Apology Cost Is More Strongly Associated With Perceived Sincerity Than Forgiveness

Yohsuke Ohtsubo[1,*], Miyu Higuchi[2]

[1]The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[2]Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan

*Author for correspondence (yohtsubo@l.u-tokyo.ac.jp)

This study aimed at replicating a previously reported pattern: costly apologies are perceived as more sincere than non-costly apologies even when the transgression was unintentional, while costly apologies do not foster forgiveness more than non-costly apologies when the transgression was unintentional. We conducted a vignette study with a 2 (apology cost: costly vs. non-costly) × 3 (intention: no vs. ambiguous vs. malicious intention) between-participants factorial design. We failed to replicate the aforementioned pattern. Instead, we found that costly apologies promote not only perceived sincerity but also forgiveness in all three intention conditions. In addition, there were two notable patterns. First, the effect of apology cost was stronger for perceived sincerity than for forgiveness in all three intention conditions. Second, both perceived sincerity and forgiveness decreased as the intentional nature of the transgression was described more clearly.

Introduction

Reconciliation among former opponents is ubiquitous among both animal and human societies (e.g., Arnold et al., 2010; McCullough et al., 2013). Human reconciliation processes consist of transgressors' apologies (Ohtsubo & Yagi, 2015) and victims' forgiveness (Burnette et al., 2012). Although apology and forgiveness are conceptually distinct, evolutionary psychological studies have shown that they both are facilitated by a common psychological variable: the relationship value of the partner (see de Waal, 2000, for the valuable relationship hypothesis).

Based on the signaling game analysis, Ohtsubo and Watanabe (2009) maintained that just saying "I'm sorry" is cheap talk and cannot be credible. Apologies, however, can be costlier than just saying sorry. For example, transgressors may offer compensation or cancel important business to prioritize apologies. These costly

forms of apologies are credible because only transgressors who highly value the endangered relationship have an incentive to incur such costs. If transgressors do not value the relationship, they should simply let the relationship be dissolved without incurring any cost. Therefore, the costly apology model, consistent with the valuable relationship hypothesis, posits that relationship value facilitates the transgressor in making a costly apology. A complementary prediction is that costly apologies communicate the apologizers' sincere intention (i.e., how much they value the relationship) to recipients of apologies.

Although a typical outcome variable in apology studies is forgiveness, the signaling model posits that the function of costly apologies is to facilitate mental state inferences (i.e., perceiving sincere intention). Therefore, it predicts that the effect of apology cost is stronger for perceived sincerity than for forgiveness. In fact, a recent functional magnetic resonance imaging (fMRI) study (Ohtsubo et al., 2018) revealed that costly apologies, as compared with non-costly apologies and no apologies, more strongly engaged the theory-of-mind network, which is activated when one perceives someone's social/communicative intention (Ciaramidaro et al., 2007).

If the effect of apology cost on forgiveness is secondary, there may be a case where costly apologies, as compared with non-costly apologies, may not necessarily foster forgiveness. Ohtsubo and Watanabe (2009) conducted vignette studies (Studies 1 and 2) in which the transgressor's awareness of their behavior's harmful impact is ambiguous. The results showed that costly apologies fostered both perceived sincerity and forgiveness. However, in their Appendix, they reported a preliminary study that used vignettes explicitly describing the absence of the transgressor's awareness of their behavior's harmful impact. Under this condition, a transgression cannot be intentional because the transgressor is unable to anticipate its consequences (Malle & Knobe, 1997). Interestingly, when the unintentional nature of transgressions was transparent, participants were willing to forgive the transgressor regardless of whether their apology was costly or not, although they still perceived costly apologies to be more sincere than non-costly apologies. A stylized pattern observed in Ohtsubo and Watanabe's Studies 1 and 2 (ambiguous intention) and Appendix (no intention) is depicted in Figure 1.

