

Denials and apologies: pathways to reconciliation

Tatiana Iwai

Department of Business Administration, Insper, São Paulo, Brazil, and

João Vinicius França Carvalho

Department of Accounting and Actuarial Science, USP, São Paulo, Brazil

Abstract

Purpose – This paper aims to examine how verbal responses (denials vs apologies) following a trust violation in cooperative relationships influence reconciliation by changing attributions of responsibility for the transgression and transgressor's perceived integrity. Additionally, the moderating role of perceived sincerity of the response is examined.

Design/methodology/approach – Two experimental studies were conducted with 465 participants. Hypotheses were tested using ordinary least squares (OLS) regressions and moderated serial mediation analyses with bootstrapping procedures.

Findings – In the occurrence of integrity-based trust violations, denials are more effective than apologies to repair trust. The positive indirect effects of these verbal responses on reconciliation are explained by a two-part mediating mechanism (attribution of responsibility followed by transgressor's perceived integrity). Additionally, when responses are perceived as highly credible, denials are much more effective in deflecting blame than apologies.

Research limitations/implications – This study contributes to the literature on trust repair by examining when and why managers' verbal responses to breaches of trust may be more or less effective in restoring cooperative relationships.

Practical implications – Managers must be aware that their perceived integrity following a breach of trust is influenced by the level of responsibility taken. Therefore, they should choose wisely which defensive tactics (apologies or denials) to use.

Social implications – As trust plays a central role in many cooperative relationships, choosing an appropriate response after a transgression is critical to solving conflicts both within and between organizations.

Originality/value – This work contributes to the reconciliation literature by uncovering the underlying cognitive mechanisms and boundary conditions by which different verbal responses influence reconciliation.

Keywords Reconciliation, Trust violations, Apologies, Denials

Paper type Research paper

Introduction

Trust plays a central role in developing and maintaining many collaborative relationships (Fulmer & Gelfand, 2012). However, despite its importance, there are many situations in which trust is violated, leading to a broad array of negative



consequences ranging from lower cooperation (Holtgrave, Nienaber, Tzafrir, & Schewe, 2019) to severe organizational level failures (Gillespie & Dietz, 2009). In an attempt to mitigate these adverse effects and facilitate reconciliation, verbal responses (e.g. apologies, excuses, denials and justifications) are usually used as repair strategies to rebuild broken trust (Lewicki & Brinsfield, 2017).

Among these verbal strategies, *apologies* and *denials* have received particular attention in the trust literature (Fuoli, van de Weijer, & Paradis, 2017; Kim, Ferrin, Cooper, & Dirks, 2004; Kim & Song, 2021). Scholars have been focusing on examining which one is more effective to repair trust and uncovering the causal mechanisms underlying that effect (Tomlinson & Mayer, 2009). The present research aims to contribute to this debate by proposing and testing a moderated serial mediation model designed to examine:

- the relationship between attribution of responsibility for a negative event and perceived integrity of the transgressor as comediators of the link between the type of verbal response and willingness to reconcile; and
- the perceived sincerity of the response as a moderator of the relationship between the type of verbal response and attribution of responsibility.

We draw on attribution theories (Kelley, 1967; Weiner, 1985) to propose that, following a transgression, denials lead to lower levels of responsibility attributions than apologies, which, in turn, influence transgressors' perceived integrity and ultimately to victims' willingness to reconcile. That is, while an apology implies acknowledging responsibility for the transgression and expressing regret for one's actions, a denial exempts transgressors from any responsibility. The resulting different levels of responsibility influence victims' perceptions of transgressors' trust-relevant qualities (i.e. perceived integrity), affecting reconciliation.

We further propose that the perceived sincerity of the response might explain why, following a transgression, trust is repaired more or less successfully. Whatever the message – taking the blame or rejecting it – its effectiveness depends on how credible it is (Lewicki & Bunker, 1996). Few studies, however, have investigated sincerity empirically, either at the individual (Basford, Offermann, & Behrend, 2014; Tomlinson, Dineen, & Lewicki, 2004) or the organizational level (Brühl, Basel, & Kury, 2018). Moreover, most of these studies focused on the perceived sincerity of an apology, while its effects on other alternative verbal responses remain unexamined.

This research provides some significant contributions to the reconciliation literature. First, although some studies have theoretically differentiated between types of verbal responses according to the level of responsibility taken (Kim *et al.*, 2004; Kim, Dirks, Cooper, & Ferrin, 2006; Lewicki & Brinsfield, 2017; Tomlinson & Mayer, 2009), little empirical research has been conducted to directly ascertain whether denials and apologies lead to varying levels of attribution of responsibility and their effects on willingness to reconcile. Second, by including the transgressor's perceived integrity as another causal mechanism in the model, this study provides the first empirical evidence of how responsibility attributions and perceived integrity function as comediators in a serial causal sequence. In other words, we show that the two mediator variables are fundamental parts of the cognitive process associated with repairing a damaged relationship.

Finally, this work expands our understanding of the boundary conditions that limit the effectiveness of verbal explanations by examining how the perceived sincerity of an explanation may influence the effects of verbal responses on the reconciliation process.

Theory and hypotheses

Trust can be defined as “a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another” (Rousseau, Sitkin, Burt, & Camerer, 1998, p. 395). Therefore, *trust* can be considered breached when victims perceive that others did not live up to their expectations (Elangovan, Auer-Rizzi, & Szabo, 2007). However, the negative perceptions resulting from a transgression are not definitive. Explanations provided following a trust violation can change or invalidate initial negative judgments about the event (Lewicki & Bunker, 1996; Tomlinson *et al.*, 2004).

Given the importance of these explanations, researchers have been particularly interested in comparing the effectiveness of two widely used verbal responses, *denial* or *apology* (Tomlinson & Mayer, 2009). Although apologies involve admitting responsibility for the transgression, they can foster forgiveness because they also show regret (Lewicki, Polin, & Lount, 2016; Lewicki & Brinsfield, 2017). By communicating an intention to avoid similar hurtful behaviors in the future, such expressions of remorse can compensate for the adverse effects of taking the blame and be very effective in restoring broken trust (Lewicki & Bunker, 1996; Tomlinson *et al.*, 2004). In contrast, by claiming that the accusation is false, denials may be effective since transgressors deflect responsibility for the event to situational factors and make any act of reparation unnecessary (Tomlinson & Mayer, 2009).

