Inferring and enforcing cooperativeness and dishonesty through rule abidance and partner choice

Jörg Gross, Sebastian Simon, Jim A.C. Everett

Institute of Psychology, Leiden University, The Netherlands and

School of Psychology, University of Kent, United Kingdom

Author Note

Correspondence can be addressed to Jörg Gross, Institute of Psychology, Leiden University, PO Box 9555, 2300 RB Leiden, The Netherlands; email: mail@joerg-gross.net

Financial support was provided by the Netherlands Science Foundation VENI Award (016.Veni.195.078), the Gratama Foundation and the Leiden University Fund to JG.

Word count (abstract and abstract): 4,077

Abstract

Rules, in the form of customs or norms, regulate social interactions, help to coordinate joint action, and allow to predict others’ actions. Since many rules demand to restrict selfishness but are not governed by formal laws or punishment, like tipping for a service, standing in line, or generously sharing resources, the question arises how rule abidance is maintained. Here we investigate partner choice as a mechanism that shapes rule abidance in groups. We show that when individuals engage in costly pro-social interactions, they use the rule following propensity of others as a cue to select interaction partners. Because rule-abiding individuals are preferred as interaction partners, rule following and pro-sociality co-emerge. Yet, when interactions between individuals are based on anti-social behavior – in our experiment cheating to secure personal benefits – individuals prefer partners that break rules. Concomitantly, the willingness to follow costly rules decreases. Resonating with previous research, our data shows that partner choice can decrease selfishness and increase costly rule following. Contrary to previous results, our results also highlight that partner choice can favor selfishness and rule breaking. Taken together, our results shed light on how rule following is enforced but also breaks down through partner choice, demonstrating how social interactions can shape rule abidance.

Keywords: cooperation, dishonesty, partner choice, rule following, norms

Inferring and enforcing cooperativeness and dishonesty through rule abidance and partner choice

Rules play an important part in social interactions. Whether in the form of customs, norms, or legal regulations, rules enable individuals to predict what others will do and help to coordinate joint action (Bicchieri, 2006; Chudek & Henrich, 2011; Gelfand, 2018). Rules also often demand to restrict selfish behavior. For example, adhering to speeding limits, not crossing a red street light, or lining up in a queue interferes with proceeding towards one’s destination or to not spend more time than strictly necessary. Previous research has shown that punishment by peers or a central authority can help to enforce rule abidance in the face of selfish temptations to break rules (Andrighetto et al., 2013; Fehr & Fischbacher, 2004; Molho, Tybur, van Lange, & Balliet, 2020). Yet, there are also rules that are not (or only imperfectly) enforced through punishment or formal sanctions, like tipping for a service or telling the truth (Cialdini & Trost, 1998). This raises the question how rule abidance can emerge and be maintained in such situations.

Here we investigate the possibility that partner choice mechanisms can enforce rule abidance in the absence of punishment. Previous research has shown that partner choice can be a powerful mechanism to sustain cooperative relationships, and that forms a critical basis of our moral sense (Alexander, 1987; Baumard, André, & Sperber, 2013). When individuals can freely choose their interaction partner, cooperators can search for like-minded cooperators and avoid defectors. In the long-run, cooperation has a competitive advantage over defection and the incentive for defectors to start cooperating themselves increases (Barclay, 2016; Debove, Andre, & Baumard, 2015; Fu, Hauert, Nowak, & Wang, 2008; McNamara, Barta, Fromhage, & Houston, 2008; J. Wang, Suri, & Watts, 2012). When choosing interaction partners, people may search for cues that allow to infer the cooperativeness of a potential interaction partner before entering an interaction (Jordan, Hoffman, Nowak, & Rand, 2016; Sommerfeld, Krambeck, Semmann, & Milinski, 2007; Tognetti, Berticat, Raymond, & Faurie, 2013; Wilson & Eckel, 2016). Accordingly, it has been argued that such mechanisms of partner choice whereby we search for cues of cooperativeness in people’s judgments can help explain the prevalence of rule-following deontological moral intuitions even when following utilitarian judgments would lead to better outcomes overall (Everett et al. 2016; 2018). In daily life, we hypothesize that one of these cues for cooperativeness is rule abidance, because when rule abidance is personally costly, it could serve as a credible signal for the willingness or ability to also exhibit costly pro-social behavior. Indeed, it has been shown that there is a link between rule following and cooperation. In a study by Kimbrough & Vostroknutov (2016), people that were willing to follow a costly rule exhibited more cooperative behavior and showed sustained cooperation in a public goods dilemma. Importantly, if (a) costly rule following is used to signal the willingness to cooperate and (b) interaction partners are chosen based on this signal, both costly cooperative actions and costly rule following may enforce each other, allowing rule abidance and cooperation to co-emerge through partner choice.