The purpose of this study is to investigate whether the apology cost × intention (ambiguous vs. transparently unintentional) interaction effect on forgiveness is replicable. Figure 1 shows the two specific predictions that we tested in this study: (1) The apology cost × intention interaction would be significant when forgiveness is the dependent variable, while it would not be significant when perceived sincerity is the dependent variable; and (2) the effect size of apology cost on forgiveness, which is operationalized as correlation between apology cost (i.e., a dummy coded variable) and forgiveness,

is significantly smaller than the effect size of apology cost on perceived sincerity (i.e., apology cost × perceived sincerity correlation) in the no intention condition but not in the ambiguous intention condition. We pre-registered these predictions in the Open Science Framework (OSF: https://osf.io/sfyzq/). In addition, for an exploratory purpose, we included an additional condition in which the transgressor's malicious intention is transparent (i.e., malicious intention condition) because Desmet et al. (2011) found the effect of overcompensation on trust recovery was attenuated when the intention of norm violation was clear (see also Struthers et al., 2008).
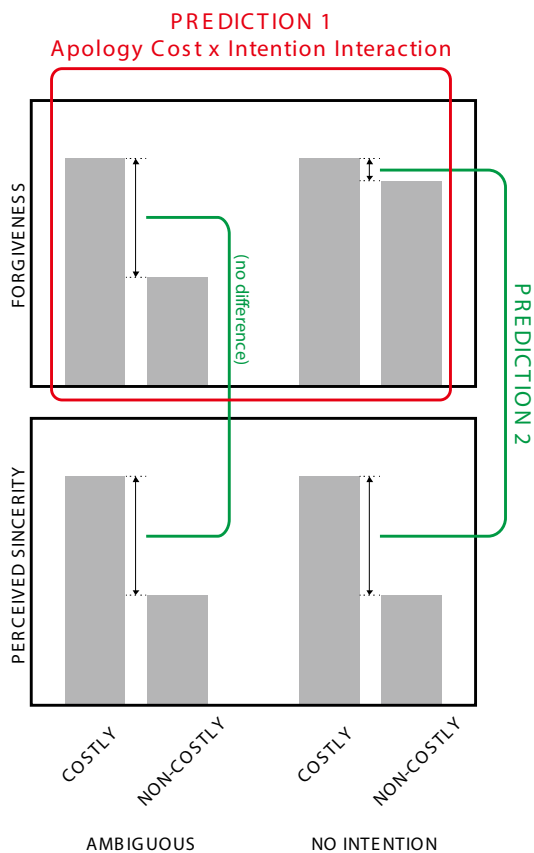

Figure 1. Two predictions tested in this study.

Methods

Participants and design

Based on the second prediction, we determined the target sample size to be 100 for each cell of the 2 × 3 conditions (i.e., a total of 600 participants) to ensure a power of .80 for a test of two dependent correlations. Participants who are currently employed workers and within the ages of 20–40 years old were recruited by a Japanese survey company, Cross Marketing. The restrictions in terms of employment status and age were applied because some scenarios described the protagonist (i.e., participant) interacting with their co-worker or boss. We had a total of 1668 accesses to our study site. Of the 1668 accesses, 790 participants (406 women, 380 men, and 4 non-reported their sex; mean age ± *SD* = 31.23 ± 5.61 years) were retained for subsequent analyses (see Section 1 in the Supplementary Material for exclusion criteria). Participants were randomly assigned to one of the six cells of a 2 (apology cost: costly vs. non-

costly) × 3 (intention: no intention vs. ambiguous intention vs. malicious intention) between-participants factorial design.

Materials

The study included four hypothetical transgression scenarios. Each scenario described an interpersonal transgression. For example, a friend (F) sent the protagonist (P) a series of consecutive text messages when P was working. P reluctantly replied to F. P's boss spotted P sending private messages during work and yelled at P. In the no intention condition, P was working on an irregular day and F did not know P was working. Therefore, F was not aware that they were sending consecutive messages when P was working. In the malicious intention condition, the scenario described that although F knew that P was working, F prioritized their own business. In the ambiguous intention condition, whether F prioritized their own business was not explicitly stated in the scenario. After reading this transgression part, participants rated how angry they would be at F and how likely they would be to end a friendship.

The second part of the scenario described F's apology: either a costly or non-costly apology. For example, in the costly apology condition, F spent more than 2 hours traveling to P's home that night to apologize as soon as possible. In the non-costly apology condition, F apologized to P the next time that F met P.

