Personality and Social
Psychology Bulletin
2022, Vol. 48(4) 596–605

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/01461672211013272
journals.sagepub.com/home/pspb

Moderators of the Liking Bias in Judgments of Moral Character

Konrad Bocian[1,2], Wieslaw Baryla[2], and Bogdan Wojciszke[2]

Abstract

Previous research found evidence for a liking bias in moral character judgments because judgments of liked people are higher than those of disliked or neutral ones. This article sought conditions moderating this effect. In Study 1 (N = 792), the impact of the liking bias on moral character judgments was strongly attenuated when participants were educated that attitudes bias moral judgments. In Study 2 (N = 376), the influence of liking on moral character attributions was eliminated when participants were accountable for the justification of their moral judgments. Overall, these results suggest that although liking biases moral character attributions, this bias might be reduced or eliminated when deeper information processing is required to generate judgments of others' moral character.

Keywords

moral judgments, moral character, attitudes, liking bias, accountability

Received August 19, 2020; revision accepted March 26, 2021

> Human judgments—even very bad ones—do not smell.
>
> —Wilson and Brekke (1994, p. 121)

Imagine that someone whom you strongly like did something wrong. Would you judge this person's behavior and moral character fairly? Affective disposition theory (Raney, 2004; Zillmann & Cantor, 1977) suggests that people like characters because their behaviors are perceived as good and moral and judge them as moral because they like them. Research confirms these suggestions by demonstrating that, on the one hand, morality is the most critical factor to liking (Hartley et al., 2016), whereas, on the other hand, liking has a profound impact on morality judgments (Bocian et al., 2018; Bocian & Myslinska-Szarek, 2021; Grizzard et al., 2020; Melnikoff & Bailey, 2018). In this article, we attempt to unpack the moderating process involved in liking influences on morality judgments. Specifically, we aim to examine the extent to which the influence of liking on moral character judgments could be moderated by two factors: education and accountability.

A large body of evidence suggests that the perception of others' moral character traits dominates impression development (Brambilla et al., 2021). Consequently, whether people perceive others as moral or not influence their willingness to help them (Pagliaro et al., 2013), impact the intensity of interpersonal mimicry (Menegatti et al., 2020), and decide whether people would engage in behavioral synchrony (Brambilla et al., 2016). However, although research has shown that perceptions of moral character traits are profoundly biased by interpersonal attitudes (Bocian et al., 2018), we know surprisingly little about the psychological processes that could explain how liking and disliking impact morality judgments. In this article, we propose that influence of liking on moral character judgments can be limited when deeper information processing is required.

Interpersonal Attitudes Bias Moral Cognition

The influence of interpersonal attitudes on moral judgments has been shown for the first time in a field experiment conducted by Bocian and Wojciszke (2014a). Researchers recruited students who did or did not have to pay a fine for overdue books; for half of the students, the librarian arbitrarily waived the fine, but for another half, she did not. The librarian's decision was judged as more moral when she broke the university rules and helped students save money than when she enforced the fine. More importantly, students' favorable moral judgments of the librarian were explained by a surge in liking toward her (Bocian & Wojciszke, 2014a). Corroborating these results, other research has also found that moral traits increase liking when morality is advantageous for a perceiver's goals, but that when immorality is goal conducive, the preference for moral traits is eliminated

[1]University of Kent, Canterbury, UK
[2]SWPS University of Social Sciences and Humanities, Sopot, Poland

Corresponding Author:
Konrad Bocian, School of Psychology, Keynes College, University of Kent, Canterbury CT2 7NZ, UK.
Email: K.Bocian-660@kent.ac.uk

or reduced (Melnikoff & Bailey, 2018). In fact, one study has found that judgments of a character's morality and likability are so profoundly tied that even orthogonal manipulation of both these factors cannot suppress the relationship between them (Grizzard et al., 2020).

Developmental studies have demonstrated that liking influences moral cognition in the early stages of social life as well. Similarity and dissimilarity to others affect infants' perceptions of harm (Hamlin et al., 2013), whereas preschoolers attribute more guilt to characters they do not like (Dumhan & Emory, 2014). Moreover, a recent study has demonstrated that young children like individuals who harm antisocial characters (vs. prosocial or neutral) and, therefore, judge their behavior as less bad (Bocian & Myslinska-Szarek, 2021). Finally, although young children display a strong aversion toward antisocial individuals, research has shown that beneficial cooperation (vs. non-beneficial) with an antisocial partner increases children's liking and preference for the antisocial partner (Myslinska-Szarek et al., 2021).

The evidence presented above suggests that interpersonal attitudes strongly influence judgments of others' behavior and moral character, although they should not and although people believe in the objectivity of their moral beliefs (Goodwin & Darley, 2008). However, only one line of research has directly investigated whether liking distorts moral cognition. Specifically, in one study, using the classic chameleon effect (see Chartrand & Bargh, 1999), participants' facial expressions were or were not mimicked by a confederate. The confederate was liked and considered more moral after mimicking the participant than when the same confederate did not copy the participant's facial expressions. In a different study, participants who believed that another person had the same (vs. different) political beliefs liked this person more and judged this person's moral character more favorably. Overall, in four experiments, the same pattern of liking influencing moral character judgments was found using three different liking induction methods (belief similarity, mimicking, and mere exposure). This suggests that subjective and inevitable interpersonal preferences (e.g., liking) strongly influence perceptions of moral character—a phenomenon described as the mere liking effect in moral cognition (Bocian et al., 2018).

Discovery of the mere liking effect, which in this article we call the liking bias, is important because perceptions of moral character dominate impression formation (Goodwin et al., 2014; Wojciszke et al., 1998) and have serious social consequences: they shape first impressions and perceived suitability for different social roles, as well as influence trust in social interactions (Everett et al., 2016). Although we have strong evidence indicating that interpersonal attitudes are a source of bias in moral cognition, the conditions reducing this bias remain unclear. We believe that the dual-process models of social cognition offer valuable suggestions toward strategies that might reduce the liking bias.

