Following Your Group or Your Morals? The In-Group Promotes Immoral Behavior While the Out-Group Buffers Against It

Marc-Lluís Vives[1], Mina Cikara[2], and Oriel FeldmanHall[1,3]

Abstract

People learn by observing others, albeit not uniformly. Witnessing an immoral behavior causes observers to commit immoral actions, especially when the perpetrator is part of the in-group. Does conformist behavior hold when observing the out-group? We conducted three experiments ($N = 1,358$) exploring how observing an (im)moral in-/out-group member changed decisions relating to justice: punitive, selfish, or dishonest choices. Only immoral in-groups increased immoral actions, while the same immoral behavior from out-groups had no effect (Experiments 1 & 2). In contrast, a compassionate or generous individual did not make people more moral, regardless of group membership (Experiments 1 & 2). When there was a loophole to deny cheating (Experiment 3), neither an immoral in-/out-group member changed dishonest behavior. Compared to observing an honest in-group member, people become more honest themselves after observing an honest out-group member, revealing that out-groups can enhance morality. Depending on the severity of the moral action, the in-group licenses immoral behavior while the out-group buffers against it.

Keywords

For millennia, humans have lived in closely knit social communities. These social groups serve multiple functions, from increasing the probability of surviving to transmitting important social information, such as moral norms (Bowles & Gintis, 1998; Gintis et al., 2005). Adherence to these moral norms (e.g., altruism, fairness, honesty) is critical for maintaining one's status as a group member. As a result, one approach to constructing a useful moral compass is to internalize the group's moral norms and apply them in a relatively automatic fashion (Bicchieri, 2005; Everett et al., 2017). However, the social world is morally mixed; ethical and unethical behaviors abound in every community. A brief glance at any history text reveals that one's in-group can behave in both morally respectable and reprehensible ways, as can the out-group. This sets up an interesting dilemma. If we take our moral cues from the in-group, and disdain the behaviors of the out-group (Stein, 2017), do we learn from an in-group member who behaves in immoral ways, or conversely, an out-group member who behaves in morally virtuous ways?

Although it is possible to dismiss the immoral in-group member, mounting evidence suggests that we often opt for another, less morally virtuous avenue—condoning immoral in-group behavior (Aquino et al., 2007). For example, after observing a dishonest in-group member, individuals are more likely to act dishonestly themselves. Observing a dishonest out-group,

however, has the opposite effect of making people less immoral (Gino et al., 2009). These findings suggest that people do not simply learn to conform to the in-group's moral behavior. Instead, they learn by a strategy which encompasses a conjunction of motives: aligning one's behaviors with an immoral in-group and opposing the behaviors of an immoral out-group.

Relying on this conjunction of motives may facilitate psychological consistency as long as the out-group behaves immorally. However, the presence of a moral out-group member poses a challenge. Moral out-groups produce conflict between not wanting to align one's self with the out-group while simultaneously trying to maintain a positive moral self-image (Aquino & Reed, 2002; Jordan et al., 2011). A solution for maintaining a positive self-image in the face of a moral out-group member is to transform the situation into a moral intergroup competition. Because people strive to be perceived as

[1] Department of Cognitive, Linguistic, Psychological Sciences, Brown University, Providence, RI, USA
[2] Department of Psychology, Harvard University, Cambridge, MA, USA
[3] Carney Institute for Brain Science, Brown University, Providence, RI, USA

Corresponding Author:
Oriel FeldmanHall, Department of Cognitive, Linguistic, Psychological Sciences, Brown University, Providence, RI 02912, USA.
Email: oriel.feldmanhall@brown.edu

positively distinct—both for themselves and the groups they belong to (Turner, 1975)—intergroup competition may fuel an individual to behave in even more morally virtuous ways relative to the moral out-group. Thus, an alternative strategy is moral competition so one is never worse than the out-group.

Current Experiments and Hypotheses

We test which strategy best characterizes people's behavior across a variety of situations by examining how observing (im)moral in-/out-group members shape one's own moral behaviors. Past work indicates that the conformist strategy best characterizes people's moral decision making (Gino et al., 2009), where an individual will follow the behaviors of an in-group member and avoid behaviors endorsed by the out-group (e.g., behave more immorally after observing a virtuous out-group member). We introduce an alternative hypothesis: People are asymmetrically influenced by a teacher's group membership depending on whether she is modeling moral versus immoral behavior. Following previous work (Gino et al., 2009), we predict people will learn to be immoral from an immoral in-group but not an out-group. We expect this pattern to be different for moral teachers, however. While individuals should act morally after observing a virtuous in-group member (Abrams et al., 1990), it is possible that observing a virtuous out-group member boosts moral behavior even further—as it is better to appear more virtuous than a moral out-group member. Across three experiments, we tested the hypothesis that group membership asymmetrically influences the contagion of moral behaviors: Moral corruption will be uniquely transmitted by in-group members, while moral virtuosity will be transmitted more by out-group members. Extending previous research that has focused on (im)moral contagion across group membership in one specific moral dimension (e.g., cheating; Gino et al., 2009), here we examine how conformity unfolds across different moral dimensions relating to justice: retributive justice (justice game, punishment; FeldmanHall et al., 2014), distributive justice (dictator game, fairness; Kahneman et al., 1986), and procedural justice (cheating game, honesty; Shalvi, Handgraaf, et al., 2011). In these paradigms, behaviors are typically considered moral when they maximize outcomes of all parties at the monetary expense of the person deciding. To ensure a particularly effective manipulation showcasing conflicting moral values between groups, we bifurcated subjects by their preferred political candidate in the 2016 U.S. presidential election. This capitalizes on the naturally occurring zero-sum structure that characterizes two-party politics, especially given that partisan membership is often accompanied by negative feelings for, and attributions of immorality to the out-group on both sides of the aisle (Finkel et al., 2020).