Prior works have found that the effectiveness of each response depends on the type of trustworthiness violated. Apologies are more effective when violations involve matters of competence, while denials are more useful for those involving integrity (Ferrin, Kim, Cooper, & Dirks, 2007; Kim *et al.*, 2004, 2006; Kim, Cooper, Dirks, & Ferrin, 2013). The rationale lies in how information about competence and integrity is processed (Kim, Diekmann, & Tenbrunsel, 2003; Kim *et al.*, 2004; Kim *et al.*, 2006). Competence-related failures have less weight than competence-related successes because while a competent person may be expected to exhibit performance at many levels (including occasional failures, according to their motivation or external constraints), one who is incompetent will only show results that are commensurate with their low level of competence.

However, information about integrity is assessed differently. While a dishonest person may act honestly or dishonestly according to the incentives and controls available, an honest person will never engage in unethical behavior. Thus, being responsible for a single integrity-related transgression conveys such an evident sign of character flaw that any benefits from an apology (remorse and an indication of future redemption) would be canceled out by their costs (accepting responsibility). Based on this reasoning, we hypothesize:

- H1.* Denials are associated with a greater willingness to reconcile than apologies in integrity-based trust violations.

Given the different effects of denials and apologies on trust repair, one interesting issue relates to the underlying cognitive mechanisms that might explain these effects. Attribution models (Kelley, 1967; Weiner, 1985) can be particularly useful as they can shed light on how verbal responses can alter victims' interpretation of a transgression (Elangovan *et al.*, 2007; Kim, Dirks, & Cooper, 2009; Tomlinson & Mayer, 2009).

In the aftermath of a trust violation, verbal responses provided primarily seek to influence the attribution of responsibility for the episode. However, as Weiner (1985, p. 13) argues, responsibility attributions are “not all-or-none inference, but vary in magnitude and degree.” Therefore, choosing to respond with an apology or a denial essentially means choosing what level of responsibility and culpability to take for the transgression.

By taking the blame with an apology, transgressors acknowledge greater responsibility for the transgression. In contrast, when transgressors deny responsibility and displace the blame onto some external factor over which they had little or no control, they can be totally (or at least partially) exonerated from the accusation (Elangovan *et al.*, 2007). Moreover, by presenting a plausible alternative cause for the negative outcome, denials may discount the first explanation of blame and thus increase victims' willingness to reconcile (Kelley, 1973). Therefore, the argument goes that the type of verbal response used, whether denial or apology, will lead to different levels of attribution of responsibility, which will influence victims' willingness to reconcile (Fuoli *et al.*, 2017; Myers, 2016). As such, responsibility attribution acts as a mediator of the relationship between verbal responses and reconciliation. Thus, we hypothesize:

H2. Attribution of responsibility mediates the relationship between the type of response and willingness to reconcile.

The effectiveness of any verbal response depends on the victim's perception that the transgressor deserves a second chance. Sincere explanations can help mitigate negative reactions following a transgression and increase victims' willingness to reconcile (Lewicki & Bunker, 1996). However, contrary to previous research that had an implicit understanding that sincerity has an unvarying influence on the effectiveness of any type of verbal response (Risen & Gilovich, 2007; ten Brinke & Adams, 2015), we argue that there is an interactive effect between perceived sincerity and type of response provided.

Extant research has argued that when transgressors apologize, their responsibility for the event is undisputable (Fuoli *et al.*, 2017). That happens because apologies are usually taken as credible and conclusive evidence of culpability and, in some cases, may even be used as formal evidence in legal proceedings (Myers, 2016). Prior work has found that even when an apology is false or has been coerced, people continue to be skeptical and reluctant to believe that an individual will own up to a transgression they did not commit (Kassin & Wrightsman, 1980). On the other hand, when transgressors respond with denials, in the absence of any unequivocal evidence of blame, victims may give them the benefit of the doubt (Bradford & Garrett, 1995). Therefore, perceptions of sincerity matter more for denials than for apologies. That is, the effectiveness of denials in reducing responsibility depends on the victim's assessment of how credible they are. Thus, we hypothesize:

H3a. Perceived sincerity of the response moderates the relationship between the type of response and attribution of responsibility, so that the less sincere the response is perceived to be, the smaller the difference between the effects of denials and apologies on responsibility attributions.

H3b. Perceived sincerity of the response moderates the indirect effect of type of response on willingness to reconcile via attribution of responsibility.

Prior work on trust repair (Tomlinson & Mayer, 2009) has theorized that the level of trust among individuals after a violation depends on the capacity of verbal responses to restore some important qualities of the other's perceived trustworthiness (e.g. competence, benevolence or integrity) damaged by the transgression. That is, the beliefs one holds about the other shape and influence their intentions to act in a particular way toward that other (Holtgrave *et al.*, 2019). In line with this, Kim *et al.* (2004) found support for the indirect effect of verbal response on trust via perceptions of the transgressor's trustworthiness. Based on this reasoning, we hypothesize the following:

H4. Perceived integrity mediates the relationship between the type of response and willingness to reconcile.

Although attributions of responsibility and perceived trustworthiness (e.g. perceived integrity) are both considered influential in trust judgments, they have yet to be studied in combination. Moreover, their interrelationship has not been examined empirically yet to the best of our knowledge. We, therefore, theorize that the effects of denials and apologies on trust repair are explained by a two-part mediating mechanism (attribution of responsibility followed by perceived integrity). In the occurrence of a trust violation, the verbal response provided leads to a certain attribution of responsibility for the event. Such attribution shapes how victims perceive the other, influencing their willingness to reconcile. Consistent with it, prior work has found that transgression intentionality shapes the impression formed about transgressors, which, in turn, influences victims' forgiveness (Struthers, Eaton, Shirvani, Georghiou, & Edell, 2008). Based on this reasoning, we hypothesize the following (as shown in Figure 1):

H5. The relationship between the type of response and willingness to reconcile is sequentially mediated by the attribution of responsibility and perceived integrity, and the perceived sincerity of the response moderates this indirect effect.

