Costly group apology communicates a group’s sincere “intention”

Ohtsubo, Yohsuke ; Matsunaga, Masahiro ; Himichi, Toshiyuki ; Suzuki, Kohta ; Shibata, Eiji ; Hori, Reiko ; Umemura, Tomohiro ; Ohira, Hideki

(Citation)
Social Neuroscience,15(2):244-254

(Issue Date)
2019-11-23

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90006578

Costly Group Apology Communicates a Group's Sincere "Intention"

Yohsuke Ohtsubo (Kobe University) Masahiro Matsunaga (Aichi Medical University)

Toshiyuki Himichi (Kochi University of Technology)

Kohta Suzuki Eiji Shibata Reiko Hori Tomohiro Umemura

(Aichi Medical University)

Hideki Ohira (Nagoya University)

Author Note

Yohsuke Ohtsubo, Department of Psychology, Graduate School of Humanities, Kobe University, Japan.

We are grateful to Haruo Isoda, Akira Ishizuka, and Tamotsu Kawai (Brain and Mind Research Center, Nagoya University, Aichi, Japan) for their technical support. This research was supported by the Japan Society for the Promotion of Science KAKENHI grant to Y.O. (15H03447).

Correspondence concerning this article should be addressed to Yohsuke Ohtsubo, Graduate School of Humanities, Department of Psychology, Kobe University, Kobe, 657-8501, Japan. E-mail: yohtsubo@lit.kobe-u.ac.jp

Abstract

Groups, such as governments and organizations, apologize for their misconduct. In the interpersonal context, the forgiveness-fostering effect of apologies is pronounced when apologizing entails some cost (e.g., compensating damage, cancelling a favorite activity to prioritize the apology) because costly apologies tend to be perceived as more sincere than non-costly apologies (e.g., merely saying "sorry"). Since groups lack a mental state (e.g., sincere intention), this could arguably render a group apology ineffective. This research investigated the possibility that people ascribe intention to group agents and that offering a costly group apology is an effective means of fostering perceived sincerity. A vignette study (Pilot Study) showed that costly group apologies tend to be perceived as more sincere than non-costly group apologies. A subsequent functional magnetic resonance imaging study revealed that costly group apologies engaged the bilateral temporoparietal junction and precuneus more so than non-costly group apologies and no apology did. The bilateral temporoparietal junction and precuneus have been implicated in the reasoning of social/communicative intention. Therefore, these results suggest that although a group mind does not exist, people ascribe a mental state (i.e., sincere intention) to a group especially when the group issues a costly apology after committing some transgression.

Keywords: costly apology, group apology, crisis communication, mental state inference

Groups apologize as frequently as individuals do (if not more frequently). For instance, governments issue political apologies for past injustices/wrongdoings (e.g., Blatz & Philpot, 2010; Blatz, Schumann, & Ross, 2009; Minow, 2002), and organizations and companies apologize to their customers and stockholders for various reasons, ranging from illegal actions to decline in corporate performance (e.g., Bisel & Messersmith, 2012; Bundy, Pfarrer, Short, & Coombs, 2017; Koehn, 2013). Nevertheless, empirical research on collective/corporate apology (henceforth referred to as "group apology") is relatively sparse compared with research on interpersonal apology. Although group and interpersonal apologies have many commonalities (Blatz et al., 2009), they differ in some critical ways. For example, Koehn (2013) pointed out that group apologies are characterized by a lack of interiority (i.e., no particular mind is assumed to be behind the apologies because groups do not have minds in the same sense that individuals do). Accordingly, Koehn argued that the expression of remorse (an instance of a mental state) is important for interpersonal apologies, while reparation is central to group apologies. However, empirical studies have shown not only that reparation facilitates interpersonal forgiveness (Desmet, De Cremer, & van Dijk, 2010; Komiya, Ohtsubo, Oishi, & Mifune, 2018) but also that effective group apologies often include expressions of remorse (Blatz et al., 2009). The latter finding appears paradoxical because a group's expression of remorse would be meaningless unless observers presume that the group has a mental state (interiority).

Previous studies examining the effect of costly apologies have underscored the role of mental state inference in the context of interpersonal apologies. It is well established that costly apologies (or costly forms of conciliatory acts) tend to be perceived as more sincere than non-costly and no apologies and, thus, they more effectively promote victim forgiveness (Bottom, Gibson, Daniels, & Murnighan, 2002; Ho, 2012; Ohtsubo & Watanabe, 2009; Ohtsubo et al., 2012; Schweitzer, Hershey, & Bradlow, 2006). Although the notion of costly apology is often

erroneously equated with compensation or reparation, it covers a wider range of conciliatory acts, some of which are non-compensatory acts, such as cancelling a favorite activity to make an immediate apology (Ohtsubo & Watanabe, 2009; Ohtsubo & Yagi, 2015). Therefore, the effect of a costly apology is not reducible to the recovery of damages. According to game theoretic models of costly apology (Ho, 2012; Martinez-Vaquero, Han, Pereira, & Lenaerts, 2015; O'Connor, 2016; Ohtsubo & Watanabe, 2009), the effect lies in its signaling property—the function of an apology is to signal the perpetrator's valuation of the relationship with the victim, and the cost of making an apology guarantees the honesty of the apology (i.e., only those who highly value the relationship with their victim are willing to incur a substantial cost to restore the relationship). According to this signaling model, the cost can take any form (e.g., financial cost, physical cost, time cost) insofar as perpetrators who do not value the relationship with their victim find it prohibitive: It becomes obvious to the eye of the beholder that those who incur a prohibitively high cost to make an apology should sincerely *intend* to restore the relationship.

In a recent functional magnetic resonance imaging (fMRI) study (Ohtsubo et al., 2018), participants imagined receiving either a costly apology, non-costly apology, or no apology from a friend. The results showed that when participants imagined receiving a costly apology, the medial prefrontal cortex (MPFC), bilateral temporoparietal junction (TPJ), and precuneus, along with some other regions, were more strongly activated than when participants imagined receiving a non-costly apology or no apology (for additional evidence of the association between the angular gyrus [part of the TPJ] and forgivability, see Strang, Utikal, Fischbacher, Weber, & Falk, 2014). These regions are typically implicated in cognitive empathy (e.g., perspective taking) and intention processing (e.g., Ciaramidaro, Becchio, Colle, Bara, & Walter, 2014; den Ouden, Frith, & Frith, 2005; Farrer & Frith, 2002; Ruby & Decety, 2001; Saxe & Kanwisher, 2003; Van Overwalle & Baetens, 2009). Interestingly, in recent studies, communicative intention expressed

via non-linguistic modalities (e.g., gestures, prosodies) as well as linguistically expressed communicative intention equally recruited these areas of the brain—sometimes called the intention processing network (e.g., Enrici, Adenzato, Cappa, Bara, & Tettamanti,2011; Hellbernd & Sammler, 2018; Tettamanti et al., 2017). This modality-independent property of the intention processing network seems especially relevant to the costly apology model because it relies on the assumption that a cost that communicate sincere intention may take any form, and thus may be perceived by different modalities. Regardless of which modality is used to perceive the cost, the aforementioned evidence suggests that it can recruit the brain's intention processing network.

Although there is sufficient evidence of the intention processing network's relevance to the interpersonal apology context, whether it extends to the group apology context remains unclear. As Koehn (2013) pointed out, collective entities, such as nations, governments, organizations, and companies, do not have minds in the same sense that individuals do (see also Allport, 1924). Therefore, the effect of costly apology, which owes its function to communicating the apologizer's sincere intention, is not readily generalizable to the group apology context. Nevertheless, there is suggestive evidence of the effectiveness of costly group apology. First, Koehn (2013) cited many examples of successful corporate apologies, many of which included some cost, such as the CEO spending a long period of time listening to victims' complaints. Second, people ascribe mental states, such as intention, to jointly acting groups (O'Laughlin & Malle, 2002; Tanibe, Hashimoto, Tomabechi, Masamoto, & Karasawa, 2019). In other words, although the notion of a group mind is scientifically unsound and fallacious, it seems viable in folk understandings of group behaviors. Third, a recent fMRI study revealed that reading sentences including an individual agent's mental state and a group's mental state (e.g., "a group wants...," "a group knows...") engaged the same brain regions (MPFC, right TPJ, precuneus), thus suggesting that people attribute mental states not only to individual agents but also to groups

(Jenkins, Dodell-Feder, Saxe, & Knobe, 2014). Fourth, in a recent fMRI study on group apology (Kim, Kralik, Yun, Chung, & Jeong, 2019), activity in the angular gyrus (part of TPJ) was positively associated with the forgivability of an apologizing group. However, in Kim et al.'s study, other intention processing regions were not associated with forgivability.