General Method

Participants

In Experiments 1–3, 1,358 participants (34% female, average age = 32.12) were recruited from Amazon Mechanical Turk (MTurk; see Online Supplement for limitations using online platforms). All methods were approved by Brown University Institutional Review Board. Informed consent was obtained from participants before the experiment. Participants were paid US$3.00 and an additional bonus from one randomly realized trial. Before data collection, we defined a $d = .4$ as our smallest effect size of interest (Lakens, 2014), which requires 100 participants per cell with 80% power (Brysbaert & Stevens, 2018), resulting in at least 400 participants per experiment. All participants were excluded if they failed to identify their group assignment. Based on previous work, we expected 20% of participants to fail this question and thus we collected at least 500 participants per experiment, each of which lasted 20–30 min.
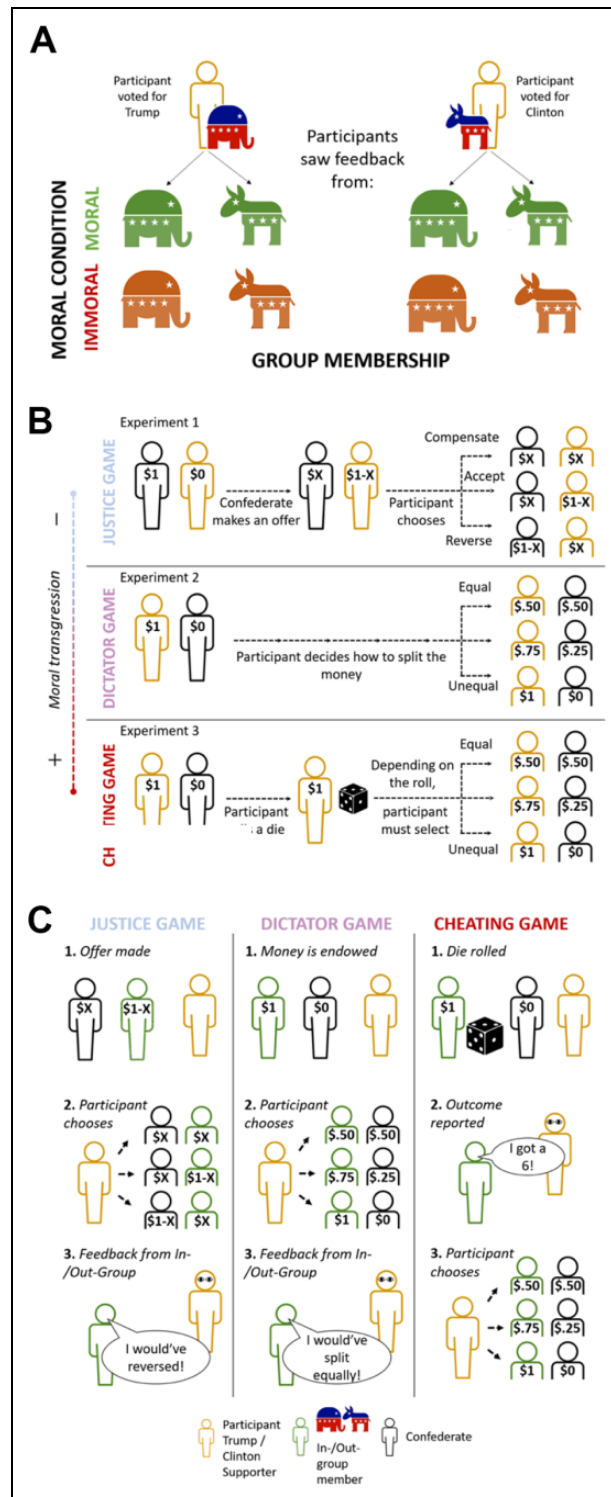
Experimental Design

At the beginning of each experiment, participants reported who they voted for in the 2016 U.S. election: Trump, Clinton, or neither (see Figure 1A). Participants who responded "neither" were excluded from analysis. To justify the existence of teams, participants were told that the Trump and Clinton teams were competing against each other in a later problem-solving task (Cikara et al., 2014; Everett et al., 2017). Participants were then randomly assigned to interact with a Clinton or Trump supporter. Different behavioral economic games were used in each experiment. We tested the extent to which moral contagion is influenced in a variety of moral violations (Figure 1B): retributive justice (concerns regarding punishing moral transgressions, justice game), distributive justice (concerns regarding the allocation of goods, dictator game), and procedural justice (concerns regarding transgressions of judicial procedures, cheating game). It was stressed that the outcome from these games was unconnected to the later problem-solving competition.

The structure of the economic games had three phases (FeldmanHall et al., 2018). First, in the baseline phase, participants played without any social influence. Other players were not affiliated with either political party. During this phase, participants could earn extra money. Second, in the learning phase, rather than passively observing the preferences of the group member, participants actively decided on behalf of either a Clinton or a Trump supporter. On every trial, participants learned what the Clinton or Trump supporter would have chosen. Participants were randomly assigned to learn from a moral or immoral player in a between-subjects 2 (group-membership: in-/out-group) × 2 (moral signal: moral/immoral) design (Figure 1A). Finally, in the transfer phase, participants decided again for themselves. As before, the other players were unaffiliated and participants could earn extra money. Comparing the transfer phase to the baseline phase allowed us to analyze to what extent decisions were influenced by the observation of a moral/immoral, in-/out-group member.

