ethically and morally appropriate each of these actions is. To submit your answer, please check one of the boxes "very appropriate", "somewhat appropriate", "somewhat inappropriate", and "very inappropriate" for each course of action. Now take a look at the table silently and **please do not fill in anything yet**.

If this were one of the situations for this study, you would indicate for each of the above possible courses of action to what extent you perceive the action as "appropriate and consistent with morally or ethically right behavior" or "inappropriate and not consistent with morally or ethically right behavior." As a reminder, appropriate behavior means behavior that most people would consider "right" or "ethi-

Person A ...

	very appropriate	somewhat appropriate	somewhat inappropriate	very inappropriate
...takes the wallet				
...leaves the wallet where it is				
...gives the wallet to the counter				
...asks other guests if the wallet belongs to them				

cal."

Instructions and payment

In the following experiment, several situations will be described that all involve decisions that "Person A" has to make. Imagine that these are decisions that Person A has to make as a participant in an experiment. Each of the situations will be described to you on the screen. You will then provide your answers by placing a checkmark in the corresponding cell.

At the end of the experiment, we will randomly select one of the situations. For this situation, we will also randomly select one of the possible actions that the participant could choose. So we will randomly select both the situation and the possible action. For the chosen action, we will determine which answer was given by the most people who are here today. **If you gave the same answer as the majority of other participants in today's study, you will receive an additional €5 on top of the show-up fee.** This amount will be paid to you in cash at the end of the experiment, together with the show-up fee. Suppose we have chosen the option "to leave the wallet" in the situation shown. If your answer was "somewhat inappropriate" and this is also the answer given by the majority of other participants in today's session, you will receive €5 in addition to the show-up fee.

Comprehension Questions

Assuming that you would later have to evaluate the actions of the example situation mentioned above in the experiment. Assume further that your evaluation of the possible actions is as follows: Taking the

wallet is "very inappropriate". Asking other guests if the wallet belongs to them is "somewhat appropriate". Leaving the wallet where it is is "somewhat inappropriate". Handing the wallet in at the counter is "very appropriate".

a) Tick the corresponding answers in the table. (text box)

b) Assuming that the action "Handing the wallet in at the counter" is determined for payment by random selection, and most other participants in today's study have checked "very appropriate" for this action. How much additional € would you earn in addition to the show-up fee? Enter the amount in the box. (text box)

c) How many € would you earn in addition to the show-up fee if most other participants in today's study checked "somewhat appropriate" for the action described in b)? Enter the amount in the box. (text box)

Please answer the comprehension questions and wait for someone to come and check them.

[On Screen]

In the following, we explain you the three decision situation. Press OK to continue. [The dice decision situation of Person A is explained.]

In the decision situation described above, there are six different individual situations depending on which dice has been randomly marked. In the following, these situations are shown to you. For each of these individual situations, you should indicate whether you think it is "appropriate" or "inappropriate" to click the "1", "2", "3", "4", "5" or "6" button in the selection bar in this situation (and thus receive the number of the button paid out in €). As a reminder, by appropriate behavior, we mean behavior that most people would say is "correct" or "ethically right". The different individual situations will be shown to you on the screen one after the other and in random order. **Your payment will be determined as follows**: If the dice decision situation is selected as relevant for payment at the end of the experiment, a single situation will be selected randomly again. From this single situation, a question will then be randomly selected (for example, "How appropriate or inappropriate is it to click the digit 3 when the digit 1 is marked?"). If your answer to this question matches the answer most frequently given by other participants today, you will receive **€5** in addition to the appearance fee.

[The dot and the traffic light decision situation of Person A are explained (random order).] In the decision situation described above, Person A is shown 12 pictures. The difficulty level of all pictures corresponds to the example picture shown above, and Person A has unlimited time to view the pictures. On 4 pictures, there are more dots on the left side. That is, on 8 pictures, there are more dots on the right side. You should now indicate whether you think it is "appropriate" or "inappropriate" to click the "right" button a total of "1 time", "2 times", "3 times", "4 times", "5 times", "6 times", "7 times", or "8 times" for these 8 pictures. As a reminder, by appropriate behavior, we mean behavior that most people would say is "correct" or "ethically right". **Your payment will be determined as follows:** At the end of the experiment, one question will be randomly selected from all the questions (for example, "How appropriate or inappropriate is it to click 'right' on 6 out of 8 pictures?"). If your answer to a question matches the answer most frequently given by the other participants today, you will receive **€5** in addition to the appearance fee.

The figure of the participant stops at 5 traffic lights. The waiting time at each traffic light is 5 seconds. Now you are asked to indicate how "appropriate" or "inappropriate" it is to click the "WALK" button "once", "twice", "three times", "four times" or "five times" in total before the traffic light turns green at these 5 traffic lights. As a Reminder: By appropriate behavior, we mean behavior that most people would say is "correct" or "ethically right". **Your payment will be determined as follows**: At the end of the experiment, one question will be randomly selected from all the questions (for example, "How appropriate or inappropriate is it to click the 'WALK' button 3 times at the 5 traffic lights where the figure stops before the light turns green?"). If your answer to this question matches the answer given by the majority of the other participants today, you will receive **€5** .

