Can I Still Trust You? Late Trust Violations Hinder Subsequent Cooperation

Megan A. Gerges  
Department of Psychology and Neuroscience, Duke University

ABSTRACT. Trust is necessary to build and maintain relationships, making the ability to rebuild trust after it has been broken also important. This study explored how the timing of a trust violation (whether it came early in a long-term interaction, before a relationship is established, or later, after a cooperative relationship had been established) impacts subsequent behavioral and attitudinal trust, based on whether the interaction partner is an ingroup or outgroup member (using political ideology). A sample of 208 U.S. participants played 20 rounds of a trust game with either an ingroup member or an outgroup member and were randomly assigned to experience an early trust violation or a late trust violation. Late trust violations were more detrimental than early trust violations. However, there was no significant difference in trusting behaviors and attitudes between ingroup and outgroup pairings, nor a significant difference in how the timing of the trust violation impacted subsequent behavioral trust between ingroup and outgroup pairings. Nonetheless, subsequent trusting attitudes were higher for ingroup pairings after an early trust violation compared to a late one. Moreover, behavioral trust in the first round, before any trust violation had occurred, was higher among ingroup partners, indicating the presence of an ingroup effect, which was nullified over the course of the longer-term interaction. Ultimately, these results suggest that it is important to establish a pattern of trustworthy behavior over time to build relationships. Furthermore, an ingroup bias does not preclude similar kinds of cooperation between outgroup members.

Keywords: trust, trust violations, cooperation, ingroup, outgroup
can benefit all parties involved, it heavily relies on mutual trust (Romano et al., 2021).

The Impacts of the Timing of Trust Violations
This study sought to explore the impacts of trust violations and how they impact subsequent trusting behavior (i.e., cooperation) between strangers. Previous work has demonstrated that, when a trust violation occurs, it influences later cooperative behavior (Kuwabara et al., 2014; Lount et al., 2008). However, there are competing models and theories regarding whether, in long-term partnerships comprising a series of repeated interactions, trust violations that occur earlier or later in that partnership are more detrimental to later cooperative behavior. Some of the prior work has argued that both the history and context of a relationship influence how trust violations impact individuals. For instance, Morrison and Robinson (1997) suggested that trust violations committed by organizations with a history of integrity and good faith—thus, trust violations that are late in a long-term interaction—provoke greater betrayal than those committed by organizations with a history of bad behavior, because the former scenario creates elevated expectations. Other theories posit that individuals begin with little trust, which must be built through trustworthy behavior (Blau, 1964; Rempel et al., 1985; Rousseau & McLean Parks, 1993). Finally, Sherif and Hovland (1961) argued that, when trust is violated early on, cooperative behavior that occurs thereafter seems relatively and disproportionately more positive. These theories and models suggest that early trust violations occurring in the context of a long-term interaction will be less harmful to subsequent trust than late trust violations. This is because, earlier in an interaction, expectations for being treated justly or trustworthy are relatively low, making trust violations less surprising and hence, less hurtful. Contrastingly, when there is a history of trustworthy early in an interaction, a greater expectation of subsequent trustworthiness develops, creating stronger feelings of betrayal when a later trust violation occurs (Morrison & Robinson, 1997).

Other empirical studies using U.S. participants have instead found that, in long-term partnerships comprising a series of repeated interactions, trust violations that occur earlier in that partnership are more detrimental to later cooperative behavior than those that occur later, after a relationship has been solidly established (Kuwabara et al., 2014; Lount et al., 2008). Various theories support this argument as well. For example, in some circumstances, individuals tend to trust others even before the interaction begins (Kim et al., 2004; McKnight et al., 1998), and violating those trust expectations damages these interpersonal impressions (Lount et al., 2008; Rink & Ellemers, 2007). Other theories suggest that individuals trust others to allow others to demonstrate trustworthiness (Weber et al., 2004). Therefore, a lack of trustworthy behavior early on will be more noticeable and disproportionately affect whether one chooses to trust in the future (Lount et al., 2008). This, combined with the notion that negative first impressions are more impactful than positive first impressions (Skowronski & Carlston, 1989), suggests that early trust violations will reduce trust and cooperation to a greater extent than will later trust violations. This study aimed to arbitrate between these two groups of theories, with a nondirectional hypothesis to test the two arguments:

Hypothesis 1: The timing of a trust violation will differentially impact subsequent levels of trust in long-term interactions.

The Potential Impacts of an Ingroup Bias
One consideration that may be relevant in explaining and understanding the impacts of trust violations and the impacts of the timing of trust violations on subsequent trusting behavior is social identity, or a person’s sense of belonging to particular groups. Although individuals enact social behavior, group categorizations based on similarities within each group can have psychological effects that influence social interactions (Tajfel, 1979). For example, many people categorize themselves as either politically conservative or liberal with other politically conservative and liberal people, respectively, being considered a member of their “ingroup” (and vice versa). This could create psychological and behavioral differences in interactions based on whether someone is interacting with a member of their political ingroup or outgroup.

