Research Article

Psychological Science
2019, Vol. 30(6) 880–892

DOI: 10.1177/0956797619837948
www.psychologicalscience.org/PS

I May Not Agree With You, but I Trust You: Caring About Social Issues Signals Integrity

Julian J. Zlatev
Harvard Business School, Harvard University

Abstract

What characteristics of an individual signal trustworthiness to other people? I propose that individuals who care about contentious social issues signal to observers that they have integrity and thus can be trusted. Critically, this signal conveys trustworthiness whether or not the target and the observer hold the same view on the issue. Five studies (N = 3,817) demonstrated the predicted effect of caring on integrity-based trust (Studies 1, 2, 3a, 3b, and 4)—even in cases of strong disagreement—across a variety of issues (Study 1) and when behavioral outcomes with real stakes were used (Studies 3a and 3b). This effect largely results from a perception of low-caring targets as particularly untrustworthy (Study 2). Additionally, participants trusted targets with staunchly opposing views about an issue even though they simultaneously disliked them (Study 4). These findings have important implications for how people form impressions of others and speak to potential interventions to help mitigate the growing ideological divide.

Keywords

morality, social cognition, social perception, cooperation, social behavior, open data, open materials, preregistered

Received 7/23/17; Revision accepted 1/23/19

The willingness to trust other people is critical in modern society. Trust is an essential element in solving coordination problems (De Cremer, Snyder, & Dewitte, 2001), sustaining cooperation (Brewer, 2008), and even encouraging macroeconomic growth (Zak & Knack, 2001). Accurately assessing the trustworthiness of other people can result in large rewards, whereas errors in evaluations of trust can have grave consequences. As a result, people are attuned to a variety of signals when deciding whether to trust another individual, including the target's previous behavior (Levine & Schweitzer, 2015), group membership (Foddy, Platow, & Yamagishi, 2009), and status level (Lount & Pettit, 2012).

The present research examined a novel signal that observers use to decide whether a target can be trusted: how one feels about social issues. Opinions about a social issue can be broken down into two components: (a) one's stance on the issue (e.g., whether someone is pro-choice or pro-life) and (b) the extent to which one cares about the issue (e.g., how important someone thinks the issue of abortion is, regardless of that person's specific stance). The relationship between trust and one's stance on an issue is relatively straightforward; people typically demonstrate a preference for others who hold similar beliefs or ideological stances (Byrne, 1969; Chambers, Schlenker, & Collisson, 2013; Snijders & Keren, 2001).

The relationship between trust and the degree to which one cares about an issue, however, is more complicated. On the one hand, caring about an issue might affect trust differently depending on whether or not the target and observer hold the same view. Caring targets with whom observers agree may be trusted more than apathetic ones, whereas caring targets with whom observers disagree may be trusted less than apathetic ones. People tend to form strong negative impressions of individuals who disagree with them on moralized issues (Skitka, 2010), and people who care strongly enough about an issue to moralize it are more entrenched

Corresponding Author:
Julian J. Zlatev, Harvard University, Harvard Business School, Soldiers Field Rd., Boston, MA 02163
E-mail: jzlatev@hbs.edu

in their view (Rozin, 1999; Skitka, Bauman, & Sargis, 2005). Thus, targets who care strongly about an opposing view may be seen as even more dissimilar to oneself (Chambers et al., 2013) and consequently trusted less (Foddy et al., 2009).

On the other hand, the signal inherent in caring about a social issue might transcend specific disagreements, acting as an indicator of the target's moral character. Moral character is a fundamental trait used in impression formation (Goodwin, Piazza, & Rozin, 2014). Thus, to the extent that caring about a social issue signals one's underlying morality (Kreps & Monin, 2014; Van Zant & Moore, 2015), it should also lead to greater trust in that person. Furthermore, people form more positive impressions of more emotional prosocial actors (Barasch, Levine, Berman, & Small, 2014) and trust individuals more when they demonstrate higher self-control (Righetti & Finkenauer, 2011). This work suggests that principled beliefs about social issues should be positively related to perceptions of a target's morality and trustworthiness.

There is also evidence that this effect should persist even when the observer disagrees with the target. Politicians who profess moral views on an issue and subsequently change their mind are viewed as hypocritical even among observers who disagreed with the politician's initial view (Kreps, Laurin, & Merritt, 2017). Relatedly, people trust religious individuals more than atheists even when the target subscribes to a different religion (Edgell, Gerteis, & Hartmann, 2006; Gervais, Shariff, & Norenzayan, 2011). This suggests that it is holding a belief in and of itself, rather than the specific content of the belief, that engenders trust.

Although research has catalogued a number of different bases of dyadic trust (Butler & Cantrell, 1984), two have emerged as particularly important cues in assessing a target's trustworthiness: integrity (i.e., a target's honesty or virtuousness) and benevolence (i.e., a target's kindness or positive intentions; Levine & Schweitzer, 2015; Mayer, Davis, & Schoorman, 1995). Integrity in particular has been linked to moral character (McFall, 1987; Uhlmann, Pizarro, & Diermeier, 2015). Importantly, integrity implies that one sticks to one's principles, even in the face of outside pressure (Mayer et al., 1995). The more strongly people feel about social issues, the less likely they are to change their minds (Krosnick & Petty, 1995) and, thus, the more principled they should appear. Therefore, targets who care about social issues should be seen as having greater integrity, which should translate into increased integrity-based trust. Differences in benevolence-based trust may also exist, but predictions regarding this dimension of trust are less clear.

Overview of Studies

This article reports the results of five studies, four of which were preregistered (Study 3a was not). Study 1 demonstrates that people think targets who care about a variety of important social issues are more trustworthy than those who do not, whether they agree or disagree with the target's stance. Study 2 examined perceptions of trustworthiness across a wider range of caring. Studies 3a and 3b provide behavioral evidence for the effect with real money at stake. Finally, Study 4 examined behavioral signals of caring (e.g., volunteering for a cause) and explored the relationship between trust and liking.

Following recent recommendations, I included at least 50 participants per condition in all studies reported here (Simmons, Nelson, & Simonsohn, 2013) and at least 100 participants per condition when examining interactions (Simmons, Nelson, & Simonsohn, 2018). All measures obtained in each study are included in the text or the Supplemental Material available online, and all sample sizes were determined a priori. For every study, all data exclusions (if any), all manipulations, and all measures are reported either in the main text or in the Supplemental Material.

Study 1

Study 1 examined the hypothesis that targets who care more about social issues are perceived as more trustworthy than targets who care less about them. A preliminary study can be found in the Supplemental Material.

Method

Participants. A total of 1,007 participants (558 female; age: $M = 33.43$ years, $SD = 10.95$) were recruited from an online subject pool. I set a sample-size target of 1,000 participants in advance.

