Conforming with peers in honesty and cooperation ☆

Ozan Isler [a,b,*], Simon Gächter [c,d,e,**]

[a] *School of Economics and Finance, Queensland University of Technology, Brisbane 4000, Australia*
[b] *Center for Behavioral Economics, Society and Technology, Brisbane 4000, Australia*
[c] *School of Economics, University of Nottingham, Nottingham NG7 2RD, UK*
[d] *CESifo Munich, 81679, Germany*
[e] *IZA Bonn, 53113, Germany*

ABSTRACT

Peer observation can influence social norm perceptions as well as behavior in various moral domains, but is the tendency to be influenced by and conform with peers domain-general? In an online experiment ($N = 815$), we studied peer effects in honesty and cooperation and tested the individual-level links between these two moral domains. Participants completed both honesty and cooperation tasks after observing their peers. Consistent with the literature, separate analysis of the two domains indicated both negative and positive peer influences in honesty and in cooperation, with negative influences tending to be stronger. Behavioral tests linking the two domains at the individual-level revealed that cooperative participants were also more honest—a link that was associated with low Machiavellianism scores. While standard personality trait measures showed no links between the two domains in the tendency to conform, individual-level tests suggested that conformism is a domain-general behavioral trait observed across honesty and cooperation. Based on these findings, we discuss the potential of and difficulties in using peer observation to influence social norm compliance as an avenue for further research and as a tool to promote social welfare.

1. Introduction

Peers shape perceptions of social norms as well as individual moral behavior. These influences can be direct, with increased pressure to conform due to the social and emotional proximity of peers as well as the material interests and strategic motivations involved in interacting with them. In addition to these direct effects, the information received via the observation of peers can indirectly influence moral behavior by allowing inferences about the prevalence of social norm compliance and social preferences in one's environment. While the existence of these latter, purely informational peer in-

fluences has previously been shown across various moral domains, including honesty and cooperation (e.g., Rauhut, 2013; Thöni and Gächter, 2015), it remains unknown whether the tendency to be influenced by peers—namely, *conformism*—is a domain-general behavioral trait. To provide causal insights on this question, we experimentally tested the effects of receiving information on the moral behavior and social preferences of peers—what we refer to here as *peer observation*—and studied the individual-level links between the domains of honesty and cooperation.

Participants in our experiment completed two independent tasks involving peer observation, one about honesty and another about cooperation. In the *honesty task*, participants played forty rounds of the Mind Game (Jiang, 2013). In each round of this task a participant is asked to think of a number and claim a match if the number thought of coincides with the number randomly drawn by the computer. While the Mind Game provides an opportunity to lie discreetly for a monetary reward, average match-claims above chance indicate the extent of dishonesty at the treatment-level. After the first twenty rounds, participants were paired for estimating and exchanging information about each other's match claims, thereby providing measures of expected and actual peer dishonesty. The next twenty rounds allowed us to observe the effects of peer information on honesty in the absence of any strategic and material incentives to influence one's peers.

Several studies have tested the effects of peer observation on honesty and found both positive and negative influences (Diekmann et al., 2015; Gino et al., 2009; Kroher and Wolbring, 2015), including an early study by Rauhut (2013) that comes closest to our setup. Rauhut (2013) used a repeated die-rolling task (Fischbacher and Föllmi-Heusi, 2013) to measure the effect of receiving information about group-level peer behavior on individual honesty. While Rauhut (2013) did not find a clear direct effect of information, a more nuanced picture emerged when comparing participants with different initial expectations such that those who underestimated group-level dishonesty became more dishonest and those who overestimated group-level dishonesty became more honest. However, this design did not allow clear causal identification of peer effects because all participants received the same group-level information (and because individual expectations were endogenous). While we build on the insightful analysis of this study by considering the role of expectations, participants in our experiment are randomly matched to identify the causal effect of peer observation.

Our *cooperation task* is based on a modified version of the Prisoner's Dilemma game. The standard version of the game generates a social dilemma by asking participants to choose to "share" a monetary endowment for group-benefit at net personal cost or to "keep" it for personal benefit. The modified version of this game is designed to measure cooperation preferences using the strategy method (Fischbacher et al., 2001) by eliciting reactions to each possible action of the other player in an incentive compatible manner. Like the setup in the honesty task, participants were then anonymously paired with an independent peer. After peer pairing, participants estimated and received information about their peer's cooperation preferences. Finally, to test for the effects of peer observation, participants were allowed to revise their initial reports of cooperation preferences.

Most experiments studying peer observation in social preferences and cooperation found strong effects on cooperation behavior (e.g., Falk et al., 2013; Gächter et al., 2013; Isler et al., 2021; Thöni and Gächter, 2015). It is now well-established that the main driver of these effects is a widespread and stable preference for conditional cooperation—the willingness to cooperate in social dilemmas to the extent that others will cooperate as well (e.g., Chaudhuri, 2011; Croson, 2007; Fehr and Schurtenberger, 2018; Fischbacher et al., 2001; Frey and Meier, 2004; Isler et al., 2021; Thöni and Volk, 2018; Volk et al., 2012). What is less known in the domain of cooperation—and what we study here—is whether and how peer observation affects the *preference* for cooperation itself. Thöni and Gächter (2015) used a gift-exchange game with independent payoffs among pairs of participants and showed that, even in the absence of strategic and material incentives, information exchange about effort choices among peers negatively (but not positively) affected each other's gift exchange. The authors combined these behavioral observations with measures of expectations about peer effort choice to infer potential changes to underlying cooperation preferences. We build on this design by eliciting (with the strategy method) a preference profile, that is, cooperation behavior for each possible expectation at the individual level. This allows us to directly observe the effects of peer observation on cooperation preferences.

Our experiment is designed to test for the pure informational effect of peer observation on honesty behavior and cooperation preferences. Participants were randomly matched to observe their peers while excluding any strategic or material dependence on them. Peer anonymity further excluded any influence that can result from the social or emotional proximity of peers. We also elicited three control conditions to check for baseline changes in honesty and cooperation preferences as well for the influence of peer matching and for the influence of making estimations about peers.

Our experimental design also allows behavioral tests of individual-level links between the two moral domains because all participants complete both the honesty and the cooperation task. While conformism has been experimentally studied in both honesty (e.g., Fosgaard et al., 2013) and cooperation (e.g., Bardsley and Sausgruber, 2005; Carpenter, 2004; Falk et al., 2013; Fatas et al., 2018), it remains unknown whether the individual tendency to be influenced by peers is domain-general. We test whether participants who prefer cooperation also act more honestly, and test whether conformism in cooperation is associated with conformism in honesty. This novel behavioral trait measure of conformism controls for behavioral spillovers between the two domains because both honesty and cooperation preferences are measured twice—once before and once after peer observation—so that any informational carryover effects between the two domains are picked up in the initial measurement in each domain. Finally, we test whether cooperation, honesty, and conformism in these domains can be explained by various standard psychological measures of personality traits known to influence moral behavior (Gunnthorsdottir et al., 2002; Hilbig and Zettler, 2015; Thielmann et al., 2020; Volk et al., 2012; Zettler et al., 2013).

The separate analysis of the two domains showed, consistent with the literature, both negative and positive peer effects in both honesty and cooperation, with negative influences tending to be stronger than the positive ones (e.g., Bicchieri et al., 2022; Rauhut, 2013; Thöni and Gächter, 2015). Behaviorally linking the two domains at the individuallevel showed that participants with stronger cooperative preferences tended to behave more honestly (for a similar finding see Biziou-van-Pol et al., 2015). While standard psychological measures of personality traits identified Machiavellianism as undermining both honesty and cooperation, which again is in line with previous findings (Gunnthorsdottir et al., 2002; Hilbig and Zettler, 2015; Zettler et al., 2013), these self-reported scale measures failed to show any significant links between the two moral domains in terms of conformism. In contrast, our individual-level behavioral analysis revealed a novel link between honesty and cooperation, suggesting that conformism might be a domain-general trait.

Because social norm perceptions are often formed by the observation of individual peers in everyday interactions and increasingly on social media, peer observation can in principle be harnessed to influence social norm compliance. In light of our findings, we discuss in our concluding section whether and how individual peer observation can be used to promote social welfare—what we refer to as *peer-nudging*—and suggest future avenues of research for developing such policy tools.