Yet, the connection between cooperation and rule abidance could also reverse (Melnikoff & Bailey, 2018). People sometimes collude to lie and bend rules for personal profit, as high-profile corruption scandals like the Diesel or Wirecard scandal illustrate. Indeed, recent studies have highlighted that cooperation is sometimes aimed at breaking societal rules (Gross, Leib, Offerman, & Shalvi, 2018b; Weisel & Shalvi, 2015); that people coordinate lies if covering the truth is mutually beneficial (Soraperra et al., 2017; Ścigała, Schild, Heck, & Zettler, 2018; Weisel & Shalvi, 2015; Wiltermuth, 2011); and that people that are willing to cheat for personal profit selectively search for partners that are willing to cheat themselves (Gross et al., 2018b). In an environment that provides opportunities to jointly profit from violating norms like honesty, people may, hence, strategically become ‘rule breakers’ to signal their value as a ‘partner in crime.’

Taken together, this means that abiding to costly rules is enforced or breaks down through partner choice, depending on the local incentives of the social environment. To test these conjectures, we experimentally let participants choose interaction partners for a social interaction that is either based on pro-social giving or breaking the rule of honesty. We hypothesized that costly rule following is used to signal cooperativeness and that rule abiding partners are more often chosen when interactions require costly cooperative actions. In contrast, when social interactions are rewarded on breaking a particular societal rule, the opposite should be the case: People should signal their willingness to break rules, revealing how environmental contingencies and partner choice can foster or deteriorate rule following in groups.

# Methods

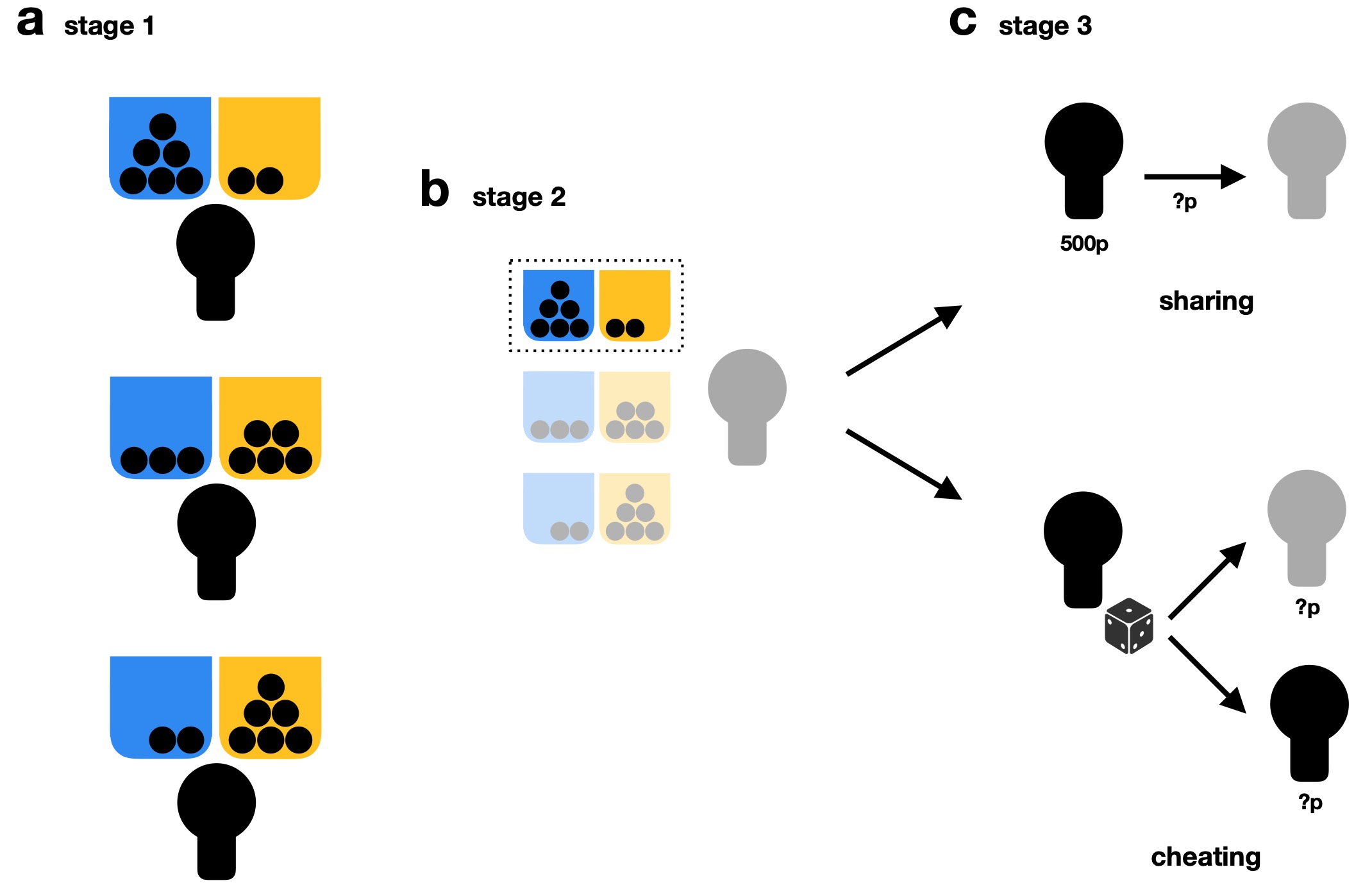
### Sample. We invited 192 participants for a study on decision making (148 female, *M*age = 23, *SD*age = 4.3). The study received ethics approval from the Psychology Ethics Board of Leiden University. Subjects provided written informed consent and were debriefed upon completion of the studies. The experiment did not involve any deception.