The present study examined whether costly group apologies would be perceived as more sincere than non-costly group apologies. In Pilot Study, participants rated the sincerity of a series of hypothetical group apology vignettes. If participants perceived costly group apologies as sincere, it implied that they assumed the presence of a mind behind the group apologies. However, it was possible that their use of the adjective "sincere" could be merely metaphorical. Therefore, we followed this up with an fMRI study to test whether costly group apology vignettes would activate the brain regions (i.e., MPFC, bilateral TPJ, precuneus) that more strongly responded to costly interpersonal apology vignettes than non-costly apology and no apology vignettes in the previous study (Ohtsubo et al., 2018).

Pilot Study

Methods

Pilot Study involved 108 Japanese undergraduate students (60 males and 48 females; ages 18–22 years, mean ± *SD* = 19.07 ± 1.04). We wrote 15 group transgression scenarios (e.g., *There have been quite a few accidents where electric appliances produced by Company A have caught fire*; *Country A has not provided a sufficient amount of food to refugees*) (see Appendix Table A1 for English translations of the 15 scenarios and Supplementary Materials Table S1 for the original Japanese scenarios). For each scenario, we wrote three types of reaction scenarios: costly apology (e.g., *Company A immediately issued an apology, offered compensation, and announced that it would offer an exchange of the same type of appliance*), non-costly apology (e.g., *Company A apologized for the accidents where its electric appliances caught fire*), and no

apology (e.g., *Company A has not yet apologized because the matter is still under investigation*). Using the 45 scenarios (15 transgressions × 3 reactions), we created three versions of a questionnaire, each of which contained 5 no apology scenarios, 5 non-costly apology scenarios, and 5 costly apology scenarios. There was no overlap of the 15 transgression scenarios in each questionnaire. Each transgression scenario was followed by questions assessing (i) the seriousness of the transgression. After responding to this seriousness question, participants read one of the reaction scenarios (no apology vs. non-costly apology vs. costly apology) and rated the group's (ii) blameworthiness, (iii) sincerity, and (iv) likelihood of committing the same transgression again (rated on a 7-point scale ranging from 1 = "not at all" to 7 = "very much"). Hereafter, these items are referred to as (i) seriousness, (ii) blameworthiness, (iii) sincerity, and (iv) likelihood of further harm. Of these four items, seriousness was a check for random assignment (i.e., the score should not vary across the three conditions because it was rated before participants saw the reaction scenarios) and sincerity was central to the hypothesis because it referred to the group's mental state. The three versions of the questionnaire were randomly distributed to participants.

Results and Discussion

Within each questionnaire, five scenarios in the same group reaction condition were collapsed for the analyses. Therefore, all participants had three scores (corresponding to the costly, non-costly, and no apology conditions) for each of the four items. The four item scores were separately submitted to a one-way analysis of variance (ANOVA) with apology type as the within-subject independent variable. Specifically, the data were analyzed by R version 3.6.0 (https://www.r-project.org/) using lmerTest, MuMIn, and emmeans packages (all codes, data, and outputs are available in the Open Science Framework: https://osf.io/zw6g7/). Participants considered the 15 transgressions as highly serious (rated slightly lower than 6 on the 7-point

scale; see Figure 1A). Mean seriousness did not vary across the three conditions: mean (±*SD*) seriousness scores were 5.80 ± 0.67, 5.82 ± 0.72, and 5.83 ± 0.71 in the no apology, non-costly apology, and costly apology conditions, respectively, $F(2, 201) = 0.15$, $p = .863$, $R^2 = .00$. Therefore, we considered the random assignment successful.

We then tested the effect of group reaction on the three dependent variables. As predicted, group reactions influenced the ratings of blameworthiness, $F(2, 204) = 123.78$, $p < .001$, $R^2 = .36$ (Figure 1B); sincerity, $F(2, 205) = 350.75$, $p < .001$, $R^2 = .67$ (Figure 1C); and likelihood of further harm $F(2, 204) = 137.62$, $p < .001$, $R^2 = .39$ (Figure 1D). A series of post hoc tests revealed that costly apology significantly mitigated participants' blame (3.94 ± 1.05) compared with no apology (5.54 ± 0.73) and non-costly apology (4.89 ± 0.84); costly apology significantly increased perceived sincerity (4.84 ± 0.88) compared with no apology (2.16 ± 0.64) and non-costly apology (3.64 ± 0.78); and costly apology reduced the perceived likelihood of further harm (4.43 ± 0.82) compared with no apology (5.82 ± 0.64) and non-costly apology (5.07 ± 0.68).

A closer look at Figures 1B, 1C, and 1D reveals that the effect of group reaction was largest for sincerity. In fact, the effect sizes measured by R^2 indicate that the largest effect was associated with perceived sincerity—group reaction accounted for almost two thirds of the variance in perceived sincerity, while group reaction accounted for slightly more than one third of the variance in blameworthiness and perceived likelihood of further harm. However, some might argue that the difference in effect sizes was owing to the presence (or absence) of apology, but not to the presence of cost. To counteract this possible criticism, we conducted comparable analyses including only the costly and non-costly apology conditions. Although the omission of the no apology condition substantially reduced the effect sizes, the largest effect size was still associated with perceived sincerity, $F(1, 103) = 180.39$, $p < .001$, $R^2 = .34$; followed by

blameworthiness, $F(1, 102) = 136.80$, $p < .001$, $R^2 = .20$; and likelihood of further harm, $F(1, 102) = 64.04$, $p < .001$, $R^2 = .15$. These results indicate that the presence of an apology cost had the largest effect on the inference of a group's mental state. In other words, participants seemed to spontaneously ascribe a certain mental state (sincere intention) to group agents. Confirming this hypothesis with self-report measures, we then conducted an fMRI study to test whether costly group apologies would engage the brain regions (i.e., bilateral TPJ, precuneus, MPFC) that were found to be responsive in communicative intention readings tasks.

fMRI Study

Methods

Participants, Ethical Approval, and Data Availability. Participants were 31 healthy Japanese university students (21.32 ± 1.81 years old). However, six participants were discarded from the data analyses (one female and one male due to left-handedness, one female due to excessive head movement [> 3 mm], two females due to mechanical errors, and one female due to misunderstanding of the task). Accordingly, data from 25 participants (16 females and 9 males) were retained in the data analyses. Before participating in the imaging study, participants signed an informed consent form.

This study was approved by the ethical review board at the first author's institute. All data and R codes used in the behavioral data analyses are available in the Open Science Framework (https://osf.io/zw6g7/). The fMRI data are available upon request from the corresponding author.

Task and Procedure. For the imaging study, we selected 10 of the 15 transgression scenarios (see Appendix). The procedure was basically the same as that of Ohtsubo et al.'s (2018) fMRI study on costly apology, but the scenarios were different: the previous study included interpersonal transgression and apology scenarios, while the present study involved group

transgression and apology scenarios.

In the imaging task, participants were first presented a transgression scenario, followed by one of the group reaction scenarios (i.e., either costly apology, non-costly apology, or no apology). Participants then indicated the forgivability of the focal group using a Visual Analogue Scale (VAS) slider. The two poles of the slider were denoted as "not at all forgivable" (converted to 0) and "completely forgivable" (converted to 100). This procedure, depicted in Figure 2, was repeated 30 times. Participants were presented three types of apology scenarios (i.e., 10 costly apology scenarios, 10 non-costly apology scenarios, and 10 no apology scenarios), which were treated as a within-subject factor. Three functional imaging runs (each consisting of 10 trials and lasting for about 6 minutes) were performed for each participant. The order of the 30 scenarios was pseudo-randomized (the same transgression scenarios did not appear more than once in each imaging run) with participants receiving either a forward (scenarios 1–30) or reversed (scenarios 30–1) version to mitigate possible order effects. Moreover, the three group reaction scenarios were spread across the 30 trials in a well-balanced manner (e.g., 4 costly apology, 3 non-costly apology and 3 no apology scenarios in the first run of the forward version).

fMRI Data Acquisition and Analyses. Functional neuroimaging was conducted using a 3-Tesla MRI scanner (Verio; Siemens Ltd., Erlangen, Germany) at the Brain and Mind Research Center, Nagoya University, Japan. Each participant's head was immobilized within a 32-element phased-array head coil. Imaging was performed using an echo-planar imaging (EPI) gradient-echo sequence (echo time [TE] = 30 ms, repetition time [TR] = 2,500 ms, field of view [FOV] = 192 × 192 mm^2, flip angle = 80º, matrix size = 64 × 64, 39 slices, slice thickness = 3 mm, total number of volumes = 148). A whole-brain, high-resolution T1-weighted anatomical magnetization-prepared rapid-acquisition gradient echo (MP-RAGE) MRI was also acquired for each participant (TE = 1.98 ms, TR = 1,800 ms, FOV = 256 × 256 mm^2, flip angle = 9º, matrix

size = 256 × 256 pixels, and slice thickness = 1 mm).