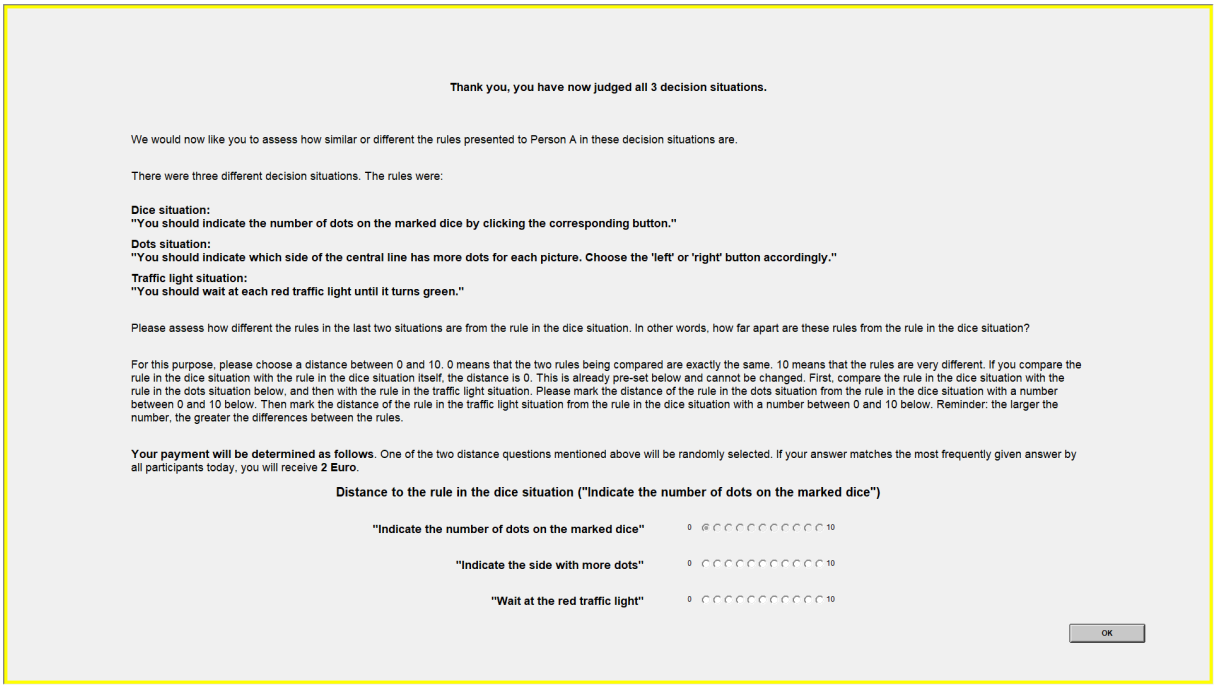


Figure A.A.1: Screenshot Distance Ratings

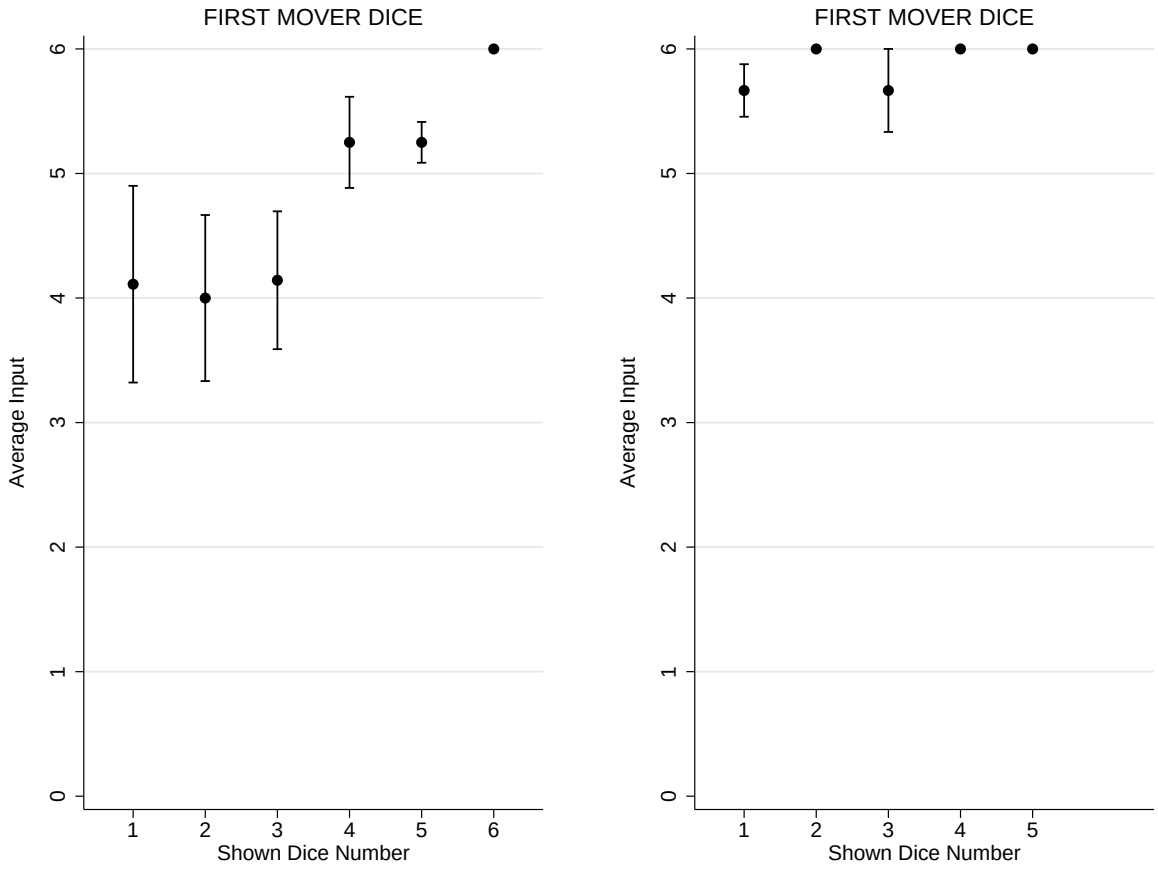
Distance Ratings (screenshot)

B Study 1: First Movers

Behavior of First Movers. This section summarizes the data of the two first-mover session. In Figure A.B.1 a), we show for each dice which number subjects have reported on average. Figure A.B.1 (b) depicts the reported dice numbers for only those subjects who have cheated. When subjects cheat most of them do so to the fullest extent (they report a six). This is line with Kajackaite and Gneezy (2015) and Gneezy, Kajackaite and Sobel (2018) who observe that partial lying is rare in a setting where cheating is completely anonymous and payouts are double blind.[44]

Selection of first-mover Data shown to Observers: Using the DICE-game as the first-mover task has the advantage that we can communicate information about rule-compliance in a neutral manner by just reporting which dice was marked for a former participant and which number she afterward clicked. Yet

[44] In this case, there is no point of trying to disguise a lie by not reporting the maximal outcome, as, for instance, in Fischbacher and Föllmi-Heusi (2013).