After reading the apology scenario, participants rated perceived sincerity (3 items × 4 scenarios; Cronbach's α = .92), forgiveness (5 items × 4 scenarios; α = .94), perceived exploitation risk (1 item × 4 scenarios; α = .68), and F's valuation of the relationship (1 item × 4 scenarios; α = .77). Sample items are as follows: "How sincere do you think the friend's apology is?" (perceived sincerity), "How much are you willing to forgive the friend?" (forgiveness), "How likely do you think the friend will cause the same trouble again?" (perceived exploitation risk), and "How much do you think the friend values the relationship with you?" (valuation). The order of the 10 items was randomized. These items were followed by three manipulation check items. An attention check item (i.e., requiring participants to choose a specific option) was embedded in the items associated with two of the four scenarios. All materials including the four scenarios (in Japanese) are available from OSF (https://osf.io/sfyzq/).

Procedure

Participants took part in the study online. It was implemented by the Qualtrics (https://www.qualtrics.com). Participants first provided their sex, age, and employment status. Those who did not meet the inclusion criteria (i.e., employed workers with an age range between 20 and 40 years) were not allowed to proceed to the main study. The order of the four scenarios was randomized. For each scenario, the first phase presented the transgression scenario and assessed participants' anger at the transgressor. The second phase presented the apology scenario and assessed perceived sincerity, forgiveness, perceived exploitation risk, and valuation. In addition, manipulation check items were included. This two-phase assessment was repeated four times. Participants were compensated for their participation by Cross Marketing.
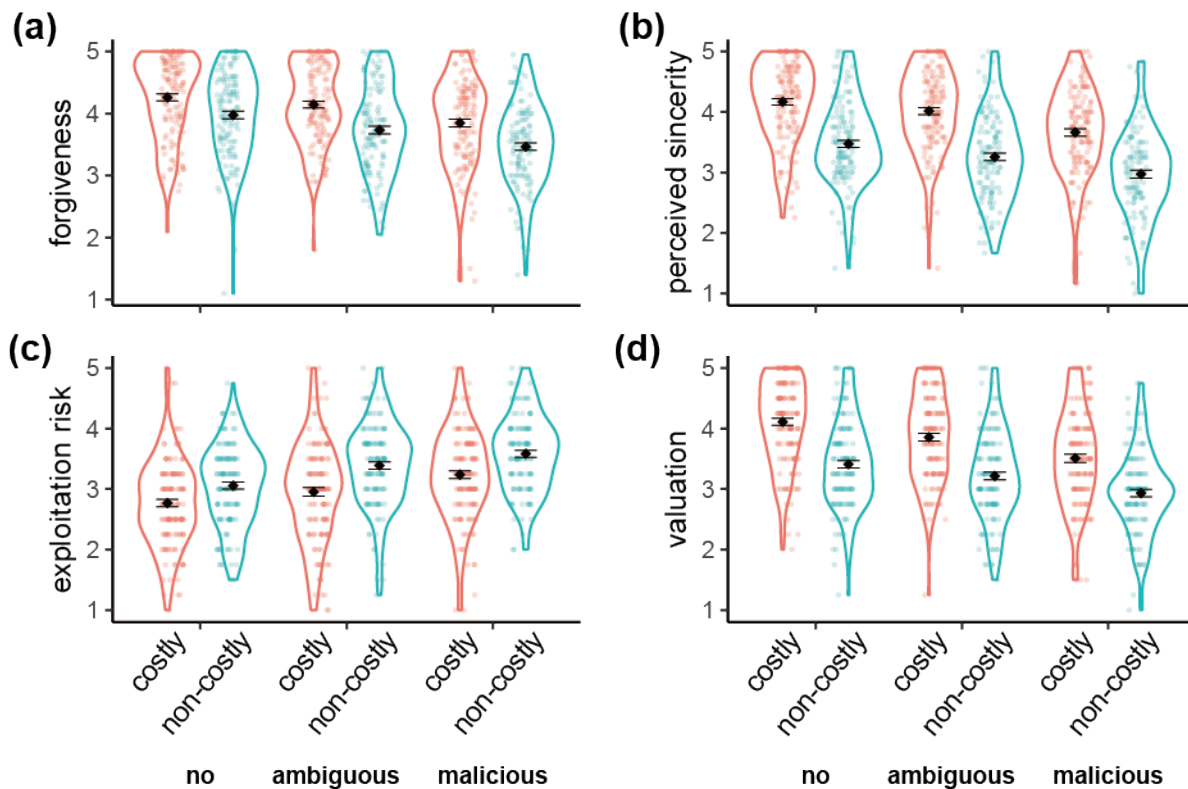
Figure 2. Distributions of (a) forgiveness, (b) perceived sincerity, (c) exploitation risk, and (d) valuation as a function of apology cost and intention.