To analyze the functional images, we used Statistical Parametric Mapping (SPM) software (SPM12 revision 6225; Wellcome Trust Centre for Neuroimaging, London, UK) implemented in MATLAB 2014b (MathWorks Inc., Massachusetts). The first four volumes of each fMRI run were discarded due to unsteady magnetization. After all volumes were realigned, differences in slice timing within each image volume were corrected. The reference image was the center of the volume. The whole-brain 3D MP-RAGE volume was co-registered with the EPI volumes and normalized to the MNI T1 image template (ICBM152) using a non-linear basis function. Subsequently, normalization parameters were applied to all of the EPI volumes. The normalized EPI images were then spatially smoothed in three dimensions using an 8-mm full-width at half-maximum Gaussian kernel. After carrying out the realignment processes, we checked for head movements (> 3 mm) during the experimental run. Task-related activation was statistically evaluated on a voxel-by-voxel basis using the general linear model at the individual level to generate contrast images. The transgression (10 s), group reaction (10 s), and forgivability rating phases (5 s) were separately modelled by a block design convolved with a canonical hemodynamic response. The transgression and rating phases were considered covariates of no interest to partial out their contribution to brain activation in single participant analyses.

Using the contrast images related to the group reaction phases of the three conditions (i.e., costly apology vs. non-costly apology vs. no apology), we conducted a random effects analysis at the group level with a one-way within-participant ANOVA. The statistical threshold was set at $< .001$ at the voxel level for an uncorrected p-value and $< .05$ at the cluster level (whole brain) for an FWE-corrected p-value.

Results

We first analyzed the forgivability ratings with the same set of R packages used in Pilot Study. As shown in Figure 3, forgivability was highest when the focal groups issued a costly apology (74.12 ± 14.33), followed by a non-costly apology (51.74 ± 13.36) and no apology (11.28 ± 16.67). A one-way ANOVA with apology type as the within-subject independent variable showed a significant effect of apology type on forgivability, $F(2, 48) = 136.63$, $p < .001$, $R^2 = .75$.

Having confirmed the effect of costly group apology in the self-reported forgivability rating, we proceeded to analyze the fMRI data. Using the subtraction method, we examined which brain regions were more active in response to the costly apology vignettes than to the non-costly apology and no apology vignettes. Given the results of Ohtsubo et al.'s (2018) study, we were primarily interested in the following two comparisons: "Costly Apology > Non-costly Apology" and "Costly Apology > No Apology." The statistical threshold was set at an uncorrected $p < .001$ at the voxel level and a familywise error (FWE) corrected $p < .05$ at the cluster level (whole brain). As shown in Table 1 (upper panel) and Figure 4A, compared with non-costly apology, costly apology was associated with significantly greater blood-oxygen-level dependent (BOLD) responses in the bilateral TPJ and precuneus, along with some other unpredicted regions, such as the orbitofrontal cortex (OFC). Similarly, as shown in Table 1 (middle panel) and Figure 4B, compared with no apology, costly apology was associated with significantly greater BOLD responses in a relatively large cluster that involved the precuneus and bilateral TPJ. Interestingly, as shown in Table 1 (bottom panel), in the comparison of "Non-costly Apology > No Apology," neither the precuneus nor bilateral TPJ was associated with significant greater BOLD responses (in response to this non-significant result, we re-ran the same analysis with a relatively relaxed threshold; the results are reported in Supplementary Materials). This result suggests that the costliness of group apologies, but not the mere presence of apology, was

responsible for the significant activation of the precuneus and bilateral TPJ.

General Discussion

The results of this research support that costly group apology is perceived as more sincere than non-costly group apology. This finding may be either unsurprising or surprising depending on the perspective one takes. Common observations tell us that groups voluntarily offering compensation (a form of costly apology) after committing some transgression are perceived more positively than groups not offering compensation (Koehn, 2013; Minow, 2002). However, Pilot Study uncovered that the effect size was unusually large—the costliness of the apology itself (i.e., costly vs. non-costly apology comparison) accounted for one third of the variance in perceived sincerity. Moreover, one might surmise that the effect of costly group apology could be subsumed to the effect of compensation, but not sincere intention. Nonetheless, the fMRI study revealed that the bilateral TPJ and precuneus were more strongly activated in response to costly group apologies vis-à-vis non-costly group apologies and no apologies. These brain regions have been implicated in cognitive empathy and intention processing (e.g., Ciaramidaro et al., 2014; den Ouden et al., 2005; Enrici et al., 2011; Farrer & Frith, 2002; Hellbernd & Sammler, 2018; Ruby & Decety, 2001; Saxe & Kanwisher, 2003; Tettamanti et al., 2017; Van Overwalle & Baetens, 2009). This result suggests that the effect of costly group apology is mediated by perceived sincere intention (i.e., a mental state ascribed to group agents). Therefore, this study not only generalized the effectiveness of costly apology from the interpersonal apology context to the group apology context but also revealed the mediating role of mental state attribution to group agents (Jenkins et al., 2014; O'Laughlin & Malle, 2002).

Lack of MPFC Activation

There was one noticeable difference between the present study and Ohtsubo et al.'s (2018) interpersonal apology study. Unlike costly interpersonal apology, costly group apology

was not associated with significant activation of MPFC. There are several possible explanations for the absence of MPFC activation. First, this difference might be partly attributable to the participants' perspective. In Ohtsubo et al.'s study, participants adopted the victim perspective and imagined that they themselves would receive a hypothetical apology from the transgressor. In contrast, in the present study, participants adopted a third-party perspective—that is, none of the hypothetical group apologies were directed at the participants themselves. Since the MPFC is often implicated in self-referential processing (e.g., Denny, Kober, Wager, & Ochsner, 2012; Northoff et al., 2016; van der Meer, Costafreda, Aleman, & David, 2010), the third-person perspective may have been responsible for the non-significant MPFC activation in this study. Parenthetically, it is noteworthy that although the precuneus is also often implicated in self-referential processing, costly group apologies engaged the precuneus. However, the precuneus is often implicated in third-party perspective-taking as well (e.g., Cavanna & Trimble, 2006; Ruby & Decety, 2001). Van Overwalle and Baetens (2009) also suggested the possibility that the precuneus might retrieve from the long-term memory social situations closely approximating the current situation to facilitate more adequate mental state inferences.

The second possibility is associated with the absence of apology recipients in the group scenarios. The group apology scenarios used in this study did not explicitly describe the apology recipients. Walter et al. (2004) showed that scenarios describing interactions between two agents (i.e., both the sender and recipient of communicative intention) activated the MPFC, whereas scenarios describing a single individual did not. Therefore, the absence of an explicit description of apology recipients may account for the absence of MPFC activation. The third possibility is associated with the dissimilarity between the group agents and participants themselves. Carrington and Bailey (2008) suggested that the MPFC's involvement (especially the ventral MPFC) in theory-of-mind reasoning may be mediated by "like me" social comparisons. In the

present study, participants (university students) might not have readily compared themselves to group agents, such as a country and company. Regarding the fourth possibility, Van Overwalle and Baetens' (2009) noted that mentalizing tasks involving shape motions, such as Heider and Simmel's (1944) stimulus depicting interactions of geometric shapes, do not systematically recruit the MPFC (and precuneus). Neither group agents nor geometric shapes are biological entities. As Van Overwalle (2009) point out, the MPFC might be more closely associated with enduring trait inferences than temporal intention inferences; thus, people may not spontaneously attribute traits to non-biological entities.