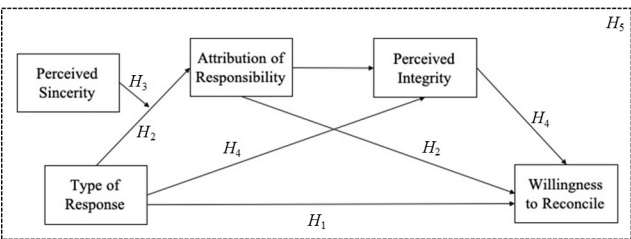
Research overview

We conducted two experimental studies to test our hypotheses. In both, we used hypothetical scenarios to control for possible confounding effects and provide greater internal validity to our results (Donovan & Priester, 2017). Study 1 examined *H1*, *H2*, *H3a* and *H3b*. Study 2 aimed to replicate Study 1 findings and test our full moderated serial mediation model (*H4* and *H5*). Each study explores one specific context (scenario) and uses different samples (difference in demographics), adding unique value to establish the robustness of the predicted relationships and providing greater generalizability to our results.

Study 1

Sample. We recruited 264 US-based participants from Amazon Mechanical Turk (MTurk) to participate in our study for \$0.75. We used two qualifications to screen participants in MTurk: participants were required to be located in the USA and have a minimum of 95% approval rate in previous tasks. Prior work has shown that data gathered from MTurk workers are as reliable as other gathered using more traditional methods in behavioral research (Hulland & Miller, 2018). Thirteen (4.92%) participants missed at least one question

Figure 1.
Moderated serial
mediation model of
the relationship
between type of
response and
willingness to
reconcile



checking for attention and were excluded from the analysis. When all data were included in the analysis, results remained the same. The final sample was 251 participants (52.19% women, $M_{\text{age}} = 41.35$, $M_{\text{work experience}} = 23.9$ years, 66.53% employed full-time).

Procedure. Once participants consented to participate in the study, they read a scenario specially created for this study to manipulate both types of verbal responses following a trust violation. Participants were asked to imagine themselves as a manager of *Bionic*, an industrial design company that has joined a partnership with Peter, the owner of another company. Participants were then informed of a breach of trust between them: Peter was going to reduce the number of his employees assigned to their joint project by 40%, yielding negative consequences for *Bionic*. Participants were then randomly assigned to one of two conditions (denial or apology). In the **denial** condition, participants read that Peter denied that it was his fault, claiming that a last-minute change of plans had happened due to factors beyond his control. According to him, “he had lost some of his most qualified workers because they were hired by a former employee who started a new business.” In the **apology** condition, participants read that Peter admitted he had decided to allocate some employees and resources to a new project offered by another business partner. According to him, “it was an unmissable business opportunity. He said he regretted not having kept his word and deeply apologized.” After reading the scenario, participants answered questions regarding the variables of our model, followed by a manipulation check (91% of the participants correctly indicated whether Peter had denied or apologized, supporting the effectiveness of our manipulation).

Measures.

Attribution of responsibility. To measure attribution of responsibility, we used four items adapted from [Struthers, Miller, Boudens, and Briggs \(2001\)](#) and [Russell’s \(1982\)](#) causal dimension scale to fit our scenario (1 = strongly disagree to 5 = strongly agree). A sample item is: “Peter could be considered responsible for this negative episode.” The Cronbach’s alpha was 0.74.

Perceived sincerity. On a five-point Likert scale (1 = strongly disagree to 5 = strongly agree), we measured perceived sincerity using four items adapted from [Basford et al. \(2014\)](#) and [Hornsey et al. \(2020\)](#) to fit the story of our scenario. A sample item is: “Peter was sincere about his reasons for reducing the number of employees assigned to the project.” The Cronbach’s alpha was 0.85.

Willingness to reconcile. Three items from [Tomlinson et al. \(2004\)](#) were used to assess willingness to reconcile on a seven-point scale (1 = very unlikely to 7 = very likely). A sample item is: “To what degree are you willing to let Peter try to reconcile the relationship with you, given Peter’s action?” The Cronbach’s alpha was 0.87.

Results. We conducted confirmatory factorial analyses to test the independence of our measures. Results indicated that the three-factor model fit the data well (including willingness to reconcile, responsibility attributions and perceived sincerity): [$\chi^2(41, N = 251) = 99.08$, comparative fit index (CFI) = 0.95, normed fit index (NFI) = 0.92, Tucker Lewis index (TLI) = 0.94, root mean square error of approximation (RMSEA) = 0.075, standardized root mean square residual (SRMR) = 0.047]. This model yielded better fit indices than the two other alternative models: one-factor model: [$\chi^2(44, N = 251) = 600.60$, CFI = 0.536, NFI = 0.52, TLI = 0.40, RMSEA = 0.22, SRMR = 0.16]; two-factor model: [$\chi^2(43, N = 251) = 322.09$, CFI = 0.77, NFI = 0.74, TLI = 0.70, RMSEA = 0.16, SRMR = 0.15]. Additionally, the chi-squared test showed that the three-factor model was significantly better than the one-factor [$\Delta\chi^2(3, N = 251) = 501.52, p < 0.001$] or the two-factor models [$\Delta\chi^2(2, N = 251) = 223.01, p < 0.001$]. These results confirm that the variables used represent separate constructs.

Supporting *H1*, a one-way ANOVA revealed that denials ($M = 4.03$, $SD = 1.25$) were associated with greater willingness to reconcile than apologies ($M = 3.18$, $SD = 1.33$),

[$F(1,250) = 24.87, p < 0.001, \eta^2 = 0.09$]. Supporting *H2*, which proposed the mediating role of attributions of responsibility, a simple mediation analysis (Hayes' PROCESS macro, Model 4) showed that denials (dummy = 0) lead to lower responsibility attributions than apologies (dummy = 1) ($B = 0.68, SE = 0.08, p < 0.001$), which, in turn, lead to greater willingness to reconcile ($B = -0.88, SE = 0.12, p < 0.001$). The indirect effect was significant [$B = -0.60, SE = 0.10, 95\% CI = (-0.81, -0.41)$].