Dual-process Models in Social and Moral Cognition

Dual-process models distinguish two different modes of information processing: automatic and controlled. While the first process is fast, effortless, and unconscious, the second process is slow, analytical, and might be recruited when needed (Greene et al., 2008). Research has demonstrated that automatic processing contributes to errors in decision-making (Epley et al., 2004; Gilovich & Savitsky, 1999), social judgments (Kruger, 1999; Van Boven et al., 2000), justice judgments (Messick & Sentis, 1979; Thompson & Loewenstein, 1992), and moral judgments (Cushman et al., 2010).

It has been argued that, in social cognition, these errors are mostly produced by egocentrism, which is an automatic perspective taken in social judgment. This is because people experience the world directly, in a fast and effortless manner, whereas taking others' perspectives requires effort, cognitive resources, and motivation (Epley & Caruso, 2004; Moore & Loewenstein, 2004). In moral cognition, errors are mostly produced by intuitive and automatic processes, which are usually affect-laden (Haidt, 2007). In other words, people's moral judgments often resemble instant perceptions rather than deliberate inferences, and the effect of these perceptions on moral judgment is often mediated through affective experience.

Similar to other kinds of evaluations, moral judgments are frequently based on evaluative feelings of good–bad or like–dislike about an action or a person (Bocian et al., 2018; Haidt, 2007). For example, research has demonstrated that moral judgments can emerge instantly (in a quarter of a second—Decety & Cacioppo, 2012) or that people need approximately 250 ms to decide whether something is right or wrong (Van Berkum et al., 2009). This evidence confirms that moral judgments, like any other judgments, can be generated automatically (Bargh, 1994), and therefore suggests that the automatic side of social and moral judgments makes them prone to different biases. In fact, theories of judgment and decision-making suggest that automatically activated associations (i.e., misleading intuitions) are among the two general sources of bias (Morewedge & Kahneman, 2010).

Can We Reduce the Impact of Liking on Attributions of Moral Character?

Helping people to debias their moral judgments that are contaminated by impressions, feelings, and attitudes might be challenging because people do not have access to them or control over them (Nisbett & Wilson, 1977; Wilson et al., 2002). Nevertheless, the Wilson and Brekke (1994) model of contamination describes how people may protect their minds from unwanted influences. For example, to avoid biased moral judgments, people must be motivated to correct them.

One way of doing so could be to help them detect a potential source of bias because people have poor access to the processes by which their own judgments are formed (Nisbett & Wilson, 1977). For example, past research has demonstrated that people are unaware that their judgments are biased by automatic and egocentric interpretations (Bocian & Wojciszke, 2014b; Wojciszke & Bocian, 2018) and, in consequence, act upon them, such as by trusting a cheater whom they like (Bocian et al., 2016).

Although some studies have found that forewarning people about specific biases (e.g., the halo effect) has no effect (Wetzel et al., 1981), different studies have demonstrated that forewarning can be effective in eliminating judgment biases (Schul, 1993). In addition, research has demonstrated that making people aware of their current mood canceled the mood-as-information effect (Messner & Wänke, 2011), and educating people about cognitive biases led them to more rational clinical decision-making (Hershberger et al., 1997). Different studies have shown that asking participants to avoid potential bias in a particular social category had a small but reliable impact on reducing bias in that category (Axt et al., 2018). Therefore, we tested whether educating people that interpersonal attitudes bias moral judgments would facilitate mental decontamination of their judgments.

As people are concerned about maintaining their identities as moral people (Aquino & Reed, 2002) and are also concerned whether they would generate persuasive justifications for their judgments (Kuhn, 1992; Shafir et al., 1993), accountability may be yet another condition debiasing moral judgments. Accountability refers to the situation when we expect to be called upon to justify our beliefs, feelings, or actions to others. Thus, accountability might make judgments more straightforward, rational, and relatively free of bias because people switch to more effortful and self-critical information processing when faced with the necessity to justify their judgments or decisions to others (Lerner & Tetlock, 1999).

Research has shown that increased cognitive effort among accountable participants decreases susceptibility to biases, such as the fundamental attribution error (Tetlock, 1985), overconfidence (Siegel-Jacobs & Yates, 1996), and oversensitivity to the order in which information appears (Schadewald & Limberg, 1992). Furthermore, when accountability is introduced, participants shift their decisions in a less self-serving direction, showing greater activation in the orbitofrontal cortex (Hughes & Beer, 2012). This activation of the orbitofrontal cortex suggests that deeper processing of desirable and undesirable information may result in positivity bias reduction (Hughes & Zaki, 2015). Therefore, it is possible that being accountable for moral judgments directs people into deeper information processing. On this account, we examined whether expectations of being called upon to justify one's moral judgments would reduce the influence of interpersonal attitudes on moral character attributions.

Overview of These Studies

The aim of this study was twofold. First, we investigated whether manipulation of liking would influence attributions of moral character. We hypothesized that participants would like the target person more when the target person would display similar (vs. dissimilar) sociopolitical views than participants or when their facial expressions would be mimicked (vs. not mimicked). In turn, in the control conditions, we expected to observe more favorable judgments of moral character for the target who had the same sociopolitical views than participants (vs. dissimilar) or for the target who mimicked the participant's facial expressions (vs. not mimicked). This hypothesis's rationale is based on past research, which has shown that liking influences moral character perceptions independently of how liking was created (Bocian et al., 2018).