**Experimental Design.**Participants were assigned to groups of four (*n* = 48 groups). In each group, one participant was randomly assigned the role of the ‘partner selector.’ The other three participants were assigned the role of ‘potential partners.’ Each round consisted of three stages. In the Stage 1, each potential partner was confronted with a simple rule following task (Figure 1a, Kimbrough & Vostroknutov, 2018). In this task, each potential partner received 15 virtual balls and had to decide, one by one, whether to assign each ball to either a blue or yellow bucket. For each ball they put in the blue bucket they received 5 points, whereas for each ball they put in the yellow bucket they received 15 points. Points, the currency of the experiment, were converted to money at the end of the experiment (see *Procedure* below). In the instructions, it was explained to all participants that ‘the rule is to place each ball in the blue bucket.’ Hence, the task confronted the three potential partners with a conflict between following the rule or maximizing their own profit: the rule was to place in the blue bucket, but placing the ball in the yellow bucket lead to them receiving more points, and therefore money. Importantly, we did not enforce this rule through economic sanctions or any formal punishment, making this a similar rule to societal norms that are voluntarily followed (or not). The behavioral rule following task has been shown to correlate with personality traits like need for structure (Gross & De Dreu, 2017), cooperation in social dilemma situations, and respecting norms like trust, pro-sociality, and honesty (Gross & De Dreu, 2020; Kimbrough & Vostroknutov, 2016; Kimbrough, Miller, & Vostroknutov, 2014). Resonating with the idea that rules create a conflict between restricting behavior and maximizing own benefits (Pfister, Wirth, Schwarz, Steinhauser, & Kunde, 2016; Pfister, Wirth, Weller, Foerster, & Schwarz, 2018), a recent brain stimulation study also showed that rule following is causally linked to the right lateral prefrontal cortex, a brain region that has been associated with value-based cost-benefit decisions (Gross, Emmerling, Vostroknutov, & Sack, 2018a).

In Stage 2, all participants including the partner selector were informed about how many balls each potential partner allocated according to the rule. Then, the partner selector selected interaction partners for Stage 3 (Figure 1b). The partner selector had to select at least one partner to interact with, but could also select up to three interaction partners . Selecting partners was costly (like in everyday life) whereby the partner selector received an endowment of 450 points in each round and had to pay 150 points for each selected partner, keeping the rest of the points for herself.

Each selected partner moved to Stage 3. What happened in this stage differed depending on the condition of the experiment (Figure 1c). In the ‘sharing’ condition, selected partners engaged in a simple social exchange task: the dictator game (Kahneman et al. 1986; Forsythe et al. 1994). Each selected partner received 500 points and had to choose how to split these 500 points between herself and the partner selector. She could choose between keeping the 500 points to herself or give between 50 and 250 points (in steps of 50) to the selector. Hence, each selected partner had six options on how to allocate these points, ranging from keeping everything (selfish option) to splitting the points equally (most pro-social option). After each selected partner made their decision, the partner selector was informed about the outcome and the group moved to the next round. Given the cost of entering a social exchange for partner selectors, chosen partners had to give at least 200 points to the partner selector. The difficulty for partner selectors was hence to select those potential partners that would be willing to share at least 40% of their points with them to prevent a net loss.