This work tests whether the conflicting models and studies on the timing of trust violations can be explained by ingroup bias—the extent to which participants feel they are part of a group or share an identity with their partner (and hence feel a connection to). Although some models predict that early trust violations will hinder subsequent trust to a greater extent, and others predict the opposite, what seems to underlie them all is the idea that the higher the level of trust, the greater the feelings of hurt or betrayal from a trust violation. The conflict between these theories stems from disagreements about when one’s levels of trust are highest. As it relates to the ingroup bias, an abundance of literature has found that people generally favor, trust, expect trustworthiness from, and cooperate more with ingroup members compared to outgroup members (Balliet et al., 2014; Balliet et al., 2018; Fischer & Derham, 2016; Mullen et al., 1992; Romano et al., 2021; Yamagishi et al., 1999). Therefore,

Hypothesis 2: Participants will (a) entrust their
resources (demonstrate greater behavioral trust) and (b) demonstrate greater trusting attitudes with ingroup members (those with whom they share a social identity) more often than outgroup members (those with whom they do not share a social identity).

It was also predicted that,

**Hypothesis 3:** There would be lower levels of trust between ingroup partners that experience an early trust violation compared to a late trust violation, because of the greater trust placed in ingroup members prior to the start of an interaction, making an immediate betrayal feel particularly offensive, especially because individuals expect to be treated favorably by ingroup members (Yamagishi et al., 1999). It is possible that ingroup members are more willing to forgive each other in order to maintain positive expectations about one another or avoid cognitive dissonance. However, there is a distinction between being more willing to forgive an ingroup member for wrongs they have committed in general and wrongs that have been committed against oneself, which would feel especially hurtful given the positive expectations one typically has of those in their ingroup. Therefore, it remains more plausible that an early trust violation will be detrimental to future trust.

Lastly, it was anticipated that:

**Hypothesis 4:** Participants will show less (a) behavioral trust and (b) attitudinal trust in their outgroup partners if they experience a late trust violation, compared to an early trust violation, because though there is less trust placed in outgroup members prior to the start of the interaction, a history of cooperativeness and trust would have been built by the time a late trust violation occurred.

**Political Affiliation as the Instance of Ingroup Bias**
Political ideology in the United States, which places conservatives on the right side of the political spectrum and liberals on the left side of the political spectrum, was chosen as the salient social identity and examined whether the effects of trust violations would be impacted based on whether participants were partnered with someone who shared their political ideology or with someone who did not. Because there are differences among different countries in the meaning of different political orientations, this study limited respondents to the United States.

Political ideology was chosen because, as it is a real group categorization (as opposed to artificially created ingroups), there may be legitimate differences in how trust violations impact politically conservative and liberal participants. For example, past research has suggested that those who identify as politically conservative have a preference for stability, hierarchy, and structure, and they tend to be more resistant to change because of a greater dislike for uncertainty and threats (Carney et al., 2008; Jost et al., 2003). Furthermore, conservative individuals have been found to be more sensitive to danger and hence more likely to create and maintain defense mechanisms against their environments (Hetherington & Weiler, 2018). Finally, there is some evidence that there is a greater association between individualism, pro-self values, self-reliance, and conservatism (Balliet et al., 2018; Chirumbolo et al., 2016; Sheldon & Nichols, 2009; Van Lange et al., 2012).

In contrast, liberal individuals tend to be more prosocial, which is associated with higher levels of cooperation, and they are more willing to cooperate with strangers, regardless of group membership (Romano et al., 2021). Furthermore, liberal individuals tend to be more concerned about equality of outcome than politically conservative people (Balliet et al., 2018), and they tend to be less conscientious (Hetherington & Weiler, 2018). Hetherington and Weiler (2018) suggested that those lower in conscientiousness are likelier to embrace norm violators rather than shun them (as those higher in conscientiousness may be more prone to do). Therefore, this study controlled for political affiliation in supplementary models, in addition to the main models.

**The Present Study**
To test these various hypotheses about how trust violations impact subsequent trust based on timing and ingroup/outgroup pairings, this study recruited a sample of United States participants, who identified as either politically liberal or politically conservative, to play a 20-round, two-player repeated trust game (Haselhuhn et al., 2015). In this game, participants interacted with a pre-programmed player in which they had the opportunity, in each round, to place their trust in their partner with the expectation and hope that their partner would reciprocate that trust. Specifically, participants could either give their tokens to their partner or keep them for themselves. Their trust was violated by the pre-programmed player either early in the interaction (within the first two rounds) or later in the interaction (during rounds 10 and 11) to test participants’ subsequent behavioral trust reactions (whether they would continue to trust their tokens to their partners) and subsequent trusting attitudes.

**Methods**

**Participants**
Following Duke University’s Campus IRB approval (protocol #2022-0353), participants were recruited to participate in the study via Mechanical Turk (MTurk), Amazon’s crowdsourcing website for participant recruitment. All participants were required to be in the United States and identify as politically liberal or conservative.
The final sample consisted of 208 participants, 124 of whom identified as politically liberal (59.62%) and 84 of whom identified as politically conservative (40.38%). In addition, there were 85 (40.87%) women, 121 (58.17%) men, and 2 (0.96%) who preferred not to say. The average age of the participants was 40.3 years ($SD = 12.34$, min = 20, max = 72). Participants with a household income of less than $75,000 represented 67.79% of the sample, and participants with a household income of $75,000 or more represented 29.81% of the sample. 2.4% of participants did not disclose their income.