2. Methods

2.1. Data collection

We collected data online from 815 participants (age: $M = 35.9$, $SD = 11.6$; female: 53.1%) who completed the experiment using LIONESS Lab (Giamattei et al., 2020). Participants were randomly allocated to the treatment or control conditions and data were collected simultaneously from the Amazon Mechanical Turk (www.mturk.com, 61.0%) and Prolific (www.prolific.co, 39.0%) survey platforms. Participants were either US (77.4%) or UK (22.6%) residents. The distribution of participants in the four experimental conditions was similar across survey platforms (χ^2 test, $p = .687$) and countries of residence (χ^2 test, $p = .453$). The experiment lasted about 20 min. Participants earned on average \$3.50 USD, including a participation payment of \$0.50, amounting to an hourly wage of about \$10.50. The experimental materials are available in the Supplementary Information, and the dataset and the analysis code are available at the Open Science Framework project site (https://osf.io/c84n5/).
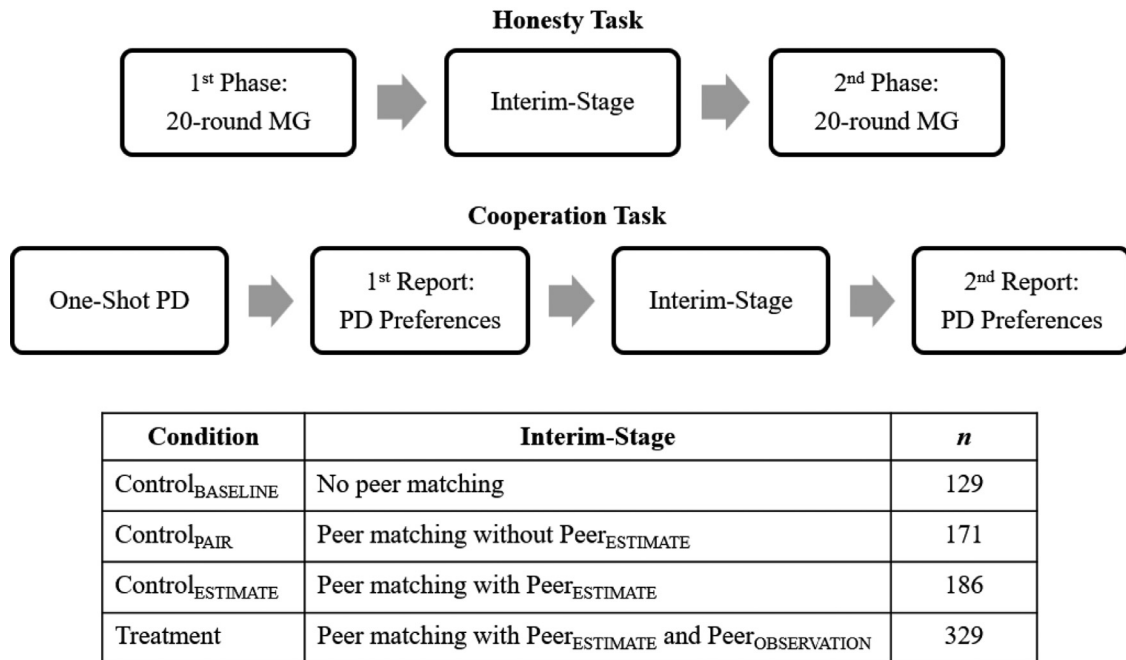
2.2. Design

In all conditions, participants completed two main parts including both the Honesty Task and the Cooperation Task, followed by a survey. The two main parts were completed under either (i) a treatment condition where pairs of participants (i.e., peers) observed each other or (ii) under one of three control conditions where there was no information exchange between participants (see Fig. 1 for an overview). One of the main parts was randomly selected to determine any additional earnings.

The Honesty Task. There were two phases. In each phase, participants played 20 rounds of a Mind Game (Jiang, 2013). In each round, participants were first asked to think of a number between 0 and 9. Afterwards, a number between 0 and 9 was randomly chosen by the computer and displayed on the next screen. Participants were asked to report a "match" if the two numbers were the same and claim a 50-cent reward. The total claims across the 20 rounds determined the potential earnings of the participant from each phase. In between the two phases was the Interim-Stage, which depended on the experimental condition, and included either peer observation or a control (see below for details). If the honesty task was chosen to determine a participant's additional earnings, then the total claims in either the 1st or the 2nd Phase were randomly selected for pay-out.

The Cooperation Task. We study cooperation using a one-shot two-person binary Prisoner's Dilemma game with \$1 endowments. If both players choose to "keep" their endowments, then each player earns \$1 from the game. If both players choose to "share" their endowments, then each player earns \$2. If one player "shares" while the other "keeps", the player who "keeps" earns \$3 while the one who "shares" does not earn any monetary rewards. After receiving instructions, participants had to correctly answer two control questions on the same screen about the Prisoner's Dilemma before they could proceed with the cooperation task.

Participants first played the standard Prisoner's Dilemma game by providing a single unconditional choice. Then they completed the 1st Report: using the strategy method to measure cooperation preferences—an approach based on Fischbacher et al. (2001)—participants had to make two conditional choices of either "keep" or "share" for each of the other player's possible choices of "keep" or "share". Next was the Interim-Stage, which included either peer preference observation in the treatment or a control (see below for details). Any peer matching in the Interim-Stage was separate from the pair matching for 1st Report, and participants knew that. Participants then completed the 2nd Report, which allowed them to revise their initial preference profiles from the 1st Report, if they wished to do so. If the cooperation task were chosen to determine participant's additional earnings, then either the standard Prisoner's Dilemma game, the 1st Report or the 2nd Report was selected for payout. If one of the reports were selected, then the unconditional choice of one's pair in the standard Prisoner's Dilemma game was combined with the conditional choices of the participant to determine the payout.

Overview of Main Tasks and Experimental Conditions

Honesty Task

| 1st Phase: 20-round MG | ⇒ | Interim-Stage | ⇒ | 2nd Phase: 20-round MG |

Cooperation Task

| One-Shot PD | ⇒ | 1st Report: PD Preferences | ⇒ | Interim-Stage | ⇒ | 2nd Report: PD Preferences |

Condition	Interim-Stage	n
Control$_{BASELINE}$	No peer matching	129
Control$_{PAIR}$	Peer matching without Peer$_{ESTIMATE}$	171
Control$_{ESTIMATE}$	Peer matching with Peer$_{ESTIMATE}$	186
Treatment	Peer matching with Peer$_{ESTIMATE}$ and Peer$_{OBSERVATION}$	329

Fig. 1. Overview of main tasks and experimental conditions. Participants were assigned to either the treatment condition or one of three controls. All participants completed both the honesty task and the cooperation task. The order of tasks was counterbalanced in the treatment condition. In the control conditions, participants first completed the honesty task and then the cooperation task. The honesty task included two phases of 20 rounds of a Mind Game (MG), and the cooperation task consisted of two reports that elicited cooperation preferences for a one-shot Prisoner's Dilemma (PD). In the treatment condition, pairs of participants were randomly matched as peers twice—once in the Interim-Stage of the honesty task and once in the Interim-Stage of the cooperation task—to estimate and exchange information about each other's behavior in the 1st Phase of the honesty task and preferences in the 1st Report of the cooperation task. In the control conditions, there was no exchange of information between participants. In C$_{BASELINE}$, there was no peer matching, no peer estimation, and no peer observation. In C$_{PAIR}$, participants were matched with peers, but they neither made peer estimates nor observed peers. In C$_{ESTIMATE}$, participants were matched with peers and made peer estimates without peer observation.

Treatment. Each participant in the treatment condition ($n = 329$) was matched with two different peers, once for the honesty task and once for the cooperation task.[1] In the Interim-Stage of each task, participants estimated what their peer did (their total claims in the 1st Phase of the honesty task) or prefers (their complete preference profile in the Prisoner's Dilemma) before observing their peer's actual behavior or preference. Hence, the information received by peer observation is naturally and randomly varied, constituting the experimental manipulation. To minimize any task order effects of peer observation, participants either completed the honesty task first and the cooperation task second or vice versa. Any spillovers between the two moral domains are further controlled by design, as both honesty behavior and cooperation preferences are elicited twice, in two consecutive phases for the former and in two consecutive reports for the latter. Any carryover effect of one domain on the other would be picked up by the initial measurement in the other domain. Hence, the carryover would not influence the outcome measures that are defined as the difference in two consecutive measurements in each domain.