In the ‘cheating’ condition, selected partners engaged in a simple cheating task (Fischbacher & Föllmi-Heusi, 2013). Each selected partner had to throw a six-sided die and was instructed to report the outcome to the computer. The number they reported had consequences for their payoff as well as the payoff of the partner selector. Specifically, they both received 0 points for reporting to have thrown a ‘one’, 50 points for a ‘two’, 100 points for a ‘three’, up to 250 points each for reporting to have thrown a ‘six’. Analogous to the ‘sharing’ condition, the partner selector had to choose partners that at least reported a ‘five’ to break even with the cost of selecting a partner. High numbers can be guaranteed if partners are willing to misreport their die-roll outcome and cheat (i.e. violating the norm of honesty). Whereas partner selectors had to select pro-social partners in the sharing condition, they profited most from selecting partners that were consistently willing to lie in the cheating condition. We chose a die-rolling tasks as a frequently-used method to investigate cheating behavior, where lying can be detected from significant deviations from the probability of each number being thrown (i.e., 16.7%). Earlier work showed that lying in die roll tasks correlates with unethical behaviors in other domains, such as not paying for public transport (Dai, Galeotti, & Villeval, 2017), being absent from work without a reason (Hanna & Wang, 2017), not returning undeserved pay (Potters & Stoop, 2016), misbehaving in school (Cohn & Maréchal, 2017), and diluting milk with water in a dairy market (Kröll & Rustagi, 2016).



*Figure 1.* Experimental design.Three participants in the role of ‘potential partners’ had to decide how to allocate balls between a blue and a yellow bucket. Each ball assigned to the blue bucket yielded 5 points, while each ball assigned to the yellow bucket yielded 15 points. Participants were given a rule to assign each ball to the blue bucket to create a conflict between abiding by a rule and maximizing own profit (**a**). The ‘partner selector’ (grey) observed the decisions of the potential partners and had to select at least one partner to interact with in the next stage (illustrated by the dotted line) (**b**). Stage 3 differed depending on the condition. In the sharing condition, each selected partner received 500 points and could decide how many points to give to the partner selector. In the cheating condition, selected partners had to privately roll a die and report the outcome to the computer. The higher the number they reported, the more money she and the partner selector earned.

To investigate how rule following changes depending on the incentives of the two different environments, in a repeated measures design groups were confronted with each condition (cheating vs. sharing) in blocks of 15 rounds. The order of blocks was counterbalanced across groups. Between rounds, partner selectors could not track the identity of potential partners to ensure that partner selection was only based on the revealed propensity to follow rules and not reputation or direct reciprocity. Importantly, this also means that a rational-selfish player in the role of a partner should always keep everything to herself in the sharing condition to maximize her profits. Since it is in the best interest for any rational-selfish player to follow this strategy, rule following should not carry any credible information, since any chosen partner would select the most selfish option in stage 3. Likewise, in the cheating condition, it is in the best interest for a rational-selfish player to report the highest number (‘6’) to maximize own profit. Since every rational-selfish player would follow this strategy, rule following should again not carry any credible information for partner selection. In other words, based on rational-choice theory, we should expect that rule following does not provide any information for the partner selector. Potential partners should, hence, violate the rule to maximize their profit regardless of the condition. Yet, if potential partners have social preferences (i.e. care about the welfare of the partner selector) and care about honesty, rule following can emerge as a credible signal, since pro-social participants may be willing to forgo profit by abiding by rules to signal trustworthiness in the sharing condition while selfish participants should be less willing to forgo the gains from rule breaking. We therefore also assessed individual-level social preferences in a separate task (see below).

**Procedure and Payment.**The experiment was performed in individual cubicles to ensure anonymity. After giving informed consent, participants received instructions on the computer screen. Participants had to correctly answer a set of comprehension questions to ensure that they understood the rules of the experiment. After the main task, participants filled out the incentivized social value orientation (SVO) slider measure (Murphy, Ackermann, & Handgraaf, 2011). In this measure of social preferences, participants decided how to allocate points between themselves and an unknown other person. Points can be allocated self-servingly or pro-socially (sacrificing points to benefit the other person), allowing to estimate participant’s social preferences. Finally, they answered demographic questions. Participants were compensated for their participation with a flat payment of 6.50€ on top of a variable payment depending on their decisions in the experiment. Specifically, two rounds of each condition were randomly selected and the earned points converted to euro. 100 points were worth 1€. Further, participants received the payment from the SVO measure. Participants earned on average 8,67€. The experiment lasted around one hour.

**Statistical Analyses.** Because individual data-points (decisions) were clustered in individuals and groups, we analyzed data on the group-level to compare observations that satisfy the assumption of independence. To investigate within-group behavior, we fitted multilevel regression models as implemented in the lmer package in R. In the regression models, we controlled for the order of experimental blocks and used the Satterthwaite's degrees of freedom method to derive p-values (Kuznetsova, Brockhoff, & Christensen, 2017).