Demographic Questions
At the start of the study, participants were asked to fill out demographic information, specifically about their age, gender, income bracket, as well as questions about their political orientation. To measure political orientation, participants indicated their level of agreement with three statements on a scale from 1 (strongly agree) to 7 (strongly disagree). The first statement was, "When it comes to politics, I consider myself politically conservative," and the second was, "When it comes to politics, I consider myself politically liberal." The third statement asked "How would you describe your political orientation?” rated on a scale from 1 (extremely left) to 7 (extremely right). A score was calculated based on the participants' answers to determine whether they were liberal or conservative. The first statement was reverse coded to produce a total score where 1 was equivalent to very liberal and 7 was equivalent to very conservative. A score greater than 4 indicated that the participant identified as conservative, and a score less than 4 indicated that the participant identified as liberal ($M = 3.45$, $SD = 2.26$, min = 1, max = 7). A Cronbach’s alpha value of $\alpha = .95$ was found.

The Study Design
To test the hypotheses, the study utilized a 2 (early or late trust violation) x 2 (ingroup or outgroup pairing) between-subjects study design using a repeated, 20-round, two-player trust game adapted from Haselhuhn and colleagues (2015), programmed in z-Tree version 3.4.2 and z-Tree unleashed (Duch et al., 2020; Fischbacher, 2007).

The two players in this trust game were the "sender" and the "receiver." The sender represented the “truster,” (the person placing their trust in their partner), and the receiver represented the “trustee” (the person receiving the trust from their partner). The terms "sender" and "receiver" were used to avoid using biased terminology. As demonstrated in Figure 1, the sender had 6 tokens at the beginning of each decision round and could make the binary decision to either (a) send those 6 tokens to the receiver or (b) keep them for themselves. If they kept the tokens, the round ended with the sender earning 6 tokens and the receiver earning 0. If the sender sent the 6 tokens, it tripled, and the receiver received 18 tokens. The receiver then could send half of the tokens back to the sender (9 tokens) or keep all 18 tokens for themselves. Thus, the sender had the opportunity to earn 0, 6, or 9 tokens, and when they did, the receiver would earn 18, 0, or 9 tokens, respectively.

It was best that the sender trust their tokens to the receiver if the receiver would behave in a trustworthy manner (resulting in 9 tokens earned, rather than 6 tokens kept by the sender at the end of the round).

\[ Given \ research \ suggests \ that \ gender \ may \ impact \ overall \ levels \ of \ trust \ and \ trustworthiness \ (Haselhuhn \ et al., 2015; \ Sutter \ et al., 2009; \ van \ den \ Akker \ et al., 2020; \ Wu \ et al., 2020), \ this \ study \ controlled \ for \ gender \ in \ the \ analysis. \]
tokens by keeping them). However, if the receiver would not be trustworthy, it was in the best interest of the sender to keep their tokens (earning 6 tokens rather than 0 tokens). Likewise, it was best for the receiver if they were trusted and did not behave in a trustworthy manner (earning 18 tokens, rather than earning 0 by not being trusted or 9 by being trusted and sending tokens back to the sender). However, the receiver had to consider the possibility that if they acted in an untrustworthy manner, the sender may choose to stop sending tokens, which would result in the receiver earning 0 tokens instead of 9. Therefore, for both participants to earn 9 tokens, the sender had to place their trust in the receiver and the receiver had to send tokens back. The possible earnings in each round for the sender and the receiver are shown in Figure 2.

Participants were told that they would be randomly assigned to one of these two roles and that another participant would be randomly assigned to the other role. In reality, all participants were assigned to be the sender (the “truster”), and a computer simulated the receiver. Furthermore, they were not told how many rounds they would be playing to avoid endgame effects.

Procedure
This experiment was conducted virtually, and study sessions were posted on MTurk, where participants could enroll in the study either on the half hour or the hour. The study was designed to be live and completed in one sitting, to reduce suspicion that the interaction partner was not a real human. Participants were directed to a Qualtrics link where a tracker ensured that participants were located in the United States and were not using a VPN/VPS/proxy to mask their location. If they successfully passed this check, participants entered their own chat box with the researcher to receive a live study link for the session, which then directed them to the study itself.

First, participants had to read and agree to the informed consent screen. After consenting, participants were asked to fill out the demographic and political orientation questions. Afterward, they were directed to instructions on how to play the game. Each participant answered three quiz questions about how the game worked to test their understanding. Anyone who got less than two quiz questions correct was dropped.

Participants were told that both players would make decisions simultaneously, such that players would learn what their partner would have done regardless of whether the sender sent or kept the 6 tokens. It was necessary for decisions to be simultaneous to ensure that participants experienced a trust violation and saw the receiver’s cooperative behavior in the other rounds. For example, if the decisions were not simultaneous and the sender did not send tokens in a round, they would not have seen that the receiver would also not have sent tokens back. Therefore, the sender would never have truly had their trust violated because they would not have known that the receiver would not have acted in a trustworthy way.