Analysis

In each experiment, we conducted a mixed-effect regression analysis fitting fixed and random (subject-specific) slopes for

Figure 1. Experimental task structure. *Note.* (A) The experimental manipulation is represented. In a full between-subjects design, participants were assigned to the Clinton or Trump team depending on their vote on the 2016 presidential election. After this, half of them were assigned to observe a moral Trump or Clinton supporter, and the other half an immoral Trump or Clinton supporter. (B) The games used in the baseline and transfer phases where participants made decisions for their own monetary gain. Games reflect mild moral transgressions (the justice game: Experiment 1) to more severe moral transgressions (the cheating game: Experiment 3). In the transfer phase—which followed the Learning phase—participants played the game again (as in the baseline phase) so we could capture how different moral signals might influence behavior. The dictator game served as the midpoint (Experiment 2). (C) The learning phase is depicted for each experiment. Participants always made decisions on behalf of an in-/out-group member. For the justice and dictator games, participants received feedback after their decision. For the cheating game, participants were told the outcomes of the roll die by a group member, and then participants chose the outcome.

each variable when needed, as well as random intercepts for each participant (Barr et al., 2013). Since we did not have any a priori predictions about how the moderating effect size for group membership would differ between observing an immoral individual compared to a moral individual, two analyses were run for each experiment, one for participants in the immoral condition and another for participants in the moral condition. Code and data are available on OSF (https://osf.io/fnm3w/).

Experiment 1: Justice Game

Although punishing transgressors disincentivizes individuals from breaking norms, it can also be perceived as vindictive. For example, victims prefer to compensate themselves than punish the perpetrator (FeldmanHall et al., 2014) and seem to do so because punishing the perpetrator is perceived as immoral (Heffner & FeldmanHall, 2019). In Experiment 1, we examined whether deciding to punish (or compensate) is enhanced by witnessing a vindictive (or compassionate) in-/out-group member, with the hypothesis that immoral behavior (vindicative punishment) will only be transmitted within groups, and moral behavior (compensation) will be transmitted both within and across groups.

Method

Participants

While 689 participants were recruited, 138 were excluded according to our exclusion criteria. This led to a final sample of 551 subjects (37% female, average age = 32.23, $SD = 9.25$).

Procedure

After being assigned to their team, participants played the justice game. In this game, a confederate can propose any division of a US\$1 endowment to participants. Participants can then reapportion the money by choosing from the following options: (1) punish by reversing the split, (2) accept the split as it is, or (3) compensate themselves by increasing their payout to match confederate's payout. On each trial, participants could always choose between these three options (see Figure 1B). To ensure that participants perceived offers as unfair, they were told that that half of the time the realized trial would be paid out according to confederate's split and half the time according to their decision.

In the baseline phase, confederates offered unfair splits to participants ranging from slightly unfair (US\$.60, US\$.40) to highly unfair (US\$.90, US\$.10). In the learning phase, participants decided for an in-/out-group member that either provided vindictive feedback (immoral condition, desiring to punish the perpetrator 90% of the time) or compassionate feedback (moral condition, desiring to compensate themselves and not punish 90% of the time)—a between-subjects design (Figure 1C). Participants played 80 trials in the learning phase (20 for each type

of offer). Finally, in the transfer phase, participants responded to offers from other confederates for 20 trials.
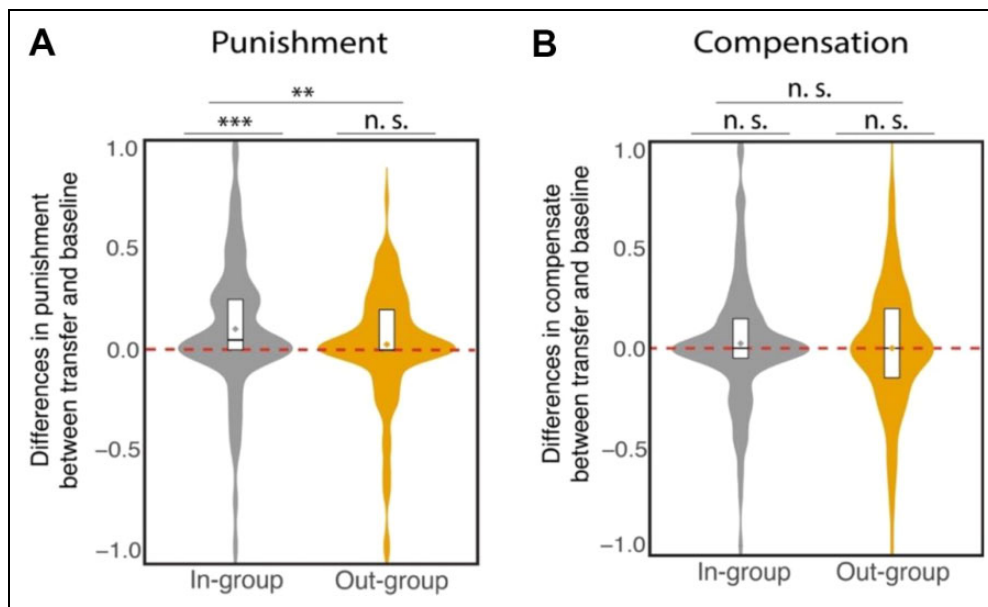
Results

To explore whether group membership had an effect on transmitting moral or immoral behaviors, we ran two mixed-effects regressions where the probability of punishing (=1, elsewise 0 = compensate/accept) was modeled as a function of phase (baseline/transfer) and group membership (in-group/out-group). Participant was treated as a random intercept. The same analysis pipeline was run for the moral condition, except that we predicted participants' probability to compensate (=1, elsewise 0 = punish/accept). Across experimental conditions, participants preferred to compensate 53.8% of the time in the baseline phase (accept = 20.5%, punish = 25.6%).

Dovetailing with previous work (FeldmanHall et al., 2018), participants became significantly more punitive after observing a punitive individual, as indexed by a significant effect of phase predicting punishment, $\beta = 0.81 \pm 0.12$, 95% CI [0.56, 1.05], $p < .001$. However, this effect was moderated by a significant interaction with group membership: Participants only became more punitive after observing a punitive in-group member but not a punitive out-group member, $\beta = -0.57 \pm 0.19$, 95% CI [−0.94, −0.20], $p = .002$. There was no main effect of group membership, $\beta = -0.46 \pm 0.35$, $p = .19$.