(a) AVERAGE INPUT. (b) AVERAGE INPUT (CHEATERS ONLY)

Figure A.B.1: AVERAGE INPUT IN FIRST-MOVER SESSIONS (MEANS AND STANDARD ERRORS).

this game has some indeterminacy about first movers' intentions: whenever a six is drawn, monetary incentives are aligned with truth-telling so that the information about how a person behaved in this situation does not allow to update about her intrinsic willingness to follow the game's rule. We therefore decided to collect two first-mover sessions, enabling us to show dice numbers from one to five only. Specifically, we created one hypothetical first-mover session by removing the sixes from one first-mover session and replacing them with the corresponding observations from the same seats of the other session. Whenever there was no six drawn for a given seat, we decided to take those observations such that the distribution of dice numbers and the frequencies of cheating by dice number (and thus the overall cheating rate) come as close as possible to those of the two first-mover session. In figure A.B.2, we show the distribution of marked dice over the two first-mover session and the distribution of dice in our hypothetical session. There are indeed no significant differences between these two (excluding sixes)

according to the Epps-Singleton test implementation by Goerg and Kaiser (2009) ($p = 0.88$).

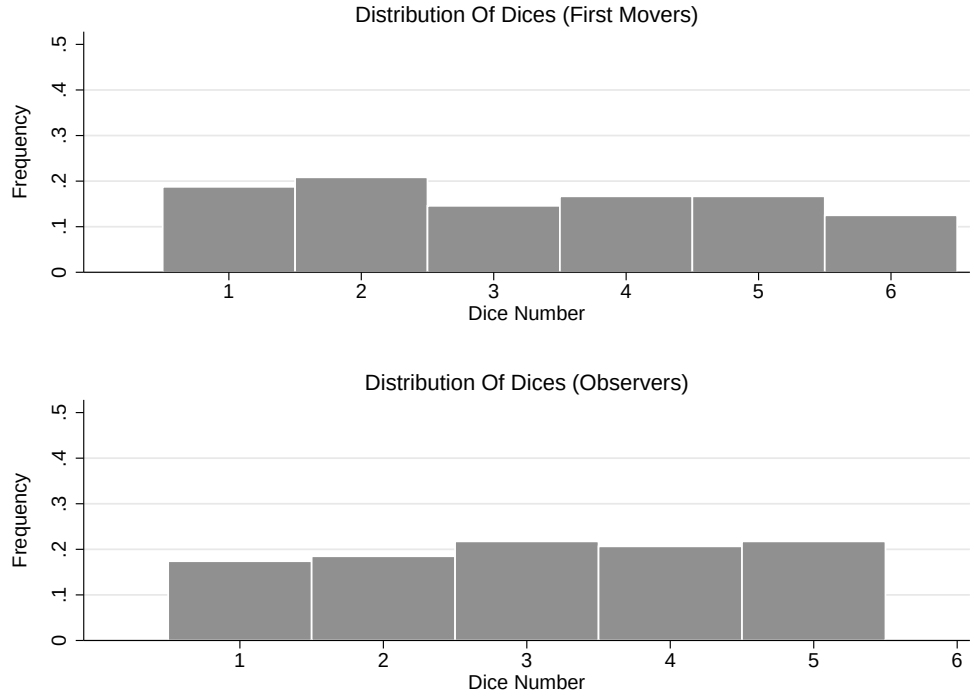


Figure A.B.2:
ACTUAL DISTRIBUTION OF FIRST MOVERS AND DISTRIBUTION OF DICE SHOWN TO OBSERVERS.

In Figure A.B.3, we compare the distribution of drawn dice used in the hypothetical between first movers who violated the norm and those who did not. Again, there are no significant differences using the Epps-Singleton test ($p = 0.555$), ruling out that observers are shown systematically different dice numbers when being matched with a cheater and when not.

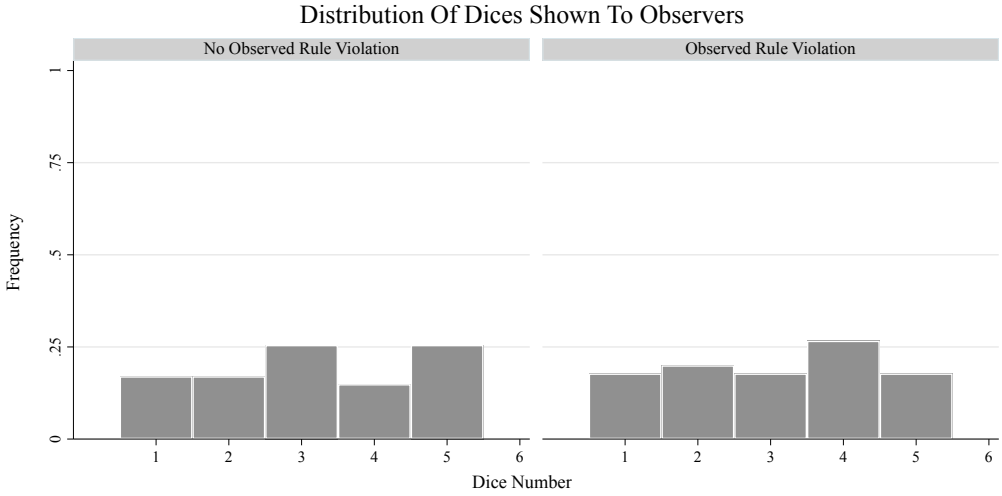


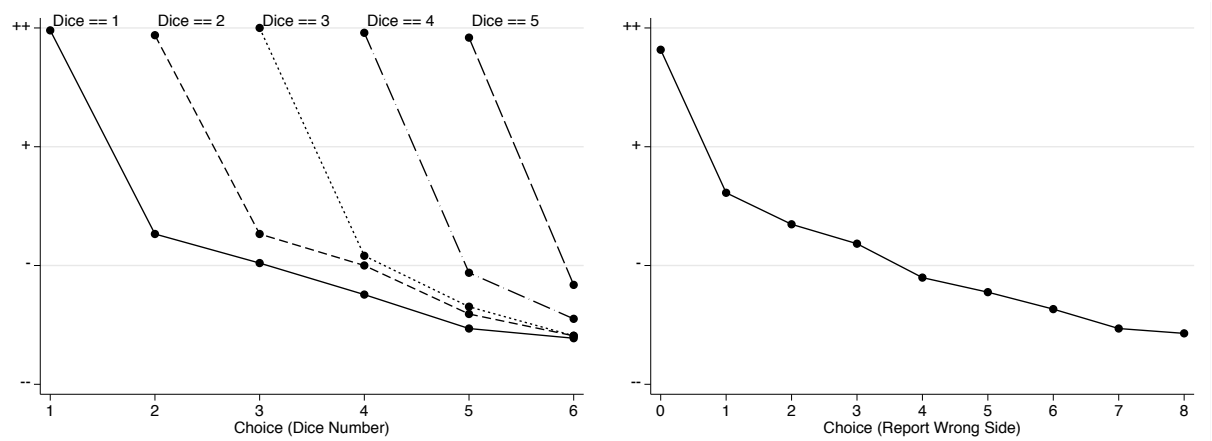
Figure A.B.3:
DISTRIBUTION OF DICES FOR FIRST MOVERS WHO DID NOT AND DID VIOLATE THE NORM