Activation of OFC

In the present study, costly group apology activated the OFC. This region was also activated by costly apology in the interpersonal apology context (Ohtsubo et al., 2018). Interestingly, a slightly different type of forgivability judgment has been shown to activate the OFC. In their fMRI study, Farrow et al. (2001) presented participants with a series of paired transgression scenarios (e.g., shoplifting from a supermarket vs. shoplifting from newsagents) and asked them to decide which scenario was more forgivable. Interestingly, in Farrow et al.'s study, the OFC was more strongly activated by this forgivability judgment task than the other social judgment task. Contrary to these findings, Kim et al. (2019) found that the OFC was more active when participants read less forgivable corporate accounts of their transgressions. Taking these findings together, although the OFC might somehow be involved in calculating forgivability, further studies are needed to draw a strong conclusion regarding the OFC–forgivability judgment relationship.

Limitations and Future Directions

It is worth emphasizing the spontaneous nature of the mental state inference observed in the present study. A previous study investigating mental state inferences of group agents included

mental state vocabularies (e.g., want, know) in the stimuli (Jenkins et al., 2014). In contrast, in the present fMRI study, participants were not explicitly asked to make any mental state inferences. They were simply presented with a series of transgression and group apology scenarios and were asked to rate the forgivability of each group agent. Therefore, it appears that when faced with costly apologies issued by group agents, people spontaneously attribute a mental state (i.e., sincere intention) to the group agents. The spontaneous nature of the task in this study created several limitations. First, we could not analyze whether perceived sincerity was positively associated with BOLD responses in the bilateral TPJ and precuneus. Although we analyzed whether the forgivability score, instead of perceived sincerity, would be positively associated with BOLD responses in these regions, the result was unclear—it was positively associated with BOLD responses in the bilateral TPJ, but not in the precuneus and MPFC (see Supplementary Materials).

Second, and relating to the first limitation, there is room for an alternative interpretation of the observed activation pattern. The right TPJ and precuneus (or their neighboring regions) have been implicated in the retrieval of social memory (e.g., Cavanna & Trimble, 2006; Patel, Sestieri, & Borbetta, 2019; Van Overwalle & Baetens, 2009; Wang et al., 2017). Costly apologies, regardless of whether they were issued by an individual agent or a group agent, may be more memorable than non-costly apologies, and hypothetical scenarios involving costly apologies may promote the retrieval of those memories. Third, it is noteworthy that none of the self-report measures (see Figure 1) showed a comparable pattern with the fMRI data—"costly apology" > "non-costly apology" ≈ "no apology." The valence of the three types of group reactions may account for the fMRI data—only costly apologies are received positively, while the other two types of reactions are received negatively. These issues must be addressed by asking participants to report perceived sincerity and valence during the fMRI data acquisition and testing whether

these measures, in fact, would be associated with BOLD responses in the intention processing network.

Despite the above limitations, the observed similarity in the processing of costly apology in both interpersonal and group apology contexts is important for both academics and practitioners. Academically, without neuroscientific methods, there would be no easy way to ascertain similarity in interpersonal and group contexts. Therefore, this study underscores the importance of adopting a neuroscientific approach to group research in general and to the emerging field of organizational neuroscience in particular (e.g., Becker, Cropanzano, & Sanfey, 2011; Waldman, Ward, & Becker, 2017). Practically, owing to the relatively sparse research on group apology as compared with interpersonal apology, the observed similarity provides a basis for exporting abundant findings accumulated in the interpersonal apology literature to applied contexts, such as crisis communication and international peacemaking. Perhaps we cannot overemphasize the finding that whether or not group apology entails a cost accounted for one third of the variance in perceived sincerity.

Appendix

Table A1. English-Translated Scenarios (Scenarios in bold were used in the fMRI study)

1	**Transgression**	**There have been quite a few accidents where electric appliances produced by Company A have caught fire.**
	Costly Apology	**Company A immediately issued an apology, offered compensation, and announced that it would offer an exchange of the same type of appliance.**
	Non-costly Apology	**Company A apologized for the accidents where their electric appliances caught fire.**
	No Apology	**Company A has not yet apologized because the matter is still under investigation.**
2	**Transgression**	**Foreign objects have been found in bread produced by Company A.**
	Costly Apology	**Company A apologized and stopped its production line to conduct a thorough investigation.**
	Non-costly Apology	**Company A apologized for the inconvenience to those who complained.**
	No Apology	**Company A has not yet apologized, saying that it is still considering how to proceed.**
3	**Transgression**	**Convenience Store A has not made overtime payments to its part-time employees.**
	Costly Apology	**Company A officially apologized and offered the unpaid overtime payments and some compensation.**
	Non-costly Apology	**Company A apologized for the unpaid overtime payments.**
	No Apology	**Company A has not yet apologized, saying that it is still investigating the total amount of unpaid overtime.**

(Table A1 *cont'd*)

4	Transgression	Country A has not provided a sufficient amount of food to refugees.
	Costly Apology	Country A issued an apology and promised to penalize the relevant department and improve refugees' conditions.
	Non-costly Apology	Country A apologized for the inhuman treatment of refugees.
	No Apology	Country A has not yet apologized, claiming that there is no confirmed evidence that it actually occurred.
5	Transgression	Company A sold apartments with insufficient earthquake resistance.
	Costly Apology	Company A immediately apologized to the residents and started aseismic reinforcement works.
	Non-costly Apology	Company A apologized and explained that it had not committed intentional forgery.
	No Apology	Company A has not yet issued an apology, saying that it is still conducting internal investigations.
6	Transgression	It has been detected that an official department formed a back-scratching relationship in the disposal of department properties.
	Costly Apology	The department immediately apologized and conducted an external audit.
	Non-costly Apology	The department apologized by saying that there will be no such incidents in the future.
	No Apology	The department has not apologized, claiming that it was not a back-scratching relationship.
7	**Transgression**	**A school failed to take appropriate action in response to bullying of which it was aware.**
	Costly Apology	**The school apologized immediately, disclosing the results of external inspections.**
	Non-costly Apology	**The school apologized by admitting that it had fallen behind in its actions.**
	No Apology	**The school has not apologized, claiming that it was not aware of the bullying.**

(Table A1 *cont'd*)

8	**Transgression**	**It has been detected that a country's athletes were doping.**
	Costly Apology	**The country publicly apologized and voluntarily returned medals.**
	Non-costly Apology	**The country apologized, stating that its method was wrong.**
	No Apology	**The country has not apologized, denying involvement in doping.**
9	**Transgression**	**A medical error that occurred at Hospital A caused sequelae in its patient.**
	Costly Apology	**Hospital A immediately apologized and accepted an external inspection team.**
	Non-costly Apology	**Hospital A apologized to the patient and their family.**
	No Apology	**Hospital A has not apologized, claiming that the focal operation is associated with a low success rate.**
10	**Transgression**	**It has been revealed that retiring officials at a particular official department have customarily taken lucrative positions in private organizations.** * For Japanese participants, it is self-evident that this custom is illegal.
	Costly Apology	**The department apologized and immediately disclosed the documents required for an investigation.**
	Non-costly Apology	**The department apologized, admitting that its custom was unjustifiable.**
	No Apology	**The department has not apologized, claiming that its custom was not a clearly illegal practice.**
11	Transgression	Company A had new employees work (under the guise of training) before their formal date of employment.
	Costly Apology	Company A immediately apologized and decided to make payments to the new employees.
	Non-costly Apology	Company A apologized, promising that it would take every action to prevent a recurrence.
	No Apology	Company A has not apologized, claiming that pre-employment training is necessary.