Second, we tested whether liking influences on moral character attributions would be reduced by two different moderators related to deeper information processing: (a) education, and (b) accountability. Specifically, we hypothesized that because both moderators could incite more reflective and less intuitive information processing, they would attenuate—if not eliminate—the impact of liking on moral character judgments. We tested our predictions in two studies.

In Study 1 (preregistered), we convinced participants that the target person has similar or dissimilar sociopolitical views to their own, and then we measured how much participants liked the target. Later, we educated (or not) participants that liking can bias moral judgments and asked them to ignore earlier induced liking in their judgments of the target's moral character. We examined whether cueing participants about liking as a source of bias (vs. not) would moderate the influence of liking on moral character judgments.

In Study 2, participants' facial expressions were or were not mimicked by the target person. Next, half of the participants were informed that the experimenter would interview them at the end of the study and ask them to explain their moral judgments. The other half of the participants were informed that all of their answers would be anonymous and confidential. Furthermore, participants indicated how much they liked the target person and judged the target's person moral character. We investigated whether being accountable (vs. not) would moderate the effect of liking, induced by mimicry, on the target's moral character judgments.

Study 1

In Study 1, we aimed to investigate whether educating participants that liking biases moral character judgments would reduce the impact of liking on moral character attributions. First, we convinced participants that the target person had the same or opposite sociopolitical views. Second, we explained to participants (or not in the control condition) how the similarity and dissimilarity of sociopolitical views might impact their attitudes. Then, we asked them to ignore how they felt about the target person while judging the target's moral character. Based on previous evidence (Bocian et al., 2018), we assumed that similarity (vs. dissimilarity) would result in higher (vs. lower) liking scores of the target and more favorable (vs. less favorable) judgments of the target's moral character. We also predicted that this main effect would be reduced among the educated participants.

In this article, we report all measures, all manipulations, and any data exclusions. Any additional measures not included in the main analyses are reported in online supplemental material. All studies have been approved by the relevant research ethics committee. This study was preregistered at aspredicted.org/blind.php?x=ug7cg5.

Method

Participants and procedure. We used G*Power (Faul et al., 2007) and effects found in previous studies (see Bocian et al., 2018, Studies 1 and 2) to calculate the sample for the main effect. This analysis yielded a total sample size of 54 participants (27 per condition). Because we expected at least a 50% attenuation of the main effect, following recommendations proposed by Roger Giner-Sorolla (2018), we estimated that our sample size, with a power of .95, should be 756. In the end, we recruited 800 U.K. participants (200 per condition) using Prolific Academic to participate in an online study about social cognition. Eight participants were excluded from the data analysis because they answered incorrectly on the screening question.[1] Therefore, we analyzed data from 792 British participants (537 women, M_{age} = 41.14 years, *SD* = 10.82). Based on a sensitivity power analysis, this sample size provides .80 power to detect an effect size of $f^2 = 0.10$.

For the impression manipulation, we used the method proposed by Bocian et al. (2018, Study 1) by adjusting the sociopolitical issues to the U.K. landscape (see online supplemental material for the pilot study results). First, we asked participants to answer whether they agree or disagree with eight different sociopolitical statements (e.g., "Great Britain should leave the EU as soon as possible, even without a Brexit deal," "Human activity is responsible for climate change"). Afterward, we told participants that a special algorithm would draw another participant's questionnaire (the target person). Based on the random manipulation, participants saw either that the other participant's six answers were the same as their own answers (the similar views condition) or that the six answers were the opposite (the dissimilar views condition).

Next, participants indicated how much they liked the target person. Furthermore, in the education condition, we told participants that previous studies have shown that similarity and dissimilarity of beliefs shape whether people like or dislike someone else and that people often judge other people they like as being moral and people they do not like as being immoral. Therefore, because we wanted to avoid this bias in our study, participants should try to make their decisions about the target's moral character independently from their positive or negative attitude. In the control condition, this information was omitted. Next, participants judged the moral character of the target person.

Measures

Attitude judgments of the target person were measured with two items: "I like this person" and "I would like to meet this person in the future." Participants indicated to what extent they agree with each of the statements using a 7-point scale ranging from 1 (*definitely not*) to 7 (*definitely yes*; $\alpha = .74$, $M = 3.75$, $SD = 1.36$).

Moral character judgments of the target person were measured with eight scenarios presenting various unethical behaviors (Dubois et al., 2015), for example, "Acting against the company policy" or "Engaging in software piracy" (see online supplemental material for a full description of the scenarios). Participants judged how likely the target person would engage in each of the eight behaviors, using a 7-point scale ranging from 0 (*not at all likely*) to 6 (*highly likely*; $\alpha = .79$, $M = 2.62$, $SD = 0.89$).

Results

Attitude judgments. When sociopolitical views of the target were similar to participants' views, they liked the target ($M = 4.59$, $SD = 1.00$) and disliked the target when the target's sociopolitical views were dissimilar ($M = 2.92$, $SD = 1.13$), $F(1, 778) = 484.51$, $p < .001$, $\omega_p^2 = .38$, 95% confidence interval (CI) = [.33, .43]. Neither the education manipulation, $F(1, 778) = .44$, $p = .507$, $\omega_p^2 = -.00$, 95% CI = [.00, 1.00], nor the interaction, $F(1, 788) = .08$, $p = .783$, $\omega_p^2 = -.00$, 95% CI = [.00, 1.00], was significant.

Moral character judgments. As predicted, participants' moral character judgments were limited by the education manipulation, $F(1, 788) = 15.24$, $p < .001$, $\omega_p^2 = -.02$, 95% CI = [.00, .04] (see Figure 1). In the control condition, participants judged the target with similar sociopolitical views as more moral ($M = 3.00$, $SD = 0.88$) than the dissimilar one ($M = 2.17$, $SD = 0.91$), $t(398) = 9.26$, $p < .001$, $d = 0.93$, 95% CI = [0.72, 1.13]. However, this difference was 2.1 times smaller when participants were educated about the potential influence of attitudes on their moral judgments: similar sociopolitical views ($M = 2.80$, $SD = 0.69$) versus dissimilar sociopolitical views ($M = 2.43$, $SD = 0.85$), $t(379.853) = 4.63$, $p < .001$, $d = 0.47$, 95% CI = [0.27, 0.67].

