LEBS
Vol. 12 No. 1 (2021) 7–11
Letters on Evolutionary Behavioral Science

Checking Boxes for Making an Apology: Testing the Valuable Relationships Hypothesis by a New Method

Ryo Oda[1,*], Kai Hiraishi[2]

[1]Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan

[2]Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan

[*]Author for correspondence (oda.ryo@nitech.ac.jp)

A reconciliation signal may have evolved because it can reduce the uncertainty of defection, which might be caused by an error in a repeated prisoner's dilemma situation. Ohtsubo and Watanabe (2009) proposed the "costly apology model" and argued that transgressors should make apologies in a costly way to convey sincerity to victims. Ohtsubo and Yagi (2015) showed that people are more likely to make a costly apology to valuable partners than to less valuable ones. The cost of apology was measured by the strength of willingness to suffer some inconvenience (e.g., cancellation of an important meeting) for it. Such a method of measuring costs is, however, dependent on culture, context, and the participants' personal situations, and can only measure costs indirectly. We attempted to replicate the results of Ohtsubo and Yagi (2015) by using the checkbox method as a measurement of cost of apology. Among the factors expected to affect the number of checkboxes checked, the primary factor was the instrumentality of the friend, which replicated the results of Ohtsubo and Yagi (2015). Because the participants paid real costs of effort and time for fictitious mistakes, the costly apology might be triggered quickly and intuitively by a heuristic.

Keywords
valuable relationships hypothesis, costly apology model, costly signaling theory, reconciliation, measuring method

Introduction

Repairing damaged relationships is essential for many primate species living in groups. The valuable relationships hypothesis (De Waal, 2000) states that reconciliation is an adaptation to maintain beneficial social relationships. It is useful not only when one has inflicted harm on a relationship but also when one is the victim of an unintended harmful action. If your partner erroneously defected in a repeated prisoner's dilemma situation, it is not a good idea to terminate the reciprocal relationship (Ohtsubo, 2015). The problem is that it is difficult to know the actor's intention and, thus, we would expect signals that convey "true" and "benign" intention to have evolved. For instance, Cheney et al. (1995) suggested that the grunts produced by baboons (*Papio cynocephalus ursinus*) might function as a reconciliation signal.

A typical reconciliation signal in humans is apology. However, if the person receiving the apology sees it as insincere, reconciliation will not be achieved; rather, it could even worsen the relationship (e.g., Zechmeister et al., 2004). Ohtsubo and Watanabe (2009) proposed the "costly apology model" and argued that transgressors should make apologies in a costly way (see also Ohtsubo, 2015). They relied on the signaling theory, which states that communication signals acquire reliability only when there is a cost to the sender in its production. There are two types of costs: qualitative and general. The qualitative one is represented by Zahavi's handicap theory, which argues that individuals with a high quality (e.g., physical strength) can live with an adaptively disadvantageous handicap just because they have higher quality, while low-quality individuals cannot afford the costly handicap (Zahavi & Zahavi, 1997). Thus, the handicapping trait serves as an honest signal of the individual's quality (Grafen, 1990). On the other hand, general cost is the same for all senders, and senders that will benefit more are more likely to signal (Laidre & Johnstone, 2013). If an apology is associated with a cost, it works as a general handicap signal of sincerity. Apologies without cost, on the other hand, will not be trusted. Indeed, Ohtsubo and Watanabe (2009) demonstrated, with vignette and behavioral experiments, that participants perceived more sincerity from costly apologies. Notably, self-punishing apologies that do not bring any benefit to the victim had the same effect as compensatory apologies such as buying a lunch. The effect was confirmed to be robust with undergraduate samples from seven countries with various religious backgrounds (Ohtsubo et al., 2012).

The other key prediction from the costly apology model was that a person makes a costly apology only when they value the damaged relationship. Specifically, the more one values the relationship, the more cost they will pay to make the apology. Ohtsubo and Yagi (2015) tested this prediction with vignette studies. They asked participants to imagine that they had committed an interpersonal transgression against one of their real friends. The cost of apology was measured by the strength of their willingness to suffer some inconvenience (e.g., cancellation of an important meeting to make an apology as soon as possible). The results supported the prediction; participants reported that they were willing to make a more costly apology to more valuable partners than to less valuable ones.

There are two weaknesses with Ohtsubo and Yagi's (2015) methodology. One is that, how much each option

doi: 10.5178/lebs.2021.82
Received 24 February 2021.
Accepted 04 March 2021.
Published online 17 March 2021.

costs depends on cultural, contextual, and personal situations. They prepared an apology cost item such as, "I will buy my friend some gift later as an apology," and measured willingness to take these actions on a 4-point Likert scale. What people think of as a "gift" may vary from person to person, and the meaning of sending a gift to a friend might vary from culture to culture. Because of the untranslatability of the materials, it is almost impossible to conduct a multicultural direct replication of their study while such a study is strongly expected because of the problems in arguing about "human nature" based on a finding from a limited, biased, homogeneous sample (e.g., a WEIRD-western, educated, industrialized, rich, and democratic sample; Cheon et al., 2020; Henrich et al., 2010). The other problem is that their method cannot directly measure the cost associated with the apology.