Procedure. Each participant was assigned to 1 of 20 conditions in a 2 (participant–target agreement: agree, disagree) × 2 (target caring: high, low) × 5 (issue type: capital punishment, abortion, gun control, animal testing, physician-assisted suicide) between-subjects design. Participants were told that they would read information about another participant who, in an earlier study, had been asked for his or her thoughts about a number of issues.[1]

After seeing a brief description of the issue and the target's response, participants were then asked for their impressions of the target on three dimensions—integrity, moral character, and benevolence (described in the Measures section). Participants were then asked whether

Table 1. Mean Participant Ratings of High- and Low-Caring Targets for Each Issue in Study 1

	Integrity of target		Benevolence of target		Moral character of target	
Issue	High care	Low care	High care	Low care	High care	Low care
Capital punishment	4.58 (1.04)	3.83 (1.21)	4.58 (1.06)	3.77 (1.08)	4.75 (1.31)	3.99 (1.30)
Abortion	4.46 (1.22)	3.76 (1.24)	4.16 (1.35)	3.52 (1.19)	4.54 (1.50)	3.76 (1.41)
Gun control	4.14 (1.29)	3.78 (1.28)	4.24 (0.96)	3.97 (1.04)	4.37 (1.20)	3.94 (1.24)
Animal testing	4.71 (1.26)	3.78 (1.26)	4.60 (1.36)	3.52 (1.29)	4.75 (1.62)	3.73 (1.40)
Physician-assisted suicide	4.35 (1.27)	3.82 (1.30)	4.17 (1.37)	3.71 (1.34)	4.39 (1.59)	3.75 (1.32)

Note: Values in parentheses are standard deviations.

they themselves thought the issue should be legal or illegal and how strongly they cared about the issue, using a 5-point Likert-type scale from *not at all* to *extremely*. Participant–target agreement was calculated by matching each participant's opinion on whether the issue should be legal or illegal to the opinion of the target he or she read about. If both the participant and the target thought the issue should be legal or both thought the issue should be illegal, then that was coded as agreement. If one of them thought the issue should be legal and the other thought it should be illegal, that was coded as disagreement. Finally, participants responded to demographics questions.

Measures.

Perceived integrity-based trust. Perceived integrity was measured using three items ($\alpha = .83$) adapted from the work of Levine and Schweitzer (2015): "This person has a great deal of integrity," "I can trust this person's word," and "This person cares about honesty and truth." All items used a 7-point Likert-type scale ranging from *strongly disagree* to *strongly agree*.

Perceived benevolence-based trust. Perceived benevolence was measured using three items ($\alpha = .82$) adapted from the work of Levine and Schweitzer (2015): "This person is kind," "This person is nice," and "This person is selfish" (reverse-coded). All items used a 7-point Likert-type scale ranging from *strongly disagree* to *strongly agree*.

Perceived moral character. Perceived moral character was measured using two items ($r = .80$) adapted from the work of Levine and Schweitzer (2014): "This person is moral" and "This person is ethical." Both items used a 7-point Likert-type scale ranging from *strongly disagree* to *strongly agree*.

Results

All aggregate results reported below include dummy codes for each issue. Results are also broken down by issue when applicable. All effects held when analyses controlled for the strength of participants' own feelings about each issue. Table 1 displays means and standard deviations across conditions for each issue.

Perceived integrity-based trust. There was a significant main effect of caring on participants' ratings of targets' integrity when all five issues were combined, $\beta = 0.51$, $t(1001) = 8.31$, $p < .001$, 95% confidence interval (CI) = [0.39, 0.63], and for each individual issue; targets who cared more about the issue were seen as higher in integrity than targets who cared less (see Fig. 1). This was the case when looking at targets whom participants agreed with, $\beta = 0.60$, $t(997) = 7.39$, $p < .001$, 95% CI = [0.44, 0.76], as well as targets whom participants disagreed with, $\beta = 0.43$, $t(997) = 5.20$, $p < .001$, 95% CI = [0.27, 0.59]. Further analyses found that these results were not being driven by participants' own level of caring about the issue (for details, see the Supplemental Material).

There was also a significant main effect of agreement on participants' ratings of targets' integrity when all five issues were combined, $\beta = 0.64$, $t(999) = 10.58$, $p < .001$, 95% CI = [0.52, 0.76], and for each individual issue; targets who agreed with participants were seen as higher in integrity than targets who disagreed. There was no interaction between targets' caring and participant–target agreement when all five issues were combined, $\beta = 0.17$, $t(997) = 1.51$, $p = .13$, 95% CI = [−0.05, 0.40], or for any individual issue (all ps > .13).

Perceived benevolence-based trust. There was a significant main effect of caring on participants' ratings of targets' benevolence when all five issues were combined, $\beta = 0.52$, $t(1001) = 8.45$, $p < .001$, 95% CI = [0.40, 0.63], and for each individual issue except gun control, which was marginal ($p = .056$). There was also a significant main effect of agreement on participants' ratings of targets' morality when all five issues were combined, $\beta = 0.72$, $t(999) = 12.27$, $p < .001$, 95% CI = [0.61, 0.84], and for each individual issue. Finally, there was a significant interaction

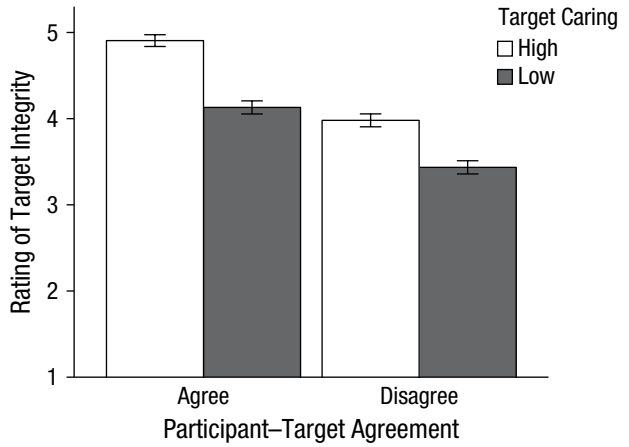

Fig. 1. Mean participant rating of targets' integrity as a function of participant–target agreement and how much the targets cared about the issues (Study 1). Error bars represent standard errors of the mean.

between caring and participant–target agreement when all five issues were combined, $\beta = 0.30$, $t(997) = 2.69$, $p = .007$, 95% CI = [0.08, 0.52]. However, no issue alone had a significant interaction; physician-assisted suicide ($p = .069$) and animal testing ($p = .054$) were each marginal.

Perceived moral character. There was a significant main effect of caring on participants' ratings of targets' morality when all five issues were combined, $\beta = 0.50$, $t(1001) = 8.20$, $p < .001$, 95% CI = [0.38, 0.62], and for each individual issue. There was also a significant main effect of participant–target agreement on participants' ratings of targets' morality when all five issues were combined, $\beta = 0.69$, $t(999) = 11.58$, $p < .001$, 95% CI = [0.57, 0.81], and for each individual issue except gun control, which was marginal ($p = .092$). There was a marginal interaction between caring and agreement when all five issues were combined, $\beta = 0.22$, $t(997) = 1.91$, $p = .056$, 95% CI = [–0.01, 0.44]. The only individual issue with a significant interaction between caring and agreement was physician-assisted suicide, $t(192) = 2.42$, $p = .016$.

Discussion

Study 1 examined impressions of caring versus apathetic targets across five important, and often contentious, social issues. Targets who cared more about a social issue were perceived as higher in integrity regardless of whether participants and targets agreed or disagreed about the issue.

Study 2

Although Study 1 provided evidence of a difference in perceptions of trustworthiness among high-caring targets compared with low-caring targets, it is not clear from this study where the change in trust was taking place (e.g., Mullen & Monin, 2016). In particular, it may be that people find low-caring targets particularly untrustworthy, high-caring targets particularly trustworthy, or both. Study 2 was designed to address this.