Moderated mediation analysis. One limitation of the manipulation used in Study 1 is its relevance to moral judgments. Given that ideology is bound with morality (e.g., Graham et al., 2009), we might expect that political ideology manipulation used in Study 1 also could shape the judgments of

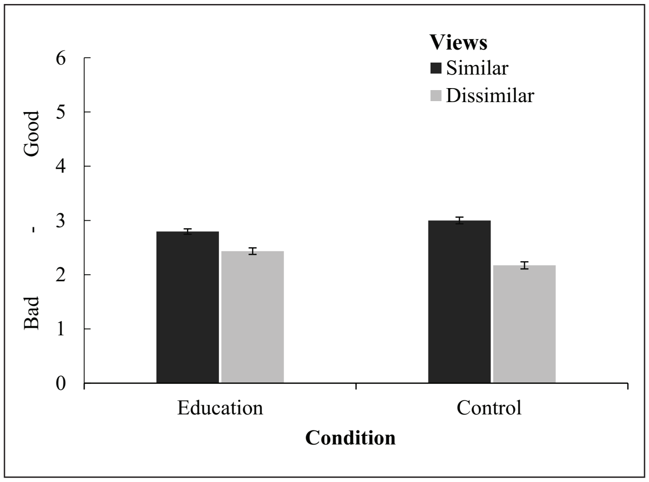

Figure 1. Mean moral character judgments as a function of education manipulation and sociopolitical views manipulation.
Note. Higher scores indicate a better moral character. Error bars represent standard error.

moral character. This issue could be addressed with a moderated mediation analysis in which the liking manipulation is entered as an independent variable, the liking judgments as a mediator, and the moral character judgments as a dependent variable, with the debiasing manipulation serving as a moderator of the link between the mediator and dependent variable. Models like this would display any direct, unmediated effect of the liking manipulation on the moral character judgments separately from the moderated indirect effect through liking.[2] Thus, we run a moderated mediation Model 15 in PROCESS macro proposed by Hayes (2013) because liking for the actor was measured before the moderator manipulation.

The indirect effect of the liking manipulation on the moral character was moderated by the education manipulation, $B = -0.16$, $SE = 0.05$, 95% CI = [−0.25, −0.07]. The indirect effect of the liking manipulation on the moral character, through measured liking, was significantly stronger in the control condition, $B = 0.24$, $SE = 0.04$, 95% CI = [0.17, 0.31], than in the education condition, $B = 0.08$, $SE = 0.03$, 95% CI = [0.02, 0.14]. The conditional direct effect of liking manipulation on the moral character was significant in the control condition, $B = 0.18$, $SE = 0.03$, 95% CI = [0.08, 0.27], but nonsignificant in the education condition, $B = 0.10$, $SE = 0.05$, 95% CI = [−0.00, 0.20]. Therefore, this result confirms that a significant part of the manipulation of liking on the moral character was mediated by the measured liking of the target person.

Discussion

The results of Study 1 provided support for our hypothesis that making people aware about the liking bias would reduce its impact on moral character judgments. As predicted, participants judged the target person's moral character more favorably when they learned that the target person had similar views as their own and less favorably when the target's views were dissimilar. However, the impact of liking on judgments of moral character was strongly attenuated when participants were educated about the potential influence of a positive or negative attitude on their moral character judgments.

In Study 1, education only attenuated the liking bias. Therefore, in Study 2, we sought to investigate conditions that would eliminate the influence of liking on moral character judgments. We assumed that accountability might debias moral character judgments because the necessity to justify judgments or decisions to others demands switching to more effortful and self-critical information processing (Lerner & Tetlock, 1999).

Study 2

In Study 2, we examined whether manipulation of accountability would eliminate the influence of liking on moral character judgments. We announced (or not) to participants that they would have to justify their judgments regarding the target person at the end of the experiment. Afterward, we induced (or not) a positive attitude toward the target by mimicking (or not) participants' facial expressions by the target. We predicted that a positive attitude would produce more favorable moral character evaluations of the target. We also assumed that the introduction of accountability would eliminate the impact of liking on moral character judgments.

Method

Participants and procedure. Because we planned to run the experiment in the laboratory, we estimated the target sample size to be $N = 30$, assuming a power of .80, two-tailed. We expected a 50% attenuation in the accountability condition, so we increased the sample size 14 times (as suggested by Giner-Sorolla, 2018), which resulted in a target of 420 participants. In the end, we managed to recruit 376 Polish participants from the university pooling sample (259 women; $M_{age} = 25.30$ years, $SD = 7.90$). Based on a sensitivity power analysis, this sample size provides .80 power to detect an effect size of $f^2 = 0.14$. An additional 23 participants were recruited but were excluded from the data analysis because they either guessed that the target's task was to induce a positive attitude or that the target's mimic behavior was previously recorded.[3]

In the accountability condition, we used the procedure introduced by Tetlock and Boettger (1989, Study 1). Before the experiment started, we informed the participants that a researcher would later conduct interviews with the participants to understand what type of information people use to form impressions of others. We told the participants that the interview would be recorded, so they would have to sign a specific consent. In the no accountability condition, the participants