H3a predicted that the perceived sincerity of the response moderates the relationship between the type of response and attributions of responsibility for the transgression. Using Hayes' PROCESS macro (Hayes, 2017, Model 1), we found that the interaction term was significant [$B = 0.48, SE = 0.09, 95\% CI = (0.30, 0.65)$]. Thus, *H3a* is supported.

Figure 2 illustrates the pattern of interaction. Simple slope analysis revealed that when perceived sincerity is low (one SD below the mean), the type of response has mild effects on responsibility attributions – denials lead to slightly lower levels of responsibility attributions than apologies (simple slope = 0.38, $t = 3.89, p < 0.001$). In contrast, when perceived sincerity is high (one SD above the mean), the difference is much more salient – denials lead to much lower responsibility attributions than apologies (simple slope = 1.12, $t = 11.32, p < 0.001$).

H3b stated that perceived sincerity moderates the indirect effect between the type of response and willingness to reconcile through attributions of responsibility. To test it, we used Hayes' PROCESS macro (Model 7) with bootstrapping procedure (10,000 resamples) to calculate bias-corrected confidence intervals. We examined the conditional indirect effect of type of response on willingness to reconcile via responsibility attributions ($X*W \rightarrow M1 \rightarrow Y$) at three levels of perceived sincerity: 1 SD below the mean [$B = -0.34, SE = 0.10, 95\% CI = (-0.53, -0.15)$]; the mean [$B = -0.66, SE = 0.11, 95\% CI = (-0.88, -0.45)$]; and 1 SD above the mean [$B = -0.99, SE = 0.15, 95\% CI = (-0.67, -0.14)$]. The index of moderated mediation was significant [$B = -0.25, SE = 0.17, 95\% CI = (-0.90, -0.41)$]. Taken together, these results support *H3b*.

Study 2

Sample. Study 2 used a sample of Brazilian working adults recruited via a snowball sampling procedure to participate in an online survey. The final sample was 214

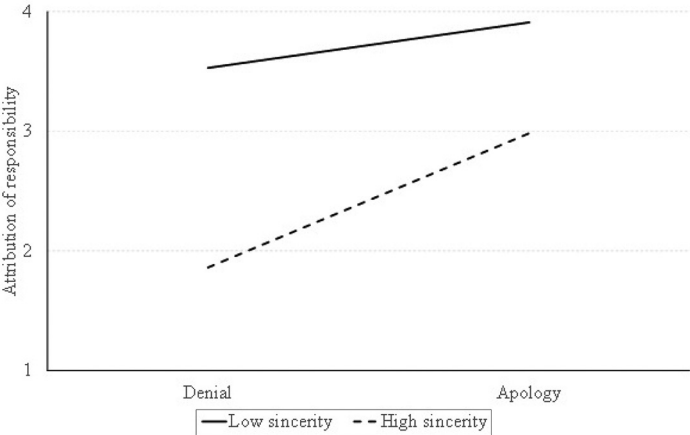


Figure 2.
Plot of interaction for
perceived sincerity
and type of response
predicting attribution
of responsibility

participants (61.70% women, $M_{\text{age}} = 32.07$, $M_{\text{work experience}} = 11.86$ years, 94.40% have at least an undergraduate degree).

Procedure. After completing the consent form, participants read a scenario adapted from Tomlinson *et al.* (2004). The original scenario depicted a deal negotiated between two small business owners. In this context, a violation in a negotiated agreement was described. We used this same background information across conditions and adapted the scenario to fit our two conditions: apology and denial.

In the scenario, participants were told to assume the role of the owner of a printing company that supplied point-of-sale advertising materials. One of the customers was Peter's company. Participants read that Peter had recently ordered a big new batch of labels. However, some days later, Peter decided to reduce the order substantially.

Similar to Study 1, participants were randomly allocated to one of two conditions: denial or apology. In the **denial** condition, participants read that Peter denied that it was his fault: "the reduction in the order had been unexpected and due to factors beyond his control since he had experienced a drop-off in business." In the **apology** condition, participants read that Peter admitted that he had decided to order part of the labels from another company: "Peter said that the price they were charging was much lower. He said he now regretted not having kept the whole order and deeply apologized." After reading the scenario, participants answered questions regarding the variables of our model, followed by a manipulation check (81% of the participants correctly indicated whether Peter had denied or apologized, supporting the effectiveness of our manipulation).

Measures. The measures of attribution of responsibility and willingness to reconcile were the same from Study 1.

Perceived integrity. We used Mayer and Davis (1999) scale to measure perceived integrity. Five items on a five-point Likert scale (1 = strongly disagree to 5 = strongly agree) were used. A sample item is: "I have never had to wonder whether Peter will stick to his word." The Cronbach's alpha was 0.77.

Results. We conducted confirmatory factorial analyses to ascertain the distinctiveness of the study variables. Results indicated that a four-factor model (including attribution of responsibility, perceived sincerity, perceived integrity and willingness to reconcile) yielded a better fit [$\chi^2(98, N = 214) = 150.90$, CFI = 0.95, NFI = 0.88, TLI = 0.94, RMSEA = 0.005, SRMR = 0.006] than the other alternative models: one-factor model: $\chi^2(104, N = 214) = 627.49$, CFI = 0.53, NFI = 0.49, TLI = 0.46, RMSEA = 0.15, SRMR = 0.14; two-factor model: $\chi^2(103, N = 214) = 590.67$, CFI = 0.565, NFI = 0.52, TLI = 0.49, RMSEA = 0.15, SRMR = 0.14; and three-factor model: $\chi^2(101, N = 214) = 543.08$, CFI = 0.61, NFI = 0.56, TLI = 0.53, RMSEA = 0.14, SRMR = 0.13.