We additionally tested whether there were any differences in choosing the punitive option between transfer and baseline within the same group (0 = *no change between phases*). Results revealed that observing a punitive in-group did make participants significantly more punitive; t test: $t(154) = 3.96$, $p < .001$; $M = 0.10$, CI [0.05, 0.16], while observing a punitive out-group did not exert any behavioral change; $t(124) = 1.13$, $p = .26$; $M = 0.03$, CI [−0.02, 0.08], Figure 2A. Effectively, observing a punitive out-group member fails to provoke any change at all.

In contrast, we found no evidence that observing a compassionate in-group versus out-group member made participants endorse the compensate option more often (neither phase, group membership, or its interaction predicted decisions to compensate; phase: $\beta = 0.21 \pm 0.13$, $p = .13$, in-/out-group: $\beta = -0.26 \pm 0.37$, $p = .49$, interaction: $\beta = -0.20 \pm 0.18$, $p = .26$; Figure 2B).

Results indicate that desiring to punish another only seems to be transmitted by in-groups. On the other hand, we failed to find any transfer effects for compensatory behavior. This is in line with previous research showing that punitive preferences are more susceptible to social influence than decisions to compensate (FeldmanHall et al., 2018). However, punishing a perpetrator is often considered the moral response (Heffner & FeldmanHall, 2019). This would leave enough moral wiggle room to disregard any cues from the compassionate out-group. Accordingly, we next examined another moral norm, fairness, testing the degree to which these results are observed when the moral signal associated with breaking the norm is less morally ambiguous.

Figure 2. Results Experiment 1. *Note.* Violin plots depict the density of the distribution of the differences between Transfer and Baseline for choices to punish or compensate. The point within the boxplot depicts the mean. 0 on the y-axis indicates no change in behavior pre-versus postexposure to the "teacher." (A) Participants became more punitive after observing a punitive in-group member but not a punitive out-group member. (B) Participants did not modify their compensation rates after observing a compassionate in- or out-group member. $**p < .01$. $***p < .001$.

Experiment 2: Dictator Game

Individuals and companies who partake in behaviors that perpetuate unfairness are perceived as immoral (Camerer, 2003; Heffner & FeldmanHall, 2019; Kahneman et al., 1986). In Experiment 2, we directly tested fairness concerns by exploring decisions to divide a monetary endowment in the dictator game (Engel, 2011), positing that people will demonstrate asymmetrical conformity effects depending on whether the unfair/fair individual is an in-/out-group. Specifically, selfish behavior will only be transmitted when observing an unfair in-group, while generous behavior will be more strongly transmitted by observing a virtuous out-group.

Method

Participants

We recruited 583 participants. Around 185 were excluded based on our criteria, leading to a final sample of 398 (38% female, average age = 31.54, $SD = 8.42$).

Procedure

After being assigned to their team, participants played the dictator game. In this dictator game (Engel, 2011), participants were endowed with US$1 and had to unilaterally decide how to divide it with a confederate who had no bargaining power. Participants could split US$1 between themselves and a confederate by choosing between the following options: (1) an unfair split (participant: $1, confederate: $0), (2) a mildly

unfair split (participant: US$.75, confederate: US$.25), or (3) a fair split (participant: $.50, confederate: $.50; see Figure 1B).

As in the previous experiment, participants dictated the monetary splits for four trials. In the learning phase, participants decided how to split the money between an in-/out-group member and a confederate. The in-/out-group member either provided selfish feedback (immoral condition, desiring to split the money unfairly 90% of the time) or generous feedback (moral condition, desiring to split the money fairly 90% of the time)—a between-subjects design (Figure 1C). Participants played 40 trials in the learning phase. Finally, in the transfer phase, participants dictated for 10 trials again. At the end of the experiment, participants judged how fair the in-/out-group member they interacted with was on a scale from 0 (= *not very fair*) to 10 (= *very fair*; see Online Supplement).

Results

As in Experiment 1, we ran two mixed-effects logistic regressions. In the immoral condition, we modeled participants' decisions to behave selfishly (*unfair split* = 1), in contrast with choosing to behave more generously (*elsewise* = 0). Choice was modeled as a function of phase (baseline/transfer) and group membership (in-group/out-group) with participant modeled as a random intercept and phase as a random slope. For the moral condition, we used the same analysis pipeline but modeled participants' probability of choosing the generous split. Across experimental conditions, participants preferred to split the money fairly in the baseline phase, endorsing the fair split
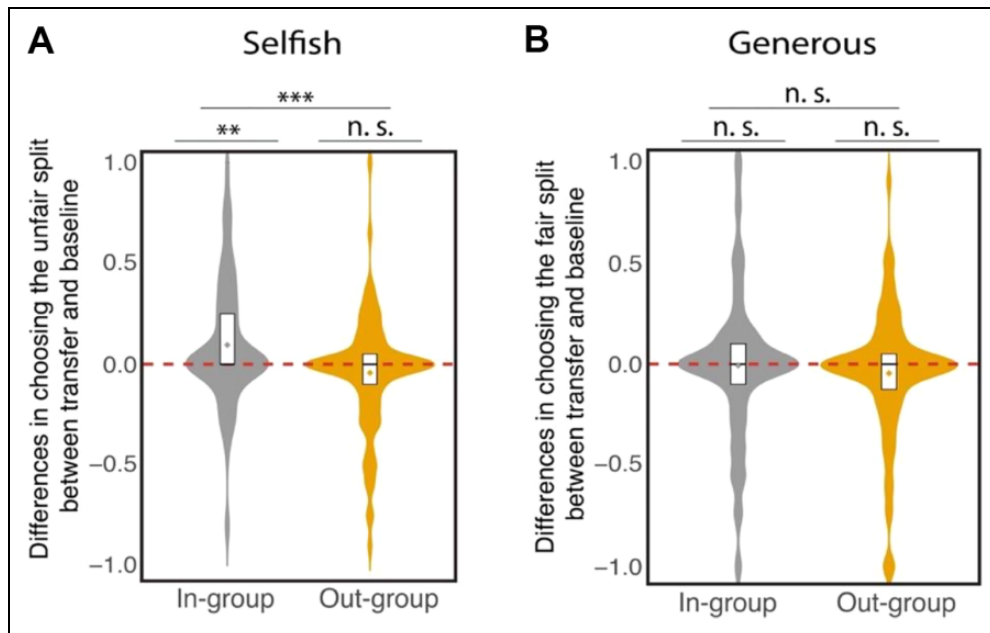
Figure 3. Results Experiment 2. *Note.* Violin plots depict the density of the distribution of the differences between transfer and baseline in choosing the unfair or fair split. The point within the boxplot depicts the mean. 0 on the *y*-axis indicates no change in behavior pre- versus postexposure to the "teacher." (A) Participants became more unfair after observing an unfair in-group member, but not an unfair out-group member. (B) Participants did not modify their fairness after observing a generous in- or out-group member. **$p < .01$. **$p < .001$.