C Study 1: Additional Tables and Figures

Table A.C..1: Summary Statistics Norm Violation by Observers

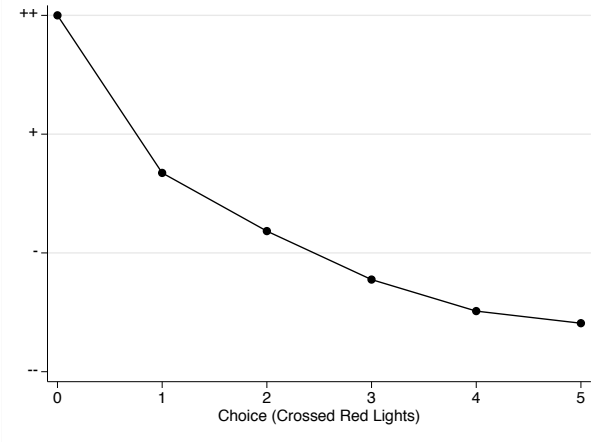
Treatment	N	Obs.	Freq. Norm Violation Mean	Sd	p-Value
Dice-Dice					
first-mover: Correct	47	47	0.64	0.49	χ^2: $p = 0.024$
first-mover: Rule Violation	45	45	0.84	0.37	
Dice-Dot					
first-mover: Correct	47	376	0.52	0.44	Mw: $p = 0.024$
first-mover: Rule Violation	47	376	0.72	0.42	
Dice-Light					
first-mover: Correct	46	230	0.55	0.46	Mw: $p = 0.511$
first-mover: Rule Violation	46	230	0.61	0.46	

Notes: One decision in Dice-Dice, multiple decisions in Dice-Dot (12×) and Dice-Light (5×). Dice-Dot data includes only decisions with incentives against correct reporting (8×).



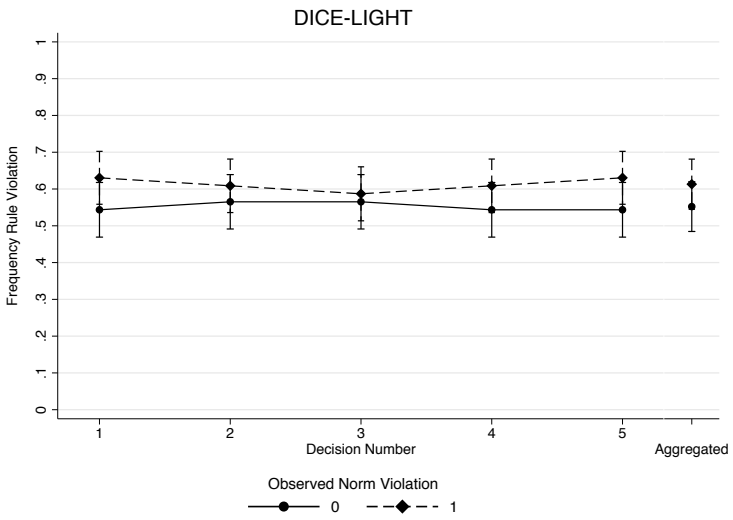
(a) DICE NORM RATINGS (MEAN)

(b) DOT NORM RATINGS (MEAN)



(c) LIGHT NORM RATINGS (MEAN)

Figure A.C.4: APPROPRIATENESS RATINGS FOR DIFFERENT ACTIONS IN THE VARIOUS GAMES.
The figure displays the mean answers from the social norm elicitation experiment on a scale from 1, indicated by "--" ('Very socially inappropriate'), to 4, represented by "++" ('Very socially appropriate'). Panel (a) shows the ratings for reporting a certain dice number, considering all possible drawn dice numbers in the DICE-game. Panel (b) shows the mean ratings for the DOT-game, encompassing all available actions (from 0 to 5 times reporting the wrong side to having more dots). Panel (c) displays the mean ratings for actions in the LIGHT-game (0 to 5 times crossing red traffic lights).



(b) DICE-LIGHT Rule violation defined as crossing a red light

Figure A.C.5: FREQUENCIES OF RULE VIOLATIONS (MEANS AND STANDARD ERRORS).

B Supplementary Material Study 3

A Study 3: Distance Elicitation

Analogously to Experiment 1 we run an incentivized survey to confirm that subjects indeed perceive the relevant norms on work effort, steeling, and tax evasion as different. Again, the survey was conducted with individuals that did not participate in any of the other experiments or sessions. Two sessions (N = 44) were conducted at the laboratory of the TU Berlin and subjects were invited via the local ORSEE-system (Greiner, 2015). The survey was programmed in z-Tree (Fischbacher, 2007). The sessions were conducted as an add-on to a different experiment on a different topic.

In the experiment, participants were given descriptions of three different decision situations. Afterwards they were asked to evaluate how far the norms that apply to these situations are from each other. Specifically, subjects were supposed to judge the distance of the norms that apply to work effort and steeling at the workplace to the norm that applies to tax evasion, using a rating scale from 0 to 10 (where 0 means the norms are exactly the same and 10 that they are maximally distant). It is important to note that in contrast to the lab experiments we do not work with injunctive norms and do not describe what the appropriate behaviour norm is, leaving this to the participants (see the instructions below). Again, respondents are paid to match the modal response provided by other participants' distance ratings. Subjects received a show-up fee of €4 and for each norm an additional €2 if their reported distance matched the mode in the session. Again, payments were made using a double blind mechanism.

Figure B.A.1 gives the average distance ratings. Subjects report that the norms applying to stealing and work effort are not the same as the norms that apply to tax evasion (both with $p<0.001$, Wilcoxon signed-rank test). However, the distance between tax evasion and stealing is smaller than the distance between tax evasion and work effort ($p<0.001$).