Study 2 replicated Study 1 results. Consistent with *H1*, compared to apologies ($M = 4.58$, $SD = 1.15$), denials ($M = 5.00$, $SD = 0.91$) were associated with greater willingness to reconcile [$F(1,212) = 8.72$, $p = 0.0035$, $\eta^2 = 0.0395$]. We also found support for *H2*, as a simple mediation analysis (Hayes' PROCESS macro, Model 4) revealed that denials (dummy = 0) lead to lower responsibility attributions than apologies (dummy = 1) ($B = 0.51$, $SE = 0.09$, $p < 0.001$), which, in turn, lead to greater willingness to reconcile ($B = -0.40$, $SE = 0.10$, $p < 0.001$). The indirect effect was significant [$B = -0.21$, $SE = 0.08$, 95% CI = $(-0.37, -0.08)$]. Additionally, in line with *H3a*, perceived sincerity showed a moderating effect on the type of response-responsibility attributions link [$B = 0.36$, $SE = 0.12$, 95% CI = $(0.12, 0.61)$], as shown in Table 1. Using Hayes' PROCESS macro (Model 7) with bootstrapping procedure (10,000 resamples), results also supported *H3b*, as the conditional indirect effect of the moderated mediation ($X*W \rightarrow M1 \rightarrow Y$) was significant at all levels of perceived sincerity: 1 SD below the mean [$B = -0.14$, $SE = 0.07$, 95% CI = $(-0.30, -0.03)$];

Table 1.
Study 2: PROCESS
results –
unstandardized OLS
regression

Variables	B	SE	p-value
<i>Direct effect on responsibility attribution ($R^2 = 0.4478$)</i>			
Constant (Denial)	4.4534	0.2632	0.0000
Type of response (Apology)	−0.4247	0.3733	0.2566
Perceived sincerity	−0.4148	0.0960	0.0000
Type of response*Perceived sincerity	0.3616	0.1248	0.0042
<i>Direct effect on perceived integrity ($R^2 = 0.5816$)</i>			
Constant (Denial)	4.1997	0.1848	0.0000
Type of response (Apology)	−0.1717	0.0755	0.0240
Attribution of responsibility	−0.4593	0.0531	0.0000
<i>Direct effect on willingness to reconcile ($R^2 = 0.4820$)</i>			
Constant (Denial)	3.2256	0.6230	0.0000
Type of response (Apology)	−0.0889	0.1388	0.5224
Attribution of responsibility	−0.0614	0.1122	0.5849
Perceived integrity	0.7427	0.1250	0.0000
<i>Structural paths (Model 83)</i>			
	Spotlight (W)	Indirect effects Estimates (SE)	95%CI with bias correction
X*W → M1 → Y	−1SD (2.18)	−0.0224 (0.04)	[−0.1160;0.0676]
	Mean (2.95)	−0.0395 (0.08)	[−0.1986;0.1055]
	−1SD (3.73)	−0.0566 (0.11)	[−0.2955;0.1427]
	Difference (±1SD)	−0.0342 (0.72)	[−0.2052;0.0829]
X → M2 → Y	–	−0.1275 (0.06)	[−0.2551;−0.0172]
X*W → M1 → M2 → Y	−1SD (2.18)	−0.1245 (0.06)	[−0.2495;−0.0322]
	Mean (2.95)	−0.2196 (0.06)	[−0.3562;−0.1114]
	−1SD (3.73)	−0.3147 (0.09)	[−0.5027;−0.1636]
	Difference (±1SD)	−0.1902 (0.08)	[−0.3583;−0.0533]
X → Y	–	−0.0889 (0.14)	[−0.3625;0.1846]

the mean [$B = -0.26$, $SE = 0.09$, 95% $CI = (-0.46, -0.10)$]; and 1 SD above the mean ($B = -0.37$, $SE = 0.14$, 95% $CI = (-0.67, -0.14)$]. The index of moderated mediation was significant [$B = -0.15$, $SE = 0.07$, 95% $CI = (-0.31, -0.03)$].

We conducted another simple mediation analysis with a bootstrapping procedure (Hayes' PROCESS, Model 4) to test $H4$, which predicted that perceived integrity mediates the relationship between the type of response and willingness to reconcile. Supporting our hypothesis, the bootstrapping procedure (with 10,000 resamples) provided evidence for this simple mediating effect [$B = -0.32$, $SE = 0.08$, 95% $CI = (-0.50, -0.16)$].

$H5$ posited that the relationship between the type of response and willingness to reconcile is sequentially mediated by the attribution of responsibility and perceived integrity, and this effect is moderated by perceived sincerity at the first stage so that the sequential indirect effect is strengthened when perceived sincerity is high. We tested the moderated serial mediation model using Hayes' PROCESS macro (Model 83) with 10,000 samples. Table 1 shows that the conditional indirect effect of this path ($X*W \rightarrow M1 \rightarrow M2 \rightarrow Y$) was significant at all levels of perceived sincerity. The index of the moderated serial mediation was significant [$B = -0.12$, $SE = 0.05$, 95% $CI = (-0.23, -0.04)$].

From the spotlight analysis in Table 1, it is worth noting that the structural path $X*W \rightarrow M1 \rightarrow Y$ ($H3b$), which had previously presented significant effects, turned into nonsignificant under the full moderated serial mediation model. These nonsignificant

effects result from including both mediators (M1 and M2) in the model. Furthermore, the effects of verbal responses on willingness to reconcile via attribution of responsibility are no longer significant because the relevant indirect effects occur sequentially through responsibility attributions and perceived integrity. Taken together, these results provide support for the structural path ($X \cdot W \rightarrow M1 \rightarrow M2 \rightarrow Y$) proposed by *H5*.

General discussion

Given the numerous negative consequences of a trust violation on collaborative business relationships and the various positive implications of reconciliation, it is crucial to focus on the psychological mechanisms by which trust repair operates. Therefore, this study investigated the causal mechanisms and boundary conditions by which different verbal responses influence reconciliation.

Across two experimental studies using different procedures (manipulations) and samples (difference in demographics: a sample of US-based workers from MTurk in Study 1 and a sample of Brazilian working adults in Study 2) to add robustness to our findings, we found convergent evidence for our moderated serial mediation model. Studies 1 and 2 showed a uniform finding that trust is repaired more successfully when transgressors deny responsibility for integrity-based trust violations than when they confess and show remorse for having done so. Interestingly, compared to apologies, the greater effectiveness of denials on trust repair is due to lower responsibility attributions for the transgression and, subsequently, higher perceived integrity of the transgressor.