# Results

**Cooperativeness and Dishonesty of Selected Partners.**Selected partners were overwhelmingly pro-social in the sharing condition (Fig. 2a). Only 16.4% of the decisions were selfish, keeping the 500 points entirely, while 47.4% of the decisions were maximally pro-social, by distributing the points equally. To put the decisions of selected partners into perspective: Based on a meta-analysis by (Engel, 2011), around 36% of participants decide to keep everything for themselves while only 29.8% of participants divide resources equally or give more to the recipient in one-shot dictator games.

In the cheating condition, participants reported to have thrown a six in 44.2% of the decisions, securing maximum profit for themselves and the partner selector. Only in 5% of the decisions, participants reported to have thrown a one (Fig. 2b). Average die roll reports significantly deviated from the expectation of honest reporting (one-sample t-test, average group report vs. chance expectation, *t*(47) = 12.64, *p* < 0.001), providing evidence that people did lie in the experiment.

**

*Figure 2.*Decisions of selected partners in the sharing condition (a) and cheating condition (b).

**Partner Selection.**Across both conditions, partner selectors successfully selected partners that would make decisions that benefitted themselves. In 67% and 77% of the interactions in the sharing and cheating condition, respectively, partner selectors chose a partner that made a decision that at least covered the cost of interacting with them. Partner selectors were also selective in choosing partners. In 42% of the rounds they only chose one partner to interact with, while they chose all three partners only in 25% of the rounds.

Across conditions, partner selectors, however, selected partners differently. In the sharing condition, partner selectors favored potential partners that followed more rather than less rules (Fig. 3). For every ball that a potential partner assigned according to the rule, her odds to be selected as an interaction partner increased by 7.3% (logistic multilevel regression, log odds of rule following in the sharing condition, *b* = 0.070, 95% CI: [0.023, 0.119], *p* = 0.004). In contrast, for every ball that a potential partner assigned according to the rule in the cheating environment, her chance to be selected as a partner *decreased* by 0.7% (logistic multilevel regression, log odds of rule following × cheating condition, *b* = -0.078, 95% CI: [-0.012, -0.144], *p* = 0.021). In other words, we found evidence that partner selectors preferred rule abiding partners when they relied on a partner that would fairly distribute a given amount of points (sharing condition), but rule breaking partners when they needed a partner to cheat for them (cheating condition).



*Figure 3.* Partner selection.Average rule following of selected partners (blue line) and not selected partners (black line) in the sharing and cheating condition. Error bars show the standard error of the mean.

**Behavioral Adaption of Potential Partners.**Potential partners, in turn, strategically adapted their rule following propensity according to the environmental incentives. In 30 out of the 48 groups, potential partners followed less rules in the cheating compared to the sharing condition (Fig. 4a, comparison of change in rule following, Wilcoxon Signed Rank Test, *Z* = 805, *p* = 0.025). At the individual level, potential partners decreased their rule following propensity in the cheating compared to the sharing condition (Fig. 4b, multilevel regression, cheating vs. sharing condition, *b* = -1.258, 95% CI: [-1.895, -0.620], *p* < 0.001). Overall, participants followed the rule 57% of the times on average in the sharing condition, which reduced to 49% in the cheating condition (Fig. 4c).



*Figure 4.* Change in rule following.Each bar shows the difference in rule following in the cheating compared to the sharing condition for one group (Inset: difference for each individual participant) (**a**). Average rule following across cheating and sharing condition (**b**). Error bars show the standard error of the mean.

Partner selection and changes in rule following of potential partners reinforced each other. The more the partner selector changed her selection criteria (i.e. the difference between selecting vs. not selecting a partner across sharing and cheating environment), the more potential partners changed their rule following strategy (Pearson correlation, *r* = 0.29, 95% CI: [0.000, 0.531], *p* = 0.050).

**Dynamics over Rounds.**Rule following did not significantly change over rounds and remained rather stable in the sharing condition (multilevel regression, round effect in the social condition, *b* = -0.035, 95% CI: [-0.097, 0.0277], *p* = 0.278). In comparison, rule following decreased over rounds in the cheating condition (multilevel regression, round × cheating condition, *b* = -0.0657, 95% CI: [-0.1263, -0.005], *p* = 0.034) – rule following deteriorated over time in the cheating environment.

**Earnings.**Discriminating and preferentially selecting rule followers vs. rule violators across conditions paid off for partner selectors. Selecting partners that followed more rules significantly increased earnings of partner selectors in the sharing condition (multilevel regression, rule following of selected partners, *b* = 1.916, 95% CI: [0.011, 3.777], *p* = 0.046), while selecting partners that followed less rules significantly increased earnings of partners selectors in the cheating condition (multilevel regression, rule following of selected partners × cheating condition, *b* = -3.257, 95% CI: [-5.689, -0.833], *p* = 0.009).