Method

Participants. A total of 1,006 participants (543 female, 2 other; age: $M = 35.73$ years, $SD = 11.54$) were recruited from an online subject pool. I set a sample-size target of 1,000 participants in advance. Because of a programming error, 10 participants received a treatment that did not make sense given the instructions (for details, see the Supplemental Material) and were thus excluded, leaving a total of 996 participants.

Procedure. All participants received information about an individual's thoughts on the issue of capital punishment. Unlike in Study 1, it was clear to participants that this was a hypothetical individual named Jamie, rather than a real participant. Participants were told either that Jamie thinks capital punishment should be legal or that Jamie thinks it should be illegal. Participants were then told, "On a scale of 1 (*not at all important*) to 100 (*extremely important*), Jamie thinks the issue of capital punishment is a ___." Participants were then shown a randomly generated number between 1 and 100. Accompanying this number was a description of what it represented: "In other words, Jamie cares [very little/a little/moderately/quite a bit/very much] about this issue." The number determined which label was displayed: 1 to 20 was "very little," 21 to 40 was "a little," 41 to 60 was "moderately," 61 to 80 was "quite a bit," and 81 to 100 was "very much."

After seeing this description, participants answered questions about Jamie's integrity and benevolence (described in the Measures section). Participants then answered whether they themselves thought capital punishment should be legal or illegal and how strongly they cared about the issue (using a 5-point Likert-type scale from *not at all* to *extremely*). Participant–target agreement was calculated in the same way as in Study 1. Finally, participants responded to demographics questions.

Measures. Perceived integrity was measured using the same three items ($\alpha = .82$) from Study 1, and perceived benevolence was measured using the same three items ($\alpha = .60$) from Study 1.

Results

Results were analyzed in two different ways. First, the number representing how much the target cared about the issue (between 1 and 100) was treated as a

continuous measure to examine how ratings of integrity and benevolence differed as a function of the entire spectrum of caring about the issue. Second, these ratings of caring were divided among three buckets: low care, medium care, and high care. Both methods of analysis provided consistent results; thus, results of the first analysis are presented below, whereas results of the second analysis are presented in the Supplemental Material.

Perceived integrity-based trust. There was a significant positive slope of target caring on ratings of perceived integrity, $\beta = 0.20$, $t(994) = 6.32$, $p < .001$, 95% CI = [0.14, 0.26]; participants rated the target as having more integrity when the target cared more about capital punishment. This positive slope was significant among participants who agreed with the target, $\beta = 0.24$, $t(982) = 5.71$, $p < .001$, 95% CI = [0.16, 0.32], as well as participants who disagreed with the target, $\beta = 0.14$, $t(982) = 3.01$, $p = .003$, 95% CI = [0.05, 0.23] (see Fig. 2). There was also a significant main effect of participant–target agreement on ratings of perceived integrity, $\beta = 0.32$, $t(984) = 5.12$, $p < .001$, 95% CI = [0.20, 0.45]; when the target's stance matched the participant's stance, the target was rated higher in integrity than when the two stances did not match. Finally, there was a marginal interaction between caring and agreement on perceived integrity, $\beta = 0.10$, $t(982) = 1.67$, $p = .096$, 95% CI = [−0.02, 0.22].

Plotting the results of target caring on participants' ratings of integrity suggested that this association may not be linear. As a result, I conducted post hoc tests for a curvilinear effect of caring on perceived integrity. To do so, I included a quadratic term in the model regressing integrity ratings on level of caring. An analysis of variance comparing the model containing the quadratic term with a model that did not contain it indicated that the model with the quadratic term provided a significantly better fit, $F(1, 993) = 12.85$, $p < .001$. This provides some evidence of a curvilinear effect.

However, recent work has criticized this method as being unable to appropriately test whether a relationship is U shaped (Simonsohn, 2018). Thus, I additionally ran a two-lines test (Simonsohn, 2018) to examine whether the relationship could be described as an inverted U (i.e., nonmonotonic). Results indicated that the curvilinear relationship was not consistent with an inverted-U shape. Instead, it was consistent with a monotonic shape; the average slope between caring values of 1 and 58.85 (a cutoff determined by the two-lines test) was significantly positive ($b = 0.20$, $z = 4.62$, $p < .001$), whereas the average slope between caring values of 58.85 and 100 was flat ($b = 0$, $z = -0.77$, $p = .44$; see Fig. S1 in the Supplemental Material). In other words, as caring increased beyond a particular point, there was no longer a corresponding increase in perceived integrity. However, there was no evidence that perceived integrity decreased at high levels of caring. These results cohere with the results presented in the Supplemental Material in which the continuous measure of caring was divided among low, medium, and high buckets.

Perceived benevolence-based trust. There was a significant positive slope of target caring on participants' ratings of perceived benevolence, $\beta = 0.11$, $t(994) = 3.45$, $p < .001$, 95% CI = [0.05, 0.17]; participants rated the

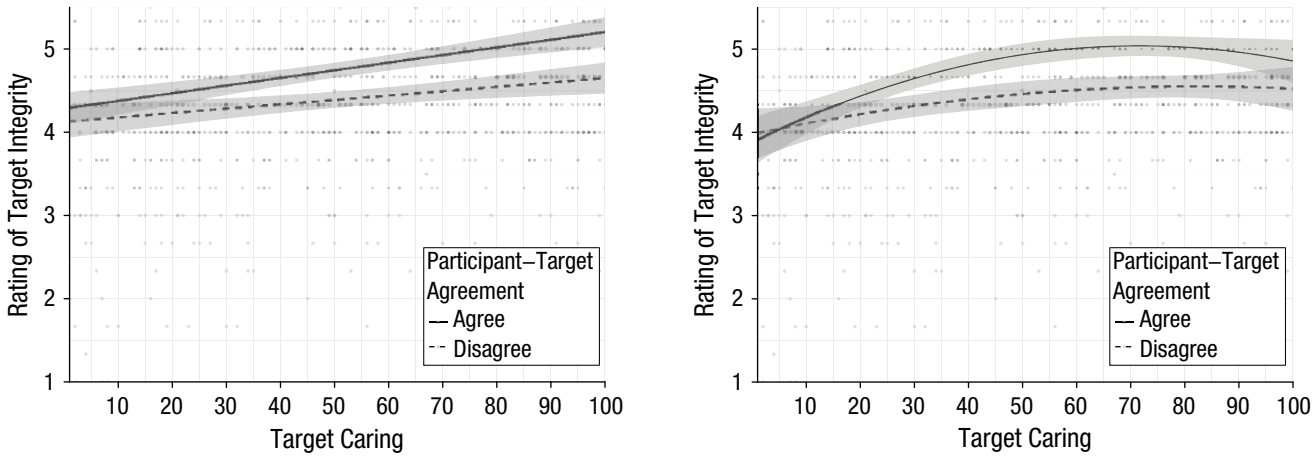

Fig. 2. Mean participant rating of the target's integrity as a function of how much the target cared about issue (as a continuous measure) and participant–target agreement (Study 2). The left panel shows results without a quadratic term, and the right panel shows results with a quadratic term included. Gray bands represent 95% confidence intervals.

target as more trustworthy when the target cared more about capital punishment. This positive slope was significant among participants who agreed with the target, $\beta = 0.14$, $t(982) = 3.18$, $p = .002$, 95% CI = [0.05, 0.22], and marginal among participants who disagreed with the target, $\beta = 0.08$, $t(982) = 1.80$, $p = .07$, 95% CI = [−0.01, 0.17]. There was also a significant main effect of participant–target agreement on ratings of perceived benevolence, $\beta = 0.29$, $t(984) = 4.52$, $p < .001$, 95% CI = [0.16, 0.41]; when the target's stance matched the participant's stance, the target was rated higher in benevolence than when the two stances did not match. Finally, there was no interaction between caring and agreement on perceived benevolence, $\beta = 0.05$, $t(982) = 0.84$, $p = .40$, 95% CI = [−0.07, 0.18].