However, when providing explanations for a transgression, the perceived sincerity of transgressors exerts a major role throughout. When responses are seen as less sincere, denials and apologies lead to more similar attributions of responsibility. However, when they are perceived as highly credible, denials are much more effective in deflecting blame for the negative episode than apologies. That is, the positive effects of denials exist and translate into greater willingness to reconcile, mainly when denials are seen as truthful and sincere.

Theoretical implications

These findings contribute to the literature on reconciliation and impression management. First, we extend existing research on the effectiveness of two verbal responses widely used to repair broken trust – denials and apologies. Although some prior work has explored the importance of both attribution of responsibility (Elangovan *et al.*, 2007; Kim *et al.*, 2006) and perceived integrity (Kim *et al.*, 2004) to the verbal response-trust repair link, the relationship between them as comediators remains underexplored. By incorporating and testing both variables as mediators in the model, we expand the serial relationship between these two critical cognitive processes that explain the effect of verbal responses on trust repair and provide the first empirical evidence that the reconciliation process is driven by a two-part mediating mechanism (responsibility attributions followed by perceived integrity).

Furthermore, by examining the moderating effect of perceived sincerity on the verbal response-reconciliation link, we challenge the idea that denials are always a better strategy for integrity-based trust violations (Kim *et al.*, 2004, 2006). According to attribution models and the discounting principle (Kelley, 1973), a potential cause of an event is minimized or discounted if other plausible causes are also present. In line with it, responding to a transgression with denial would provide an alternative cause (external to the transgressor) that could alleviate the transgressor's fault.

However, perceived sincerity is a critical boundary condition for the effectiveness of denials to repair trust violations. The effect of an explanation on the attribution of

responsibility depends on the perception of how believable this explanation is. In the case of apologizing, the veracity of the apology matter less – the transgressor still tends to be seen as guilty. However, in the case of denials, as individuals tend to discount responses that act in one's self-interest as less credible (Brühl *et al.*, 2018), the effects of denials on responsibility attributions in comparison to apologies depend more on the perceived sincerity: the less sincere denials are perceived to be, the less effective they are in exonerating the transgressor from responsibility for the alleged misconduct.

Practical implications

Our study also provides valuable insights to practitioners. Our findings hold practical guidance for the management of collaborative relationships. As trust plays a central role in business contexts, choosing an appropriate response following a transgression is critical to solve conflicts both within and between organizations and reconcile relationships damaged by actual or alleged trust violations. Managers have to be aware that their perceived integrity following a breach of trust is influenced by the level of responsibility taken. Therefore, they should choose wisely which defensive tactics (apologies or denials) to use. The responsibility attributions that each response implicitly conveys can influence their image and reputation at work.

This is particularly critical considering that integrity-related failures are assessed very harshly, and a single integrity-related transgression is highly detrimental to one's image, being capable of permanently damaging a person's reputation (Kim *et al.*, 2006). At the same time, managers should be cautious when deflecting blame for a transgression. Denials are effective only when they seem truthful. Thus, if there is any evidence of guilt or any other factor that undermines the plausibility or legitimacy of a denial, it will tend to be less effective in fostering reconciliation. Managers have to acknowledge that when a denial is barely credible, it has no advantage at all over an apology in terms of attribution of responsibility and can even convey a message of malicious intent of misleading the other person.

Finally, although we have drawn on interpersonal relationship literature to address how verbal responses influence trust repair, our findings may also shed light on trust breaches in business relationships at the organizational level. Even when violations occur in an interorganizational context, reparative actions are undertaken by specific persons in the organizations (Yu, Yang, & Jing, 2017). As Janowicz-Panjaitan and Krishnan (2009, p. 247) stated, "trust is an inherently individual-level phenomenon, which can be attributed to an organization only by virtue of it being made up of individuals." Therefore, our results contribute to the current debate on how partners in interorganizational relationships may repair violated trust and which are the most appropriate strategies for business partners to rebuild broken trust.

Limitations and future research

This work has some limitations that should be addressed. First, although scenario-based experiments allowed us to compare the effects of alternative verbal responses on reconciliation and test the underlying causal mechanisms of these effects, we acknowledge the external validity concerns associated with experimental studies (Oll, Hahn, Reimsbach, & Kotzian, 2018). For instance, although we strived to create experimental materials exhibiting a high level of mundane realism, the participants' decision costs were far lower than in a real-life breach of trust. Additionally, we measured behavioral intention (willingness to reconcile) and not actual behavior. Therefore, collecting behavioral data would be a helpful extension of the findings of this study.

Our findings also open new fronts for future research. By finding evidence of the mediating role of responsibility attributions in the relationship between verbal responses and reconciliation, our work paves the way for examining potential moderators of these relationships. For instance, Tomlinson (2011) posits how relationship dependence between two parties may affect causal attributions made for a negative outcome and consequently help or hamper trust repair. As high dependence makes a relationship breakdown less desirable (Holtgrave *et al.*, 2019), victims are more willing to make more benevolent attributions for a trust violation, which, in turn, facilitates trust repair. Consistent with it, scholars theorize that the potentially positive effect of denials for reconciliation would be boosted since victims would be more motivated to accept transgressors' repair efforts (Kim *et al.*, 2009; Weber, Malhotra, & Murnighan, 2004).

Although we used samples from different nationalities (the USA and Brazil), they are both from Western cultures. Therefore, future studies could examine whether the results found here hold true using samples with different cultural backgrounds. Few studies have examined trust repair in a cross-cultural context, and so more research is needed (Kuwabara, Vogt, Watabe, & Komiya, 2014; Maddux, Kim, Okumura, & Brett, 2011).

Finally, this study investigates the moderating effect of response sincerity on trust repair focusing on two specific types of verbal responses, denial and apology. Future work could also test whether these effects hold for other important verbal statements, such as justifications, which accept responsibility for the act but try to reframe the misbehavior as serving some sort of superordinate goal (Kim & Harmon, 2014).