(Table A1 *cont'd*)

12	**Transgression**	**Modeling Agency A forced its model to take undesired job offers.**
	Costly Apology	**Modeling Agency A publicly apologized and offered compensation to the model.**
	Non-costly Apology	**Modeling Agency A apologized to the model saying they were sorry.**
	No Apology	**Modeling Agency A has not apologized, claiming that the model voluntarily took the offers.**
13	Transgression	There was elder abuse at Nursing Home A.
	Costly Apology	Nursing Home A apologized and secured a budget to install preventive measures.
	Non-costly Apology	Nursing Home A apologized to the victim and their family.
	No Apology	Nursing Home A has not apologized, stating that it is still investigating whether the abuse, in fact, occurred.
14	**Transgression**	**Nation A's jet fighter violated Japan's airspace.**
	Costly Apology	**Nation A publicly apologized and disclosed the results of its investigation.**
	Non-costly Apology	**Nation A apologized for its unintentional violation.**
	No Apology	**Nation A has not apologized, claiming that no violation occurred.**
15	**Transgression**	**A teacher inflicted corporal punishment on students at a school.**
	Costly Apology	**The school penalized the teacher for using corporal punishment and apologized.**
	Non-costly Apology	**The school apologized and admitted that some educational activities had been excessive.**
	No Apology	**The school has not apologized, claiming that it was a necessary disciplining practice.**

References

Allport, F. H. (1924). The group fallacy in relation to social science. *Journal of Abnormal Psychology and Social Psychology*, *19*(1), 60-73. http://dx.doi.org/10.1037/h0065675

Becker, W. J., Cropanzano, R., & Sanfey, A. G. (2011). Organizational neuroscience: Taking organizational theory inside the neural black box. *Journal of Management*, *37*(4), 933-961. http://dx.doi.org/10.1177/0149206311398955

Bisel, R. S., & Messersmith, A. S. (2012).Organizational and supervisory apology effectiveness: Apology giving in work settings. *Business Communication Quarterly*, *75*(4) 425-448. https://doi.org/10.1177/1080569912461171

Blatz, C. W., & Philpot, C. (2010). On the outcomes of intergroup apologies: A review. *Social and Personality Psychology Compass*, *4*(11), 995-1007. http://dx.doi.org/10.1111/j.1751-9004.2010.00318.x

Blatz, C. W., Schumann, K., & Ross, M. (2009). Government apologies for historical injustices. *Political Psychology*, *30*(2), 219-241. http://dx.doi.org/10.1111/j.1467-9221.2008.00689.x

Bottom, W. P., Gibson, K., Daniels, S. E., & Murnighan, J. K. (2002). When talk is not cheap: Substantive penance and expressions of intent in rebuilding cooperation. *Organization Science*, *13*(5), 497-513. https://doi.org/10.1287/orsc.13.5.497.7816

Bundy, J., Pfarrer, M. D., Short, C. E., & Coombs, W. T. (2017). Crises and crisis management: Integration, interpretation, and research development. *Journal of Managemen*t, *43*(6), 1661-1692. http://dx.doi.org/10.1177/0149206316680030

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, *30*(8), 2313-2335. http://dx.doi.org/10.1002/hbm.20671

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain: A Journal of Neurology*, *129*(3), 564-583. http://dx.doi.org/10.1093/brain/awl004

Ciaramidaro, A., Becchio, C., Colle, L., Bara, B. G., & Walter, H. (2014). Do you mean me? Communicative intentions recruit the mirror and the mentalizing system. *Social Cognitive and Affective Neuroscience*, *9*(7), 909-916. http://dx.doi.org/10.1093/scan/nst062

Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *24*(8), 1742-1752. http://dx.doi.org/10.1162/jocn_a_00233

den Ouden, H. E. M., Frith, U., Frith, C. & Blakemore, S.-J. (2005). Thinking about intentions. *NeuroImage*, *28*(4), 787-796. https://doi.org/10.1016/j.neuroimage.2005.05.001

Desmet, P. T. M., De Cremer, D., & van Dijk, E. (2010). On the psychology of financial compensation to restore fairness transgression: When intentions determine value. *Journal of Business Ethics*, *95*, 105-115. https://doi.org/10.1007/s10551-011-0791-3

Enrici, I., Adenzato, M., Cappa, S., Bara, B. G., & Tettamanti, M. (2011). Intention processing in communication: A common brain network for language and gestures. *Journal of Cognitive Neuroscience*, *23*(9), 2415-2431. https://doi.org/10.1162/jocn.2010.21594

Farrer, C. & Frith, C. D. (2002). Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency. *NeuroImage 15*(3), 596-603 https://doi.org/10.1006/nimg.2001.1009

Farrow, T. F. D., Zheng, Y., Wilkinson, I. D., Spence, S. A., Deakin, J. F. W., Tarrier, N., Griffiths, P. D., & Woodruff, P. W. R. (2001). Investigating the functional anatomy of empathy and forgiveness. *NeuroReport*, *12*(11), 2433-2438. http://dx.doi.org/10.1097/00001756-

200108080-00029

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, *57*, 243-259. http://dx.doi.org/10.2307/1416950

Hellbernd, N., & Sammler, D. (2018). Neural bases of social communicative intentions in speech. *Social Cognitive and Affective Neuroscience*, *13*(6), 604-615. http://dx.doi.org/10.1093/scan/nsy034

Ho, B. (2012). Apologies as signals: With evidence from a trust game. *Management Science*, *58*(1), 141-158. http://dx.doi.org/10.1287/mnsc.1110.1410

Jenkins, A. C., Dodell-Feder, D., Saxe, R., & Knobe, J. (2014). The neural bases of directed and spontaneous mental state attributions to group agents. *PLoS ONE*, *9*(8), e105341. https://doi.org/10.1371/journal.pone.0105341

Kim, H., Kralik, J. D., Yun, K., Chung, Y.-A., & Jeong, J. (2019). Neural correlates of public apology effectiveness. *Frontiers in Human Neuroscience*, *13*:00229. doi:10.3389/fnhum.2019.00229

Koehn, D. (2013). Why saying "I'm sorry" isn't good enough: The ethics of corporate apologies. *Business Ethics Quarterly*, *23*(2), 239-268. https://doi.org/10.5840/beq201323216

Komiya, A., Ohtsubo, Y., Oishi, S. & Mifune, N. (2018). Providing compensation promotes forgiveness for replaceable, but not irreplaceable, losses. *Journal of Social Psychology*, *158*(6), 680-693. https://doi.org/10.1080/00224545.2017.1395792

Martinez-Vaquero, L. A., Han, T. A., Pereira, L. M. & Lenaerts, T. (2015). Apology and forgiveness evolve to resolve failures in cooperative agreements. *Scientific Reports*, *5*, 10639 https://doi.org/10.1038/srep10639

Minow, M. (2002). *Breaking the cycles of hatred*. Princeton, NJ: Princeton University Press.

Northoff G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006).

Self-referential processing in our brain—A meta-analysis of imaging studies on the self. *NeuroImage*, *31*(1), 440-457. https://doi.org/10.1016/j.neuroimage.2005.12.002

O'Connor, C. (2016). The evolution of guilt: A model-based approach. *Philosophy of Science*, *83*(5), 897-908. https://doi.org/10.1086/687873

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. http://dx.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. http://dx.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. http://dx.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 perpetrators perspective. *Evolution and Human Behavior*, *36*(3), 232-239. http://dx.doi.org/10.1016/j.evolhumbehav.2014.11.008

O'Laughlin, M. J., & Malle, B. F. (2002). How people explain actions performed by groups and individuals. *Journal of Personality and Social Psychology*, *82*(1), 33-48. http://dx.doi.org/10.1037/0022-3514.82.1.33

Patel, G. H., Sestieri, C., & Corbetta, M. (2019). The evolution of the temporoparietal junction and posterior temporal sulcus. *Cortex*. https://doi.org/10.1016/j.cortex.2019.01.026

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action:

A PET investigation of agency. *Nature Neuroscience*, *4*(5), 546-550. https://doi.org/10.1038/87510

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *NueroImage*, *19*(4), 1835-1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes*, *101*(1), 1-19. http://dx.doi.org/10.1016/j.obhdp.2006.05.005

Strang, S., Utikal, V., Fischbacher, U., Weber, B., & Falk, A. (2014). Neural correlates of receiving an apology and active forgiveness: An fMRI study. *PLoS ONE*, *9*(2), e87654. https://doi.org/10.1371/journal.pone.0087654

Tanibe, T., Hashimoto, T., Tomabechi, T., Masamoto, T., & Karasawa, K. (2019). Attributing mind to groups and their members on two dimensions. *Frontiers in Psychology*, *10*, 840. https://doi.org/10.3389/fpsyg.2019.00840

van der Meer, L., Costafreda, S., Aleman, A., & David, A. S. (2010). Self-reflection and the brain: A theoretical review and meta-analysis of neuroimaging studies with implications for schizophrenia. *Neuroscience and Biobehavioral Reviews*, *34*(6), 935-946. http://dx.doi.org/10.1016/j.neubiorev.2009.12.004

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*(3), 829-858. http://dx.doi.org/10.1002/hbm.20547

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*(3), 564-584. http://dx.doi.org/10.1016/j.neuroimage.2009.06.009

Waldman, D. A., Ward, M. K., & Becker, W. J. (2017). Neuroscience in organizational behavior.