Again, post hoc analyses suggested that the model including a quadratic term was a significantly better fit than the model that did not include it, $F(1, 993) = 8.90$, $p = .003$, suggesting that the effect of caring on perceptions of benevolence is also curvilinear. A two-lines test indicated that the curvilinear relationship was not consistent with an inverted-U shape. The average slope between caring values of 1 and 57.0 (a cutoff determined by the two-lines test) was significantly positive ($b = 0.14$, $z = 3.24$, $p = .001$), whereas the average slope between caring values of 57.0 and 100 was flat ($b = -0.07$, $z = -1.41$, $p = .16$).

Relative effect of caring on perceived benevolence and integrity. To assess relative differences between caring about an issue and perceptions of the target's integrity versus benevolence, I ran a mixed-effects analysis examining the effect of the caring manipulation on ratings of both perceived integrity and perceived benevolence, treating participant as a random effect. This analysis revealed that there was a significantly greater difference in ratings of integrity between when the target showed high caring and when the target showed low caring than the difference in ratings of benevolence, $\beta = 0.10$, $t(994) = 3.80$, $p < .001$, 95% CI = [0.05, 0.16]. In other words, the gap between the perceived integrity of the target who cared more and the perceived integrity of the target who cared less was larger than the gap between the two levels of the target's perceived benevolence.

Discussion

Study 2 replicated the primary effect from Study 1 using a continuous measure of caring, finding a positive relationship between a target's level of caring and trustworthiness. Additionally, Study 2 provided evidence that this is primarily because of a perception that low-caring targets are particularly untrustworthy.

Studies 3a and 3b

The goal of Studies 3a and 3b was to obtain a behavioral measure of trusting behavior using the rely-or-verify game, a validated measure of integrity-based trust (Levine & Schweitzer, 2015). Study 3a provided an exploratory test of the primary hypothesis, whereas Study 3b was a preregistered direct replication of Study 3a.

Method

Participants. A total of 1,201 participants (573 female, 2 other; age: $M = 33.78$ years, $SD = 12.92$; 1 participant did not complete the demographics questions) were recruited from an online subject pool. I set a sample-size target of 400 participants for Study 3a and 800 participants for Study 3b in advance. Participants who failed the four comprehension checks about the rely-or-verify game (described in the Procedure section) were excluded. This exclusion criterion was preregistered in Study 3b. These items were asked prior to either manipulation to prevent differential attrition by condition. I decided to exclude participants on the basis of these items because it was essential that participants fully understood the rules of the rely-or-verify game to interpret their choices in it. Three hundred fifty-two participants were excluded on the basis of this criterion, leaving a total sample of 842 participants for analysis.

Procedure. Each participant was randomly assigned to one of four conditions in a 2 (participant–target agreement: agree, disagree) × 2 (target caring: high, low) between-subjects factorial design. All participants played the rely-or-verify game with an individual who had answered questions regarding their feelings about capital punishment (from the same additional sample described in the Method for Study 1). The rely-or-verify game proceeded as follows.

Player 1 received information about whether the amount of money in a jar of coins was odd or even. Player 1 could then send a message to Player 2 that conveyed either accurate or inaccurate information about the amount of money in the jar. Player 2 then decided whether to rely on Player 1's message or to verify whether it was accurate. Both Player 1's and Player 2's payoff structures depended on whether Player 1 sent accurate or inaccurate information and whether Player 2 decided to rely on or to verify this information.

If Player 1 sent accurate information and Player 2 relied on it, then Player 1 received \$0.75 and Player 2 received \$1.50. If Player 1 sent accurate information and Player 2 verified it, then Player 1 received \$0.50

and Player 2 received $1.00. If Player 1 sent inaccurate information and Player 2 relied on it, then Player 1 received $1.50 and Player 2 received $0. If Player 1 sent inaccurate information and Player 2 verified it, then Player 1 received $0 and Player 2 received $1.00. Thus, Player 2 should rely on Player 1 if he or she believes that Player 1 is sending accurate information and should verify Player 1's information if he or she believes that Player 1 is sending inaccurate information.

All participants were assigned the role of Player 2 and were matched with a Player 1 who had previously participated in the study. Participants were first given instructions about how to play the rely-or-verify game and were then asked to answer four comprehension checks about the game adapted from Levine and Schweitzer (2015). These comprehension checks appeared on the same page as the instructions, so participants could check on the answers by reading the instructions above them. On the next page, participants were shown Player 1's response to a question about whether capital punishment should be legal or illegal. Player 1's response indicated both whether he or she cared a lot or a little about the issue and whether he or she thought that capital punishment should be legal or illegal.

On the following page, participants were again shown the instructions for the rely-or-verify game so that they could use them as a reference for their decision. They were also told the contents of Player 1's message, that is, whether the amount of money in the jar was odd or even. Participants then indicated whether they would rely on Player 1's message or verify it. On the following page, participants answered questions about the other player's morality, integrity, and benevolence (described in the Measures section). Participants then answered whether they themselves thought capital punishment should be legal or illegal and how strongly they cared about the issue using a 5-point Likert-type scale ranging from *not at all* to *extremely*. Participant–target agreement was calculated in the same way as in Study 1. Finally, participants responded to demographics questions.

Measures. Trusting behavior was measured using participants' responses in the rely-or-verify game. Participants were coded as having trusted the other player if they relied on the message and as having not trusted the other player if they verified the message. Perceived integrity was measured using the same three items ($\alpha = .85$) from Study 1, perceived benevolence was measured using the same three items ($\alpha = .85$) from Study 1, and perceived moral character was measured using the same two items ($r = .77$) from Study 1.

Results

Results from Studies 3a and 3b are pooled and reported below including dummy codes for each study. Results from each individual study are reported in the Supplemental Material.

Trusting behavior. There was a significant main effect of targets' caring on participants' trusting behavior ($b = 0.42$, $z = 2.83$, $p = .005$, 95% CI = [0.13, 0.72]); targets who cared more about capital punishment were trusted more than targets who cared less. This difference was significant among participants who disagreed with the target ($b = 0.61$, $z = 2.54$, $p = .011$, 95% CI = [0.14, 1.09]) but was not significant among participants who agreed with the target ($b = 0.30$, $z = 1.54$, $p = .12$, 95% CI = [−0.08, 0.68]). However, because of the lack of a significant interaction between caring and agreement (see below), this difference should be interpreted with caution.

There was also a significant main effect of participant–target agreement on participants' trusting behavior ($b = 0.41$, $z = 2.68$, $p = .007$, 95% CI = [0.11, 0.71]); targets who agreed with participants were trusted more than targets who disagreed. There was no interaction between caring and agreement ($b = 0.31$, $z = 1.02$, $p = .31$, 95% CI = [−0.29, 0.92]; see Fig. 3).