References

- Basford, T. E., Offermann, L. R., & Behrend, T. S. (2014). Please accept my sincerest apologies: Examining follower reactions to leader apology. *Journal of Business Ethics*, 119(1), 99–117, <https://doi.org/https://doi.org/10.1007/s10551-012-1613-y>.
- Bradford, J. L., & Garrett, D. E. (1995). The effectiveness of corporate communicative responses to accusations of unethical behavior. *Journal of Business Ethics*, 14(11), 875–892, <https://doi.org/https://doi.org/10.1007/BF00882067>.
- Brühl, R., Basel, J. S., & Kury, M. F. (2018). Communication after an integrity-based trust violation: How organizational account giving affects trust. *European Management Journal*, 36(2), 161–170, <https://doi.org/https://doi.org/10.1016/j.emj.2017.08.001>.
- Donovan, L. A. N., & Priester, J. R. (2017). Exploring the psychological processes underlying interpersonal forgiveness: The superiority of motivated reasoning over empathy. *Journal of Experimental Social Psychology*, 71, 16–30, <https://doi.org/https://doi.org/10.1016/j.jesp.2017.02.005>.
- Elangovan, A. R., Auer-Rizzi, W., & Szabo, E. (2007). Why don't I trust you now? An attributional approach to erosion of trust. *Journal of Managerial Psychology*, 22(1), 4–24, <https://doi.org/https://doi.org/10.1108/02683940710721910>.
- Ferrin, D. L., Kim, P. H., Cooper, C. D., & Dirks, K. T. (2007). Silence speaks volumes: The effectiveness of reticence in comparison to apology and denial for responding to integrity- and competence-based trust violations. *Journal of Applied Psychology*, 92(4), 893–908, <https://doi.org/https://doi.org/10.1037/0021-9010.92.4.893>.
- Fulmer, C. A., & Gelfand, M. J. (2012). At what level (and in whom) we trust: trust across multiple organizational levels. *Journal of Management*, 38(4), 1167–1230, <https://doi.org/https://doi.org/10.1177/0149206312439327>.
- Fuoli, M., van de Weijer, J., & Paradis, C. (2017). Denial outperforms apology in repairing organizational trust despite strong evidence of guilt. *Public Relations Review*, 43(4), 645–660, <https://doi.org/https://doi.org/10.1016/j.pubrev.2017.07.007>.

- Gillespie, N., & Dietz, G. (2009). Trust repair after an organization-level failure. *Academy of Management Review*, 34(1), 127–145, <https://doi.org/https://doi.org/10.5465/AMR.2009.35713319>.
- Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis* (2nd ed.), New York, NY: The Guilford Press.
- Holtgrave, M., Nienaber, A.-M., Tzafrir, S. S., & Schewe, G. (2019). Cooperation in the face of conflict: Effects of top managers' trust beliefs in their firms' major suppliers. *British Journal of Management*, 28, 1467–8551, <https://doi.org/https://doi.org/10.1111/1467-8551.12350>.
- Hornsey, M. J., Wohl, M. J. A., Harris, E. A., Okimoto, T. G., Thai, M., & Wenzel, M. (2020). Embodied remorse: Physical displays of remorse increase positive responses to public apologies, but have negligible effects on forgiveness. *Journal of Personality and Social Psychology*, 119(2), 367–389, <https://doi.org/https://doi.org/10.1037/pspi0000208>.
- Hulland, J., & Miller, J. (2018). Keep on Turkin? *Journal of the Academy of Marketing Science*, 46(5), 789–794, <https://doi.org/https://doi.org/10.1007/s11747-018-0587-4>.
- Janowicz-Panjaitan, M., & Krishnan, R. (2009). Measures for dealing with competence and integrity violations of interorganizational trust at the corporate and operating levels of organizational hierarchy. *Journal of Management Studies*, 46(2), 245–268, <https://doi.org/https://doi.org/10.1111/j.1467-6486.2008.00798.x>.
- Kassin, S. M., & Wrightsman, L. S. (1980). Prior confessions and mock juror verdicts. *Journal of Applied Social Psychology*, 10(2), 133–146, <https://doi.org/https://doi.org/10.1111/j.1559-1816.1980.tb00698.x>.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *NE symposium on motivation* (Vol. 15, pp. 192–238). Lincoln: University of NE Press.
- Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, 28(2), 107–128, <https://doi.org/https://doi.org/10.1037/h0034225>.
- Kim, P. H., & Harmon, D. J. (2014). Justifying one's transgressions: How rationalizations based on equity, equality, and need affect trust after its violation. *Journal of Experimental Psychology: Applied*, 20(4), 365–379, <https://doi.org/https://doi.org/10.1037/xap0000030>.
- Kim, P. H., Cooper, C. D., Dirks, K. T., & Ferrin, D. L. (2013). Repairing trust with individuals vs. groups. *Organizational Behavior and Human Decision Processes*, 120(1), 1–14, <https://doi.org/https://doi.org/10.1016/j.obhdp.2012.08.004>.
- Kim, P. H., Diekmann, K. A., & Tenbrunsel, A. E. (2003). Flattery may get you somewhere: The strategic implications of providing positive vs. Negative feedback about ability vs. ethicality in negotiation. *Organizational Behavior and Human Decision Processes*, 90(2), 225–243, [https://doi.org/https://doi.org/10.1016/S0749-5978\(02\)00522-8](https://doi.org/https://doi.org/10.1016/S0749-5978(02)00522-8).
- Kim, P. H., Dirks, K. T., & Cooper, C. D. (2009). The repair of trust: A dynamic bilateral perspective and multilevel conceptualization. *Academy of Management Review*, 34(3), 401–422, <https://doi.org/https://doi.org/10.5465/AMR.2009.40631887>.
- Kim, P. H., Dirks, K. T., Cooper, C. D., & Ferrin, D. L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence- vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes*, 99(1), 49–65, <https://doi.org/https://doi.org/10.1016/j.obhdp.2005.07.002>.
- Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: The effects of apology versus denial for repairing competence- versus integrity-based trust violations. *Journal of Applied Psychology*, 89(1), 104–118, <https://doi.org/https://doi.org/10.1037/0021-9010.89.1.104>.
- Kim, T., & Song, H. (2021). How should intelligent agents apologize to restore trust? Interaction effects between anthropomorphism and apology attribution on trust repair. *Telematics and Informatics*, 61, 101595, <https://doi.org/https://doi.org/10.1016/j.tele.2021.101595>.