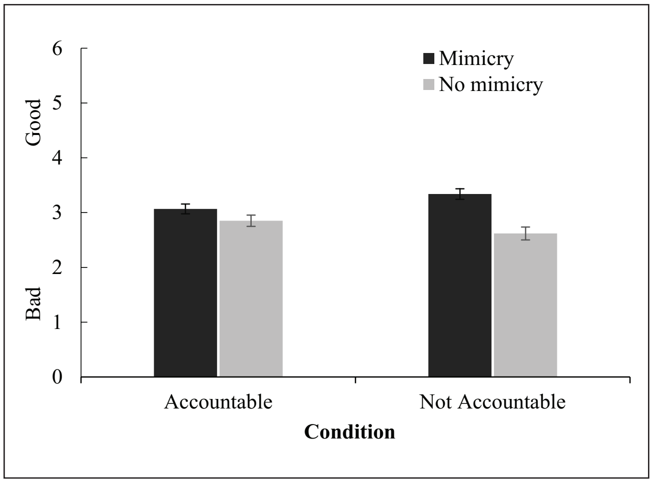

Figure 2. Mean moral character judgments as a function of accountability and mimicry manipulation.
Note. Higher scores indicate a better moral character. Error bars represent standard error.

were informed that their judgments and evaluations about others would remain entirely confidential and anonymous.

For the attitude manipulation, we used a computer-based method of positive attitude induction that involves mimicking participants' facial expressions. Past studies proved that this method is an effective manipulation, evoking a positive attitude toward the mimicking person (Kulesza et al., 2015). We told participants that they would participate in a live interaction with another person (the target) via video chat (in fact, the participants observed a prerecorded female person). The participants' task was to express facially different basic emotions (e.g., surprise, sadness, and happiness) to the person visible on the screen (the target). Furthermore, we told participants that the person visible on the screen would guess what emotion they expressed. We randomly allocated participants to one of two conditions. In the mimicry condition, the target was expressing the emotion in question, thereby creating the mimicry effect (e.g., when the participant was asked to show happiness, the target immediately smiled back). In the no-mimicry condition, the target person kept their face still, not expressing any emotions. Next, participants indicated how much they liked the target person and judged the target's moral character.

Measures

Attitude judgments of the target person were measured with seven items: "I like this person," "I would like to meet this person in the future," "I think we would quickly make good contact with this person," "I feel sympathy for this person," "I have the impression that this person would understand my feelings well," "This person makes me feel warm," and "I think this person is nice." Participants indicated to what extent they agree with each of the statements, using a 7-point scale ranging from 1 (*definitely not*) to 7 (*definitely yes*) (α = .94, M = 4.45, SD = 1.36).

Moral character judgments of the target person were measured with the same eight scenarios as used in Study 1 (α = .82, M = 3.00, SD = 1.02).

Results

Attitude judgments. When participants' facial expressions were mimicked, they liked the target person more (M = 5.34, SD = 0.90) than when they were not mimicked (M = 3.57, SD = 1.14), $F(1, 372) = 277.73$, $p < .001$, $\omega_p^2 = .42$, 95% CI = [.35, .50]. Neither the main effect of the accountability, $F(1, 372) = 3.11$, $p = .079$, $\omega_p^2 = .01$, 95% CI = [.00, .03], nor the interaction, $F(1, 372) = 1.12$, $p = .291$, $\omega_p^2 = .00$, 95% CI = [.00, .01], was significant.

Moral character judgments. Correspondingly, with our hypothesis, the mimicry impact on participants' moral character judgments was limited by the accountability manipulation, $F(1, 372) = 6.09$, $p = .014$, $\omega_p^2 = .01$, 95% CI = [.00, .05] (see Figure 2). In the no accountability condition, participants judged the target person as more moral after their facial expressions were mimicked (M = 3.34, SD = 0.98) and less moral when they were not mimicked (M = 2.62, SD = 1.12), $t(190) = 4.73$, $p < .001$, d = .62, 95% CI = [0.39, 0.97]. However, in the accountability condition, this effect was eliminated: moral character judgments between the mimicry and no-mimicry conditions did not differ significantly (M = 3.07, SD = 0.85 vs. M = 2.85, SD = 0.97), $t(182) = 1.60$, $p = .111$, d = .22, 95% CI = [−0.05, 0.53].

Moderated mediation analysis. Similar to Study 1, one could argue that mimicry manipulation used in Study 2 could directly influence perceptions of moral character. For example, we have evidence that facial and emotional mimicry facilities trust, prosocial behavior, and affective empathy (Duffy & Chartrand, 2017) and, therefore, may have shaped the way we perceive other people's moral character. Again, we used the PROCESS macro proposed by Hayes (2013), but we employed Model 59 because liking for the target was measured after the moderator manipulation.

The indirect effect of the liking manipulation on the moral character was moderated by the accountability manipulation, $B = -0.21$, $SE = 0.10$, 95% CI = [−0.41, −0.003]. The indirect effect of the liking manipulation on the moral character through measured liking was significantly stronger in the control condition, $B = 0.51$, $SE = 0.08$, 95% CI = [0.36, 0.68], than in the accountability condition, $B = 0.30$, $SE = 0.07$, 95% CI = [0.18, 0.45]. The conditional direct effect of the liking manipulation on the moral character was nonsignificant in the control condition, $B = -0.15$, $SE = 0.08$, 95% CI = [−0.32, 0.02], but significant in the accountability condition, $B = -0.20$, $SE = 0.08$, 95% CI = [−0.36, −0.03]. Again, this result demonstrates that the manipulation of liking on the moral character was mediated by the measured liking of the target person.

Discussion

Study 2 replicated the results of Study 1 by demonstrating the liking bias in moral character attributions. In line with our predictions, participants' judgments regarding the target's moral character depended on the mimicry manipulation. When participants' facial expressions were mimicked, they judged the target's moral character more favorably and less favorably when their facial expressions were not mimicked. However, this effect was only observed when participants were not obliged to justify their judgments. When participants were accountable for their moral evaluations, the impact of liking on moral character judgments was eliminated.