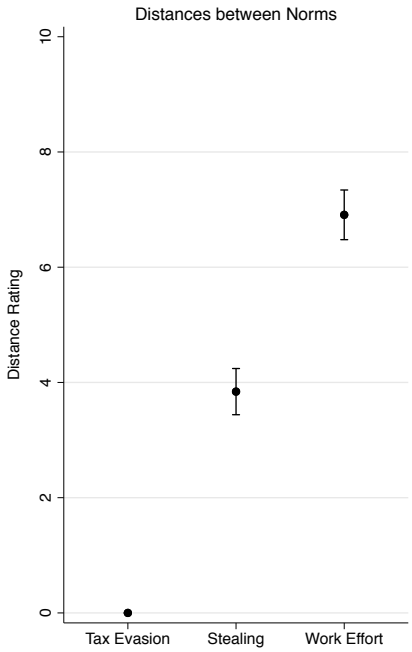


Figure B.A.1: Distance to the norm that apply to the situation of tax evasion

Instructions Distance Elicitation

Our behavior is regularly influenced by social norms. Social norms are generally accepted rules that determine what behavior a society expects from individuals. Social norms determine what people perceive as acceptable behavior in a decision-making situation. For example, an individual is faced with the decision of whether to throw their trash on the street or not. The social norm in this case could be: "One should not throw trash on the street (but in a garbage can)". In this study, you will examine the social norms in the following three decision-making situations:

- Decision situation 1 (tax evasion): Someone faces the decision of whether to evade taxes or not.
- Decision situation 2 (stealing at the workplace): Someone faces the decision of whether to steal office materials at work or not.
- Decision situation 3 (work effort): Someone faces the decision of whether to make an effort when typing data at work or not.

First, think about each of the three decision-making situations: What behavior is generally expected by society here? What is therefore the social norm or rule that determines what behavior is perceived as socially acceptable?

You should then assess how much the three norms in the decision-making situations differ. In other words, how far apart the norms are from each other. Choose a distance between 0 and 10 below for this purpose. 0 means that two norms are exactly the same. 10 means that two norms are very different. If we compare the norm of decision situation 1 with the norm that applies to decision situation 1—i.e., the same norm twice—the distance is 0 (this is already preset below and cannot be changed). Specifically, you should assess the following questions: How far apart is the norm that determines behavior in decision situation 1 (tax evasion) from the norm that determines behavior in decision situation 2 (stealing at the workplace)? How far apart is the norm that determines behavior in decision situation 1 (tax evasion) from the norm that determines behavior in decision situation 3 (work effort)?

To do this, please first mark the distance between the norms of decision situations 1 and 2 below with a number between 0 and 10. Then please mark the distance between the norms of decision situations 1 and 3 below with a number between 0 and 10. As a reminder: the larger the number, the greater the differences between the norms.

B Study 3: Screenshots

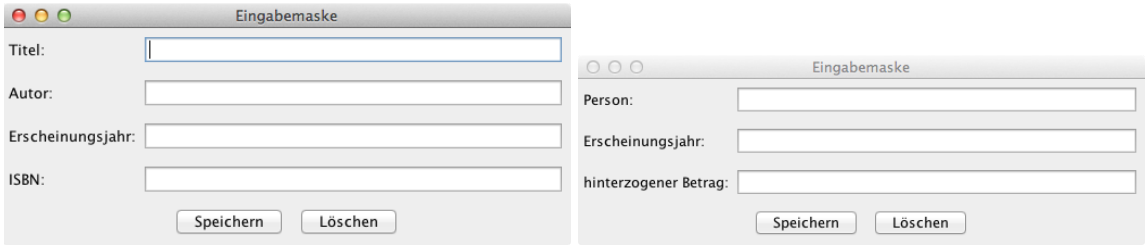


Figure B.B.1: SCREENSHOTS
LEFT: Input mask used for book entries.
RIGHT: Input mask used for tax evasion entries

C Study 3: Experimental Protocol

Workers were recruited through job postings that requested assistance with a data entry task. The job was advertised as a one-time opportunity limited to three hours, with an hourly wage of €10 (see Appendix E for a copy of the announcement). In total, we received about 600 job applications, most of which were from students. The participants were then randomly assigned to a treatment. We are confident that the composition of the participants did not change throughout the experiment, as we did not observe a significant correlation between theft incidences and the running session number ($r = -0.0235$, $p = 0.817$, biserial correlation) or work output and the running session number ($r = 0.035$, $p = 0.731$, Spearman correlation). We held three sessions each day: one in the morning (8:30 am-11:30 am), one at noon (12:00 pm-3:00 pm), and one in the afternoon (3:30 pm-6:30 pm). Treatments were alternated during these sessions.

Upon arrival, the participants were given an envelope containing €30. After signing the receipt, we asked them to keep the money and proceeded to explain the tasks. We informed them that they would need to perform two tasks. Firstly, they were introduced to the book entry task. To ensure they understood the task, they had to enter two books on their own. Afterwards, if the participants were in the newspaper or biography treatments (Evasion-Newspaper, Neutral-Bio, or Tax Evasion-Bio), we introduced them to the second task.

In the "Evasion-Newspaper" task, we informed them that for a research project they need to collect the following information (if provided): the name of the person the article deals with, the year of publication, and the amount evaded. We used the same task description in the "Tax Evasion-Bio" task. In the "Neutral-Bio" task, we required participants to collect the following information (if provided): the name of the person the article deals with, the year of publication, the name of the current spouse, the year of marriage, and the number of children. In the "Neutral-Data Entry" task, the participants did not perform a second task but started working on the book entries directly.

In all four treatments the experimenter returned after 25 Minutes. In the newspaper and biography treatments the experimenter then started the book entry task, while in “Neutral - Data Entry” the experimenter asked whether any questions occurred. Subjects in all treatments were made aware at the beginning of the experiment that the experimenter would return after 25 Minutes. After this the experimenter left and all subjects were working on the book entry task for the following 150 minutes. After 150 minutes the experimenter returned and the subjects left.

A protocol was used by the research assistants who handled the interaction with the participants. Everything was kept the same between treatments except for the second task (item 4 in the protocol), this is where we applied the different treatments. In the following we provide the experimental protocol.

Protocol

- Do not report anything about former participants
- Do not give numbers on how many books are expected. If asked, reply as many as possible
- Always stick to this protocol!