47.11% of the time (mildly unfair splits $= 19.53\%$ and unfair splits $= 33.29\%$).

We conceptually replicate our findings from Experiment 1. Participants became more selfish after observing a selfish individual as indexed by a significant main effect of phase, $\beta = 0.81 \pm 0.23$, 95% CI [0.36, 1.25], $p < .001$, and this effect was modulated by group membership, as denoted by a significant interaction with group membership, $\beta = -1.17 \pm 0.31$, 95% CI [−1.78, −0.56], $p < .001$. There was no main effect of group membership, $\beta = 0.66 \pm 0.43$, $p = .13$. In addition, we tested whether there were any differences in choosing the unfair split between transfer and baseline (0 = *no change between phases*). Results revealed that observing a selfish in-group did make participants significantly more selfish: $t(106) = 3.21$, $p = .002$; $M = 0.10$, CI [0.04, 0.15], while observing a selfish out-group did not exert any behavioral change: $t(96) = -1.53$, $p = .13$; $M = -0.04$, CI [−0.10, 0.01], Figure 3A.

As before, we observed no increase in choosing the generous option after observing another generous individual, and no influence of group membership nor its interaction with phase (phase: $\beta = 0.11 \pm 0.24$, $p = .64$; in-/out-group: $\beta = 0.13 \pm 0.37$, $p = .74$; interaction: $\beta = -0.29 \pm 0.35$, $p = .42$, Figure 3B).

Put simply, selfish behavior increased only after observing a selfish in-group member, not after observing a selfish out-group member—even though that meant foregoing a monetary loss. However, receiving a monetary windfall may have led to an endowment effect, such that participants felt it was optional (not normative) to generously split the money. This may explain why there was a transmission of selfish but not generous behaviors. In Experiment 3, we increased the severity of

the immoral action by modifying the paradigm such that the advantageous unequal split was associated with an even more immoral trait: dishonesty.

Experiment 2: Cheating Game

Honesty is considered one of the most important dimensions for evaluating a person's moral character (Anderson, 1968). Nevertheless, dishonesty is present in almost every society, and it escalates when others promote dishonest behavior (Gino & Galinsky, 2012; Weisel & Shalvi, 2015). In laboratory paradigms, a majority of participants are willing to lie about cheating (Mazar et al., 2008). This allows us to test the moral contagion hypothesis in a context in which participants behave more immorally than the previous experiments (where moral behavior was the norm). We hypothesized that cheating will only be transmitted by in-group members, while honesty will be more strongly transmitted by observing honest out-group members.
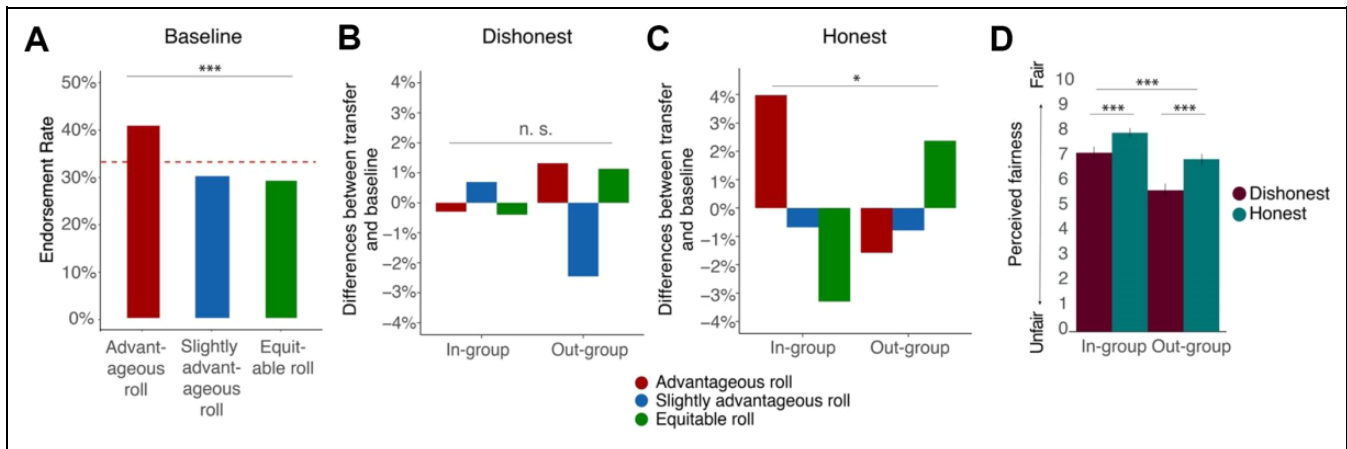
Method

Participants

While we recruited 646 participants, 237 were excluded following our exclusion criteria. This led to a final sample of 409 participants (32% female, average age $= 32.53$ $SD = 9.66$).