- Kuwabara, K., Vogt, S., Watabe, M., & Komiya, A. (2014). Trust, cohesion, and cooperation after early versus late trust violations in two-person exchange. *Social Psychology Quarterly*, 77(4), 344–360, <https://doi.org/https://doi.org/10.1177/0190272514546757>.
- Lewicki, R. J., & Brinsfield, C. (2017). Trust repair. *Annual Review of Organizational Psychology and Organizational Behavior*, 4(1), 287–313, <https://doi.org/https://doi.org/10.1146/annurev-orgpsych-032516-113147>.
- Lewicki, R. J., & Bunker, B. B. (1996). Developing and maintaining trust in work relationships. In R. M. Kramer, & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 114–139). Thousand Oaks, CA: Sage Publications, Inc. <https://doi.org/https://doi.org/10.4135/9781452243610.n7>.
- Lewicki, R. J., Polin, B., & Lount, R. B. (2016). An exploration of the structure of effective apologies. *Negotiation and Conflict Management Research*, 9(2), 177–196, <https://doi.org/https://doi.org/10.1111/ncmr.12073>.
- Maddux, W. W., Kim, P. H., Okumura, T., & Brett, J. M. (2011). Cultural differences in the function and meaning of apologies. *International Negotiation*, 16(3), 405–425, <https://doi.org/https://doi.org/10.1163/157180611X592932>.
- Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, 84(1), 123–136, <https://doi.org/https://doi.org/10.1037/0021-9010.84.1.123>.
- Myers, C. (2016). Apology, sympathy, and empathy: The legal ramifications of admitting fault in U.S. public relations practice. *Public Relations Review*, 42(1), 176–183, <https://doi.org/https://doi.org/10.1016/j.pubrev.2015.10.004>.
- Oll, J., Hahn, R., Reimsbach, D., & Kotzian, P. (2018). Tackling complexity in business and society research: The methodological and thematic potential of factorial surveys. *Business & Society*, 57(1), 26–59, <https://doi.org/https://doi.org/10.1177/0007650316645337>.
- Risen, J. L., & Gilovich, T. (2007). Target and observer differences in the acceptance of questionable apologies. *Journal of Personality and Social Psychology*, 92(3), 418–433, <https://doi.org/https://doi.org/10.1037/0022-3514.92.3.418>.
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393–404, <https://doi.org/https://doi.org/10.5465/AMR.1998.926617>.
- Russell, D. (1982). The causal dimension scale: A measure of how individuals perceive causes. *Journal of Personality and Social Psychology*, 42(6), 1137–1145, <https://doi.org/https://doi.org/10.1037/0022-3514.42.6.1137>.
- Struthers, C. W., Eaton, J., Shirvani, N., Georghiou, M., & Edell, E. (2008). The effect of preemptive forgiveness and a transgressor's responsibility on shame, motivation to reconcile, and repentance. *Basic and Applied Social Psychology*, 30(2), 130–141, <https://doi.org/https://doi.org/10.1080/01973530802209178>.
- Struthers, C. W., Miller, D. L., Boudens, C. J., & Briggs, G. L. (2001). Effects of causal attributions on coworker interactions: A social motivation perspective. *Basic and Applied Social Psychology*, 23(3), 169–181, https://doi.org/https://doi.org/10.1207/S15324834BASP2303_3.
- ten Brinke, L., & Adams, G. S. (2015). Saving face? When emotion displays during public apologies mitigate damage to organizational performance. *Organizational Behavior and Human Decision Processes*, 130, 1–12, <https://doi.org/https://doi.org/10.1016/j.obhdp.2015.05.003>.
- Tomlinson, E. C. (2011). The context of trust repair efforts: Exploring the role of relationship dependence and outcome severity. *Journal of Trust Research*, 1(2), 139–157, <https://doi.org/https://doi.org/10.1080/21515581.2011.603507>.
- Tomlinson, E. C., Dineen, B. R., & Lewicki, R. J. (2004). The road to reconciliation: Antecedents of victim willingness to reconcile following a broken promise. *Journal of Management*, 30(2), 165–187, <https://doi.org/https://doi.org/10.1016/j.jm.2003.01.003>.

- Tomlinson, E. C., & Mayer, R. C. (2009). The role of causal attribution dimensions in trust repair. *Academy of Management Review*, 34(1), 85–104, <https://doi.org/https://doi.org/10.5465/AMR.2009.35713291>.
- Weber, J. M., Malhotra, D., & Murnighan, J. K. (2004). Normal acts of irrational trust: Motivated attributions and the trust development process. *Research in Organizational Behavior*, 26, 75–101, [https://doi.org/https://doi.org/10.1016/S0191-3085\(04\)26003-8](https://doi.org/https://doi.org/10.1016/S0191-3085(04)26003-8).
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92(4), 548–573, <https://doi.org/https://doi.org/10.1037/0033-295X.92.4.548>.
- Yu, Y., Yang, Y., & Jing, F. (2017). The role of the third party in trust repair process. *Journal of Business Research*, 78, 233–241, <https://doi.org/https://doi.org/10.1016/j.jbusres.2017.01.015>.

Author contributions are as follows: Iwai, Tatiana – corresponding author: conceptualization (Lead), data curation (Equal), formal analysis (Equal), investigation (Equal), methodology (Equal), validation (Equal), visualization (Equal), writing – original draft (Lead), writing – review and editing (Lead). Carvalho, João Vinícius França: conceptualization (Supporting), data curation (Lead), formal analysis (Equal), investigation (Equal), methodology (Equal), software (Lead), validation (Equal), visualization (Equal), writing – original draft (Supporting), writing – review and editing (Supporting).

Corresponding author

Tatiana Iwai can be contacted at: tatianai@insper.edu.br

Associate editor: Silvia Pereira de Castro Casa Nova