1. **Preparations**
    1. Treatment: pick up two binders
    2. Keys
    3. Envelope with money
    4. Replacement pens and post-it notes
    5. Check input mask to match task
        * If the input-mask does not match, copy the correct program to the desktop.
2. **Introduction: organizational stuff at the beginning**
    1. Payment and receipt
        * Ask to come in: "Thank you for coming. You can drop your stuff over there." Make participant aware of the envelope with the €30 on the table. Ask them to sign a receipt with name, address, and signature. After participant has signed receipt. "You can pocket the money right away."
    2. One-time job
        * After participant has signed receipt, emphasize that this is a one-time job today. We do not have an open permanent RA position. Also students are not repeatedly invited for this job.
    3. Participant can leave the rom anytime, e.g., restroom, coffee....
    4. Sketch two tasks "In the next three hours we want you to do some cataloging tasks: Books and Articles. I will first explain the book task, which will take most of your time."
3. **Explain book task**
    1. Sketch: from left to right, computer, book trolley. As many as possible
        * Background information: We need to catalog some books for the library. Your task is to enter the data from the books in this shelf into the computer. From left to right, then place them in the book trolley (2 rows per level). Do as many books as you can.
    2. Explain input-mask. Go over the handout with the example. Type in first example
        * Point at title, author, publication year, ISBN. Also explain the input-mask.

        - Change input-fields with tabulator
        - Save data entry with click on the button
        - If program is closed accidentally, just open it again with a double on the icon on the desktop
    - Show the handout and go over the details
        - Do not enter sub-title
        - Spelling is not case-sensitive
        - Author always in the order: *first name, second name*, separate authors with a semicolon *first name, second name*; *first name, second name*
        - For collection of essays: use editor as author
        - If dissertation: use author and not publisher
        - Publication date: use the most recent date
        - Enter ISBN without hyphen
        - If ISBN and ISSN: use ISBN
        - If two ISBN: just select one
        - If you cannot find information, enter "?" in corresponding field
    - Enter first example
    - Show other examples, but do not type into computer
    - After book is done, put it on the library trolley
    - "Are there any questions?" Give some time and allow for questions
3. Participant should enter two books
    - Example books: Mankiw and Franz: "To check if you understood everything you should enter two books" Explain that if fields remain empty they are still saved. Leave student alone during the input. Respond only to inquiries.
    - While student enters data: prepare the article-folders

Any questions left? Okay, before you start with the book task we want you to do a different task: We need you to enter data from articles. You will do this task first and the input-mask is very similar to the first. Roughly 25 minutes. Then you switch to the book task. I will now explain how the article task works.

4. Explain second task 'article task'. Depends on treatments

<u>Treatment: Neutral-Bio</u>

- Put folder with articles on table, put hand-out next to it
- Info: For a research project, please search for the following information in the articles and enter the information into the input-mask: Name of the celebrity, year of publication of the article, name of spouse, year of marriage number of kids.
- Go through the details with the help of the handout
    - Name the person you think this article is mostly about
    - Names always in order: *First Name, Second Name*
    - Year of publication for the article, not year of last update

    - Name of current spouse in the order: *First Name, Second Name*
    - Year of marriage
    - Number of kids
    - If you cannot find some of the information, please enter "?" into the corresponding field

Treatment: Tax Evasion-Bio

- Put folder with articles on table, put hand-out next to it
- Info: For a research project, please search for the following information in the articles and enter the information into the input-mask: Name of the celebrity, year of publication of the article, evaded amount.
- Go through the details with the help of the handout
    - Name the person you think this article is mostly about
    - Names always in order: *First Name, Second Name*
    - Year of publication for the article, not year of last update
    - evaded amount - depending on the article, according to the charges, suspicion, or verdict. (Explain difference between evaded amount and not declared income)
    - Evaded amount always with currency, e.g., DM or €
    - Round amounts to nearest reasonable number
    - If you cannot find some of the information, please enter "?" into the corresponding field

Treatment: Tax Evasion-News

- Put folder with articles on table, put hand-out next to it
- Info: For a research project, please search for the following information in the articles and enter the information into the input-mask: Name of the celebrity, year of publication of the article, evaded amount.
- Go through the details with the help of the handout
    - Name the person you think this article is mostly about
    - Names always in order: *First Name, Second Name*
    - Year of publication for the article, not year of download
    - evaded amount - depending on the article, according to the charges, suspicion, or verdict. (Explain difference between evaded amount and not declared income)
    - Evaded amount always with currency, e.g., DM or €
    - Round amounts to nearest reasonable number
    - If you cannot find some of the information, please enter "?" into the corresponding field

Treatment: Neutral - Data Entry

- Announce that you return after roughly 25 minutes

In all treatments

- Announce that you return after roughly 25 minutes

D Study 3: Example of Newspaper/Biography treatments

SPIEGEL ONLINE

23. Oktober 2002, 07:34 Uhr

Prozessbeginn

Boris Becker wegen Steuerhinterziehung vor Gericht

Ex-Tennisstar Boris Becker muss sich von heute an vor dem Landgericht München I wegen des Vorwurfs der Steuerhinterziehung verantworten. Er war Anfang der Neunziger im Steuerparadies Monaco gemeldet, soll aber hauptsächlich in München gelebt haben.

München - Die Staatsanwaltschaft wirft dem dreifachen Wimbledonsieger vor, dem deutschen Fiskus rund 1,5 Millionen Euro Steuern vorenthalten zu haben. Die Münchner Steuerfahndung fand in jahrelangen Ermittlungen heraus, dass Becker von 1991 bis 1993 seinen Wohnsitz zwar in dem Steuerparadies Monaco gehabt, aber überwiegend im Münchener Nobelstadtteil Bogenhausen gelebt habe.

Die Anklageschrift gegen Boris Franz Becker - so lautet sein vollständiger Name - umfasst 46 Seiten. Die Akten mit den Ermittlungsergebnissen füllen allein 15 Aktenordner. Darüber hinaus liegen dem Verfahren mehr als 1000 Aktenordner mit Beweismaterial zu Grunde, das auch aus mehreren Hausdurchsuchungen stammt.

Die 4. Strafkammer mit der Vorsitzenden Richterin Huberta Knöringer hat für das Verfahren vorerst drei Verhandlungstage angesetzt. Mit Spannung wird erwartet, ob Becker - falls die Vorwürfe ganz oder teilweise zutreffen - gleich zu Prozessbeginn ein Geständnis ablegen wird. Damit könnte er nach Angaben aus Juristenkreisen ein langwieriges Verfahren vermeiden und auf eine Bewährungsstrafe hoffen. Bei einem Schuldspruch droht dem 34-Jährigen Gefängnis.