Controls. Three additional conditions were used to provide baseline measures of honesty and cooperation, to control for any direct effect of the act of peer pairing, and to control for any direct effect of the act of making estimations about one's peer. Participants in the control conditions completed first the honesty task and then the cooperation task without observing peers. Specifically, participants in C$_{ESTIMATE}$ ($n = 186$) were paired online with peers and made estimates about them without observing them afterward. In C$_{PAIR}$ ($n = 171$), participants were also paired with peers, but differently from C$_{ESTIMATE}$, they neither made estimates about them nor observed them afterward. In C$_{BASELINE}$ ($n = 129$), participants were not paired with peers, no estimates about peers were elicited, and there was no peer observation. Hence, comparing C$_{PAIR}$ to C$_{BASELINE}$ isolates the effect of being paired with a peer; comparing C$_{ESTIMATE}$ to C$_{PAIR}$ isolates the additional effect of

---

[1] In addition to the two peers, participants were matched with a pair in the cooperation task. Hence, participants in the treatment conditions were matched with three different participants: (1) a peer in the honesty task and (2) a peer in the cooperation task using online matching in the Interim-Stages, and (3) a pair in the Prisoner's Dilemma using passive matching after the study was completed. To isolate the pure informational effect of peer observation on cooperation preferences (i.e., to exclude strategic and pecuniary motivations), participants in the treatment conditions were informed that their peer in the cooperation task was not the same person as their pair in the Prisoner's Dilemma.

making estimates about peers; and comparing the treatment condition to $C_{ESTIMATE}$ isolates the additional effect of actually observing peers.

Survey. Participants finally completed a survey, eliciting various personality trait measures and demographic information. The following personality measures were included in the survey: the MACH-IV scale of Machiavellianism (Christie and Geis, 1970) was elicited because it has been associated with lack of reciprocity (Gunnthorsdottir et al., 2002) and dishonesty (Hilbig and Zettler, 2015; Sutton and Keogh, 2001); the short version of the Iowa-Netherlands Social Comparison Orientation Scale (INCOM-S) was elicited to measure the tendency to rely on social information for making personal comparisons (Gibbons and Buunk, 1999), which should in principle be associated with conformism; the Rational-Experiential Inventory (REI-10), composed of the "Need for Cognition" and "Faith in Intuition" subscales, was elicited to measure tendencies to rely on reflection and intuition (Epstein et al., 1996), since these thinking styles are associated with cooperation (Isler et al., 2021) and honesty (Capraro et al., 2019; Volk et al., 2012); and the Ten-Item Personality Inventory (TIPI) of the Big-Five model (Gosling et al., 2003), including the "Extraversion", "Agreeableness", "Conscientiousness", "Emotional Stability", and "Openness to Experience" subscales, was elicited to have comprehensive measures of personality and because "Agreeableness" has been associated with cooperativeness (Volk et al., 2012). Four attention check items were added throughout the survey and average pass rate across the four items was 98.8%.

3. Results I: peer effects in honesty and cooperation

3.1. Honesty

Behavior in the honesty task was measured twice, in two phases of twenty-round Mind Games (see Fig. 1). A population of completely honest players of the Mind Game is expected to claim on average four matches across the forty rounds. Indicating substantial dishonesty, claims across the forty rounds were higher than this complete honesty threshold in all conditions (one-sample t-tests, $ps < 0.001$), with an overall average of 10.8 claims.

There was no peer observation in the control conditions, which provide reference levels of change in honesty between the two phases. In the treatment conditions, participants first estimated and then observed peer honesty (i.e., peer's total number of match claims at the end of 1st Phase). We refer to the observed peer behavior as $Peer_{OBSERVATION}$ and to its estimate as $Peer_{ESTIMATE}$. To study changes in honesty between the two phases, we estimated $\triangle_{CLAIMS}$ for each participant, defined as the difference in total claims made between the 2nd Phase and the 1st Phase.

Controls. In the control conditions, there was no discernable difference in claims between the two phases. In one-sample t-tests, $\triangle_{CLAIMS}$ were indistinguishable from zero in the control without peer pairing ($C_{BASELINE}$: $M = -0.11$, $p = .560$, $d = 0.05$) as well as in the controls with peer pairing (C_{PAIR}: $M = 0.08$, $p = .679$, $d = 0.03$) and with estimation of peer behavior ($C_{ESTIMATE}$: $M = 0.35$, $p = .058$, $|d = 0.14$).[2] Therefore, we pool the control conditions and use the overall average $\triangle_{CLAIMS}$ ($M = 0.13$) as the control benchmark for the rest of the analysis.
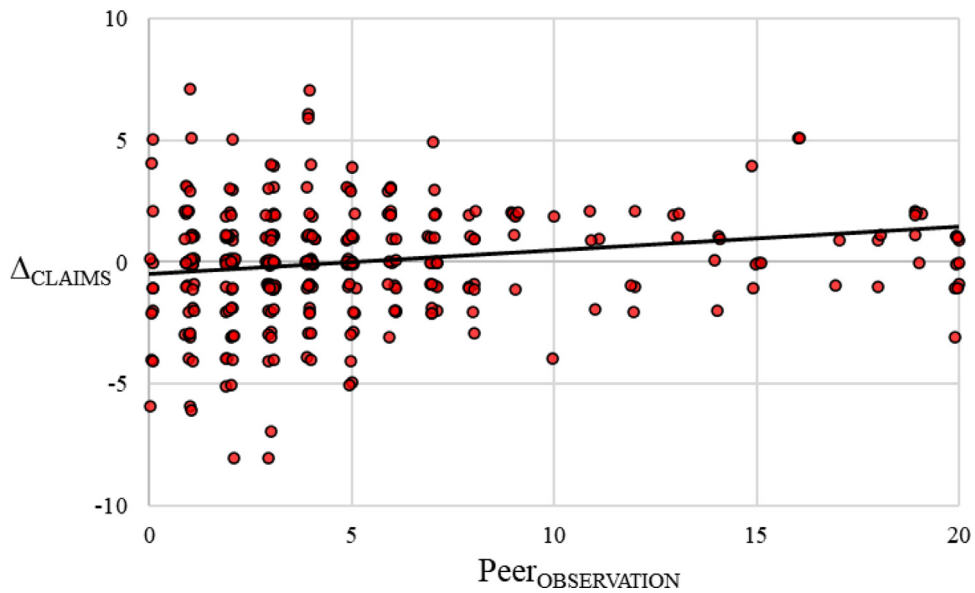
Effect of peer observation. Fig. 2 depicts the relationship between the claims made by peers in the 1st Phase that were observed by the participants in the treatment condition (i.e., $Peer_{OBSERVATION}$) and change in the claims made by the participants between the 2nd Phase and the 1st Phase ($\triangle_{CLAIMS}$). A simple linear regression shows a significant positive effect of $Peer_{OBSERVATION}$ on $\triangle_{CLAIMS}$ ($\beta = 0.10$, $p < .001$), indicating an overall increase of 2 claims as $Peer_{OBSERVATION}$ ranges from 0 to 20 claims.[3] The magnitude of this effect is similar to the average effect size of $\beta = 0.12$ found in a meta-analysis of peer effects (Herbst and Mas, 2015). The average $Peer_{OBSERVATION}$ was 5.72 claims such that receiving above-average $Peer_{OBSERVATION}$ resulted in positive $\triangle_{CLAIMS}$ ($M = 0.66$) significantly above the control benchmark (independent-samples t-tests: $p = .031$, $d = 0.23$). On the other hand, receiving below-average $Peer_{OBSERVATION}$ resulted in negative $\triangle_{CLAIMS}$ ($M = -0.28$) significantly below the control benchmark ($p = .038$, $d = 0.17$). In short, peer observation significantly influenced claims made in the 2nd Phase, increasing dishonesty with implied peer dishonesty, and decreasing dishonesty with implied peer honesty.

Role of expectations. We categorize the observation of peer behavior as *expected* if $Peer_{ESTIMATE} = Peer_{OBSERVATION}$ and *unexpected* otherwise. Unexpected peer honesty is when $Peer_{ESTIMATE} > Peer_{OBSERVATION}$ and unexpected peer dishonesty is when $Peer_{ESTIMATE} < Peer_{OBSERVATION}$. Most participants in the treatment condition experienced either unexpected peer honesty (58.1%) or unexpected peer dishonesty (32.8%), while only 9.1% observed expected levels of peer dishonesty. The influence of these different experiences on $\triangle_{CLAIMS}$ is depicted in Fig. 3.