Perceived integrity-based trust. There was a significant main effect of caring on participants' ratings of targets' integrity, $\beta = 0.56$, $t(839) = 8.48$, $p < .001$, 95% CI = [0.43, 0.69]; targets who cared more about capital punishment were seen as higher in integrity than targets who cared less. This was the case both when participants agreed with the target, $\beta = 0.56$, $t(837) = 6.51$, $p < .001$, and when they disagreed with the target, $\beta = 0.55$, $t(837) = 5.54$, $p < .001$.

There was also a significant main effect of agreement on participants' ratings of targets' morality, $\beta = 0.38$, $t(839) = 5.49$, $p < .001$, 95% CI = [0.24, 0.51]. There was no interaction between caring and agreement, $\beta = 0.02$, $t(837) = 0.15$, $p = .88$, 95% CI = [−0.24, 0.28].

Perceived benevolence-based trust. There was a significant main effect of caring on participants' ratings of targets' benevolence, $\beta = 0.49$, $t(839) = 7.37$, $p < .001$, 95% CI = [0.36, 0.62]; targets who cared more about capital punishment were seen as more benevolent than targets who cared less. This was the case both when participants agreed with the target, $\beta = 0.48$, $t(837) = 5.45$, $p < .001$, and when they disagreed with the target, $\beta = 0.50$, $t(837) = 5.05$, $p < .001$.

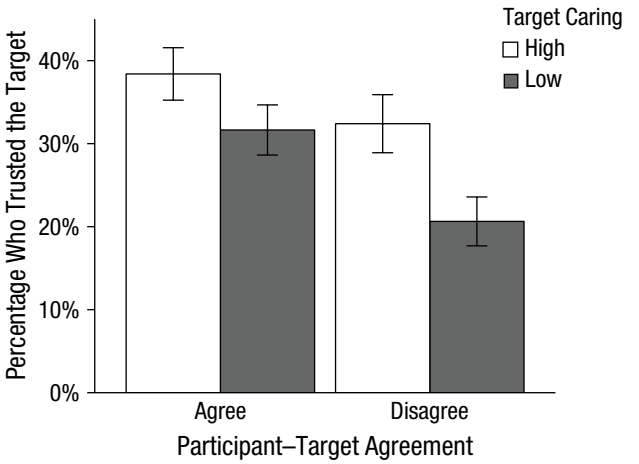

Fig. 3. Percentage of participants who trusted the target in a rely-or-verify game as a function of participant–target agreement and how much the target cared about capital punishment (Studies 3a and 3b). Error bars represent standard errors of the mean.

There was also a significant main effect of agreement on participants' ratings of targets' benevolence, $\beta = 0.38$, $t(839) = 5.73$, $p < .001$, 95% CI = [0.24, 0.51]. There was no interaction between caring and agreement, $\beta = 0.02$, $t(837) = 0.19$, $p = .85$, 95% CI = [−0.24, 0.28].

Perceived moral character. There was a significant main effect of caring on participants' ratings of targets' moral character, $\beta = 0.54$, $t(839) = 8.12$, $p < .001$, 95% CI = [0.41, 0.67]; targets who cared more about capital punishment were seen as higher in moral character than targets who cared less. This was the case both when participants agreed with the target, $\beta = 0.54$, $t(837) = 6.36$, $p < .001$, and when they disagreed with the target, $\beta = 0.52$, $t(837) = 5.33$, $p < .001$.

There was also a significant main effect of agreement on participants' ratings of targets' moral character, $\beta = 0.51$, $t(839) = 7.62$, $p < .001$, 95% CI = [0.38, 0.64]; targets who agreed with participants were seen as higher in moral character than targets who disagreed. There was no interaction between caring and agreement, $\beta = 0.03$, $t(837) = 0.20$, $p = .839$, 95% CI = [−0.23, 0.28].

Relative effect of caring on perceived benevolence and integrity. To assess relative differences between caring about an issue and participants' perceptions of the target's integrity versus benevolence, I ran a mixed-effects analysis examining the effect of the caring manipulation on ratings of both perceived integrity and perceived benevolence, treating participant as a random effect. This analysis revealed that there was a significantly greater difference in ratings of integrity between the high-caring and low-caring targets than the difference in ratings of benevolence, $t(840) = 2.17$, $p = .031$. In other words, the gap between the perceived integrity of targets who cared more and targets who cared less was larger than the gap between the perceived benevolence of these two groups of targets.

Discussion

This study provided behavioral evidence that participants trust targets who care more about social issues, even in cases in which the participant and target disagree on the issue.

Study 4

The goals of this final study were threefold. First, this study examined a behavioral signal of caring: volunteering on behalf of a social cause. In many ways, this represents a more conservative test of the primary hypothesis. Because volunteering is a more concrete step that one can take to benefit a cause, contact with a target who has taken this step in the perceived wrong direction could be particularly acrimonious.

Second, the specific stances of targets in previous studies were always known. In contrast, Study 4 compared an individual who cares about the issue and has taken a stance with an individual who does not care and has not taken a stated stance. This provided a different baseline from that in previous studies; namely, rather than comparing caring within a given stance, I compared targets who cared about a specific stance with apathetic targets who held no professed stance.

Third, this study additionally explored the relationship between trust and liking. Whereas previous work has demonstrated that liking a target increases perceptions of that target's trustworthiness (Brooks, Dai, & Schweitzer, 2014), it is not clear whether this is a necessary component of trust. In particular, one might consider someone to be high in integrity—and subsequently trust him or her—without feeling particularly close to that individual; this may especially be the case when the participant and target disagree.

Method

Participants. A total of 603 participants (303 female, 2 other; age: $M = 34.93$ years, $SD = 11.71$) were recruited from an online subject pool. I set a sample-size target of 600 participants in advance.

Procedure. Participants were given descriptions of two individuals, Taylor and Jamie. Which name was associated with which individual was counterbalanced. The caring individual was described as follows: "[Taylor/Jamie] has strong feelings about the issue of abortion and volunteers

regularly for an organization that advocates for [pro-choice/pro-life] legislation." Which organization was described was randomly determined for each participant. The apathetic individual was described as follows: "[Taylor/Jamie] does not have strong feelings about the issue of abortion and does not volunteer at all for any pro-life or pro-choice organizations." Participants then read about the trust game (Berg, Dickhaut, & McCabe, 1995), which is played as follows.

Player 1 starts with $1 and has the option to either keep the $1 or pass the $1. If Player 1 chooses to keep the $1, then Player 1 receives $1 and the game ends. If Player 1 chooses to pass the $1, then the money grows to $3 and is given to Player 2. Player 2 can then choose to either keep the $3 or return $1.50. If Player 2 keeps the $3, then Player 2 receives $3 and the game ends. In this case, Player 1 receives nothing. If Player 2 returns $1.50, then the $3 is split evenly, and each player receives $1.50. In the current study, participants were told to imagine that they were Player 1 and then answered questions about what they thought each individual would do as Player 2. Participants then answered questions about the perceived integrity and benevolence of each individual and about how much they liked each individual (described in the Measures section).