General Discussion

In this research, we sought to replicate the past results that demonstrated the influence of liking on moral character judgments, and we investigated conditions that could limit this influence. We demonstrated that liking elicited by similarity (Study 1) and mimicry (Study 2) biases the perceptions of another person's moral character. Thus, we corroborated previous findings by Bocian et al. (2018), who found that attitudes bias moral judgments. More importantly, we showed conditions that moderate the liking bias. Specifically, in Study 1, we found evidence that forewarning participants that liking can bias moral character judgments weaken the liking bias two times. In Study 2, we demonstrated that the liking bias was eliminated when we made participants accountable for their moral decisions.

By systematically examining the conditions that reduce the liking influences on moral character attributions, we built on and extended the past work in the area of moral cognition and biases reduction. First, while past studies have focused on the impact of accountability on the fundamental attribution error (Tetlock, 1985), overconfidence (Tetlock & Kim, 1987), or order of information (Schadewald & Limberg, 1992), we examined the effectiveness of accountability in debiasing moral judgments. Thus, we demonstrated that biased moral judgments could be effectively corrected when people are obliged to justify their judgments to others. Second, we showed that educating people that attitudes might bias their moral judgments, to some extent, effectively helped them debias their moral character judgments. We thus extended the past research on the effectiveness of forewarning people of biases in social judgment and decision-making (Axt et al., 2018; Hershberger et al., 1997) to biases in moral judgments.

Limitations, Implications, and Future Directions

We acknowledge that our work has some limitations that might warrant future research. Some evidence from studies on facial and emotional mimicry (Duffy & Chartrand, 2017) or ideology (Graham et al., 2009) suggests that the manipulations of liking used in the present studies are relevant to making moral judgments. To address this issue, we run separate moderated mediation analysis for each of our experiments. Overall, the indirect effects of the control conditions showed that moral character judgments were mediated by the liking toward the target. In other words, these results confirm, to some extent, that liking influences the perception of the target's moral character when people do not control this bias. Nevertheless, future studies could use different manipulations of liking that are devoid of any relevance to morality. For example, past studies have shown that people are attracted to others whose faces they have repeatedly seen before (Bocian et al., 2018) or whose surname shares the same letters as their own surname (Jones et al., 2004).

Furthermore, our education manipulation could be more persuasive by additionally making participants aware of the magnitude of liking bias and its consequences for future relationships (e.g., liking influences trustworthiness) as proposed by the Wilson and Brekke (1994) model of contamination. However, research has demonstrated that merely asking people to avoid bias in their judgments effectively induces a small reduction of that bias (Axt et al., 2018), and the results of Study 1 confirmed this effect. Nevertheless, our manipulation in Study 1 used a combination of both strategies, education and bias avoiding. Therefore, future studies could try to disentangle these strategies to investigate their effectiveness independently.

Another question that might be answered in future research regards the amount of deliberate and intuitive thinking about moral judgments. Past research has suggested that attitudes bias judgments of moral character and behavior because of intuitive processing (see Bocian et al., 2020 for review); however, these assumptions were not tested directly. The present research results suggest that the effectiveness of our debiasing factors might depend on the amount of cognitive effort people had to put into correcting their biased judgments. Therefore, future research using manipulations such as cognitive load, time pressure, or priming could establish to what extent liking influences on moral judgments are driven by deliberate and intuitive thinking. In fact, most of the research that has investigated the role of deliberate thinking on moral judgments has so far focused either on individual differences or on abilities in cognitive style (Landy & Royzman, 2018).

Finally, future research should answer the question of the mechanism underpinning the influence of liking on moral character judgments. We recognize that the present research did not directly explain how tested moderators reduced (Study 1) or eliminated (Studies 2 and 3) the influence of liking on moral character judgments. Instead, based on the premises of dual-process models in social (Moore & Loewenstein, 2004) and moral cognition (Cushman et al., 2010; Epley & Caruso, 2004), we assumed that judgments biased by liking might be corrected when people would have to put more cognitive effort into expressing them. However, as we do not have any direct data supporting this proposition (e.g., reaction times), these assumptions remain only hypothetical. Clearly, further research is needed on mediators of the liking bias.

Our work might contribute to understanding why people correct their biased moral beliefs only sporadically. We have ample evidence that factors such as personal (Bocian & Wojciszke, 2014a) and group interests (Bocian et al., 2021) or attitudes (Bocian et al., 2018) bias moral judgments because these factors trigger egocentric evaluations, which appear to people as objective, impartial, and morally right. People are not aware that egocentric evaluations bias their moral judgments because these evaluations are fast, do not require effort or resources to operate, and sometimes are strategically motivated by social and personal relationships (Bocian et al., 2020). For example, people protect close others by justifying (Weidman et al., 2020) or judging leniently (Lee & Holyoak, 2020) their harmful behavior. In light of such a strong bias, we found that only the necessity of justifications freed moral judgments from liking bias, probably because people are concerned about their moral identities (Aquino & Reed, 2002), while a threat to the self leads to less self-serving decision-making (Hughes & Beer, 2012). Future research would do well to investigate whether these factors can eliminate the impact of liking bias on moral judgments.

Finally, the present work has some implications of a practical nature. An increasing number of contemporary societies experience cultural wars concerning abortion, capital punishment, same-sex marriages, immigration, or climate change. Because those questions get easily moralized, the accompanying beliefs become moral convictions, that is, attitudes that their holders experience as grounded in fundamental right and wrong. Compared with other attitudes, moral convictions instigate stronger emotions, are less amenable to change, and more resistant to procedural solutions for related conflicts, leading to more extreme actions (Skitka et al., 2021). No wonder that the resulting moral disputes generate more heat than light. The present data suggest how to shed more light—by asking for justification of moral judgments. The necessity of justifications makes judgments less biased and less extreme. Therefore, including justifications into moral discourse seems to be a good way to make the discourse less divisive and more constructive.