Results

All data analyses were implemented by R (R Core Team, 2019). The test of differences in two correlations was conducted using the "cocor" package (Diedenhofen & Musch, 2015). The data and code used in the subsequent analyses are available in OSF (https://osf.io/sfyzq/).

We first confirmed that both intention and apology cost manipulations were successful. Participants in the no intention condition rated the transgressor's awareness (i.e., whether the transgressor was aware of the potential harmful effect) lower than those in the other two conditions. Participants in the malicious intention condition rated the transgressor's prioritization of their own business higher than those in the other two conditions. Participants in the costly apology condition rated the costliness of the transgressor's apology higher than those in the non-costly apology condition. We report details of the manipulation checks in Section 2 of the Supplementary Material.

Prediction 1

We conducted a series of four 2 (apology cost) × 3 (intention) analyses of variance (ANOVAs) with perceived sincerity, forgiveness, perceived exploitation risk, and valuation as dependent variables. The latter two dependent variables were included for exploratory purposes.

As for forgiveness, the main effects of apology cost and intention were significant: $F(1, 784) = 55.18$, $p < .001$, $\eta_p^2 = .066$ and $F(2, 784) = 29.76$, $p < .001$, $\eta_p^2 = .072$, respectively. However, contrary to Prediction 1, the interaction effect was not significant: $F(2, 784) = 0.64$, *ns*. Post-hoc tests (Tukey's HSD test) revealed that the

effect of apology cost on forgiveness was significant in all three levels of the intention condition at the .01 level. Parenthetically, the three levels of the intention condition were significantly different from each other at the .01 level: it was highest in the no intention condition and lowest in the malicious intention condition (Figure 2a).

As for perceived sincerity, the main effects of apology cost and intention were significant: $F(1, 784) = 211.25$, $p < .001$, $\eta_p^2 = .212$ and $F(2, 784) = 33.95$, $p < .001$, $\eta_p^2 = .084$, respectively. The interaction effect was not significant: $F(2, 784) = 0.20$, *ns*. Post-hoc tests (Tukey's HSD test) revealed that the effect of apology cost was significant in all three levels of the intention condition at the .01 level, and that the three levels of the intention condition were significantly different from each other at the .01 level: being highest in the no intention condition and lowest in the malicious intention condition (Figure 2b).

We observed a comparable pattern for perceived exploitation risk (Figure 2c) and valuation (Figure 2d). Details of these results are reported in Section 3 of the Supplementary Material. The prediction depicted in Figure 1 is based on a 2 (apology cost) × 2 (intention: no vs. ambiguous intention) factorial design. Therefore, we ran additional analyses with this 2 × 2 design, but the predicted interaction was not significant for forgiveness. These results are reported in Section 4 of the Supplementary Material.