A “trust violation” was operationalized as two rounds in a row of not receiving tokens back based on past research, as one round may be interpreted as a fluke rather than a violation of trust (Haselhuhn et al., 2015; Komorita & Mechling, 1967; Kuwabara et al., 2014; Lount et al., 2008). The preprogrammed receiver violated trust by deciding not to return tokens in rounds 1 and 2 (“early trust violation”) or rounds 10 and 11 (“late trust violation”; Kuwabara et al., 2014). These rounds were chosen to provide enough time between an early trust violation and a late trust violation to detect a possible difference between the two, while also providing enough distance between the end of the late trust violation and the last five rounds so that an effect could be measured. It additionally aligned with existing literature (Kuwabara et al., 2014; Lount et al., 2008). In every other round, the ostensible receiver was programmed to behave trustworthy, sending tokens back.

In the instructions, participants had been told that they and the other participants had been given an identifying letter that would be displayed alongside their political affiliation based on a calculation of their political orientation survey answers. After participants read the instructions, participants were told, “[y]ou have been assigned to the role of Sender and you have been paired with Receiver, “Participant W ­Liberal/Conservative [which was colored blue if Participant W was liberal and red if Participant W was conservative], with whom you will be paired with throughout the duration of the study.” Although participants were all assigned to be the sender, the political affiliation of whom they were partnered with was randomly assigned. They had also been told that the study was interested in decision-making as it related to the distribution of valuable resources between themselves and another participant.

For every decision screen (asking whether they wanted to send tokens) and profit screen (showing their earnings for each round), their interaction partner’s political ideology was included on the screen to increase ingroup/outgroup salience. Furthermore, random
late trust violations hinder subsequent cooperation

TABLE 1
Summary Results of the Decision in the First Round Based on Ingroup/Outgroup Pairings

<table>
<thead>
<tr>
<th>Political Affiliation</th>
<th>Ingroup</th>
<th>Outgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td>Liberal</td>
<td>60</td>
<td>1.00</td>
</tr>
<tr>
<td>Conservative</td>
<td>43</td>
<td>0.95</td>
</tr>
<tr>
<td>Overall</td>
<td>103</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Note. R^2 = .02

TABLE 2
Summary Results of Average Proportion of Trust in the Last 5 Rounds for all Four Conditions

<table>
<thead>
<tr>
<th>Trust Violation</th>
<th>Ingroup</th>
<th>Outgroup</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Early</td>
<td>43</td>
<td>0.93</td>
<td>0.24</td>
</tr>
<tr>
<td>Late</td>
<td>60</td>
<td>0.77</td>
<td>0.40</td>
</tr>
<tr>
<td>Overall</td>
<td>103</td>
<td>0.84</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Note. R^2 = .08

late trust violations hinder subsequent cooperation

Poststudy Questionnaire
Following the completion of the game, to measure attitudinal trust, participants completed a short trust scale with identifying letters added (Lount et al., 2008). They were asked to answer the following questions on a 7-point scale from 1 (not at all) to 7 (“very much”): (1) “How trustworthy is Participant W-Liberal/Conservative?” (M = 5.07, SD = 1.51, min = 1, max = 7); (2) “How much would you trust Participant W-Liberal/Conservative in the future?” (M = 4.93, SD = 1.67, min = 1, max = 7); and (3) “How honest is Participant W-Liberal/Conservative?” (M = 5.03, SD = 1.55, min = 1, max = 7). Lount et al. (2008) reported the value of Cronbach’s alpha to be α = .86, and the present found the value for Cronbach’s alpha to be α = .96. An average “trust score” was created based on the answers to the three questions with a higher score indicating greater attitudinal trust.

Afterward, participants completed two items measuring their suspicion adapted from Kuwabara and colleagues (2014) to see whether they were suspicious that they had not played against another real participant (the present study found the value of Cronbach’s alpha to be α = .88). They were first asked, “Did you suspect that your partner wasn’t an actual human” on a scale from 1 (not at all) to 3 (very much) (mode = 1). If participants answered a 2 or 3, they were directed to another page that asked, “If you suspected your partner wasn’t an actual person, how much did this affect your decisions in the study?” using the same 3-point scale (mode = 1). Participants who answered 3 for both questions were dropped from analyses, following Kuwabara and colleagues (2014).

Finally, participants were debriefed. At the beginning of the study, participants were told that they would have the opportunity to earn a bonus ranging from $0 to $2 (in addition to a flat payment of $2 for completing the 20-minute study) to incentivize them to try and maximize their profit in the game. Because all participants were playing against a computer and not another participant, they were all paid the full $2 bonus.

Results
All data were analyzed using R (Version 4.2.0) and RStudio (Version 2022.07.2). The significance level (α) was set a priori at .05, two-tailed, to determine statistical significance. Furthermore, although the final sample consisted of 208 participants, the study recruited 316 participants who completed the study and who met the necessary qualifications (i.e., located in the United States and identifying as politically liberal or conservative). The final sample excluded participants who were very suspicious (i.e., those who answered that they were very suspicious that their partner was not human and reported that this suspicion “very much” impacted their responses), and participants who scored less than a two out of the three questions correct on a comprehension quiz about how the game is played or less than 75% on attention check questions during the task. Although very suspicious participants were dropped from all main effects and interaction models, models run with those suspicious individuals included did not significantly impact the results.²