Procedure

After being assigned to their team, participants played the cheating game (Shalvi et al., 2011). In this cheating game,

Figure 4. Results Experiment 3. *Note.* (A) Baseline cheating behavior is collapsed across the four experimental conditions. Participants cheated relatively often, selecting the advantageous option (red bar) more than the slightingly advantageous and equitable options. (B) Group differences between transfer and baseline phases after the observation of a cheater reveal no significant effect of group. (C) Group differences between baseline and transfer phases after observing the honest teacher. We observed a significant interaction such that participants became more honest after observing an honest out-group member and more dishonest after observing an honest in-group member. (D) Judgments of perceived fairness: In-group members are rated as more fair than out-group members regardless of whether they cheated. *$p < .05$. ***$p < .001$.

participants were asked to roll a die from a website. Each side on the die accords with a different payout combination for participants and a confederate. To parallel the outcomes in the dictator game used in Experiment 2, payout combinations from the die equated to the following: (1) An unequal advantageous split if a one or two was rolled (participant: US\$1, confederate: US\$0); (2) a mildly advantageous split if a three or four was rolled (participant: US\$.75, confederate: US\$.25); or (3) an equal split if a five or six was rolled (participant: US\$.50, confederate: US\$.50; see Figure 1B). Participants played in the baseline phase for 10 trials. During the learning phase, participants made decisions on behalf of an in-/out-group member 48 times (Figure 1C). Two types of in-/out-group members were generated for the learning phase, an immoral member who reported to the participant that they rolled a one or two 66% of the time (in which case, dishonesty is coupled with unfair selfish behavior), and a moral member who reported rolling an equal number of times all sides of the die. Finally, in the transfer phase, participants played 10 trials as Player A again. After this, participants judged how fair the in-/out-group member they interacted with was on a scale from 0 (= *not very fair*) to 10 (= *very fair*).

Results

Cheating behavior can be evaluated by contrasting the aggregate frequency of observed outcomes with the aggregate frequency that should be expected by chance (i.e., cheating at the population level). Because the analysis is conducted at the group level, we first ensured that behavior in the baseline phase was similar across the four experimental conditions (moral in/out-group and immoral in/out-group). A χ^2 test with the reported outcomes revealed no statistical differences between

groups: $\chi^2(6, N = 409) = 4.19, p = .66$. We therefore collapsed the responses across groups to analyze whether the reported outcomes deviated from the frequencies that should have been observed had participants reported the true roll of the dice (i.e., 33% chance of getting any of the two sides). Results revealed that reported outcomes deviated from chance, such that participants reported rolling a one or two (the advantageous split) 41% of the time, $\chi^2(2) = 104.17, p < .001$; Figure 4A.

Because an honest player would report rolling any of the numbers on the die with an equal frequency, our analysis focused on detecting changes in the probability to endorse any of the three options after the learning phase. Accordingly, we conducted two mixed-effect ordinal regressions (separate regressions for the (im)moral conditions) with participants coded as random intercepts and phase entered as a random slope. The dependent variable was coded as *unequal advantageous split* (= 1), *mildly advantageous split* (= 2), and *the equal split* (= 3).

In contrast with previous work (Gino et al., 2009), participants did not behave more immorally after observing an immoral individual, regardless of the individual's group membership (phase: $\beta = -0.005 \pm 0.05, p = .92$; in/out-group: $\beta = -0.05 \pm 0.10, p = .64$, interaction: $\beta = -0.008 \pm 0.08, p = .92$; Figure 4B). However, in contrast with results in Experiments 1 and 2, we found a significant effect of observing an honest individual: Participants became more dishonest in the transfer phase relative to the baseline phase, $\beta = -0.13 \pm 0.07, 95\%$ CI $[-0.26, -0.0006], p = .04$. However, this was qualified by an interaction with group membership: Participants became more dishonest after witnessing an honest in-group member, while they engaged in more honest behavior after witnessing an honest out-group member (as indexed by a significant interaction: $\beta = 0.17 \pm 0.09, 95\%$ CI $[0.01,

Table 1. Summary of Results.

Experiments	Condition			
	Moral In-Group	Moral Out-Group	Immoral In-Group	Immoral Out-Group
Experiment 1	No effect	No effect	Increases immorality	No effect
Experiment 2	No effect	No effect	Increases immorality	No effect
Experiment 3	Increases immorality	Increases morality	No effect	No effect

0.35], $p = .03$, Figure 4C). There was no significant main effect of group-membership: $\beta = -0.10 \pm 0.09$, $p = .29$.

In the cheating game, an individual who repeatedly rolled favorable numbers may be cheating or may be just lucky, an ambiguity that may exacerbate in-group biases. Indeed, this is what we found. An analysis of variance revealed that the honest individual was perceived as more fair than the dishonest one, $F(1, 405) = 18.28$, $p < .001$, partial $\eta^2 = 0.04$; Figure 4D. However, group membership explained more variance than the moral condition did: $F(1, 405) = 28.65$, $p < .001$, partial $\eta^2 = 0.07$, suggesting that group membership had an increased role in a situation that allowed plausible deniability of any cheating behavior. Indeed, an honest out-group member was judged on par with a dishonest in-group member (Figure 4D). The interaction between group membership and morality was not significant: $F(1, 405) = 0.88$, $p = .34$.