Conclusion

By systematically examining whether interpersonal attitudes bias moral character impressions, we replicated prior findings of the liking bias (Bocian et al., 2018), demonstrating that subjective preferences influence moral character judgments. Participants consistently judged the moral character of liked individuals more favorably than the moral character of neutral or disliked individuals. We also tested two ways to reduce this bias by educating people that liking biases moral character perceptions or making people accountable for their moral character judgments. Although the two moderators successfully debiased participants' moral character judgments, accountability emerged as the only one that completely freed their moral judgments from the liking bias.

Acknowledgments

The authors would like to thank Katarzyna Myslinska-Szarek, Kamila Kaffka-Skierka, and Weronika Trzemielewska for their help with data collection.

Data Availability Statement

The data are openly available at https://osf.io/khrw3/.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The preparation of this article was supported by the National Science Center (Poland) grant 2018/29/B/HS6/00658 (OPUS) awarded to Konrad Bocian.

ORCID iD

Konrad Bocian https://orcid.org/0000-0002-8652-0167

Supplemental Material

Supplemental material is available online with this article.

Notes

1. When we ran the analyses with the excluded participants, there were no changes in the observed effects.
2. We thank the anonymous reviewer for suggesting this solution.
3. When we ran the analyses with the excluded participants, there were no changes in the observed effects.

References

Aquino, K., & Reed, A., II (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*, 1423–1440.

Axt, J. R., Casola, G., & Nosek, B. A. (2018). Reducing social judgment biases may require identifying the potential source of bias. *Personality and Social Psychology Bulletin*, *45*, 1232–1251.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes; Applications* (pp. 1–40). Lawrence Erlbaum.

Bocian, K., Baryla, W., Kulesza, W. M., Schnall, S., & Wojciszke, B. (2018). The mere liking effect: Attitudinal influences on judgments of moral character. *Journal Experimental Social Psychology*, *79*, 9–20.

Bocian, K., Baryla, W., & Wojciszke, B. (2016). When dishonesty leads to trust: Moral judgments biased by self-interest are truly believed. *Polish Psychological Bulletin*, *47*, 366–372.

Bocian, K., Baryla, W., & Wojciszke, B. (2020). Egocentrism shapes moral judgments. *Social and Personality Psychology Compass*, *14*, 1–14.

Bocian, K., Cichocka, A., & Wojciszke, B. (2021). Moral tribalism: Moral judgments of actions supporting ingroup interests depend on collective narcissism. *Journal of Experimental Social Psychology*, *93*, Article 104098.

Bocian, K., & Myslinska-Szarek, K. (2021). Children's sociomoral judgements of antisocial but not prosocial others depend on recipients' past moral behaviour. *Social Development*, *30*, 396–409.

Bocian, K., & Wojciszke, B. (2014a). Self-interest bias in moral judgments of others' actions. *Personality and Social Psychology Bulletin*, *40*, 898–909.

Bocian, K., & Wojciszke, B. (2014b). Unawareness of self-interest bias in moral judgments of others' behavior. *Polish Psychological Bulletin*, *45*, 411–417.

Brambilla, M., Sacchi, S., Menegatti, M., & Moscatelli, S. (2016). Honesty and dishonesty don't move together: Trait content information influences behavioral synchrony. *Journal of Nonverbal Behavior*, *40*, 171–186.

Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. (2021). The primacy of morality in impression development: Theory, research, and future directions. *Advances in Experimental Social Psychology*, *64*.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910. https://doi.org/10.1037/0022-3514.76.6.893

Cushman, F., Young, L., & Greene, J. D. (2010). Multi-system moral psychology. In J. M. Doris & Moral Psychology Research Group (Eds.), *The moral psychology handbook* (pp. 47–71). Oxford University Press.

Decety, J., & Cacioppo, S. (2012). The speed of morality: A high density electrical neuroimaging study. *Journal of Neurophysiology*, *108*, 3068–3072.

Dubois, D., Rucker, D. D., & Galinsky, A. D. (2015). Social class, power, and selfishness: When and why upper and lower class individuals behave unethically. *Journal of Personality and Social Psychology*, *108*, 436–449.

Duffy, K. A., & Chartrand, T. L. (2017). From mimicry to morality: The role of prosociality. In W. Sinnott-Armstrong & C. B. Miller (Eds.), *Moral psychology: Virtue and character* (pp. 439–464). MIT Press.

Dumhan, Y., & Emory, J. (2014). Of affect and ambiguity: The emergence of preference for arbitrary ingroups. *Journal of Social Issues*, *70*, 81–98. https://doi.org/10.1111/josi.12048

Epley, N., & Caruso, E. M. (2004). Egocentric ethics. *Social Justice Research*, *17*, 171–187.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*, 327–339.

Everett, J. A. C., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, *145*, 772–787.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Gilovich, T., & Savitsky, K. (1999). The spotlight effect and the illusion of transparency: Egocentric assessments of how we are seen by others. *Current Directions in Psychological Science*, *8*, 165–168.

Giner-Sorolla, R. (2018, January 24). Powering your interaction. *Approaching Significance. A Methodology Blog for Social Psychology*. https://approachingblog.wordpress.com

Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, *106*, 1339–1366. https://doi.org/10.1016/j.cognition.2007.06.007

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*, 148–168.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*, 1029–1046.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*, 1144–1154.

Grizzard, M., Huang, J., Ahn, C., Fitzgerald, K., Francemone, C. J., & Walton, J. (2020). The Gordian Knot of disposition theory: Character morality and liking. *Journal of Media Psychology: Theories, Methods, and Applications*, *32*, 100–105.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, *316*, 998–1002.