An additional supplementary model that included a term for the political affiliation (conservative or liberal) of the participant was run, though this also did not alter the significance of the results nor improve the fit of an analysis of variance model (F = 1.01, p = .32). A third supplementary model that considered the participant’s political affiliation, not as a binary variable, but as a continuous one, by calculating each participant’s political orientation score also did not alter the significance of the results nor improve the fit of the model,²

²Two additional models—the first in which anyone who indicated in which anyone who indicated a 3 on the either suspicion question (rather than only those who indicated a 3 on both) was dropped, and the second in which anyone who indicated a 3 on the first question (that they were very suspicious that their partner was not human)—were also run, but they did not impact the significance of the results.
The proportion of trust in the last 5 rounds of the interaction ($\alpha = .94$, modeled after Kuwabara and colleagues (2014). The proportion of trust ranged from 0 (did not give tokens in any of the rounds) to 1 (gave tokens in all five rounds). The means and standard deviations of average trust in the last 5 rounds are listed in Table 2.

It was predicted that the timing of the trust violation would impact subsequent behavioral trust (Hypothesis 1a), or in other words, that there would be a significant difference in the average proportion of trust in the last five rounds of the interaction between those who were randomly assigned to an early trust violation compared to those randomly assigned to a late trust violation. This prediction was supported. However, the results were the opposite of Kuwabara and colleagues’ (2014) and Lount and colleagues’ (2008) findings, which had found lower levels of behavioral trust among those who experienced an early trust violation. Instead, those who experienced a late

Decision in the First Round
A logistic regression was run to look at decisions in the first round, before participants saw any of their partner’s decisions. In this first round, participants only had information about their partner’s political affiliation, and they could either trust their tokens to their partner (coded as 1) or keep the tokens for themselves (coded as 0). The summary results can be found in Table 1. For all the following results, political affiliation, age, gender, and income did not impact the significance of the results.

Overall, mean trust was 0.95 ($SD = 0.22$) collapsed across conditions. However, a significant difference emerged between participants partnered with an ingroup member and participants partnered with an outgroup member. Participants more likely to trust their tokens to their partner in the first round if they were partnered with an ingroup member ($\beta = .07$, $SE = 0.03$, $p = .03$, 95% CI [.01, .13]), yielding an odds ratio of 4.73 (see Figure 3).

Interestingly, all 60 liberal participants paired with a liberal receiver gave their tokens in the first round. Moreover, among both liberal and conservative participants, paired with either an ingroup or outgroup member, trust in the initial round was high (ranging from 91% to 100%). This suggests that participants were testing out the relationship in this first round. Participants knew that there would be many rounds (even if they did not know the exact number of rounds) and understood that there was a clear benefit to trusting their partner. Trust being measured as a binary variable might have encouraged greater trusting behavior, especially when participants were partnered with someone who they expected to be more trustworthy (an ingroup member).

Trust in the Last Five Rounds
To determine whether subsequent trust was significantly different between those who experienced an early trust violation and those who experienced a late trust violation, and between those who were partnered with an ingroup member and those who were partnered with an outgroup member, a linear regression using ingroup identity (vs outgroup) and early trust violation (vs late) was run, also testing for their interaction. The dependent variable was the average proportion of trust in the last 5 rounds of the interaction. (The group of participants who were randomly assigned to early trust violation were not significantly different demographically (age, income, gender) from the group of participants randomly assigned to a late trust violation. Nor did the group of participants who were partnered with an ingroup member significantly differ demographically from the group of participants who were partnered with an outgroup member.)

\[ F = 1.30, p = .26 \]

Finally, the proportion of trust based on the demographic information that was collected (age, gender, and income) was regressed.
trust violation were significantly less trusting of their partners in the final five rounds than those who experienced an early trust violation, ($\beta = .85, \ SE = 0.05, \ p = .001, \ 95\% \ CI \ [0.06, 0.23], f^2 = .09$; see Figure 4). Furthermore, though it was expected that participants would show greater behavioral trust with ingroup members than with outgroup members (Hypothesis 2a), it was not the case that ingroup pairings demonstrated a higher average proportion of trust ($\beta = .00, \ SE = .04, \ p = .95$).

Controlling for age and income did not impact the results, nor did they significantly predict trust ($\beta = .00, \ SE = .00$ and $\beta = -.01, \ SE = .01$ respectively)—or improve the fit of the model ($F = 2.32, \ p = .13$ and $F = 1.33, \ p = .25$). However, as shown in Figure 5, men were significantly less trusting ($M = 0.82, \ SD = 0.36$) than women ($M = 0.92, \ SD = 0.22; \beta = -.11, \ SE = 0.04, \ p = .01, \ 95\% \ CI \ [-0.19, -0.02]$) which is consistent with some prior work (Haselhuhn et al., 2015).4

Thus, overall, only in the timing condition was there a significant difference in the average proportion of trust in the last five rounds of the interaction and it was those who experienced a late trust violation that were significantly less trusting than those who experienced an early trust violation.