General Discussion

Choosing compassion over punishment, generosity over selfishness, and honesty over dishonesty is the by-product of many factors, including virtue-signaling, norm compliance, and self-interest. There are times, however, when moral choices are shaped by the mere observation of what others do in the same situation (Gino & Galinsky, 2012; Nook et al., 2016). Here, we investigated how moral decisions are shaped by one's in- or out-group—a factor known to shift willingness to conform (Gino et al., 2009). Conceptually replicating past research (Gino et al., 2009), results reveal that immoral behaviors were only transmitted by the in-group: While participants became more punitive or selfish after observing a punitive or selfish in-group, they did not increase their immoral behavior after observing an immoral out-group (Experiments 1 and 2). However, when the same manipulation was deployed in a context where the immoral acts could not be traced, neither the dishonest in- nor out-group member produced any behavioral shifts in our subjects (Experiment 3). These results suggest that immoral behaviors are not transmitted equally by all individuals. Rather, they are more likely to be transmitted within groups than between groups. In contrast, prosocial behaviors were rarely transmitted by either group. Participants did not become more compassionate or generous after observing a compassionate or generous in- or out-group member (Experiments 1 and 2). We only find modifications for prosocial behavior when participants observe another participant behaving in a costly honest manner, and this was modulated by group membership. Witnessing an honest out-group member attenuated the degree

to which participants themselves cheated compared to participants who witnessed an honest in-group member (see Table 1 for a summary of results). Together, these findings suggest that the transmission of moral corruption is both determined by group membership and is sensitive to the degree of moral transgression. Namely, given the findings from Experiment 3, in-groups appear to license moral corruption, while virtuous out-groups can buffer against it.

The main hypothesis tested was that intergroup competition might lead to a greater increase in moral behavior after observing a moral out-group compared to a moral in-group. By structuring our experiments so that subjects could observe either in- or out-group members behaving morally or immorally, we examined whether the naturally occurring zero-sum structure of our two-party political system leads people to act more moral in the presence of a highly moral out-group member. This hypothesis was confirmed when participants observed an honest out-group but not when the out-group behaved compassionately or fairly. Although there are a few possible reasons for the observed results, the fact that across the three experiments we found behavioral differences during the baseline phase (prior to any group manipulation; see Online Supplement for baseline and transfer phase data plotted separately) may be the most straightforward explanation. In the first two experiments, participants endorsed the moral option more often than the other two, less moral options. For instance, compensation was chosen 54% of the time in Experiment 1 and the fair split was chosen 47% in Experiment 2. This was not the case in Experiment 3, where we observed a more omnipresent pattern of cheating (i.e., people overly reported the roll that was most favorable to them). That participants were already behaving fairly moral in the first two experiments likely limited the possibility of observing conformity effects in the moral condition, especially considering that moral behavior is transmitted less easily than immoral behavior (Jordan et al., 2013). Overall, our results suggest that establishing a baseline of how people behave prior to being influenced by others is important for properly identifying and quantifying behavioral changes without overestimating conformity effects.

A natural next question is why participants in Experiment 3 showed the greatest rates of immoral behavior during the baseline phase when it is arguably the most egregious moral transgression in the three experiments? One unique feature of the final experiment that sets it apart from the others is the opportunity to deny any wrongdoing. At the individual level, when people report high levels of advantageous die rolls, it is impossible to disentangle whether this was attributed to dishonest

greed or random luck. Contexts that allow people to deny any wrongdoing have been shown to boost immorality (Dana et al., 2007; De Kwaadsteniet et al., 2007). It is likely that the necessary moral wiggle room was sufficient for individuals to exhibit a dishonest behavioral pattern prior to any social influence, and thus why we observed people failing to report the true averages of their die during the baseline phase. Unlike Experiments 1 and 2, this pattern of immoral behavior gave participants the space to exhibit more honest behaviors after the group manipulation. Together, these findings provide evidence that participants might have engaged in moral intergroup competition so that they could either feel or appear morally superior to the outgroup. These results should be interpreted with caution, however, as this interaction effect was also driven by an increase in dishonesty after observing an honest in-group. Although this dovetails with vicarious moral licensing (Kouchaki, 2011), we were not expecting participants to alter their behavior in this way.

Further caution needs to be added regarding the generalizability of our results to non-virtual experiments. Nevertheless, we conceptually replicated previous findings that have been corroborated in the laboratory as well. It is also unclear to what degree our results generalize to other types of in-/out-groups. In the research reported here, we recruited groups that have large discrepancies in their moral worldviews (Zhong et al., 2008). This is known to make the out-group appear especially threatening, and therefore, more likely to generate moral disdain (Parker & Janoff-Bulman, 2013). As research shows that political group dynamics (i.e., specifically overly negative meta-perceptions) replicate across 26 countries and extend to nonpolitical intergroup contexts (Ruggeri et al., 2020), there is good reason to believe these results would extend to most group contexts where the "other side" is moralized.

Finally, we found a strong dissociation between the influence of an honest out-group and participants' judgments of their honest behavior. Even though participants became more honest after observing an honest out-group member, they failed to positively judge such moral virtuosity. Instead, an outgroup member's honesty was perceived as on par with a dishonest in-group member, conceptually replicating the general phenomenon that people perceive in-group behavior as more indicative of morality than the same behaviors from an out-group member (Goldring & Heiphetz, 2020). This parallels the daily moral contortions observed in the news and on social media, where people are quick to condone immoralities committed by their own political party but easily overlook the moral acts endorsed by the opposing party.

As we have noted in our prior work (FeldmanHall et al., 2018), the old adage "we are like chameleons, we take our hue and the color of our moral character from those who are around us" contains some truth (Locke, 1824, cited in Bargh & Chartrand, 1999). However, according to the findings here, our moral character also changes color depending on who is around us. While our moral hue swiftly turns antisocial when

surrounded by immoral in-groups, it is steadier—perhaps even bending towards virtuosity—when those around us are virtuous out-groups.

ORCID iD

Oriel FeldmanHall ⓘ https://orcid.org/0000-0002-0726-3861

Supplemental Material

The supplemental material is available in the online version of the article.