Compared to the control benchmark, $\triangle_{CLAIMS}$ were higher for unexpected peer dishonesty (independent-samples t-tests: $p = .011$, $d = 0.27$) and lower for unexpected peer honesty ($p = .003$, $d = 0.25$). Although observation of expected peer behavior showed no difference from the control benchmark in $\triangle_{CLAIMS}$ ($p = .196$, $d = 0.24$), these accurate estimates nevertheless resulted in significantly positive $\triangle_{CLAIMS}$ (one-sample t-test: $p = .023$, $d = 0.44$), possibly because participants with accurate estimates expected significantly more peer dishonesty than they themselves showed in the 1st Phase (4.8 vs. 3.4 claims; paired sample t-test: $p = .004$, $d = 0.58$). The negative influence of unexpected peer dishonesty was 1.7 times larger

---

[2] As statistical effect size estimates for t-tests, we report Cohen's d. For independent sample t-tests, Cohen's d is defined as the difference between the sample means divided by the pooled standard deviation. Typically, $d = 0.20$ is considered as small, $d = 0.50$ is considered as medium, and $d = 0.80$ is considered as a large effect size (Cohen, 2013).

[3] Controlling for age, gender, country of residence, and recruitment platform in otherwise the same linear regression provides consistent results: $\beta = 0.10$, $p < .001$.

Fig. 2. Scatter plot showing the relationship between the claims made by peers in the 1st Phase that were observed by the participants ($Peer_{OBSERVATION}$) and change in the claims made by the participants between the 2nd Phase and the 1st Phase (Δ_{CLAIMS}) of the treatment condition ($n = 329$). Each dot is drawn with quarter point two-dimensional jitters to improve visibility of overlapping observations. Trend line from simple OLS regressions detailed in the main text indicates the overall positive effect of $Peer_{OBSERVATION}$ on Δ_{CLAIMS}.
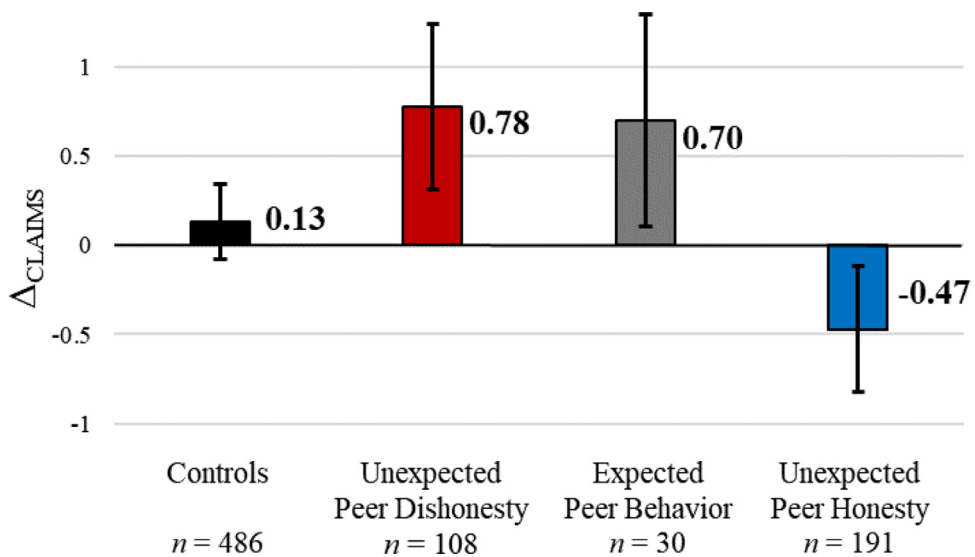
Fig. 3. Mean Δ_{CLAIMS} in the control conditions (i.e., among those who did not observe peers) and among those in the treatment condition who experienced either unexpected peer dishonesty (i.e., $Peer_{ESTIMATE} < Peer_{OBSERVATION}$), expected peer behavior ($Peer_{ESTIMATE} = Peer_{OBSERVATION}$), and unexpected peer honesty ($Peer_{ESTIMATE} > Peer_{OBSERVATION}$). Error bars indicate 95% confidence intervals.

than the positive influence of unexpected peer honesty on average. While this difference was not statistically significant (independent-samples *t*-test comparing Δ_{CLAIMS} under unexpected honesty vs. unexpected dishonesty: $p = .302$, $d = 0.12$), it's magnitude was consistent with the asymmetry in peer effects on honesty found by Rauhut (2013). These results are robust to an alternative categorization of observed peer behavior incorporating the idea that participants may consider small inaccuracies in their expectations to be negligible.[4]

---

[4] Assuming a negligible margin of error in expectations of 1 claim, 22.2% of participants in the treatment condition are categorized as having observed expected peer behavior (i.e., |$Peer_{ESTIMATE}$ - $Peer_{OBSERVATION}$| ≤ 1), 51.1% as having observed unexpected peer honesty ($Peer_{ESTIMATE}$ - $Peer_{OBSERVATION}$ > 1), and 26.8% as having observed unexpected peer dishonesty ($Peer_{ESTIMATE}$ - $Peer_{OBSERVATION}$ < -1.). Using this alternative categorization, Δ_{CLAIMS} were higher for unexpected peer dishonesty ($p = .009$, $d = 0.31$), lower for unexpected peer honesty ($p = .001$, $d = 0.29$), and not different for expected peer behavior ($p = .246$, $d = 0.15$) as compared to the control benchmark.

Table 1

Effect of peer observation on cooperation preference revision.

Peer$_{OBSERVATION}$	$\Delta_{PREFERENCE}$		
	Uncooperative Revision	No Revision	Cooperative Revision
Complete Uncooperativeness	19.1% (17)	76.4% (68)	4.5% (4)
Partial Cooperativeness	8.2% (18)	82.7% (182)	9.1% (20)
Complete Cooperativeness	0.0% (0)	95.0% (19)	5.0% (1)
Controls	10.7% (52)	85.0% (413)	4.3% (21)

Note. Table rows describe either the pool of control conditions (where there was no peer observation) or the types of peer preferences observed by participants in the treatment condition (i.e., Peer$_{OBSERVATION}$), ranging from Complete Uncooperativeness (i.e., zero "shares"), Partial Cooperativeness (one "share"), to Complete Cooperativeness (two "shares"). Table columns indicate the type of preference revision (i.e., $\Delta_{PREFERENCE}$), ranging from Uncooperative Revision (i.e., $\Delta_{PREFERENCE} < 0$), No Revision ($\Delta_{PREFERENCE} = 0$), to Cooperative Revision ($\Delta_{PREFERENCE} > 0$). The values in parentheses denote the number of observations.

3.2. Cooperation

Recall that participants first played a standard one-shot Prisoner's Dilemma and then noted their conditional preferences for cooperation in the game twice, in an initial and a final report (see Fig. 1). Each report elicited a complete preference profile, including a preferred action ("keep" or "share") for each of the two possible actions by their pair in the game ("keep" and "share"). Overall, the percentage of participants who chose to "share" was 50.8% in the standard one-shot Prisoner's Dilemma, 70.6% in the 1st Report conditional on their peers choosing to "share", and 10.4% in the 1st Report conditional on their peers choosing to "keep".[5]

The control conditions, with no peer observation, provide baseline levels of preference change between the two reports. In the treatment condition, participants observed a peer's preferences in the 1st Report (i.e., the peer's initial strategy profile) before finalizing their own preferences in the 2nd Report. We denote this information received by the participants (i.e., the number of "shares" in a peer's 1st Report, ranging from 0 to 2) as Peer$_{OBSERVATION}$ and to its estimate as Peer$_{ESTIMATE}$. We also define $\Delta_{PREFERENCE}$ to note any change in a participant's cooperation preferences. Comparing the change in the number of "shares" from the initial to the final report, if the change is positive then $\Delta_{PREFERENCE}$ indicates a "cooperative revision", if the change is negative then $\Delta_{PREFERENCE}$ indicates an "uncooperative revision", and if there is no change then $\Delta_{PREFERENCE}$ indicates "no revision".