A test for an interaction between the timing of the trust violations and whether the participant was partnered with an ingroup member was also conducted. This was looking at whether there were significant differences in trusting behavior between those who experienced an early trust violation versus a late trust violation depending on whether participants were partnered with an ingroup member or an outgroup member. Specifically, it was predicted that ingroup partners would show less behavioral trust if they experienced an early trust violation compared to a late trust violation (Hypothesis 3a) and that outgroup partners would show less behavioral trust if they experienced a late trust violation compared to an early one (Hypothesis 4a). However, the interaction effect was not significant ($\beta = .04, \ SE = .09, \ p = .63$), and therefore, both hypotheses were not supported. Given that the interaction effect was not significant, only the model with the main effects was retained.

**Trust Scale**

In addition to measuring behavioral trust by looking at the average proportion of trust in the last five rounds, an objective measure of participants’ behavior, this study also measured attitudinal trust to gauge participants’ subjective sense of trust using a trust scale (Lount et al., 2008), following the completion of the trust game. The higher the trust score, the greater the attitudinal trust. Including control terms for political affiliation, age, gender, and income did not impact the significance of any of the results.

It was expected that there would be a significant difference in trusting attitudes between those who experienced an early trust violation versus those who experienced a late trust violation (Hypothesis 1b). This was the case, and as demonstrated in Table 3 and Figure 6, those who experienced a late trust violation, on average, gave lower ratings of their partner than those who experienced an early trust violation ($\beta = 4.71, \ SE = .20, \ p = .01, \ 95\% \ CI \ [0.11, 0.95], \ f^2 = .03$). However, although it was predicted that attitudinal trust would be higher, on average, for those who were partnered with an ingroup member compared to an outgroup member (Hypothesis 2b), there was no significant difference between how participants rated ingroup partners and outgroup partners ($\beta = .05, \ SE = .21$, controlling for gender did improve the fit of the model ($F = 9.43, \ p = .002$).
Thus, only the timing condition produced significant differences in attitudinal trust and the direction of this difference (namely, that trust was significantly lower for those who experienced a late trust violation) matched the direction of the difference between early and late trust violation conditions for behavioral trust (the average proportion of tokens given by participants to their partners in the last five rounds).

A test for an interaction effect between the timing of the trust violation and whether the participant was partnered with an ingroup member for attitudinal trust, looking at whether there were significant differences in trusting attitudes between those who experienced an early trust violation versus a late trust violation based on whether participants were partnered with an ingroup member or an outgroup member, demonstrated that the timing of the trust violation especially impacted trust scores among ingroup members. Specifically, as shown in Figure 7, participants trusted their partner less when they experienced a late trust violation than when they experienced an early one ($\beta = .89, SE = .30, p < .001$), contrary to the prediction in Hypothesis 3b (which predicted the opposite). Likewise, Hypothesis 4b was also not supported, as trust scores were lower when there was a late trust violation, but not especially so for outgroup members ($\beta = .17, SE = .30, p = .57$).

Discussion

This study sought to explore how people’s behavioral and attitudinal trust is impacted by the timing of a trust violation (Kuwabara et al., 2014; Lount et al., 2008) by determining whether such responses are impacted based on whether people interact with others whom they know they share a social identity with. Specifically, participants were partnered with an ingroup member, or an outgroup member and their subsequent trusting behaviors and attitudes were measured based on whether they experienced an early or late trust violation, whom they were partnered with, and whether the impacts of the timing of trust violations on subsequent trust is itself dependent on whether one is partnered with an ingroup or outgroup member.

Although it was predicted that there would be an ingroup bias effect, both behaviorally and attitudinally, there were no significant differences in levels of trust between ingroup and outgroup pairings. Although this result on its own suggests that there was no ingroup bias, the significant difference between ingroup and outgroup pairings in the first round suggests that there was some ingroup effect; it just did not last very long. In other words, there may be biases that predispose individuals to trust an ingroup member more initially, but a history of cooperative (or uncooperative) behavior over time may be enough to overcome these initial biases. If this is true, it implies that ingroup biases are not a limiting factor in cooperative interactions and that trust between outgroup members is possible to a similar extent as between ingroup members. Future research should continue exploring this effect to understand better how experiences of trustworthy behavior overcome an ingroup bias effect. Future research should moreover test this effect with different levels of the salience of the ingroup and outgroup pairings and with different types of ingroups.

There was also no significant effect with regards to whether the effects of the timing of trust violations on subsequent behavioral trust was different based on whether one was paired with an ingroup member or an outgroup member. However, people especially trusted ingroup members more if they experienced an early trust violation compared to a late trust violation. This possibly suggests that these individuals were willing to or tried to forgive their partners early on to maintain positive

![FIGURE 7](image)

### TABLE 3

<table>
<thead>
<tr>
<th></th>
<th>Ingroup</th>
<th>Outgroup</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trust Violation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early</td>
<td>43</td>
<td>65</td>
<td>108</td>
</tr>
<tr>
<td>Late</td>
<td>60</td>
<td>40</td>
<td>100</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td>103</td>
<td>105</td>
<td>208</td>
</tr>
<tr>
<td><strong>n</strong></td>
<td>M</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td><strong>Early</strong></td>
<td>5.50</td>
<td>1.31</td>
<td>5.10</td>
</tr>
<tr>
<td><strong>Late</strong></td>
<td>4.61</td>
<td>1.57</td>
<td>4.93</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td>4.98</td>
<td>1.52</td>
<td>5.04</td>
</tr>
</tbody>
</table>