References

Abrams, D., Wetherell, M., Cochrane, S., Hogg, M. A., & Turner, J. C. (1990). Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization. *British Journal of Social Psychology*. https://doi.org/10.1111/j.2044-8309.1990.tb00892.x

Anderson, N. H. (1968). Likableness ratings of 555 personality-traits words. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/h0025907

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*. https://doi.org/10.1037//0022-3514.83.6.1423

Aquino, K., Reed, A., Thau, S., & Freeman, D. (2007). A grotesque and dark beauty: How moral identity and mechanisms of moral disengagement influence cognitive and emotional reactions to war. *Journal of Experimental Social Psychology*. https://doi.org/10.1016/j.jesp.2006.05.013

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462–479.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bicchieri, C. (2005). The grammar of society: The nature and dynamics of social norms. In *The grammar of society: The nature and dynamics of social norms*. https://doi.org/10.1017/CBO9780511616037

Bowles, S., & Gintis, H. (1998). The moral economy of communities: Structured populations and the evolution of pro-social norms. *Evolution and Human Behavior*. https://doi.org/10.1016/S1090-5138(98)00015-4

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*. https://doi.org/10.5334/joc.10

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Russell Sage Foundation.

Cikara, M., Jenkins, A. C., Dufour, N., & Saxe, R. (2014). Reduced self-referential neural response during intergroup competition predicts competitor harm. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2014.03.080

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*. https://doi.org/10.1007/s00199-006-0153-z

De Kwaadsteniet, E. W., Van Dijk, E., Wit, A., De Cremer, D., & De Rooij, M. (2007). Justifying decisions in social dilemmas: Justification pressures and tacit coordination under environmental uncertainty. *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/0146167207307490

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*. https://doi.org/10.1007/s10683-011-9283-7

Everett, J. A. C., Ingbretsen, Z., Cushman, F., & Cikara, M. (2017). Deliberation erodes cooperative behavior—Even towards competitive out-groups, even when using a control condition, and even when eliminating selection bias. *Journal of Experimental Social Psychology*. https://doi.org/10.1016/j.jesp.2017.06.014

FeldmanHall, O., Otto, A. R., & Phelps, E. A. (2018). Learning moral values: Another's desire to punish enhances one's own punitive behavior. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000405

FeldmanHall, O., Sokol-Hessner, P., Van Bavel, J. J., & Phelps, E. A. (2014). Fairness violations elicit greater punishment on behalf of another than for oneself. *Nature Communications*. https://doi.org/10.1038/ncomms6306

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., & Rand, D. G. (2020). Political sectarianism in America. *Science*, *370*(6516), 533–536.

Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.2009.02306.x

Gino, F., & Galinsky, A. D. (2012). Vicarious dishonesty: When psychological closeness creates distance from one's moral compass. *Organizational Behavior and Human Decision Processes*. https://doi.org/10.1016/j.obhdp.2012.03.011

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2005). Moral sentiments and material interests: The foundations of cooperation in economic life. *Economic Learning and Social Evolution*. https://doi.org/10.1525/aa.2007.109.2.380

Goldring, M. R., & Heiphetz, L. (2020). Sensitivity to ingroup and outgroup norms in the association between commonality and morality. *Journal of Experimental Social Psychology*. https://doi.org/10.1016/j.jesp.2020.104025

Heffner, J., & FeldmanHall, O. (2019). Why we don't always punish: Preferences for non-punitive responses to moral violations. *Scientific Reports*. https://doi.org/10.1038/s41598-019-49680-2

Jordan, J. J., Mullen, E, & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/0146167211400208

Jordan, J. J., Rand, D. G., Arbesman, S., Fowler, J. H., & Christakis, N. A. (2013). Contagion of cooperation in static and fluid social networks. *PLoS One*. https://doi.org/10.1371/journal.pone.0066199

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *The Journal of Business*, *59*(S4), S285. https://doi.org/10.1086/296367

Kouchaki, M. (2011). Vicarious moral licensing: The influence of others' past moral actions on moral behavior. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/a0024552

Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*. https://doi.org/10.1002/ejsp.2023

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*. https://doi.org/10.1509/jmkr.45.6.633

Nook, E. C., Ong, D. C., Morelli, S. A., Mitchell, J. P., & Zaki, J. (2016). Prosocial conformity: Prosocial norms generalize across behavior and empathy. *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/0146167216649932

Parker, M. T., & Janoff-Bulman, R. (2013). Lessons from morality-based social identity: The power of outgroup "hate," not just ingroup "love." *Social Justice Research*. https://doi.org/10.1007/s11211-012-0175-6

Ruggeri, K., Većkalov, B., Bojanić, L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., Arroyo, P. B., Berge, M., Bjørndal, L., Bursalıoğlu, A., Bühler, V., Čadek, M., Çetinçelik, M., Clay, G., Cortijos-Bernabeu, A., Damnjanović, K., Dugue, T., Esberg, M., Serna, C. E., . . . Folke, T. (2020). *The general fault in our fault lines*. https://doi.org/https://doi.org/10.31219/osf.io/xvksa

Shalvi, S., Dana, J., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*. https://doi.org/10.1016/j.obhdp.2011.02.001

Shalvi, S., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). People avoid situations that enable them to deceive others. *Journal of Experimental Social Psychology*. https://doi.org/10.1016/j.jesp.2011.04.015

Stein, R. (2017). Trumping conformity: Urges towards conformity to ingroups and nonconformity to morally opposed outgroups. *Journal of Experimental Social Psychology*. https://doi.org/10.1016/j.jesp.2016.12.007

Turner, J. C. (1975). Social comparison and social identity: Some prospects for intergroup behaviour. *European Journal of Social Psychology*. https://doi.org/10.1002/ejsp.2420050102

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1423035112

Zhong, C. B., Phillips, K. W., Leonardelli, G. J., & Galinsky, A. D. (2008). Negational categorization and intergroup behavior. *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/0146167208315457

Author Biographies

Marc-Lluís Vives is a postdoctoral researcher at Brown University. He obtained his doctoral degree at Pompeu Fabra University and studies decision-making.

Mina Cikara directs the Harvard Intergroup Neuroscience Lab. She uses social psychological and cognitive neuroscience approaches to study how group membership and prejudice change the course of social cognition.

Oriel FeldmanHall directs the FeldmanHall Lab at Brown University. She studies social behavior using a multdisciplinary approach that merges economics, psychology and neuroscience.

Handling Editor: Eva Walther