Controls. The distribution of $\Delta_{PREFERENCE}$ was no different between the three control conditions (χ^2 test: $p = .303$), which we therefore pool and use as the control benchmark for the rest of our analysis. Across the control conditions, most participants showed stable cooperation preferences (85.0%), while among those who changed preferences, uncooperative revisions (10.7%) were significantly more frequent than cooperative revisions (4.3%) (binomial test: $p < .001$).

Effect of peer observation. Table 1 describes the relationship between Peer$_{OBSERVATION}$ and $\Delta_{PREFERENCE}$ in the treatment condition as compared to the controls that were without Peer$_{OBSERVATION}$. Compared to the controls, uncooperative preference revisions were significantly more prevalent under complete peer uncooperativeness (19.1% vs. 10.7%; χ^2 tests: $p = .025$), and cooperative preference revisions were significantly more prevalent under partial peer cooperativeness (9.1% vs. 4.3%; $p = .012$). Possibly due to the low number of observations, cooperative preference revisions were not different between complete peer cooperativeness and the control benchmark (5.0% vs. 4.3%; $p = .884$). Overall, observation of peer preferences affected participants' preference revisions positively and negatively.[6]

Role of expectations. We categorize an observed peer preference as *expected* if Peer$_{ESTIMATE}$ = Peer$_{OBSERVATION}$ and *unexpected* otherwise. Specifically, unexpected peer uncooperativeness is defined as Peer$_{ESTIMATE}$ > Peer$_{OBSERVATION}$ and unexpected peer cooperativeness is defined as Peer$_{ESTIMATE}$ < Peer$_{OBSERVATION}$. Peer preferences were as expected for 44.4% of participants, whereas unexpected peer cooperativeness (28.9%) and unexpected peer uncooperativeness (26.7%) were experienced to a similar extent. Table 2 describes the influence of these different experiences on $\Delta_{PREFERENCE}$.

Compared to the control benchmark, preference revisions were not different for expected peer preference (χ^2 tests: $p = .168$), higher in cooperative revisions for unexpected peer cooperativeness ($p = .035$), and higher in uncooperative revisions for unexpected peer uncooperativeness ($p < .001$). Unlike the negative influence of expected peer behavior in the hon-

---

[5] Preference profiles are often used to categorize participants as having one of four preferences: freeriders (i.e., those who "keep" regardless of pair's actions), unconditional cooperators (those who "share" regardless of pair's actions), conditional cooperators (those who "share" if their pair "shares" and "keep" if their pair "keeps"), and others (those who "keep" if their pair "shares" and "share" if their pair "keeps"). Although we do not analyze preference types here for concision, across both the control and treatment conditions and according to 1st Reports, 63.9% of participants were categorized as conditional cooperators, 25.6% as freeriders, 6.6% as unconditional cooperators, and 3.8% as "others".

[6] By definition, some participants in the experimental condition (27.1%) could not revise their preferences to be any less cooperative because their 1st Reports were already completely selfish (i.e., "keep" and "keep"). Likewise, 6.1% of the participants in the treatment condition could not revise their 2nd Reports to be any more cooperative because their 1st Reports were already completely cooperative (i.e., "share" and "share"). These limiting cases do not pose a problem for the analysis of peer effects, because they were symmetrically distributed (χ^2 test, $p = .624$) across the different types of Peer$_{OBSERVATION}$ received (i.e., zero, one or two "shares" in peer's 1st Report).

Table 2

The role of expectations in cooperation preference revision.

	$\Delta_{\text{PREFERENCE}}$		
Peer$_{\text{OBSERVATION}}$	Uncooperative Revision	No Revision	Cooperative Revision
Unexpected Peer Uncooperativeness	22.7% (20)	68.2% (60)	9.1% (8)
Expected Peer Preference	5.5% (8)	89.7% (131)	4.8% (7)
Unexpected Peer Cooperativeness	7.4% (7)	82.2% (78)	10.5% (10)
Controls	10.7% (52)	85.0% (413)	4.3% (21)

Note. Table rows describe either the pool of control conditions (where there was no peer observation) or the types of unexpected peer preferences observed by participants in the treatment condition, ranging from Unexpected Peer Uncooperativeness (i.e., Peer$_{\text{ESTIMATE}}$ > Peer$_{\text{OBSERVATION}}$), Expected Peer Preference (Peer$_{\text{ESTIMATE}}$ = Peer$_{\text{OBSERVATION}}$), to Unexpected Peer Cooperativeness (Peer$_{\text{ESTIMATE}}$ < Peer$_{\text{OBSERVATION}}$). Table columns indicate the type of preference revision (i.e., $\Delta_{\text{PREFERENCE}}$), ranging from Uncooperative Revision (i.e., $\Delta_{\text{PREFERENCE}}$ < 0), No Revision ($\Delta_{\text{PREFERENCE}}$ = 0), to Cooperative Revision ($\Delta_{\text{PREFERENCE}}$ > 0). The values in parentheses denote the number of observations.
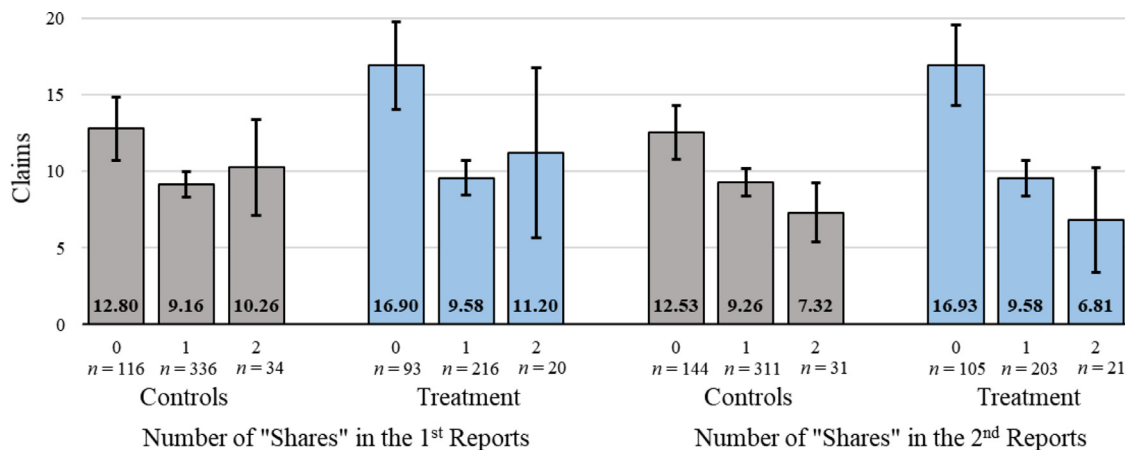
Fig. 4. Mean number of total claims in the honesty task (i.e., across the 40-rounds) in the control and the treatment conditions separated by the number of "share" decisions in the 1st and the 2nd Reports at the individual level. Error bars indicate 95% confidence intervals.

esty task, participants with accurate estimates of peer preferences tended not to revise their preferences, possibly because the distribution of these accurate estimates were not different than the distribution of the initially reported preferences of these participants ($p = .760$). The negative influence of unexpected peer uncooperativeness was 2.2 times larger than the positive influence of unexpected peer cooperativeness, which was statistically significant (22.7% vs. 10.5%; $p = .026$). The asymmetry in peer effects that is revealed when controlling for expectations is consistent with the results of Thöni and Gächter (2015), who found stronger negative peer effects in a gift-exchange game even after controlling for expectations. It is also consistent with Bicchieri et al. (2022), who observed that peers' violations of prosocial norms tended to have stronger negative effects compared to the positive effects of their norm compliance.

4. Results II: linking the two moral domains

Here, we consider the behavioral, motivational, and cognitive links between honesty and cooperation. Individual-level links between the two domains can be identified in our experimental design where participants completed both the honesty and the cooperation task. First, we test whether cooperative participants are also more honest. Next, we consider whether conformism in one domain predicts conformism in the other domain. Finally, we consider to what extent standard personality trait measures explain cooperativeness and honesty as well as conformism in the two moral domains.

Cooperativeness and honesty. Across all conditions, participants with stronger cooperative preferences tended to be more honest as well (see Fig. 4). Specifically, the total number of match claims in the honesty task showed significant negative trends with the number of "share" decisions for both the initial and the final reports in both the control and the treatment conditions (Jonckheere trend tests: $ps \leq 0.011$).