**Social Value Orientation.**While potential partners adapted their rule abidance according to the environment and the partner preferences of partner selectors, people may also differ in their sensitivity to incentives and situational demands. Indeed, participants that were classified as pro-social in the SVO task followed more rules than participants classified as selfish on average (average rule following of pro-selfs: 6.6 vs. average rule following of pro-socials: 8.5, Mann-Whitney U-test, *U* = 1659.5, *p* = 0.019). Also independent of the condition, participants who scored higher on the SVO task exhibited more rule following (multilevel regression, SVO main effect, *b* = 0.056, 95% CI: [0.0112, 0.1003], *p* = 0.015). Consequently, participants with higher social preferences had a higher chance to be selected as partners in the sharing condition (multilevel regression, SVO / sharing condition, *b* = 0.003, 95% CI: [0.0004, 0.0061], *p* = 0.027), but a lower chance to be selected as partners in the cheating condition (multilevel regression, SVO × cheating condition, *b* = -0.006, 95% CI: [-0.0100, -0.0023], *p* = 0.002). Compared to selfish types, pro-social types also reported lower numbers in the die rolling task when selected as partners (average report of pro-socials: 4.4 vs. average reports of pro-selfs: 5.0, Mann-Whitney U-test, *U* = 2986.5, *p* < 0.001). Hence, people with higher social preferences were more resilient to the change in partner preferences, followed more rules, and cheated less in the cheating condition at the cost of being selected less frequently as partners.

# Discussion

Understanding why and when people follow costly rules is important to understand the emergence and prevalence of norms in groups and how people sustain cooperative relationships. On the flipside, revealing under what circumstances people violate rules can give us important insights under what condition ethical misconduct can emerge and in which environments people tolerate or even enforce rule breaking.

Similar to cooperation dilemmas in which defection is the economically rational choice, costly rules raise the question how they are established and why they are followed in the first place. Previous research suggested and has shown that costly rule following can be maintained through punishment (Andrighetto et al., 2013; Fehr & Fischbacher, 2004), because people internalize rules (Hoffman, 1977; Peysakhovich & Rand, 2015), want to conform to what others do (Bicchieri, 2006; Cialdini & Goldstein, 2004; Henrich & Boyd, 2001), or because they attach virtue and meaning to societal rules (Cialdini, Kallgren, & Reno, 1991; Tankard & Paluck, 2016). Here we showed how rule following and rule breaking can be likewise reinforced and deteriorate by partner choice.

When searching for a pro-social partner, partner selectors preferred rule abiding over rule violating individuals which also increased the relative propensity to follow rules of potential partners. Previous research has shown that partner choice is a powerful mechanism to enforce costly cooperation (Barclay, 2016; Debove et al., 2015; Fu et al., 2008; McNamara et al., 2008), and may even form the basis of more abstract moral judgments about right and wrong (Baumard, Andre, & Sperber, 2013; Everett et al. 2016, 2018. We extend this line of research by showing that rule following is used as a cue for cooperativeness and that partner selection can enforce rule abidance, explaining how costly rules can be established and maintained even in the absence of punishment. Our results also suggest that rule following can have a strategic advantage over and beyond avoiding punishment of others by using it as a signal for cooperativeness.

On the flipside, in an environment that required unethical behavior, rule following decreased. In comparison to the sharing condition, partner selectors preferred rule violating partners when they were seeking someone that would cheat for them. Consequently, potential partners strategically adapted their behavioral cue and became more rule breaking. Depending on the ‘social game they were playing’, rule abidance was sustained or broke down. This shows that partner choice does not only promote pro-social behavior but can also foster rule breaking, selfishness, and attaining benefits through illegitimate means (see also Melnikoff & Bailey, 2018).

Importantly, based on rational-choice theory, rule following should not carry any credible information for partner choice and, hence, costly rule following should not have emerged regardless of the environment. Instead, our data shows that (a) costly rule following is indeed used to signal the willingness to cooperate or cheat, (b) interaction partners are chosen based on this signal, and (c) rule following and cooperation enforce each other through partner choice. We also provided evidence that this dynamic is possible due to individual differences in social preferences. Not all people were willing to opportunistically change their behavior to be increase their chances to be chosen as partners. Participants scoring high on social preferences were more resilient in their rule abidance at the cost of being selected less frequently in the cheating environment. If chosen as partners, they did not cheat as much as their fellow selfish group members (see also Ścigała et al., 2018). In other words, rule following becomes a credible signal for partner choice due to existing interindividual difference in social preferences. It also highlights that personality characteristics influence the degree to which people react to changes in the incentives of the social environment.