Participants next indicated their own view on the issue of abortion ("Do you consider yourself to be more pro-life or pro-choice?" using a 4-point Likert-type scale ranging from *definitely pro-life* to *definitely pro-choice*) and how strongly they cared about the issue (using a 5-point Likert-type scale ranging from *not at all* to *extremely*). Participant–target agreement was calculated the same way as in Study 1, with one exception. Because there were four options in this item rather than two, responses were first collapsed into a dichotomous measure indicating either a pro-life or pro-choice stance. Finally, participants responded to demographics questions.

Measures. Perceived general trust of the target was measured using two items (high care: $r = .88$; low care: $r = .87$) referring to the trust game described above (adapted from Levine & Schweitzer, 2015): "I would trust [Taylor/Jamie] to RETURN money" and "I am confident that [Taylor/Jamie] would RETURN money." Both items used a 7-point Likert-type scale ranging from *strongly disagree* to *strongly agree*.

Perceived integrity was measured using the same three items (high care: $\alpha = .89$; low care: $\alpha = .86$) from Study 1, and perceived benevolence was measured using the same three items (high care: $\alpha = .72$; low care: $\alpha = .59$) from Study 1.

Liking was measured using five items (high care: $\alpha = .84$; low care: $\alpha = .80$) adapted from the measure of interpersonal attraction in the work by McCroskey and McCain (1974): "I think [Taylor/Jamie] could be a friend of mine," "It would be difficult to meet and talk with [Taylor/Jamie]" (reverse-coded), "[Taylor/Jamie] just wouldn't fit into my circle of friends" (reverse-coded), "[Taylor/Jamie] and I could never establish a personal friendship with each other" (reverse-coded), and "I would like to have a friendly chat with [Taylor/Jamie]."

Results

Results were run using mixed-effects models treating participant as a random effect (see the Results section of the Preliminary Study in the Supplemental Material for more details).

Perceived general trust. Participants trusted targets who cared about abortion more than targets who did not care about abortion, $\beta = 0.43$, $t(599.21) = 7.84$, $p < .001$, 95% CI = [0.32, 0.54]. There was also a significant interaction between caring and participant–target agreement, $\beta = 0.48$, $t(593.40) = 4.39$, $p < .001$, 95% CI = [0.27, 0.69]; the difference in trust between the caring and apathetic targets was greater for targets who agreed with the participant, $\beta = 0.67$, $t(592.86) = 8.82$, $p < .001$, 95% CI = [0.52, 0.82], than for targets who disagreed with the participant, $\beta = 0.19$, $t(593.90) = 2.39$, $p = .017$, 95% CI = [0.03, 0.34], although both remained significant (see Fig. 4).

Perceived integrity-based trust. Participants rated targets who cared about abortion as higher in integrity than targets who did not care about abortion, $\beta = 0.46$, $t(1199) = 8.10$, $p < .001$, 95% CI = [0.34, 0.57]. There was also a significant interaction between caring and agreement, $\beta = 0.81$, $t(1187) = 7.35$, $p < .001$, 95% CI = [0.60, 1.03]; the

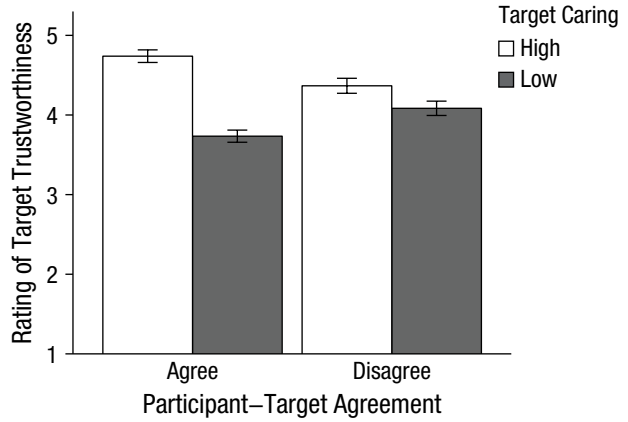

Fig. 4. Mean participant rating of the targets' trustworthiness as a function of participant–target agreement and how much the targets cared about the issue of abortion (Study 4). Error bars represent standard errors of the mean.

difference in integrity between the caring and apathetic targets was greater for targets who agreed with the participant, β = 0.85, $t(1187) = 11.14$, $p < .001$, 95% CI = [0.70, 1.00], than for targets who disagreed with the participant, β = 0.04, $t(1187) = 0.50$, $p = .62$, 95% CI = [−0.12, 0.20].

Perceived benevolence-based trust. Participants rated targets who cared about abortion as higher in benevolence than targets who did not care about abortion, β = 0.39, $t(1199) = 6.94$, $p < .001$, 95% CI = [0.28, 0.50]. There was also a significant interaction between caring and agreement, β = 0.86, $t(1187) = 7.78$, $p < .001$, 95% CI = [0.64, 1.08]; the difference in benevolence between the caring and apathetic targets was greater for targets who agreed with the participant, β = 0.81, $t(1187) = 10.60$, $p < .001$, 95% CI = [0.66, 0.96], than for targets who disagreed with the participant, β = −0.05, $t(1187) = -0.61$, $p = .54$, 95% CI = [−0.20, 0.11].

Liking. There was no main effect of caring on ratings of target liking, β = −0.05, $t(600) = -0.94$, $p = .35$, 95% CI = [−0.16, 0.06]. There was, however, a significant interaction between caring and agreement on liking, β = 1.00, $t(594) = 9.41$, $p < .001$, 95% CI = [0.79, 1.21]; participants liked the target with whom they agreed significantly more than the apathetic target, β = 0.43, $t(594) = 5.80$, $p < .001$, 95% CI = [0.28, 0.57], but liked the target with whom they disagreed significantly less than the apathetic target, β = −0.57, $t(594) = -7.48$, $p < .001$, 95% CI = [−0.72, −0.42] (see Fig. 5).

Indirect effects of integrity, benevolence, and liking. A structural equation model assessed the indirect effect of caring on trust via integrity, benevolence, and liking (for additional details, see the Supplemental Material). Results indicated a significant indirect effect of perceived integrity ($b = 0.35$, $z = 6.70$, $p < .001$, 95% CI = [0.25, 0.45]) but no significant indirect effects of benevolence ($b = 0.03$, $z = 1.17$, $p = .24$, 95% CI = [−0.02, 0.08]) or liking ($b = -0.004$, $z = -0.82$, $p = .41$, 95% CI = [−0.01, 0.01]).

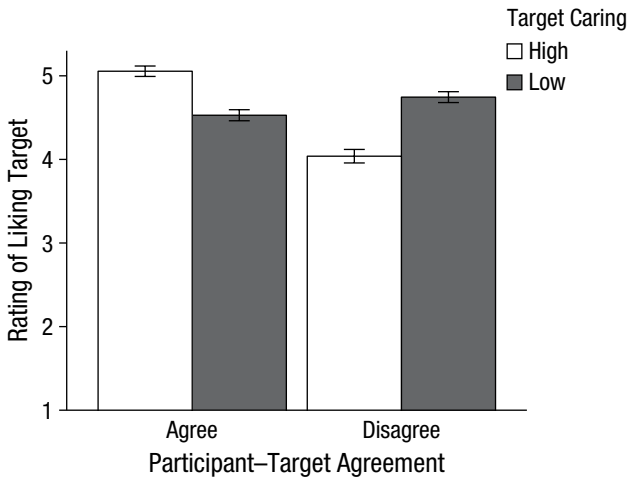

Fig. 5. Mean rating of how much participants liked targets as a function of participant–target agreement and how much the targets cared about the issue of abortion (Study 4). Error bars represent standard errors of the mean.