Conformism. We define conformism as the tendency to be influenced by peers and measure it in honesty as the absolute value of Δ_{CLAIMS} (i.e., $|\Delta_{\text{CLAIMS}}|$) in the treatment condition. Hence, conformism in honesty indicates the extent of change in behavior in the honesty task, whether positive or negative, after observing a peer. In cooperation, we categorize those participants who revised their preferences after observing a peer as showing conformism ($|\Delta_{\text{PREFERENCE}}| \neq 0$) and as not showing conformism otherwise ($|\Delta_{\text{PREFERENCE}}| = 0$). We define conformism in cooperation as a binary categorical variable because only one participant in the treatment condition (0.3% of the sample) had $|\Delta_{\text{PREFERENCE}}| > 1$. Due to the within-
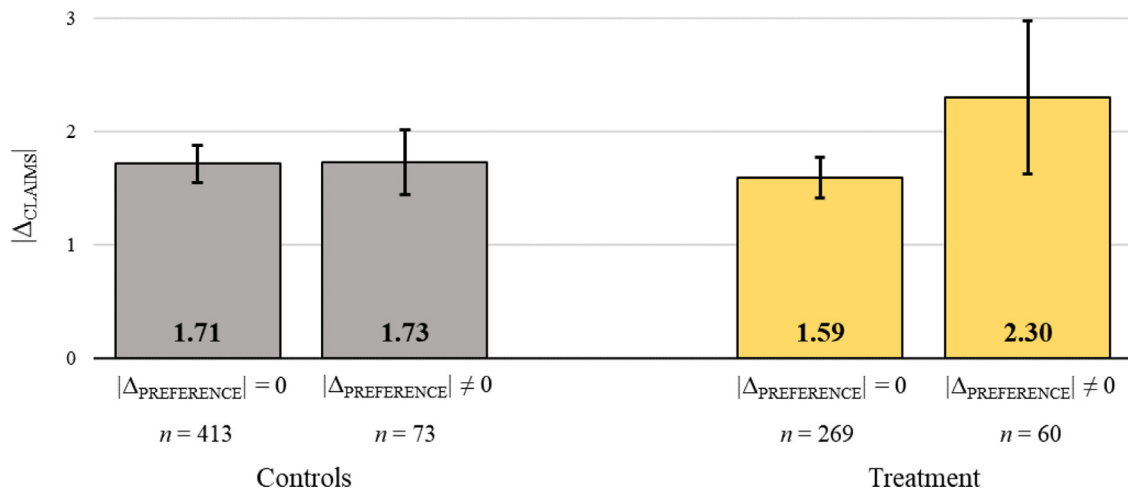
Fig. 5. Extent of conformism in honesty (i.e., $|\Delta_{CLAIMS}|$) among those who showed conformism in cooperation ($|\Delta_{PREFERENCE}| \neq 0$) vs. those who did not show conformism in cooperation ($|\Delta_{PREFERENCE}| = 0$) compared across the control conditions (i.e., among those who did not observe peers) and the treatment condition (i.e., among those observed peers).

Table 3
Role of personality traits: Spearman correlations.

Personality Trait	Honesty Task		Cooperation Task	
	Honesty (a)	Conformism (b)	Cooperativeness (c)	Conformism (d)
Machiavellianism	0.19***	0.00	-0.17**	0.04
Social Comparison Orientation	0.09	0.02	-0.02	-0.05
Need for Cognition	-0.04	-0.04	-0.03	-0.08
Faith in Intuition	0.05	0.00	0.00	0.02
Extraversion	0.03	-0.11*	-0.06	-0.03
Agreeableness	-0.01	-0.07	0.13*	-0.02
Conscientiousness	-0.01	-0.11	-0.08	-0.06
Emotional Stability	0.14*	-0.08	-0.09	-0.05
Openness to Experience	0.10	-0.04	-0.05	0.01

Note. The table describes Spearman correlation coefficients (r_S) between various measures of personality traits on the one hand and measures elicited in the honesty and cooperation tasks in the treatment condition, where participants observed peers. Task measures include (a) honesty behavior, measured as the total number of match claims in forty rounds of the Mind Game, (b) conformism in honesty, measured as $|\Delta_{CLAIMS}|$, (c) initial cooperativeness, measured as the number of "shares" in the 1st Report of the one-shot Prisoner's Dilemma, and (d) conformism in cooperation, defined as $|\Delta_{PREFERENCE}| \neq 0$. * $p < .05$, ** $p < .01$, *** $p < .001$. None of the statistically significant results survive Bonferroni corrections for multiple hypothesis testing (i.e., $ps < 0.050/9$), except for the significant results of Machiavellianism in Honesty (a) and Cooperativeness (c).

subjects design of our experiment, these measures of conformism control for possible behavioral and informational spillovers between the two domains, since any carryover of the honesty task to the cooperation task would be picked up by the initial preference measure (i.e., the number of "shares" in the 1st Report). Similarly, any carryover of the cooperation task to the honesty task would be picked up by the initial behavioral measure in the honesty task (i.e., total match claims in the 1st Phase). We test for the association between these two measures of conformism and compare them to the control benchmark to investigate whether conformism is a domain-general behavioral trait.

Our analysis suggests that conformism is a domain-general behavioral trait. Participants who showed a tendency to be influenced by peers in the cooperation domain were also more conforming in the honesty domain (see Fig. 5). In the treatment condition, conformism in honesty, measured as $|\Delta_{CLAIMS}|$, was significantly higher among those who showed conformism in cooperation than among those who were not influenced by peers: 2.30 vs. 1.59 claims, $t(327) = 2.85$, $p = .005$, $d = 0.41$. The control benchmark that excluded peer observation supported the validity of the conformism measures, since no difference in $|\Delta_{CLAIMS}|$ was found for the controls when comparing participants with $|\Delta_{PREFERENCE}| \neq 0$ and those with $|\Delta_{PREFERENCE}| = 0$: 1.73 vs. 1.71 claims, $t(484) = 0.06$, $p = .955$, $d = 0.01$.

Personality traits. Here, we explore whether various standard personality traits explain honesty, cooperation, and conformism in the treatment condition. As detailed in Table 3, the level of dishonesty, measured as the total match claims across the forty rounds of the Mind Game, was positively correlated with Machiavellianism ($p < .001$) and the Emotional Stability component of the Big-Five personality model ($p = .011$). On the other hand, initial cooperativeness, as indicated by the number of "shares" in the 1st Report of the one-shot Prisoner's Dilemma, was negatively correlated with Machiavellianism ($p = .002$) and positively correlated with the Agreeableness component of the Big-Five personality model ($p = .002$).

These results on honesty and cooperation are consistent with the literature (Gunnthorsdottir et al., 2002; Hodson et al., 2018; Volk et al., 2011, 2012). In contrast, the extent of conformism in honesty, measured as $|\Delta_{CLAIMS}|$, was significantly (and negatively) correlated with only the Extraversion component of the Big-Five personality model ($p = .047$), and conformism in cooperation, defined as $|\Delta_{PREFERENCE}| \neq 0$, was not correlated with any of the personality traits. Overall, Machiavellianism was the only underlying personality trait that survived Bonferroni corrections for multiple hypothesis testing (i.e., $ps < 0.050/9$) and that simultaneously explained both honesty and cooperativeness, while none of the personality trait measures explained conformism across the moral domains.

5. Conclusion and future research

We estimated the pure informational effects of peer observation in honesty and cooperation. Consistent with the literature, our estimates showed significant peer effects in both directions in honesty and cooperation, with negative peer influences tending to be stronger than positive peer influences.

Our study also provided behavioral tests of individual-level links between honesty and cooperation. These tests revealed that cooperative participants were more honest, and these measures were negatively associated with Machiavellianism. While standard personality trait scales, including Machiavellianism, failed to show a link in conformism between the two moral domains, our behavioral test provided evidence for the domain-generality of conformism such that participants who were influenced by peers in cooperation tended to be influenced by peers in honesty as well. Future research should conduct confirmatory tests of the domain-generality of conformism using behavioral experiments in honesty and cooperation and extend this finding to other moral domains such as generosity, trust, and punishment. Likewise, personality traits underlying conformism can be further explored, including intellectual humility which is known to correlate positively with prosociality (Krumrei-Mancuso, 2017) and honesty (Alfano et al., 2017).