Note: $R^2 = .03$

---

5 Overall, the interaction was not significant ($\beta = .72, SE = 0.42, p = .09, 95\% CI [-0.11, 1.56], f^2 = .05$).
Late Trust Violations Hinder Subsequent Cooperation

Late Trust Violations Hinder Subsequent Cooperation

expectations about their ingroup partners’ willingness to cooperate (Yamagishi et al., 1999) and avoid the cognitive dissonance of their ingroup partner acting in an uncooperative manner. However, for those assigned to a late trust violation, those positive expectations had been compounded by a growing sense of trust over rounds of cooperative behavior by their ingroup partner. This created a more salient sense of betrayal when the violation finally occurred. Thus, even if there were no significant differences in how those paired with an ingroup member acted between an early and late trust violation, there were still differences in how individuals thought about their ingroup member. Attitudes and behavior may similarly track one another, but they are not the same. It is possible that behavior, as an application of attitudinal trust, does not accurately reflect individuals’ subjective sense of how much they trust their partner, at least in certain cases. Participants might have wanted to believe that their ingroup partner was a good and trustworthy person, and this was much easier to do when trust was violated early on.

Moreover, some past research suggested that early trust violations hinder subsequent cooperation to a greater extent than late trust violations for United States participants (Kuwabara et al., 2014; Lount et al., 2008), the results here suggest the opposite. Late trust violations were more detrimental to later trusting behavior and attitudes. There are a few possibilities for why late trust violations hinder cooperation to a greater extent. First, a recency effect might have made late trust violations more memorable, especially in a task oriented toward short-term goals, although this study had the late trust violation occur in rounds 10 and 11 to allow multiple rounds of buffer time before trusting behavior was measured in the last five rounds. Second, there may be support for Morrison & Robinson’s (1997) theory, which posits that betrayal hurts more when there is a longstanding pattern of cooperativeness than in situations where there is a history of untrustworthiness. It felt worse for those who experienced nine consecutive rounds of cooperative behavior when their trust was betrayed in rounds 10 and 11 because a cooperative relationship had been built over so many rounds. That same history does not exist for those who experienced an early trust violation. Third, if it is true that individuals start with little trust that builds over time (Blau, 1964; Rempel et al., 1985; Rousseau & McLean Parks, 1993), then the early trust violation might not have “stung” as much because there was little trust at the beginning of the relationship. In fact, uncooperative behavior at the very beginning of the interaction might have made the subsequent rounds of consecutive cooperation seem relatively more positive than they otherwise might have seemed, disproportionately increasing trust by the last five rounds (Sherif & Hovland, 1961).

Still, it is interesting that these results do not align with the findings of Kuwabara and colleagues (2014) or Lount and colleagues (2008). However, there are some notable differences between the samples used in their studies and the sample used in this study. Both Kuwabara and colleagues (2014) and Lount and colleagues (2008) used university students (specifically students at an elite metropolitan university and a large midwestern university, respectively). Psychological studies have historically relied on college students as their sample, although these populations are often much more homogenous than the general public (Henrich et al., 2010; Peterson & Merunka, 2014; Reber & Smith, 2023). Colleges are also disproportionately attended by those of a higher socioeconomic status, women, and Asian American and White individuals (Reber & Smith, 2023).

In contrast, this study utilized a more diverse sample, which was mostly male, liberal, of lower socioeconomic status (with a household income of less than $75,000), and older (with an average age of around 40 years old). Although racial information was not collected, MTurk’s racial demographics are similar to those of the U.S. population (Moss & Litman, 2020). Overall, MTurk is slightly more diverse than the average internet sample but much more diverse than the average university sample (Buhrmester et al., 2011). Furthermore, Lount and colleagues (2008) had 138 participants across four timing conditions, and Kuwabara and colleagues (2014) recruited 93 participants across two timing conditions. This study recruited 208 participants across two timing conditions (and four conditions overall), a significantly larger sample of participants per condition than either of these studies. This sample size of 208 excluded those who were very suspicious and those who were not paying attention during the task to increase the quality of the sample. As a result, there was much greater internal credibility in the present study. Future research should replicate this study with special attention to the sample, particularly the representativeness and demographics of the sample, to better understand the effects that the timing of trust violations has on subsequent trust and cooperation.

Still, this experiment has certain limitations that must be addressed, and that signal a need for more research in this field. First, trust may manifest differently in different types of interaction. This study looked at a particular kind of trust game, but there are many other kinds of situations that require trust, and that should be explored. Specifically, more research should be done utilizing more real-world settings, beyond the kind of trust game in the present study. In conducting an
experiment (the gold standard for establishing causality) with a more demographically diverse sample, this study has high internal validity. Although internal validity and external validity have long been seen as trade-offs, internal validity is a necessary prerequisite to external validity (Gërxhani & Miller, 2022). Therefore, although this study used a trust game to demonstrate trusting behaviors and attitudes in the participants, it is a strong first step in testing a causal theory that future research can do using more real-world settings.