Annual Review of Organizational Psychology and Organizational Behavior, *4*, 425-444. http://dx.doi.org/10.1146/annurev-orgpsych-032516-113316

Wallter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., & Bara, B. G. (2004). Understanding intentions in social interaction: The role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience*, *16*(10), 1854-1863. http://dx.doi.org/10.1162/0898929042947838

Wang, Y., Collins, J. A., Koski, J., Nugiel, T., Metoki, A., & Olson, I. R. (2017). Dynamic neural architecture for social knowledge retrieval. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(16), E3305-E3314. https://doi.org/10.1073/pnas.1621234114

Table 1. Brain regions that were more active in the presence of costly/non-costly apologies.

Region	Peak MNI Coordinates					Cluster *p* value (FWE-corrected)
	Cluster size	x	y	z	*T*	
Costly Apology > Non-costly Apology						
Occipital Lobe	5992	12	−84	2	6.62	< .001
Orbitofrontal Cortex (OFC)	1076	−2	32	−14	5.74	< .001
Superior Temporal Sulcus (Right)	517	56	−6	−18	5.64	< .001
Precuneus	1703	−6	−50	8	5.35	< .001
TPJ (Left)	641	−40	−70	28	5.01	< .001
Parahippocampal Gyrus	551	26	−26	−4	4.98	< .001
TPJ (Right)	282	50	−58	22	4.97	.038
Costly Apology > No Apology						
Occipital Lobe/***TPJ*** (Bilateral)/ ***Precuneus***	18068	30	−92	−2	10.46	< .001
Middle Frontal Gyrus (Left)	1592	−36	20	24	6.75	< .001
Parahippocampal Gyrus	421	24	−30	−2	5.64	.022
Orbitofrontal Cortex	484	−4	32	−20	4.75	< .001
Middle Frontal Gyrus (Right)	375	38	14	26	4.60	.004
Non-costly Apology > No Apology						
Occipital Lobe/Temporal Lobe (Right)	1220	40	−56	−10	5.93	< .001
Occipital Lobe/Temporal Lobe (Left)	953	−22	−94	−10	5.81	< .001

Notes. Regions relevant to the hypothesis are in bold font. The labels reported in this table were obtained by the atlas tool in the Results section of SMP12 (by choosing “atlas > label using > neuromorphometrics”).

Figure Captions

Figure 1. Boxplots of seriousness, blameworthiness, sincerity, and likelihood of further harm (Pilot Study) as a function of apology type (no apology vs. non-costly apology vs. costly apology).

Figure 2. Time course of the experiment. Each transgression scenario was followed by a group reaction scenario, which was either a costly apology, a non-costly apology, or no apology. After observing the group reaction, participants rated the forgivability of the group. After a 10-second relaxation phase, the same procedure was repeated.

Figure 3. Boxplots of forgivability (in the fMRI study) as a function of apology type (no apology vs. non-costly apology vs. costly apology).

Figure 4. Significantly activated regions in (A) the "Costly Apology > Non-costly Apology" Comparison and (B) the "Costly Apology > No Apology" comparison. Statistical significance thresholds were set at $p < .001$ (uncorrected) at the voxel level and $p < .05$ (familywise-error-[FWE]-corrected; whole brain) at the cluster level.

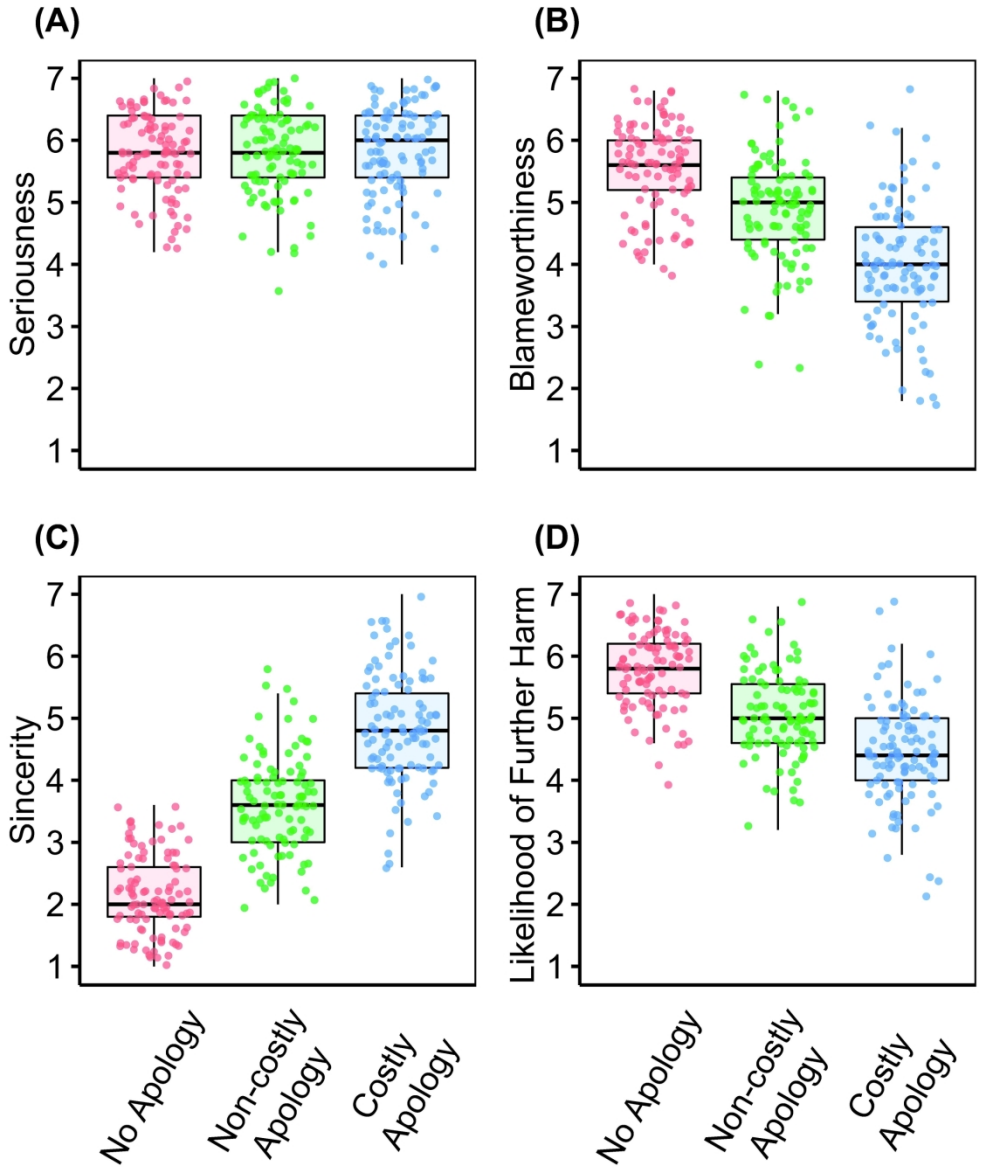


Figure 1. Boxplots of seriousness, blameworthiness, sincerity, and likelihood of further harm (Pilot Study) as a function of apology type (no apology vs. non-costly apology vs. costly apology).