These findings are relevant for the study of social norms and for their application in policy settings. A promising method for changing social norm perceptions is *norm-nudging* (for a comprehensive review see Bicchieri and Dimant, 2019). Norm-nudges often work by relaying messages that describe majority behavior in one's reference group—students in one's classroom (e.g., Damgaard and Nielsen, 2018); residents in one's neighborhood (e.g., Allcott, 2011; Ferraro and Price, 2013); or fellow citizens in one's society (e.g., Fellner et al., 2013; Gerber and Rogers, 2009; Hallsworth et al., 2017). In a famous example, messages left in hotel rooms that most guests reuse their towels were shown to increase conservation behavior (Goldstein et al., 2008). Nevertheless, evidence on the overall effectiveness of norm-nudges is mixed (Bicchieri and Dimant, 2019; Bursztyn et al., 2020; Chen et al., 2021; Dimant et al., 2020; Fosgaard et al., 2013; Silva and John, 2017), suggesting the need for new tools for harnessing the power of social norm perceptions.

While norm-nudges tend to involve group-level descriptions (e.g., reading in a letter that "most of your neighbors are tax-compliant"), in everyday social interactions we routinely acquire social information by observing *individual peers*, whether through descriptions (e.g., hearing a story about how your next-door neighbor was honest) or direct observation (i.e., witnessing your next-door neighbor telling a lie). Hence, we propose that *peer-nudging*—changing social norm perceptions through the observation of individual peers—can serve as an additional, ecologically realistic norm-nudging tool. The targeted use of peer-nudging promises to be an effective tool for disseminating social information because individual observations can act as vivid exemplars or memorable anecdotes and consequently have outsized influence on norm perceptions (Brosius, 2000; Hamill et al., 1980; Hornikx, 2007). Furthermore, the influence of peer observation on norm perceptions increases with social proximity to the peer (Bicchieri et al., 2022). For example, the above-mentioned study on norm-nudging of towel reuse by Goldstein et al. (2008) was most effective for the message stating that "the majority of guests *in this room* reuse their towels" (emphasis ours). Furthermore, the increasing prevalence of social media use in our everyday social interactions has dramatically amplified the effects of individual peer observation, as demonstrated by the recent phenomena of social media "influencers" and "echo-chambers".

Admittedly, the way we individually perceive our peers is conceptually distinct from norm perceptions. At the same time, peer observation is likely to influence the perceptions of reference group norms. Any unexpected peer information will likely result in the update of norm perceptions about one's reference group which the peer represents. For example, observing the surprisingly well-behaved manners of a peer can lead to positive behavior change through the update of the observer's initially pessimistic beliefs about group norms. In contrast, negative peer-effects on norm perceptions could occur if a peer (e.g., a citizen) who is representative of one's group (e.g., a society) unexpectedly violates a norm. Peer-nudging is flexible enough to have many applications in the field. For example, public messages can provide vivid anecdotal descriptions of individual acts of surprising prosociality to motivate norm compliance. In our setup, beliefs were elicited only prior to peer observation. Future studies should re-elicit beliefs after peer observation to study whether peer effects work through belief change.

Our results suggest that the directed use of individual peer information can be welfare enhancing in both honesty and cooperation. Furthermore, our results on conformism, as an individual trait linking honesty and cooperation, suggests that peer-nudges effective in one moral domain might be effective in other domains as well. Our analyses also indicate that negative peer effects tend to be stronger than positive ones, which suggests that any peer-nudges should be carefully piloted before implemented as behavioral public policy. Nevertheless, even these generally detrimental dynamics can potentially be harnessed for motivating non-compliance for social benefit, for example, where observations of free riding can be used to lower cooperation towards unethical goals.

Future research should test the effectiveness of anecdotal public messages involving peer-nudges in comparison to standard norm-nudges that provide group-level descriptions of norms. These public messages would use the potentially outsized influence of individual peers for positive social influence, for instance, through campaigns about "leading by example" (e.g., Gächter and Renner, 2018) or "being a role-model" (e.g., Mani and Riley, 2019; Riley, 2017). This promises to be particularly effective when disseminated by trusted leaders (the "messenger" matters – see Dolan et al., 2012) or through social media channels (e.g., Bonnevie et al., 2020). In this way, the promise of peer-nudging can go beyond the standard public policy initiatives and empower individuals to personally engage in positive behavior change. While we leave the explicit comparison of the effectiveness of standard norm-nudges and peer-nudges for future research, our findings on peer influence and conformism indicate that peer information can have predictable influences on honesty and cooperation that could be harnessed as peer-nudges. In short, peer-nudges promise to advance norm-nudging, but the pitfalls of peer effects require further tests.

Declaration of Competing Interest

None.

Data availability

Data and analysis files can be found here: https://osf.io/c84n5/

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2021.12.026.

References

Alfano, M., Iurino, K., Stey, P., Robinson, B., Christen, M., Yu, F., Lapsley, D., 2017. Development and validation of a multi-dimensional measure of intellectual humility. PLoS One 12 (8), e0182950.

Allcott, H., 2011. Social norms and energy conservation. J. Public Econ. 95 (9–10), 1082–1095. doi:10.1016/j.jpubeco.2011.03.003.

Bardsley, N., Sausgruber, R., 2005. Conformity and reciprocity in public good provision. J. Econ. Psychol. 26 (5), 664–681. doi:10.1016/j.joep.2005.02.001.

Bicchieri, C., Dimant, E., 2019. Nudging with care: the risks and benefits of social information. Public Choice 1–22.

Bicchieri, C., Dimant, E., Gächter, S., Nosenzo, D., 2022. Social proximity and the erosion of norm compliance. Games Econ. Behav. 132, 59–72. doi:10.1016/j.geb.2021.11.012.

Biziou-van-Pol, L., Haenen, J., Novaro, A., Occhipinti Liberman, A., Capraro, V., 2015. Does telling white lies signal pro-social preferences? Judgm. Decis. Mak. 10, 538–548.

Bonnevie, E., Rosenberg, S.D., Kummeth, C., Goldbarg, J., Wartella, E., Smyser, J., 2020. Using social media influencers to increase knowledge and positive attitudes toward the flu vaccine. PLoS One 15 (10), e0240828. doi:10.1371/journal.pone.0240828.

Brosius, H.B., 2000. Toward an exemplification theory of news effects. Doc. Des. 2 (1), 18–27.

Bursztyn, L., González, A.L., Yanagizawa-Drott, D., 2020. Misperceived social norms: women working outside the home in saudi arabia. Am. Econ. Rev. 110 (10), 2997–3029.

Capraro, V., Schulz, J., Rand, D.G., 2019. Time pressure and honesty in a deception game. J. Behav. Exp. Econ. 79, 93–99. doi:10.1016/j.socec.2019.01.007.

Carpenter, J.P., 2004. When in Rome: conformity and the provision of public goods. J. Soc. Econ. 33 (4), 395–408. doi:10.1016/j.socec.2004.04.009.

Chaudhuri, A., 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Exp. Econ. 14 (1), 47–83. doi:10.1007/s10683-010-9257-1.

Chen, J.C., Fonseca, M.A., Grimshaw, S.B., 2021. When a nudge is (not) enough: experiments on social information and incentives. Eur. Econ. Rev. 134, 103711.

Christie, R., Geis, F.L., 1970. Studies in Machiavellianism. Academic Press.

Cohen, J., 2013. Statistical Power Analysis For the Behavioral Sciences. Taylor & Francis.

Croson, R.T.A., 2007. Theories of commitment, altruism and reciprocity: evidence from linear public goods games. Econ. Inq. 45 (2), 199–216. doi:10.1111/j.1465-7295.2006.00006.x.

Damgaard, M.T., Nielsen, H.S., 2018. Nudging in education. Econ. Educ. Rev. 64, 313–342. doi:10.1016/j.econedurev.2018.03.008.

Diekmann, A., Przepiorka, W., Rauhut, H., 2015. Lifting the veil of ignorance: an experiment on the contagiousness of norm violations. Ration. Soc. 27 (3), 309–333. doi:10.1177/1043463115593109.

Dimant, E., Van Kleef, G.A., Shalvi, S., 2020. Requiem for a Nudge: framing effects in nudging honesty. J. Econ. Behav. Organ 172, 247–266. doi:10.1016/j.jebo.2020.02.015.

Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., Vlaev, I., 2012. Influencing behaviour: the mindspace way. J. Econ. Psychol. 33 (1), 264–277. doi:10.1016/j.joep.2011.10.009.