Moreover, this study was conducted completely online using MTurk. Although some have expressed concerns about the quality of MTurk samples, research has shown that, in addition to the greater demographic diversity, MTurk can produce high-quality data (Buhrmester et al., 2011). For example, Wright and colleagues (2021) were able to replicate, using an MTurk sample, the findings of previous research that had utilized a student sample. In addition, this study took steps to increase the quality of the sample, including the addition of attention checks. Nonetheless, individuals’ willingness to trust may be different when interacting with others face-to-face, or, at the very least, the sense of realism may be enhanced by playing the game in-person, even without seeing the other players. However, this study took significant measures to heighten the feeling of authenticity, and the results demonstrate that trust can be built—and broken—even when the interactants feel separated via their screens and the added layer of anonymity. This is especially important to consider in the modern world of endless online interactions and communications. Still, future studies should consider whether there are significant differences when conducted in-person versus online, and what the implications of those differences might be.

In addition, interactions over more extended periods may look different than those during a 20-minute experiment, especially if individuals can interact in real-time to a greater extent (such as via a chat function or face-to-face communications). With enough time and enough interactions, even a late trust violation becomes an early trust violation. This study focused on how the timing of a trust violation impacted subsequent trust within a fixed time-interval by comparing the timing conditions relative to one another. However, more prolonged interactions may also provide evidence of recovery times and recovery rates from trust violations. Because this study was conducted over 20 rounds, it did not provide a clear picture regarding the recovery times for early trust violations versus late trust violations nor a complete picture of the recovery rates. Understanding these phenomena may provide more insight into the effects of the timing of trust violations on subsequent trust. Finally, this study looked at trust as a binary variable. Participants could either trust their tokens to their partner or withhold them. Although there are some situations in which trust is a binary decision, there are other situations in which it is not. In many cases, trust is measured on a continuous scale. Future research should consider how the timing of trust violations, as well as the ingroup bias, may be impacted by the ability to trust only partially.

Ultimately, however, this experiment sheds light on a meaningful topic—trust. The results demonstrate that although early impressions may be influential, maintaining a pattern of trustworthiness and cooperativeness is ultimately more significant in maintaining healthy and mutually beneficial relationships. Even more importantly, the natural and ubiquitous categorization of individuals into ingroups and outgroups does not preclude the ability for people to trust one another or cooperate for mutually beneficial outcomes. It is possible to work together despite differences.

References
Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Exceptions that prove the rule—Using a theory of motivated social cognition to account for

Author Note. Material and data for this study can be accessed at https://osf.io/x4ahs/.
I have no known conflict of interest to disclose. This study was supported by the Charles Lafitte Foundation Program for Research in Psychology & Neuroscience at Duke University, as well as the Duke Undergraduate Research Support Office’s Independent Study Grants. Special thanks to Ashley Harrell for her mentorship, guidance, support, and encouragement during this project.
Correspondence concerning this article may be addressed to Megan Gerges. Email: megangerges@gmail.com
Psi Chi Journal of Psychological Research: 2023 Reviewers

We sincerely appreciate the hard work on the part of the following individuals who each completed at least one review in 2023. Without the assistance of such dedicated professionals, Psi Chi Journal would not be able to function.
—Steve Rouse (Editor)
Find Your True North in Psychology

Getting good grades in college is the first step to obtaining a rewarding career in psychology. The second step is membership in Psi Chi, the International Honor Society in Psychology.

Could you use some help identifying and achieving your unique calling in life?

Become a member today.
Visit https://www.psichi.org/page/become_member

Discover Your Next Job in Psychology

Thousands of psychology-related jobs are waiting for you on Psi Chi’s unique Career Center.
• Receive a free resumé evaluation
• Obtain email alerts when relevant positions are submitted
• See career insights, such as wages and needed skills
• Explore advice for interviewing, growth, and more

Get started at https://jobs.psichi.org
Publish Your Research in *Psi Chi Journal*

Undergraduate, graduate, and faculty submissions are welcome year round. Only one author (either first author or coauthor) is required to be a Psi Chi member. All submissions are free.

Reasons to submit include

- a unique, doctoral-level, peer-review process
- indexing in PsycINFO, EBSCO, and Crossref databases
- free access of all articles at psichi.org
- our efficient online submissions portal

View Submission Guidelines and submit your research at [www.psichi.org/?page=JN_Submissions](http://www.psichi.org/?page=JN_Submissions)

---

**Become a Journal Reviewer**

Doctoral-level faculty in psychology and related fields who are passionate about educating others on conducting and reporting quality empirical research are invited become reviewers for *Psi Chi Journal*. Our editorial team is uniquely dedicated to mentorship and promoting professional development of our authors—Please join us!

To become a reviewer, visit [www.psichi.org/page/JN_BecomeAReviewer](http://www.psichi.org/page/JN_BecomeAReviewer)

---

**Resources for Student Research**

Looking for solid examples of student manuscripts and educational editorials about conducting psychological research? Download as many free articles to share in your classrooms as you would like.

Search past issues, or articles by subject area or author at [www.psichi.org/journal_past](http://www.psichi.org/journal_past)

---

**Learn About Psi Chi**

Psi Chi is the International Honor Society in Psychology. Membership is primarily open to undergraduates, graduate students, transfer students, full-time and part-time faculty members, and alumni.

See membership benefits and a link to apply at [www.psichi.org/page/member_benefits](http://www.psichi.org/page/member_benefits)