210x251mm (300 x 300 DPI)

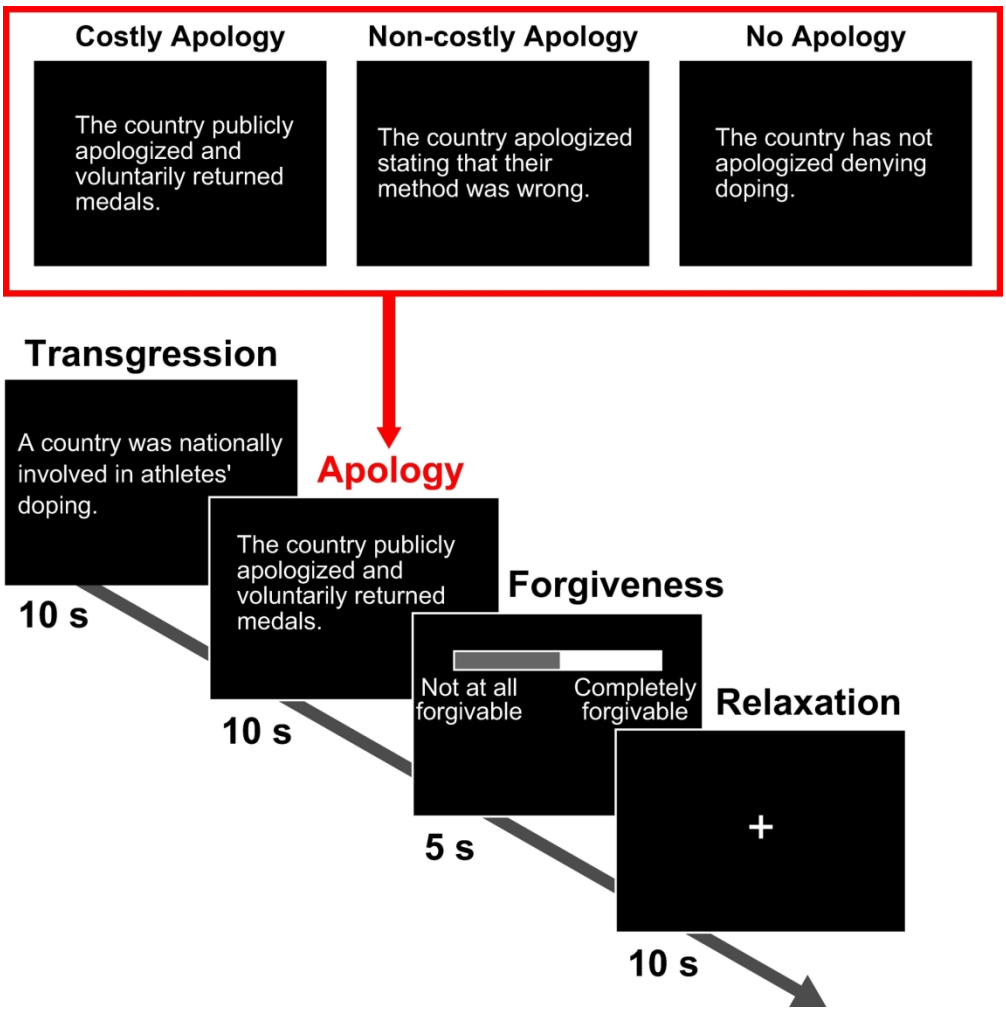


Figure 2. Time course of the experiment. Each transgression scenario was followed by a group reaction scenario, which was either a costly apology, a non-costly apology, or no apology. After observing the group reaction, participants rated the forgivability of the group. After a 10-second relaxation phase, the same procedure was repeated.

172x172mm (300 x 300 DPI)

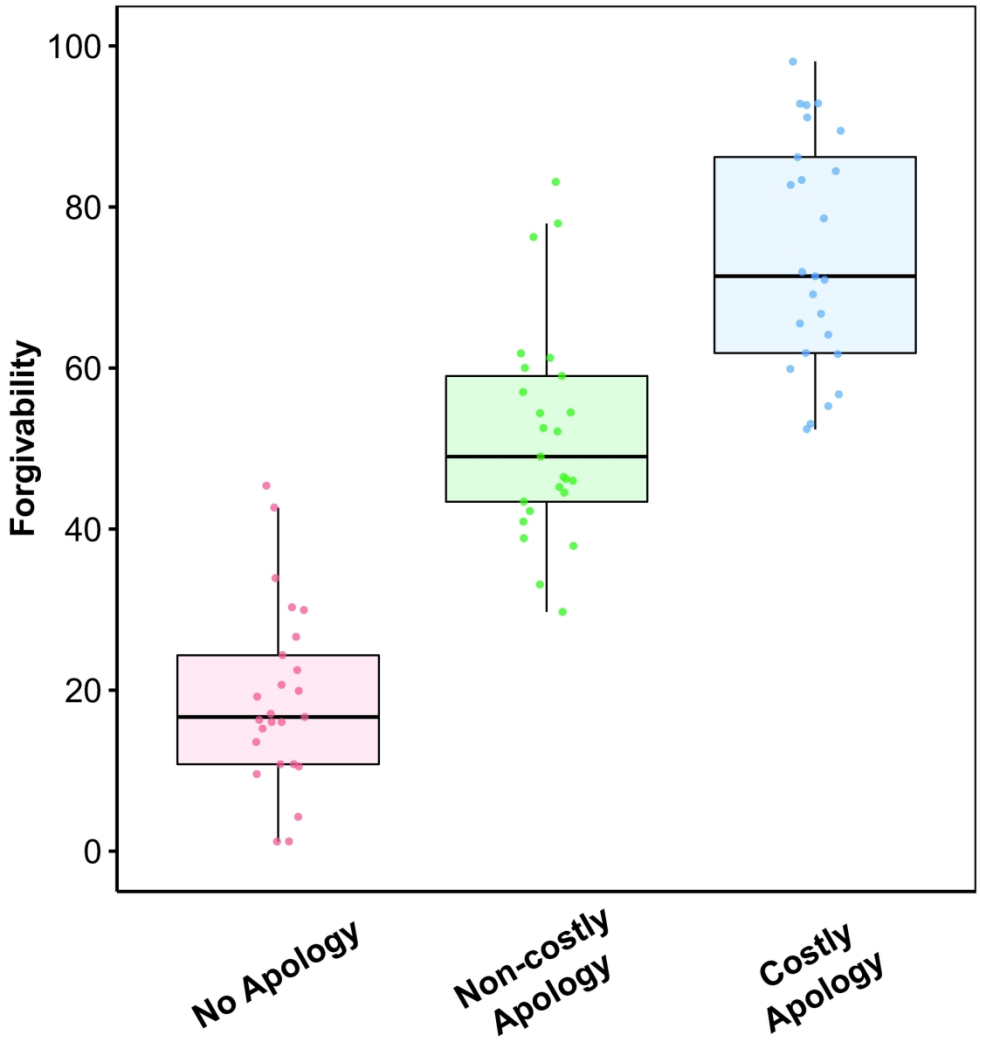
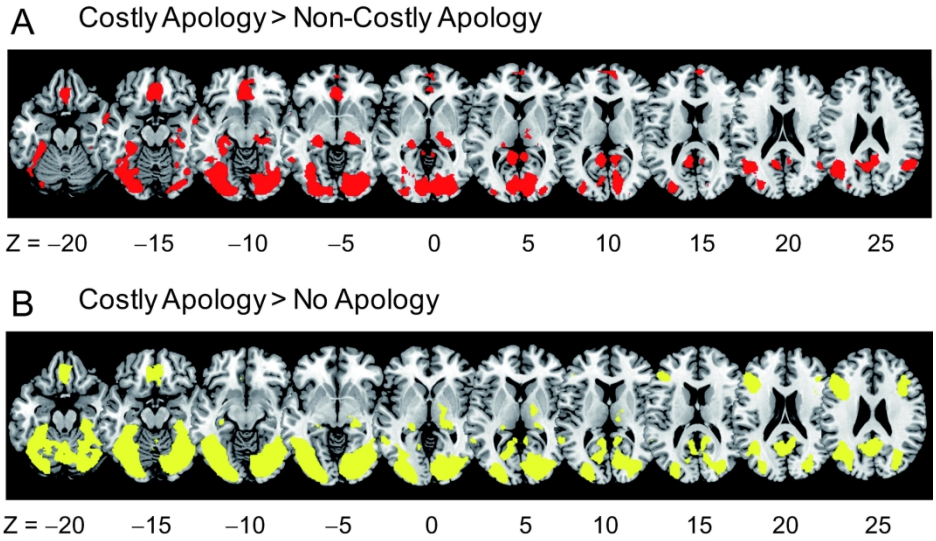


Figure 3. Boxplots of forgivability (in the fMRI study) as a function of apology type (no apology vs. non-costly apology vs. costly apology).

210x224mm (300 x 300 DPI)



Significantly activated regions in (A) the "Costly Apology > Non-costly Apology" Comparison and (B) the "Costly Apology > No Apology" comparison. Statistical significance thresholds were set at $p < .001$ (uncorrected) at the voxel level and $p < .05$ (familywise-error-[FWE]-corrected; whole brain) at the cluster level.