Prediction 2

We tested the difference between the apology cost × perceived sincerity correlation and the apology cost × forgiveness correlation for all three levels of intention condition. We employed the test of dependent correlations
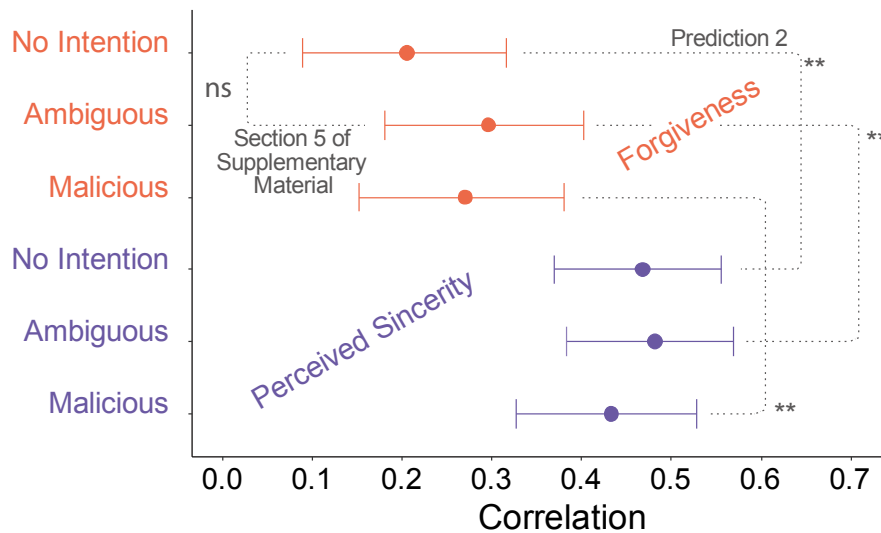
Figure 3. Correlation between apology cost and forgiveness (upper) and correlation between apology cost and perceived sincerity (lower) as a function of the intention condition.

Note. Visual inspections of Figure 3 (the upper two data points) suggest that the effect of apology cost on forgiveness was larger in the ambiguous intention condition than in the no intention condition. This is consistent with Prediction 1, which was not supported by an ANOVA. To confirm the non-significant apology cost × intention interaction effect, we directly tested the difference in these two correlations. The analysis corroborated the non-significant difference (see Section 5 of Supplementary Material). ** $p < .01$

because the two correlations shared apology cost as a common variable. As predicted, the apology cost × perceived sincerity correlation was greater than the apology cost × forgiveness correlation in the no intention condition: .47 vs. .21, Hotelling's $t(270) = 7.43$, $p < .001$. However, they were also significantly different in the other two conditions: .48 vs. .30, Hotelling's $t(259) = 4.98$, $p < .001$ and .43 vs. .27, Hotelling's $t(252) = 4.06$, $p < .001$ in the ambiguous and malicious intention conditions, respectively (Figure 3). Although the significant difference in the no intention condition was consistent with Prediction 2, the entire pattern (i.e., the effect of apology cost is stronger for perceived sincerity than forgiveness in the other two conditions as well) is contradictory with Prediction 2.

Discussion

We expected that the effect of apology cost on forgiveness would be attenuated specifically in the no intention condition (Figure 1). However, the effect of apology cost on forgiveness was attenuated in all three intention conditions. Therefore, we failed to replicate the pattern in Figure 1. Although the replication was unsuccessful, there were two notable patterns in the results. First, the effect of apology cost on perceived sincerity was stronger than its effect on forgiveness across all three levels of intentionality. Second, the main effect of intention was significant not only for forgiveness but also for perceived sincerity.

These two patterns may explain why Ohtsubo and Watanabe (2009) found the asymmetric effect of apology cost on perceived sincerity and forgiveness. As the effect of apology cost on forgiveness is generally smaller, it is more likely to fail to detect the effect (i.e., Type II error). In addition, since forgiveness and perceived sincerity tend to increase when participants consider that the transgression

was unintentional, the likelihood of the ceiling effect should be greater in the no intention condition. Thus, it is possible that Ohtsubo and Watanabe's result in their Appendix was an instance of either Type II error, the ceiling effect, or both (see also Ohtsubo et al., 2012, for the smaller effect for forgiveness in a cross-cultural study).

Although we failed replication, the two findings merit some brief discussions. First, the greater effect for perceived sincerity than for forgiveness is consistent with the main thesis of the costly apology model: the primary function of apology cost is to communicate sincere intention. Second, the significant effect of intention on perceived sincerity is noteworthy because perceived sincerity is a subjective evaluation of the apology and thus is theoretically separable from the evaluation of transgressions. This suggests that in response to the perceived intentionality of the transgression, people may adjust a threshold of sincerity. This possibility needs further scrutiny.