Epstein, S., Pacini, R., Denes-Raj, V., Heier, H., 1996. Individual differences in intuitive–experiential and analytical–rational thinking styles. J. Pers. Soc. Psychol. 71 (2), 390–405. doi:10.1037/0022-3514.71.2.390.

Falk, A., Fischbacher, U., Gächter, S., 2013. Living in two neighborhoods-social interaction effects in the laboratory. Econ. Inq. 51 (1), 563–578. doi:10.1111/j.1465-7295.2010.00332.x.

Fatas, E., Heap, S.P.H., Arjona, D.R., 2018. Preference conformism: an experiment. Eur. Econ. Rev. 105, 71–82.

Fehr, E., Schurtenberger, I., 2018. Normative foundations of human cooperation. Nat. Hum. Behav. 2 (7), 458–468. doi:10.1038/s41562-018-0385-5.

Fellner, G., Sausgruber, R., Traxler, C., 2013. Testing enforcement strategies in the field: threat, moral appeal and social information. J. Eur. Econ. Assoc. 11 (3), 634–660. doi:10.1111/jeea.12013.

Ferraro, P.J., Price, M.K., 2013. Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. Rev. Econ. Stat. 95 (1), 64–73. doi:10.1162/rest_a_00344.

Fischbacher, U., Föllmi-Heusi, F., 2013. Lies in disguise-an experimental study on cheating. J. Eur. Econ. Assoc. 11 (3), 525–547. doi:10.1111/jeea.12014.

Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Econ. Lett. 71, 397–404.

Fosgaard, T.R., Hansen, L.G., Piovesan, M., 2013. Separating will from grace: an experiment on conformity and awareness in cheating. J. Econ. Behav. Organ. 93, 279–284. doi:10.1016/j.jebo.2013.03.027.

Frey, B.S., Meier, S., 2004. Social comparisons and pro-social behavior: testing" conditional cooperation" in a field experiment. Am. Econ. Rev. 94 (5), 1717–1722.

Gächter, S., Nosenzo, D., Sefton, M., 2013. Peer effects in pro-social behavior: social norms or social preferences? J. Eur. Econ. Assoc. 11 (3), 548–573. doi:10.1111/jeea.12015.

Gächter, S., Renner, E., 2018. Leaders as role models and 'belief managers' in social dilemmas. J. Econ. Behav. Organ. 154, 321–334. doi:10.1016/j.jebo.2018. 08.001.

Gerber, A.S., Rogers, T., 2009. Descriptive social norms and motivation to vote: everybody's voting and so should you. J. Politics 71 (1), 178–191. doi:10.1017/ s0022381608090117.

Giamattei, M., Yahosseini, K.S., Gächter, S., Molleman, L., 2020. LIONESS Lab: a free web-based platform for conducting interactive experiments online. J. Econ. Sci. Assoc. 6 (1), 95–111. doi:10.1007/s40881-020-00087-0.

Gibbons, F.X., Buunk, B.P., 1999. Individual differences in social comparison: development of a scale of social comparison orientation. J. Personal. Soc. Psychol. 76 (1), 129.

Gino, F., Ayal, S., Ariely, D., 2009. Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. Psychol. Sci. 20 (3), 393–398.

Goldstein, N.J., Cialdini, R.B., Griskevicius, V., 2008. A room with a viewpoint: using social norms to motivate environmental conservation in hotels. J. Consum. Res. 35 (3), 472–482. doi:10.1086/586910.

Gosling, S.D., Rentfrow, P.J., Swann, W.B., 2003. A very brief measure of the Big-Five personality domains. J. Res. Personal. 37 (6), 504–528. doi:10.1016/ s0092-6566(03)00046-1.

Gunnthorsdottir, A., Mccabe, K., Smith, V., 2002. Using the Machiavellianism instrument to predict trustworthiness in a bargaining game. J. Econ. Psychol. 23 (1), 49–66. doi:10.1016/s0167-4870(01)00067-8.

Hallsworth, M., List, J.A., Metcalfe, R.D., Vlaev, I., 2017. The behavioralist as tax collector: using natural field experiments to enhance tax compliance. J. Public Econ. 148, 14–31.

Hamill, R., Wilson, T.D., Nisbett, R.E., 1980. Insensitivity to sample bias: generalizing from atypical cases. J. Personal. Soc. Psychol. 39 (4), 578.

Herbst, D., Mas, A., 2015. Peer effects on worker output in the laboratory generalize to the field. Science 350 (6260), 545–549.

Hilbig, B.E., Zettler, I., 2015. When the cat's away, some mice will play: a basic trait account of dishonest behavior. J. Res. Personal. 57, 72–88. doi:10.1016/ j.jrp.2015.04.003.

Hodson, G., Book, A., Visser, B.A., Volk, A.A., Ashton, M.C., Lee, K., 2018. Is the dark triad common factor distinct from low honesty-humility? J. Res. Personal. 73, 123–129.

Hornikx, J., 2007. Is anecdotal evidence more persuasive than statistical evidence? A comment on classic cognitive psychological studies. Stud. Commun. Sci. 7 (2).

Isler, O., Gächter, S., Maule, A.J., Starmer, C., 2021a. Contextualised strong reciprocity explains selfless cooperation despite selfish intuitions and weak social heuristics. Sci. Rep. 11 (1), 13868. doi:10.1038/s41598-021-93412-4.

Isler, O., Yilmaz, O., Maule, A.J., 2021b. Religion, parochialism and intuitive cooperation. Nat. Hum. Behav. 5, 512–521. doi:10.1038/s41562-020-01014-3.

Jiang, T., 2013. Cheating in mind games: the subtlety of rules matters. J. Econ. Behav. Organ. 93, 328–336. doi:10.1016/j.jebo.2013.04.003.

Kroher, M., Wolbring, T., 2015. Social control, social learning, and cheating: evidence from lab and online experiments on dishonesty. Soc. Sci. Res. 53, 311–324. doi:10.1016/j.ssresearch.2015.06.003.

Krumrei-Mancuso, E.J., 2017. Intellectual humility and prosocial values: direct and mediated effects. J. Posit. Psychol. 12 (1), 13–28.

Mani, A., & Riley, E. (2019). Social networks, role models, peer effects, and aspirations (2019/120). WIDER Working Paper. https://www.wider.unu.edu/sites/ default/files/Publications/Working-paper/PDF/wp-2019-120.pdf

Rauhut, H., 2013. Beliefs about lying and spreading of dishonesty: undetected lies and their constructive and destructive social dynamics in dice experiments. PLoS One 8 (11), e77878. doi:10.1371/journal.pone.0077878.

Riley, E. (2017). Increasing students' aspirations: the impact of Queen of Katwe on students' educational attainment. https://mbrg.bsg.ox.ac.uk/sites/default/ files/2020-01/csae-wps-2017-13.pdf

Silva, A., John, P., 2017. Social norms don't always work: an experiment to encourage more efficient fees collection for students. PLoS One 12 (5), e0177354. doi:10.1371/journal.pone.0177354.

Sutton, J., Keogh, E., 2001. Components of Machiavellian beliefs in children: relationships with personality. Personal. Individ. Differ. 30 (1), 137–148.

Thielmann, I., Spadaro, G., Balliet, D., 2020. Personality and prosocial behavior: a theoretical framework and meta-analysis. Psychol. Bull. 146 (1), 30.

Thöni, C., Gächter, S., 2015. Peer effects and social preferences in voluntary cooperation: a theoretical and experimental analysis. J. Econ. Psychol. 48, 72–88. doi:10.1016/j.joep.2015.03.001.

Thöni, C., Volk, S., 2018. Conditional cooperation: review and refinement. Econ. Lett. 171, 37–40. doi:10.1016/j.econlet.2018.06.022.

Volk, S., Thöni, C., Ruigrok, W., 2011. Personality, personal values and cooperation preferences in public goods games: a longitudinal study. Personal. Individ. Differ. 50 (6), 810–815. doi:10.1016/j.paid.2011.01.001.

Volk, S., Thöni, C., Ruigrok, W., 2012. Temporal stability and psychological foundations of cooperation preferences. J. Econ. Behav. Organ. 81 (2), 664–676. doi:10.1016/j.jebo.2011.10.006.

Zettler, I., Hilbig, B.E., Heydasch, T., 2013. Two sides of one coin: honesty–Humility and situational factors mutually shape social dilemma decision making. J. Res. Personal 47 (4), 286–295. doi:10.1016/j.jrp.2013.01.012.