208x143mm (300 x 300 DPI)

Supplementary Materials

Costly Group Apology Communicates a Group's Sincere "Intention"

Table S1. Japanese Scenarios

1	Transgression	ある会社Aが作る電化製品が燃える事故が多発した
	Costly Apology	A社はすぐに謝罪と補償をし、残りも交換すると発表した
	Non-costly Apology	A社は製品が燃え出す事故を起こしたことを謝罪した
	No Apology	A社は調査中として今のところまだ謝罪をしていない
2	Transgression	ある会社Aの作るパンに異物が混入していた
	Costly Apology	A社は謝罪し、生産ラインをとめて原因を入念に調べた
	Non-costly Apology	A社は申し出た人たちにご迷惑をかけたと謝罪した
	No Apology	A社は対応を協議中とし、まだ謝罪していない
3	Transgression	あるコンビニAがバイトに残業代を支払っていなかった
	Costly Apology	A社は公式に謝罪し、残業代の支払いと補償を約束した
	Non-costly Apology	A社は未払いの残業代があったと謝罪した
	No Apology	A社は未払い額の確認中としてまだ謝罪していない
4	Transgression	ある国Aが難民に十分な食事を与えていなかった
	Costly Apology	A国は謝罪し、担当部署の処分と遇改善を約束した
	Non-costly Apology	A国は難民に対する非人道的な扱いを謝罪した
	No Apology	A国はそのような事実は確認されないと謝罪していない
5	Transgression	ある会社Aが耐震上問題のあるマンションを販売した
	Costly Apology	A社はすぐ住民に謝罪し、補強に必要な工事を始めた
	Non-costly Apology	A社は意図的な偽装ではないと説明し謝罪した
	No Apology	A社は社内の調査を行っているとして、謝罪していない
6	Transgression	ある役所が、不要品の払い下げで癒着を指摘された
	Costly Apology	その役所はすぐに謝罪し、第三者の監査を実施した
	Non-costly Apology	その役所は今後このようなことをしないと謝罪した
	No Apology	その役所は不正な癒着ではないとして謝罪していない
7	Transgression	ある学校が、いじめの事実を知りながら対応を怠った
	Costly Apology	その学校はすぐに第三者調査の結果を公表し謝罪した
	Non-costly Apology	その学校は対応が後手にまわったとして謝罪した
	No Apology	その学校はいじめは知らなかったとして謝罪していない
8	Transgression	ある国が組織的に選手にドーピングをしていた
	Costly Apology	その国は公式に謝罪しメダルを自主的に放棄した
	Non-costly Apology	その国はやり方に問題があったと謝罪した
	No Apology	その国はドーピングではないとして謝罪していない
9	Transgression	ある病院Aで、手術ミスで患者に後遺症が残った
	Costly Apology	病院Aはすぐに謝罪し、外部調査委員会を受け入れた
	Non-costly Apology	病院Aは患者とその家族に謝罪をした
	No Apology	病院Aは成功率の低い手術だったとして謝罪していない
10	Transgression	ある役所が代々天下りをしていたことが発覚した
	Costly Apology	その役所は謝罪し、調査に必要な全資料を開示した
	Non-costly Apology	その役所は天下りはよくない慣行だったと謝罪した
	No Apology	その役所は天下りとはいえないとして謝罪していない

(Table S1 *cont'd*)

11	Transgression	ある会社Aは、入社前の学生を研修として働かせた
	Costly Apology	A社はすぐに謝罪し、学生に賃金の支払いを決めた
	Non-costly Apology	A社は再発防止につとめるとして謝罪した
	No Apology	A社は事前の研修は必要だとして謝罪していない
12	Transgression	ある事務所Aは、所属モデルに不本意な仕事をさせた
	Costly Apology	事務所Aは公的に謝罪し、モデルに補償を申し出た
	Non-costly Apology	事務所Aはモデルに申し訳なかったと謝罪した
	No Apology	事務所Aは本人も了承していたとして謝罪していない
13	Transgression	ある介護施設Aで、高齢の入居者への虐待があった
	Costly Apology	施設Aは謝罪し、再発防止策のための予算を増額した
	Non-costly Apology	施設Aは入居者とその家族に謝罪した
	No Apology	施設Aは虐待の有無を調査中として謝罪していない
14	Transgression	ある国Aの戦闘機が日本の領空に侵入した
	Costly Apology	A国は公式に謝罪し、その経緯の調査結果を公表した
	Non-costly Apology	A国は意図せずに日本領空に入ったとして謝罪した
	No Apology	A国は領空侵犯はなかったとして謝罪していない
15	Transgression	ある学校で教師が生徒に体罰をふるっていた
	Costly Apology	その学校はすぐに体罰をした教師を処分し、謝罪した
	Non-costly Apology	その学校はいきすぎた指導があったと謝罪した
	No Apology	その学校は必要な指導だったとして謝罪していない

Supplementary Analyses

"Non-Costly Apology > No Apology" Comparison with a Relaxed Threshold

As reported in the main text, the comparison of "Non-costly Apology > No Apology" revealed non-significant BOLD responses in the precuneus and bilateral TPJ. However, it is reasonable to assume that even non-costly apologies convey some conciliatory intention, and thus recruit these regions. We conducted a set of two follow-up subtraction analyses with a relaxed threshold. First, we defined the precuneus (Precuneus_R + Precuneus_L) as the region of interest (ROI) and the statistical threshold was set as $p < .005$ at the voxel level. Second, we defined the bilateral regions involving TPJ (Postcentral_L + Postcentral_R) as ROI and the statistical threshold was set as $p < .005$ at the voxel level. As show in Table S1, there were no significantly greater BOLD responses in the precuneus and bilateral TPJ. We admit that to address this comparison in a more meaningful manner, we need a replication study with a larger sample size.

Table S1. Results of the subtraction analysis of "Non-costly Apology > No Apology" with a Relaxed Threshold.

	Peak MNI Coordinates					Cluster *p* value	
Region	*Cluster size*	x	y	z	*T*	FWE-corrected	Uncorrected
Precuneus (as ROI)	2	12	−46	14	2.80	.750	.868
TPJ (as ROI)	5	50	−30	48	2.88	.774	.770
TPJ (as ROI)	1	36	−28	44	2.75	.829	.914

Parametric Modulation Analysis: Forgivability and Brain Activity

Participant's forgivability ratings associated with the all 30 scenarios were entered as covariates in the individual-level analysis. We conducted a random effects analysis at the group level using a one-sample *t*-test design. The statistical threshold was set as "< .001" at the voxel level for an uncorrected *p*-value, and "< .05" at the cluster level (whole brain) for a familywise-error-corrected (FWE-corrected) p-value. We found that BOLD responses in some brain regions (including the bilateral TPJ) were positively correlated with participants' forgivability scores (see Table S2). However, it was positively associated with neither precunues nor medial prefrontal cortex (MPFC). Such an ambiguous pattern may be attributable to the use of non-mental state variable (i.e., forgivability). We avoided to ask participants to report the "sincerity" of the apologizing groups because we did not want to artificially activate the intention processing network by forcing participants to apply the mental state inferences to the group agents. In future studies, it is desirable to have two conditions (i.e., the forgivability condition and the sincerity condition) and test whether asking participants to rate sincerity would recruit the intention processing network (i.e., the bilateral TPJ, precuneus, MPFC) more than asking them to rate forgivability. More importantly, we need to test whether the sincerity scores would be positively associated with the BOLD responses in these regions.

Table S2. Brain regions that were significantly correlated with forgivability in response to the friend's commitment signals (and signal failures) collapsing the three conditions.

Region	Peak Coordinates				Cluster size	Cluster *p*-value (FWE-corrected)
	x	*y*	*z*	*T*		
Lingual Gyrus (Right)	14	−80	−6	7.22	1809	< .001
Inferior Occipital Gyrus (Left)	−32	−94	−2	6.14	1654	< .001
Superior Frontal Gyrus (Right)	18	36	46	3.05	575	< .001
Middle Frontal Gyrus (Right)	32	22	24	5.38	253	< .001
TPJ (Angular Gyrus) (Left)	−36	−60	22	5.27	229	< .001
TPJ (Angular Gyrus) (Right)	28	−68	24	5.08	230	< .001
Middle Frontal Gyrus (Left)	−32	2	32	4.86	222	< .001

Notes. TPJ = temporoparietal junction.