Discussion

Study 4 indicated that participants trusted targets who cared about the issue of abortion more than apathetic targets, whether or not they held the same opinion. This occurred even though participants liked targets with whom they disagreed even less than apathetic targets.

Unlike in previous studies, participants did not perceive targets with whom they disagreed as higher in integrity than apathetic targets. This may have been because the target's specific stance was not specified in this study, meaning that there was still a chance that the agnostic individual agreed with them. Also unlike in previous studies, there was a significant interaction between participant–target agreement and degree of caring. Interestingly, this seemed at least partially driven by differences in assessments of the agnostic individual, even though in this study, the agnostic individual was described in exactly the same way across both conditions. This suggests that providing different comparison points had the unanticipated consequence of changing impressions of the agnostic individual, making him or her seem more trustworthy when compared with a caring target with whom participants disagreed.

General Discussion

The present work suggests that a target's level of caring and participant–target agreement represent two empirically and theoretically separable components in people's assessments of a target's trustworthiness (see Table 2). Across five studies using a variety of contentious social issues, I found evidence that people trust others who demonstrate strong feelings about social issues, even when they disagree with or dislike them. This appears to be primarily driven by low rates of trust toward individuals who care little about these issues. Caring influences both integrity-based and benevolence-based trust, but there is some evidence that it influences integrity-based trust to a greater extent. Whereas the effect of caring on perceptions of integrity did not differ significantly by agreement in Studies 1 to 3, it did in Study 4. This suggests potential boundary conditions for the effect, such as taking tangible actions to address an issue (e.g., volunteering) rather than simply stating a preference. Notably, however, no study found the potential opposing pattern of results described in the

Table 2. Effects of Target Caring on Dependent Measures in Studies 1 to 4

Dependent measure and predictor	Study 1	Study 2	Studies 3a and 3b	Study 4
General trustworthiness of target				
Main effect of caring			$b = 0.42$, $p = .005$	$\beta = 0.43$, $p < .001$
Main effect of caring when participant and target agree			$b = 0.30$, $p = .124$	$\beta = 0.67$, $p < .001$
Main effect of caring when participant and target disagree			$b = 0.61$, $p = .011$	$\beta = 0.19$, $p = .017$
Perceived integrity of target				
Main effect of caring	$\beta = 0.51$, $p < .001$	$\beta = 0.22$, $p < .001$	$\beta = 0.56$, $p < .001$	$\beta = 0.46$, $p < .001$
Main effect of caring when participant and target agree	$\beta = 0.60$, $p < .001$	$\beta = 0.27$, $p < .001$	$\beta = 0.56$, $p < .001$	$\beta = 0.85$, $p < .001$
Main effect of caring when participant and target disagree	$\beta = 0.43$, $p < .001$	$\beta = 0.14$, $p = .003$	$\beta = 0.55$, $p < .001$	$\beta = 0.04$, $p = .619$
Perceived benevolence of target				
Main effect of caring	$\beta = 0.52$, $p < .001$	$\beta = 0.14$, $p < .001$	$\beta = 0.49$, $p < .001$	$\beta = 0.39$, $p < .001$
Main effect of caring when participant and target agree	$\beta = 0.67$, $p < .001$	$\beta = 0.18$, $p < .001$	$\beta = 0.48$, $p < .001$	$\beta = 0.81$, $p < .001$
Main effect of caring when participant and target disagree	$\beta = 0.37$, $p < .001$	$\beta = 0.10$, $p = .057$	$\beta = 0.50$, $p < .001$	$\beta = -0.05$, $p = .544$

Note: Statistics for Study 2 are the results of comparing the high-care condition with the low-care condition (see the Supplemental Material available online for details).

introduction, namely, a decrease in trustworthiness of disagreeing targets as their caring increased.

The present research contributes to a number of different literatures. First, it advances work on the antecedents of trust (Kramer, 1999; Kramer & Tyler, 1996), particularly research examining the distinction between benevolence-based trust and integrity-based trust (Levine & Schweitzer, 2015). I demonstrated a novel signal of trustworthiness and identified a unique setting in which trust and liking diverge. Second, work on moralization has shown that there can be various negative repercussions when people on different sides of an issue engage with one another (Skitka, 2010). The present research provides a potential counter to these findings, demonstrating that strong feelings about an issue in certain circumstances may actually bolster trust rather than undermine it.

The present work also has practical implications. By many accounts, polarization on social issues in the United States and other countries has increased substantially in recent decades (Hare & Poole, 2014; Westfall, Van Boven, Chambers, & Judd, 2015). Whereas previous work has looked to bridge the political divide by changing the messaging around social issues (Campbell & Kay, 2014; Feinberg & Willer, 2013), changing the information that is conveyed about an individual might also prove beneficial. The present results suggest that signals that someone is apathetic about an issue can harm trust, suggesting possible routes for novel interventions.

However, more research is needed before putting this work into practice. For example, there may be a higher barrier to believing that someone with an opposing view actually does care strongly about an issue. Future research could look at how believable different signals of caring are, depending on whom those signals are coming from. It should also be noted that although caring about an issue seemed to have an effect on trust independently of whether or not there was agreement, it was still the case empirically that the highest levels of trust occurred when targets both cared about the issue and agreed with the participant about it.

Further, there are other aspects of these studies that could limit their generalizability. First, participants received information about targets only indirectly; that is, they did not interact with them directly. It may be the case that more subtle cues of trust in face-to-face situations influence people as much as, or more than, the cues used in these studies (e.g., Porter & ten Brinke, 2008). Second, although Study 1 demonstrated that this effect generalized across a broad range of social issues, there may be other issues for which this is not the case. It would be especially useful to see whether this effect persists outside of U.S. issues. Finally, these studies involved only one-off instances of trusting that were unrelated to the social issue discussed. People may be more hesitant to trust targets with whom they disagree over repeated interactions or when the matter at hand involves the central issue. I have no reason to believe

that the results depend on other characteristics of the participants, materials, or context.

Although not predicted a priori, it is interesting that caring also affects benevolence-based trust. In particular, it may be that people infer traits such as kindness more generally from caring about social issues. Future research could explore this association further. Future research should also examine whether the perceptions of trustworthiness found in these studies are actually warranted. In other words, are people who feel more strongly about social issues actually more trustworthy, or is this a stereotype that can lead to the exploitation of trusting individuals? Finally, in the current work, individuals varied in the extent to which they cared about the social issue, but they were clear about their stance. It would be interesting for future research to examine other more ambiguous stances that individuals might take regarding social issues, such as being undecided or ambivalent. Relatedly, future work should further explore why there is a particularly strong aversion to individuals who express low levels of caring. Not caring about social issues may signal selfishness, (intentional or unintentional) ignorance, or something else entirely. These different reasons imply different psychological processes that would be fruitful to unpack.

Action Editor

Leaf Van Boven served as action editor for this article.