Groups and societies differ in their extent to which rules are followed and how strongly norms guide decision making (Gaechter & Schulz, 2016; Gelfand, 2018; Gelfand et al., 2011; Herrmann, Thöni, & Gaechter, 2008). In an environment that requires pro-social actions, people were selected based on their rule abidance, while in an environment that required cheating, people were preferentially selected based on their willingness to break rules. Norm abidance on the group or societal level may hence reflect, at least to some degree, the social interactions that group members engage in. If group members frequently engage in socially deviant behavior, following costly rules may signal non-compliance and is disfavored, whereas if group members frequently engage in non-deviant pro-social exchange, following costly rules may signal cooperativeness and the willingness to restrict selfish behavior. Changing the incentives of local social interactions may, hence, increase global and general norm abidance of groups.

References

Andrighetto, G., Brandts, J., Conte, R., Sabater-Mir, J., Solaz, H., & Villatoro, D. (2013). Punish and voice: Punishment enhances cooperation when combined with norm-signalling. *PLoS ONE*, *8*(6), 1–8.

Barclay, P. (2016). Biological markets and the effects of partner choice on cooperation and friendship. *Current Opinion in Psychology*, *7*, 33–38.

Bicchieri, C. (2006). The grammar of society: The nature and dynamics of social norms. *The Grammar of Society The Nature and Dynamics of Social Norms*. New York: Cambridge University Press.

Chudek, M., & Henrich, J. (2011). Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences*, *15*(5), 218–226.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*, 591–621.

Cialdini, R., & Trost, M. (1998). Social influence: social norms, conformity, and compliance. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of Social Psychology* (pp. 151–192).

Cialdini, R., Kallgren, C., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In *Advances in experimental social psychology* (pp. 201–234).

Cohn, A., & Maréchal, M. A. (2017). Laboratory measure of cheating predicts school misconduct. *The Economic Journal*, *128*(615), 2743–2754.

Dai, Z., Galeotti, F., & Villeval, M. C. (2017). Cheating in the lab predicts fraud in the field: An experiment in public transportation. *Management Science*, *64*(3), 1081–1100.

Debove, S., Andre, J. B., & Baumard, N. (2015). Partner choice creates fairness in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *282*, 1–7.

Engel, C. (2011). Dictator games: a meta study. *Experimental Economics*, *14*(4), 583–610.

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*(4), 185–190.

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

Fu, F., Hauert, C., Nowak, M. A., & Wang, L. (2008). Reputation-based partner choice promotes cooperation in social networks. *Physical Review E*, *78*(2), 1–8.

Gaechter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, *531*(7595), 496–499.

Gelfand, M. J. (2018). Rule makers, rule breakers: How tight and loose cultures wire our world. New York: Simon and Schuster.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., et al. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, *332*, 1100–1104.

Gross, J., & De Dreu, C. K. W. (2017). Oxytocin conditions trait-based rule adherence. *Social Cognitive and Affective Neuroscience*, *12*(3), 427–435.

Gross, J., & De Dreu, C. K. W. (2020). Rule following mitigates collaborative cheating and facilitates the spreading of honesty within groups. *Personality and Social Psychology Bulletin*, *13*(4), 1–15.

Gross, J., Emmerling, F., Vostroknutov, A., & Sack, A. T. (2018a). Manipulation of pro-sociality and rule-following with non-invasive brain stimulation. *Scientific Reports*, *8*(1), 1–10.

Gross, J., Leib, M., Offerman, T., & Shalvi, S. (2018b). Ethical free-riding: When honest people find dishonest partners. *Psychological Science*, *29*(12), 1956–1968.

Hanna, R., & Wang, S.-Y. (2017). Dishonesty and selection into Public Service: Evidence from India. *American Economic Journal: Economic Policy*, *9*(3), 262–90.

Henrich, J., & Boyd, R. (2001). Why people punish defectors. *Journal of Theoretical Biology*, *208*(1), 79–89.

Herrmann, B., Thöni, C., & Gaechter, S. (2008). Antisocial punishment across societies. *Science*, *319*, 1362–1367.

Hoffman, M. L. (1977). Moral internalization: Current theory and research. In *Advances in Experimental Social Psychology* (Vol. 10, pp. 85–133). Elsevier.

Jordan, J. J., Hoffman, M., Nowak, M. A., & Rand, D. G. (2016). Uncalculating cooperation is used to signal trustworthiness. *Proceedings of the National Academy of Sciences*, *113*(31), 8658–8663.

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

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

Kimbrough, E. O., Miller, J., & Vostroknutov, A. (2014). Norms, Frames and Prosocial Behavior in Games. *Working Paper*, 1–30.

Kröll, M., & Rustagi, D. (2016). Shades of dishonesty and cheating in informal milk markets in India. *Working Paper,* 1–35.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). Lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

McNamara, J. M., Barta, Z., Fromhage, L., & Houston, A. I. (2008). The coevolution of choosiness and cooperation. *Nature*, *451*, 189–192.