Author contribution
Y.O. and M.H. conceived the research design. M.H. wrote the scenarios and developed the study platform. Y.O. and M.H. analyzed data. Y.O. drafted the manuscript and M.H. approved it.

Ethical statement
This study was approved by the ethics committee at the Graduate School of Humanities and Sociology, the University of Tokyo (UTSP–21012).

Data accessibility & program code

The materials, analyzed datasets, and R code are available in the Open Science Framework (https://osf.io/sfyzq/).

Supplementary material

Electronic supplementary materials are available online.

References

Arnold, K., Fraser, O. N., & Aureli, F. (2010). Postconflict reconciliation. In C. J. Campbell, A. Fuentes, K. C. MacKinnon, S. K. Bearder, & R. M. Stumpf (Eds.), *Primates in perspective* (2nd ed., pp. 608–625). Oxford University Press.

Burnette, J. L., McCullough, M. E., Van Tongeren, D. R., & Davis, D. E. (2012). Forgiveness results from integrating information about relationship value and exploitation risk. *Personality and Social Psychology Bulletin*, *38*(3), 345–356. https://doi.org/10.1177/0146167211424582

Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B. G., & Walter, H. (2007). The intentional network: How the brain reads varieties of intentions. *Neuropsychologia*, *45*(13), 3105–3113. https://doi.org/10.1016/j.neuropsychologia.2007.05.011

Desmet, P. T. M., De Cremer, D., & van Dijk, E. (2011). In money we trust? The use of financial compensations to repair trust in the aftermath of distributive harm. *Organizational Behavior and Human Decision Processes*, *114*(2), 75–86. https://doi.org/10.1016/j.obhdp.2010.10.006

de Waal, F. B. M. (2000). Primates—A natural heritage of conflict resolution. *Science*, *289*(5479), 586–590. https://doi.org/10.1126/science.289.5479.586

Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE*, *10*(4), Article e0121945. https://doi.org/10.1371/journal.pone.0121945

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, *33*(2), 101–121. https://doi.org/10.1006/jesp.1996.1314

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2013). Cognitive systems for revenge and forgiveness. *Behavioral and Brain Sciences*, *36*(1), 1–15. https://doi.org/10.1017/S0140525X11002160

Ohtsubo, Y., Matsunaga, M., Tanaka, H., Suzuki, K., Kobayashi, F., Shibata, E., Hori, R., Umemura, T., & Ohira, H. (2018). Costly apologies communicate conciliatory intention: An fMRI study on forgiveness in response to costly apologies. *Evolution and Human Behavior*, *39*(2), 249–256. https://doi.org/10.1016/j.evolhumbehav.2018.01.004

Ohtsubo, Y., & Watanabe, E. (2009). Do sincere apologies need to be costly? Test of a costly signaling model of apology. *Evolution and Human Behavior*, *30*(2), 114–123. https://doi.org/10.1016/j.evolhumbehav.2008.09.004

Ohtsubo, Y., Watanabe, E., Kim, J., Kulas, J. T., Muluk, H., Nazar, G., Wang, F., & Zhang, J. (2012). Are costly apologies universally perceived as being sincere? A test of the costly apology-perceived sincerity relationship in seven countries. *Journal of Evolutionary Psychology*, *10*(4), 187–204. https://doi.org/10.1556/JEP.10.2012.4.3

Ohtsubo, Y., & Yagi, A. (2015). Relationship value promotes costly apology-making: Testing the valuable relationships hypothesis from the perpetrator's perspective. *Evolution and Human Behavior*, *36*(3), 232–239. https://doi.org/10.1016/j.evolhumbehav.2014.11.008

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Struthers, C. W., Eaton, J., Santelli, A. G., Uchiyama, M., & Shirvani, N. (2008). The effects of attributions of intent and apology on forgiveness: When saying sorry may not help the story. *Journal of Experimental Social Psychology*, *44*(4), 983–992. https://doi.org/10.1016/j.jesp.2008.02.006