Im Vorfeld des Prozesses hatten Spekulationen über einen entsprechenden "Deal" der Verfahrensbeteiligten - nämlich milde Bewährungsstrafe bei Geständnis und Steuernachzahlung - immer wieder für Aufsehen gesorgt. Nach Angaben des Gerichts und der Staatsanwaltschaft gibt es eine solche Verständigung aber nicht. Eine entsprechende Absprache kann jedoch auch nach Prozessbeginn getroffen werden. Sie sind gerade in Wirtschaftsprozessen üblich, um langwierige Beweisaufnahmen abzukürzen. Nach der Rechtsprechung des Bundesgerichtshofs ist aber auch bei Absprachen zwingend zu berücksichtigen, dass die Strafe schuldangemessen ist.

Becker soll unbestätigten Informationen zufolge kurz vor Prozessbeginn knapp 1,6 Millionen Euro an das Finanzamt überwiesen haben. Der Wahlmünchner und gebürtige Leimener hat sich bisher kein Wort zu dem Prozess entlocken lassen. Für die Vorbereitung des Verfahrens hat er drei Verteidiger gewonnen: die beiden Münchner Juristen Prof. Klaus Volk und Jörg Weigell sowie Prof. Erich Samson aus Hamburg.

Samson, der an der privaten, von der "Zeit"-Stiftung getragenen Hamburger Hochschule "Bucerius Law School" unterrichtet, wird nach eigenen Angaben aber am Prozess nicht teilnehmen. Volk unterrichtet an der Münchner Ludwig-Maximilians-Universität Strafrecht, Wirtschaftsstrafrecht und Strafprozessrecht. Weigell hat sich unter anderem mit Veröffentlichungen zum internationalen Steuerrecht einen Namen gemacht.

Für Becker ist der Prozess nicht der erste Rückschlag seit dem Ende seiner Profikarriere im Jahre 1999. Im Sommer 2001 meldete Beckers Onlineportal Sportgate Insolvenz an. Auch seine Sportler-Agentur BBM hat ihre Geschäfte eingestellt. "Sportlich einst Weltklasse, geschäftlich nur Kreisklasse", kommentierte das die "Financial Times Deutschland". Zuletzt machte er Schlagzeilen mit angeblichen Verkaufsplänen seiner Villa auf Mallorca, die trotz fünfjähriger Bauarbeiten noch nicht bezugsfertig ist.

Figure B.D.2: EXAMPLE OF A NEWSPAPER ARTICLE USED IN TAX EVASION – NEWS TREATMENT. ORIGINAL VERSION.

Der Spiegel
October 23, 2002, 07:34 am
Trial Begins
Boris Becker on Trial for Tax Evasion

Former tennis star Boris Becker is set to face trial at the Munich Regional Court I on charges of tax evasion. In the early 1990s, he was registered in the tax haven of Monaco but was believed to have primarily lived in Munich.

Munich - The prosecution accuses the three-time Wimbledon champion of withholding approximately 1.5 million euros in taxes from the German treasury. After years of investigation, the Munich tax investigation department discovered that Becker had his residence in the tax haven of Monaco from 1991 to 1993 but had mainly resided in the upscale Munich district of Bogenhausen.

The indictment against Boris Franz Becker, his full name, consists of 46 pages. The investigative files alone fill 15 case folders. Additionally, the proceedings are supported by more than 1000 case folders containing evidence, including materials obtained from several house searches. The 4th Criminal Chamber, presided over by Judge Huberta Knöringer, has initially scheduled three trial days. It is eagerly anticipated whether Becker, if the allegations are wholly or partly true, will confess at the beginning of the trial. According to legal sources, this could potentially allow him to avoid a protracted legal process and hope for a suspended sentence. If found guilty, the 34-year-old faces imprisonment.

Speculations about a potential "deal" among the parties involved, namely a lenient suspended sentence in exchange for confession and tax repayment, have sparked public interest prior to the trial. However, according to the court and the prosecution, such an agreement does not exist. Nonetheless, such an arrangement could still be reached even after the trial begins. They are commonly employed in economic cases to expedite lengthy evidence gathering. However, according to the case law of the Federal Court of Justice, it is imperative to ensure that the punishment is commensurate with the culpability, even in cases of agreements.

Unconfirmed reports suggest that shortly before the trial, Becker transferred nearly 1.6 million euros to the tax office. The Munich resident and Leimen native has not uttered a word about the trial. He has enlisted three defense lawyers to prepare for the proceedings: the Munich-based legal experts Prof. Klaus Volk and Jörg Weigell, as well as Prof. Erich Samson from Hamburg.

Samson, who teaches at the private "Bucerius Law School" in Hamburg, supported by the "Zeit" Foundation, stated that he would not participate in the trial. Volk teaches criminal law, economic criminal law, and criminal procedure law at the Ludwig Maximilian University in Munich. Weigell has gained recognition through publications on international tax law, among other accomplishments.

This trial is not the first setback for Becker since ending his professional career in 1999. In the summer of 2001, Becker's online portal Sportgate filed for bankruptcy. His sports agency BBM also ceased operations. "Once a sporting world-class, only amateurish in business," commented the "Financial Times Deutschland". Recently, he made headlines with alleged plans to sell his unfinished villa in Mallorca, which has been under construction for five years.

Figure B.D.3: EXAMPLE NEWSPAPER ARTICLE USED IN TAX EVASION – NEWS TREATMENT. ENGLISH TRANSLATION.

Figure B.D.4: Example biography in Tax Evasion – Bio (bottom), Neutral – Bio (top) treatments. Original version (abbreviated).