Melnikoff, D. E., & Bailey, A. H. (2018). Preferences for moral vs. immoral traits in others are conditional. *Proceedings of the National Academy of Sciences*, *115*(4), E592–E600.

Molho, C., Tybur, J. M., van Lange, P. A. M., & Balliet, D. (2020). Direct and indirect punishment of norm violations in daily life. *Nature Communications*, *11*(1), 1–9.

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

Peysakhovich, A., & Rand, D. G. (2015). Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science*, *62*(3), 631–647.

Pfister, R., Wirth, R., Schwarz, K. A., Steinhauser, M., & Kunde, W. (2016). Burdens of non-conformity: Motor execution reveals cognitive conflict during deliberate rule violations. *Cognition*, *147*, 93–99.

Pfister, R., Wirth, R., Weller, L., Foerster, A., & Schwarz, K. A. (2018). Taking shortcuts: Cognitive conflict during motivated rule-breaking. *Journal of Economic Psychology*, 1–10.

Potters, J., & Stoop, J. (2016). Do cheaters in the lab also cheat in the field? *European Economic Review*, *87*, 26–33.

Sommerfeld, R. D., Krambeck, H.-J., Semmann, D., & Milinski, M. (2007). Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the National Academy of Sciences*, *104*(44), 17435–17440.

Soraperra, I., Weisel, O., Zultan, R., Kochavi, S., Leib, M., Shalev, H., & Shalvi, S. (2017). The bad consequences of teamwork. *Economics Letters*, *160*, 12–15.

Ścigała, K. A., Schild, C., Heck, D. W., & Zettler, I. (2018). Who deals with the devil? Interdependence, personality, and corrupted collaboration. *Social Psychological and Personality Science*, *2*, 1–9.

Tankard, M. E., & Paluck, E. L. (2016). Norm perception as a vehicle for social change. *Social Issues and Policy Review*, *10*(1), 181–211.

Tognetti, A., Berticat, C., Raymond, M., & Faurie, C. (2013). Is cooperativeness readable in static facial features? An inter-cultural approach. *Evolution and Human Behavior*, *34*(6), 427–432.

Wang, J., Suri, S., & Watts, D. J. (2012). Cooperation and assortativity with dynamic partner updating. *Proceedings of the National Academy of Sciences*, *109*(36), 14363–14368.

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences*, *112*(34), 10651–10656.

Wilson, R. K., & Eckel, C. C. (2016). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly*, *59*(2), 189–202.

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Making*, *115*(2), 157–168.

**Supplementary Information**

**Experimental Instructions and Implementation.**

**Graphical user interface, text, application

Description automatically generated**

**Fig. S1. Instructions for role assignments.**

**Graphical user interface, text, application, email

Description automatically generated**

**Fig. S2. General structure.**

**Graphical user interface, text, application

Description automatically generated**

**Fig. S3. Payment explanation.**

**Graphical user interface, text, application, email

Description automatically generated**

**Fig. S4. Instructions for stage one.**

**Graphical user interface, text

Description automatically generated**

**Fig. S5. Instructions for stage two.**

**Graphical user interface, text, application, email

Description automatically generated**

**Fig. S6. Instructions for stage three.**

**Graphical user interface, text, application, email

Description automatically generated**

**Fig. S7. Comprehension checks.**

**Graphical user interface, text, application

Description automatically generated**

**Fig. S8. Assignment to role.**

**Graphical user interface, text, application

Description automatically generated**

**Fig. S9. Reminding explanation of task at the start of experimental block stage one.**

**Graphical user interface, text, application

Description automatically generated**

**Fig. S10. Reminding explanation of task at the start of experimental block stage two.**

**Graphical user interface, text

Description automatically generated**

**Fig. S11. Reminding explanation of task at the start of experimental block stage three.**

**Graphical user interface, text, application, email

Description automatically generated**

**Fig. S12. Rule following task.**

**Graphical user interface, text, application, email

Description automatically generated**

**Fig. S13. Partner choice task.**

**Graphical user interface, text, application, email

Description automatically generated**

**Fig. S14. Die roll task.**

**Graphical user interface, application

Description automatically generated**

**Fig. S15. Dictator game.**

**Graphical user interface, text, application

Description automatically generated**

**Fig. S16. Feedback for deciders after stage two.**

**Graphical user interface, text, application

Description automatically generated**

**Fig. S17. Feedback for deciders after every round.**

**Graphical user interface, text, application, email

Description automatically generated**

**Fig. S18. Feedback for selectors after every round.**

– Screenshots of the experimental implementation / computer interface

– full regression tables