Author Contributions

J. J. Zlatev is the sole author of this article and is responsible for its content.

ORCID iD

Julian J. Zlatev https://orcid.org/0000-0002-9427-9887

Acknowledgments

I thank Jen Dannals, Nir Halevy, Dale Miller, Benoît Monin, and Sarah Stein for their help and advice, as well as members of the Design and Process of Experimental Research group and the Stanford University Morality Lab for their valuable feedback.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797619837948

Open Practices

All data, analysis code, and materials have been made publicly available via the Open Science Framework and can be accessed at osf.io/vh5jt. The design and analysis plans for all the studies except for Study 3a were preregistered on AsPredicted.org; copies of these preregistrations can be found at osf.io/vh5jt. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797619837948. This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

Note

1. To avoid deception, I ran a separate sample ($N = 30$) in which I asked participants to select among four options to describe their opinion on each of five issues (capital punishment, abortion, gun control, animal testing, and physician-assisted suicide): (a) "I care a lot about this issue and I think [issue] should be legal," (b) "I care a lot about this issue and I think [issue] should be illegal," (c) "I don't care at all about this issue and I think [issue] should be legal," or (d) "I don't care at all about this issue and I think [issue] should be illegal." Participants in Study 1 were then matched with the response of one of these previously run participants for one of the five issues.

References

Barasch, A., Levine, E. E., Berman, J. Z., & Small, D. A. (2014). Selfish or selfless? On the signal value of emotion in altruistic behavior. *Journal of Personality and Social Psychology*, *107*, 393–413. doi:10.1037/a0037207

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*, 122–142.

Brewer, M. B. (2008). Depersonalized trust and ingroup cooperation. In J. I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn Mason Dawes* (pp. 215–232). New York, NY: Psychology Press.

Brooks, A. W., Dai, H., & Schweitzer, M. E. (2014). I'm sorry about the rain! Superfluous apologies demonstrate empathic concern and increase trust. *Social Psychological & Personality Science*, *5*, 467–474. doi:10.1177/1948550613506122

Butler, J. K., & Cantrell, R. S. (1984). A behavioral decision theory approach to modeling dyadic trust in superiors and subordinates. *Psychological Reports*, *55*, 19–28.

Byrne, D. E. (1969). Attitudes and attraction. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 4, pp. 35–89). San Diego, CA: Academic Press.

Campbell, T. H., & Kay, A. C. (2014). Solution aversion: On the relation between ideology and motivated disbelief. *Journal of Personality and Social Psychology*, *107*, 809–824.

Chambers, J. R., Schlenker, B. R., & Collisson, B. (2013). Ideology and prejudice: The role of value conflicts. *Psychological Science*, *24*, 140–149. doi:10.1177/0956797612447820

De Cremer, D., Snyder, M., & Dewitte, S. (2001). 'The less I trust, the less I contribute (or not)?' The effects of trust, accountability and self-monitoring in social dilemmas. *European Journal of Social Psychology*, *31*, 93–107.

Edgell, P., Gerteis, J., & Hartmann, D. (2006). Atheists as "other": Moral boundaries and cultural membership in American society. *American Sociological Review*, *71*, 211–234. doi:10.1177/000312240607100203

Feinberg, M., & Willer, R. (2013). The moral roots of environmental attitudes. *Psychological Science*, *24*, 56–62. doi:10.1177/0956797612449177

Foddy, M., Platow, M. J., & Yamagishi, T. (2009). Group-based trust in strangers: The role of stereotypes and expectations. *Psychological Science*, *20*, 419–422. doi:10.1111/j.1467-9280.2009.02312.x

Gervais, W. M., Shariff, A. F., & Norenzayan, A. (2011). Do you believe in atheists? Distrust is central to anti-atheist prejudice. *Journal of Personality and Social Psychology*, *101*, 1189–1206. doi:10.1037/a0025882

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. doi:10.1037/a0034726

Hare, C., & Poole, K. T. (2014). The polarization of contemporary American politics. *Northeastern Political Science Association*, *46*, 411–429. doi:10.1057/pol.2014.10

Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology*, *50*, 569–598. doi:10.1146/annurev.psych.50.1.569

Kramer, R. M., & Tyler, T. R. (1996). *Trust in organizations: Frontiers of theory and research*. Thousand Oaks, CA: SAGE.

Kreps, T. A., Laurin, K., & Merritt, A. C. (2017). Hypocritical flip-flop, or courageous evolution? When leaders change their moral minds. *Journal of Personality and Social Psychology*, *113*, 730–752.

Kreps, T. A., & Monin, B. (2014). Core values versus common sense: Consequentialist views appear less rooted in morality. *Personality and Social Psychology Bulletin*, *40*, 1529–1542. doi:10.1177/0146167214551154

Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1–24). Mahwah, NJ: Erlbaum.

Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology*, *53*, 107–117.

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes*, *126*, 88–106. doi:10.1016/j.obhdp.2014.10.007

Lount, R. B., & Pettit, N. C. (2012). The social context of trust: The role of status. *Organizational Behavior and Human Decision Processes*, *117*, 15–23. doi:10.1016/j.obhdp.2011.07.005

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*, 709–734. doi:10.5465/AMR.1995.9508080335

McCroskey, J. C., & McCain, T. A. (1974). The measurement of interpersonal attraction. *Speech Monographs*, *41*, 261–266. doi:10.1080/03637757409375845

McFall, L. (1987). Integrity. *Ethics*, *98*, 5–20.

Mullen, E., & Monin, B. (2016). Consistency versus licensing effects of past moral behavior. *Annual Review of Psychology*, *67*, 363–385. doi:10.1146/annurev-psych-010213-115120

Porter, S., & ten Brinke, L. (2008). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological Science*, *19*, 508–514. doi:10.1111/j.1467-9280.2008.02116.x

Righetti, F., & Finkenauer, C. (2011). If you are able to control yourself, I will trust you: The role of perceived self-control in interpersonal trust. *Journal of Personality and Social Psychology*, *100*, 874–886. doi:10.1037/a0021827

Rozin, P. (1999). The process of moralization. *Psychological Science*, *10*, 218–221.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2013). Life after P-hacking. In S. Botti & A. Labroo (Eds.), *Advances in Consumer Research* (Vol. 41). Duluth, MN: Association for Consumer Research.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2018). False-positive citations. *Perspectives on Psychological Science*, *13*, 255–259.

Simonsohn, U. (2018). Two-lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science*, *1*, 538–555. doi:10.1177/2515245918805755

Skitka, L. J. (2010). The psychology of moral conviction. *Social & Personality Psychology Compass*, *4*, 267–281. doi:10.1111/j.1751-9004.2010.00254.x

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, *88*, 895–917. doi:10.1037/0022-3514.88.6.895

Snijders, C., & Keren, G. (2001). Do you trust? Whom do you trust? When do you trust? *Advances in Group Processes*, *18*, 129–160. doi:10.1016/S0882-6145(01)18006-9

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, *10*, 72–81. doi:10.1177/1745691614556679

Van Zant, A. B., & Moore, D. A. (2015). Leaders' use of moral justifications increases policy support. *Psychological Science*, *26*, 934–943. doi:10.1177/0956797615572909

Westfall, J., Van Boven, L., Chambers, J. R., & Judd, C. M. (2015). Perceiving political polarization in the United States: Party identity strength and attitude extremity exacerbate the perceived partisan divide. *Perspectives on Psychological Science*, *10*, 145–158. doi:10.1177/1745691615569849

Zak, P. J., & Knack, S. (2001). Trust and growth. *The Economic Journal*, *111*, 295–321.