Hamlin, J. K., Mahajan, N., Liberman, Z., & Wynn, K. (2013). Not like me = bad: Infants prefer those who harm dissimilar others. *Psychological Science*, *24*, 589–594.

Hartley, A. G., Furr, R. M., Helzer, E. G., Jayawickreme, E., Velasquez, K. R., & Fleeson, W. (2016). Morality's centrality to liking, respecting, and understanding others. *Social Psychological and Personality Science*, *7*, 648–657.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.

Hershberger, P. J., Markert, R. J., Part, H. M., Cohen, S. M., & Finger, W. W. (1997). Understanding and addressing cognitive bias in medical education. *Advances in Health Sciences Education*, *1*, 221–226.

Hughes, B. L., & Beer, J. S. (2012). Medial orbitofrontal cortex is associated with shifting decision thresholds in self-serving cognition. *Neuroimage*, *61*, 889–898.

Hughes, B. L., & Zaki, J. (2015). The neuroscience of motivated cognition. *Trends in Cognitive Sciences*, *19*, 62–64.

Jones, J. T., Pelham, B. W., Carvallo, M., & Mirenberg, M. C. (2004). How do I love thee? Let me count the Js: Implicit egotism and interpersonal attraction. *Journal of Personality and Social Psychology*, *87*, 665–683.

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, *77*, 221–232.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review*, *62*, 155–178.

Kulesza, W. M., Cislak, A., Vallacher, R. R., Nowak, A., Czekiel, M., & Bedynska, S. (2015). The face of the chameleon: The experience of facial mimicry for the mimicker and mimickee. *The Journal of Social Psychology*, *155*, 590–604.

Landy, J. F., & Royzman, E. B. (2018). The moral myopia model: Why and how reasoning matters in moral judgment. In G. Pennycook (Ed.), *The new reflectionism in cognitive psychology: Why reason matters* (pp. 70–92). Routledge.

Lee, J., & Holyoak, K. J. (2020). "But he's my brother": The impact of family obligation on moral judgments and decisions. *Memory & Cognition*, *48*, 158–170.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*, 255–275.

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

Menegatti, M., Moscatelli, S., Brambilla, M., & Sacchi, S. (2020). The honest mirror: Morality as a moderator of spontaneous behavioral mimicry. *European Journal of Social Psychology*, *00*, 1–12.

Messick, D. M., & Sentis, K. P. (1979). Fairness and preference. *Journal of Experimental Social Psychology*, *15*, 418–434.

Messner, C., & Wänke, M. (2011). Good weather for Schwarz and Clore. *Emotion*, *11*, 436–437.

Moore, D. A., & Loewenstein, G. (2004). Self-Interest, automaticity, and the psychology of conflict of interest. *Social Justice Research*, *17*, 189–202.

Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences*, *14*, 435–440. https://doi.org/10.1016/j.tics.2010.07.004

Myslinska-Szarek, K., Bocian, K., Baryla, W., & Wojciszke, B. (2021). Partner in crime: Rewarding cooperation overcomes children's aversion to antisocial others. *Developmental Science*, *24*, e13038.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*, 250–256.

Pagliaro, S., Brambilla, M., Sacchi, S., D'Angelo, M., & Ellemers, N. (2013). Initial impressions determine behaviours: Morality predicts the willingness to help newcomers. *Journal of Business Ethics*, *117*, 37–44.

Raney, A. A. (2004). Expanding disposition theory: Reconsidering character liking, moral evaluations, and enjoyment. *Communication Theory*, *14*, 348–369.

Schadewald, M. S., & Limberg, S. T. (1992). Effect of information order and accountability on causal judgments in a legal context. *Psychological Reports*, *71*, 619–625.

Schul, Y. (1993). When warning succeeds: The effects of warning on success in ignoring invalid information. *Journal of Experimental Social Psychology*, *29*, 42–62.

Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, *49*, 11–36.

Siegel-Jacobs, K., & Yates, J. F. (1996). Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes*, *66*, 1–17.

Skitka, L. J., Hanson, B. E., Morgan, S. G., & Wisneski, D. C. (2021). The psychology of moral conviction. *Annual Review of Psychology*, *72*, 347–366.

Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly*, *48*, 227–236.

Tetlock, P. E., & Boettger, R. (1989). Accountability: A social magnifier of the dilution effect. *Journal of Personality and Social Psychology*, *57*, 388–398.

Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology*, *52*, 700–709.

Thompson, L., & Loewenstein, G. (1992). Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes*, *51*, 176–197.

Van Berkum, J. J., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science*, *20*, 1092–1099.

Van Boven, L., Dunning, D., & Loewenstein, G. (2000). Egocentric empathy gaps between owners and buyers: Misperceptions of the endowment effect. *Journal of Personality and Social Psychology*, *79*, 66–76.

Weidman, A. C., Sowden, W. J., Berg, M. K., & Kross, E. (2020). Punish or protect? How close relationships shape responses to moral violations. *Personality and Social Psychology Bulletin*, *46*, 693–708.

Wetzel, C. G., Wilson, T. D., & Kort, J. (1981). The halo effect revisited: Forewarned is not forearmed. *Journal of Experimental Social Psychology*, *77*, 427–439.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*, 117–142.

Wilson, T. D., Centerbar, D. B., & Brekke, N. (2002). Mental contamination and the debiasing problem. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 185–200). Cambridge University Press.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*, 1245–1257.

Wojciszke, B., & Bocian, K. (2018). Bad methods drive out good: The curse of imagination in social psychology research. *Social Psychological Bulletin*, *13*, Article e26062.

Zillmann, D., & Cantor, J. R. (1977). Affective responses to the emotions of a protagonist. *Journal of Experimental Social Psychology*, *13*, 155–165.