Boris Becker

deutscher Tennisspieler
Geburtstag:22. November 1967 Leimen bei Heidelberg
Klassifikation: Profisportler, prominenter Sportler
Nation: Deutschland - Bundesrepublik
Internationales Biographisches Archiv 11/2009 vom 10. März 2009 (wk) Ergänzt um Nachrichten durch MA-Journal bis KW 22/2011

Herkunft
Boris Becker, kath., wurde am 22. Nov. 1967 als Sohn eines Architekten in Leimen bei Heidelberg geboren und wuchs auch dort auf. Sein Großvater war Busfahrer in Heidelberg gewesen, die Urgroßeltern hatten in Leimen eine kleine Zigarrenfabrik. Seine Mutter stammt aus dem Sudetenland.
[...]
Wirken
Der erste große Turnier-Erfolg gelang B. 1983/1984 beim "Orange Bowl" in Miami Beach/Florida, den inoffiziellen Weltmeisterschaften der Tennisjunioren, wo er sich sowohl im Einzel als auch im Doppel bis ins Finale vorkämpfte. Als 16-Jähriger nahm er 1984 erstmals am Grand-Slam-Turnier in Wimbledon teil und kam auf Anhieb bis in die dritte Runde, wurde dann jedoch durch eine Verletzung gestoppt.
[...]
Für viel mediales Aufsehen sorgte auch B.s Steueraffäre, in die er bereits in den 90er Jahren geraten war. Im Dez. 1996 hatte die Steuerfahndung in seiner Abwesenheit seine Münchener Villa durchsucht, im Juli 1998 war es dann erneut zu einer bundesweiten Aktion der Steuerfahndung gegen den Tennisstar gekommen. Erst im Okt. 2002 endete dann ein jahrelanges Verfahren gegen B., als dieser vom Münchner Landgericht wegen Steuerhinterziehung zu einer Freiheitsstrafe von zwei Jahren auf Bewährung verurteilt wurde. [...] Das Gericht befand B. für schuldig, den Finanzbehörden verschwiegen zu haben, dass er in den Jahren zwischen 1991 und 1993 neben seinem Wohnsitz in Monaco auch eine Wohnung in München unterhalten habe. Damit hatte Becker Einkommen- und Vermögenssteuern in Höhe von rund 1,7 Mio. Euro hinterzogen.
[…]

Boris Becker

deutscher Tennisspieler
Geburtstag:22. November 1967 Leimen bei Heidelberg
Klassifikation: Profisportler, prominenter Sportler
Nation: Deutschland - Bundesrepublik
Internationales Biographisches Archiv 11/2009 vom 10. März 2009 (wk) Ergänzt um Nachrichten durch MA-Journal bis KW 22/2011

Herkunft
Boris Becker, kath., wurde am 22. Nov. 1967 als Sohn eines Architekten in Leimen bei Heidelberg geboren und wuchs auch dort auf. Sein Großvater war Busfahrer in Heidelberg gewesen, die Urgroßeltern hatten in Leimen eine kleine Zigarrenfabrik. Seine Mutter stammt aus dem Sudetenland.
[...]
Wirken
Der erste große Turnier-Erfolg gelang B. 1983/1984 beim "Orange Bowl" in Miami Beach/Florida, den inoffiziellen Weltmeisterschaften der Tennisjunioren, wo er sich sowohl im Einzel als auch im Doppel bis ins Finale vorkämpfte. Als 16-Jähriger nahm er 1984 erstmals am Grand-Slam-Turnier in Wimbledon teil und kam auf Anhieb bis in die dritte Runde, wurde dann jedoch durch eine Verletzung gestoppt.
[...]

Boris Becker

German Tennis Player
Date of Birth: November 22, 1967
Birthplace: Leimen near Heidelberg
Classification: Professional athlete, prominent sports figure
Nationality: Germany - Federal Republic
International Biographical Archive 11/2009, as of March 10, 2009 (wk), supplemented with news through MA-Journal until week 22/2011

Background
Boris Becker, a Catholic, was born on November 22, 1967, in Leimen near Heidelberg as the son of an architect. He grew up in the same town. His grandfather worked as a bus driver in Heidelberg, and his great-grandparents owned a small cigar factory in Leimen. His mother is from Sudetenland.
[...]
Career
B. achieved his first major tournament success in 1983/1984 at the "Orange Bowl" in Miami Beach, Florida, the unofficial world championships of junior tennis. He reached the finals in both singles and doubles. At the age of 16, he participated in the Wimbledon Grand Slam tournament for the first time in 1984 and immediately made it to the third round but was then halted by an injury.
[…]
Significant media attention was also caused by B's tax affair, which he had become involved in during the 1990s. In December 1996, tax authorities searched his Munich villa in his absence, and in July 1998, there was another nationwide operation by the tax authorities against the tennis star. It wasn't until October 2002 that a lengthy legal process against B. finally concluded, with the Munich District Court sentencing him to a two-year suspended prison term for tax evasion. [...] The court found B. guilty of concealing from the tax authorities that, between 1991 and 1993, he maintained a residence in Munich in addition to his residence in Monaco. As a result, Becker evaded income and wealth taxes amounting to approximately 1.7 million euros.
[…]

Boris Becker

German Tennis Player
Date of Birth: November 22, 1967
Birthplace: Leimen near Heidelberg
Classification: Professional athlete, prominent sports figure
Nationality: Germany - Federal Republic
International Biographical Archive 11/2009, as of March 10, 2009 (wk), supplemented with news through MA-Journal until week 22/2011

Background
Boris Becker, a Catholic, was born on November 22, 1967, in Leimen near Heidelberg as the son of an architect. He grew up in the same town. His grandfather worked as a bus driver in Heidelberg, and his great-grandparents owned a small cigar factory in Leimen. His mother is from Sudetenland.
[...]
Career
B. achieved his first major tournament success in 1983/1984 at the "Orange Bowl" in Miami Beach, Florida, the unofficial world championships of junior tennis. He reached the finals in both singles and doubles. At the age of 16, he participated in the Wimbledon Grand Slam tournament for the first time in 1984 and immediately made it to the third round but was then halted by an injury.
[…]

Figure B.D.5: EXAMPLE BIOGRAPHY IN TAX EVASION – BIO (BOTTOM), NEUTRAL – BIO (TOP) TREATMENTS. ENGLISH TRANSLATION (ABBREVIATED).

E Study 3: Job Announcement

Leibniz Universität Hannover

Students needed for a temporary job – now!

3 hours, €30.

The **Institute for Social Policy** at the Economics Department of the Leibniz University in Hannover is looking for for reliable helpers to support the work of our librarians. Your task will be to catalog books and other documents. Basic knowledge of using a PC is necessary.

The job is limited to 3 hours. The hourly wage is €10 and will be paid in cash. We will arrange the exact time for the job on a **day between 3.12. until 21.12.2012** together with you.

Location:
Institute for Social Policy
Königsworther Platz 1 (Conti-Campus)
30167 Hannover

Interested? Send an email to the following address: **bib@sopo.uni-hannover.de**

Just put "bib" in the subject line! We will send you a form, where you can check the dates when you are available.

ISBN: 978-91-8082-171-1 (pdf)

Linnæus University