In the present study, we attempted to replicate Ohtsubo and Yagi (2015) conceptually using the newly developed "checkbox method", which can solve the aforementioned two weaknesses. In this method, participants are asked to check one box in each of 10 × 10 checkboxes on a web page to express their willingness to incur a cost. Checking the boxes is a tedious and time-consuming activity, though not necessarily an exceptionally large burden; this method allows us to measure quantitatively willingness to incur a cost. We employed the same transgression scenarios used in Study 2 of Ohtsubo and Yagi (2015), and measured relationship value by asking participants to quantify the usefulness of the friends in their work or hobby. Study 2 also examined gender, scenario, expected forgiveness, and closeness with the friend as factors; therefore, we also examined these influences. If the relationship value also affected the number of checkboxes selected, the results of Ohtsubo and Yagi (2015) would be conceptually replicated.

Methods

Participants

A total of 400 Japanese adults (200 females and 200 males, mean age: 44.8 years, range: 20−69 years) were recruited through Cross Marketing, Inc. (Tokyo, Japan), a research agency that maintains a panel of more than 2,000,000 individuals who have agreed and consented to participate in web-based online surveys. Participants were recruited based on age to ensure an equal number of participants in each experimental group (40 females and 40 males in each age group: 20−29, 30−39, 40−49, 50−59, and 60−69 years). They were allocated equally to each of two scenarios.

Composition of the questionnaire

First, participants were asked to respond either "yes" or "no" to the following questions: *Are you working or looking for a job? Do you have or are you looking for a part-time job? Do you have any hobbies? Do you volunteer or otherwise engage in other activities? Are you interested in interpersonal relationships (e.g., continuing current relationships or making new friends)? Are you married? Do you have a family?*

Participants were then asked to identify one of their best friends as a target by writing down that friend's initials and were asked questions about the relationship length (in months) and subjective closeness (7-point scale) with that best friend. Then, participants were asked to imagine that they had committed one of two types of interpersonal transgression. The transgression scenarios were the same as those used in Ohtsubo and Yagi (2015). One scenario described a situation in which participants unwittingly failed to arrive at a meeting place (no-show scenario). The other scenario described a situation in which participants unwittingly stained a book that they had borrowed from the target (stained book scenario). After reading one of the scenarios, participants were asked to indicate their willingness to apologize by clicking the checkboxes; there were 100 of these, with 10 per row (Figure 1). Each box was numbered from 1 to 100, from left to right, and it was only possible to check them in order, starting with the smallest number (from 0 as no apology at all to 100 as apologize profusely).

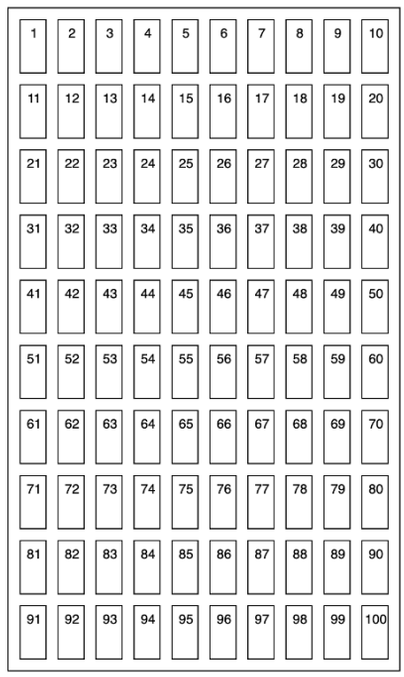

Figure 1. Image of the checkboxes.

Then participants were asked several questions regarding the nature of their friendship with that target, including the target's instrumentality. We used the measure of relationship value (i.e., perceived goal instrumentality) proposed by Smith et al. (2020). Participants were asked to rate the instrumentality of the best friend using a 7-point scale (from −3 as an extreme hindrance to +3 as an extreme help). Specifically, they rated how much of a help or hindrance their best friend was for achieving goals in the following eight domains of life: work, part-time jobs, hobbies, volunteer activities, interpersonal relationships, romantic relationships, family relationships, and other important goals. An item was removed from the calculation of the participant's mean score on this measure if the participant had responded "no" for the corresponding first question. The last section of the questionnaire included an item assessing expected forgiveness: "*How likely do you think it is that the friend will forgive you anyway even if you do not apologize for it?*" This item was accompanied by a 4-point scale ranging from 1 (not at all) to 4 (very much).

Data analysis

We considered the number of checked boxes as an ordinal scale and categorized every 10 checks into 11 levels, with none checked being level 1. We used a generalized linear model to analyze the effects of predictor variables on the rankings from 1 to 11. The predictor variables analyzed were scenario (stained book: 1, no-show: 0), gender (male: 1, female: 0), age, length of relationship, closeness, instrumentality, and expected forgiveness. The dependent variable was modeled as an ordinal scale using HAD software to produce an ordinal logit model (Shimizu, 2016).

Results

In Table 1, we list the sample size and median number of boxes checked for each rank category. Figure 2 shows the detailed distribution of the number of boxes checked. The most frequent number of boxes checked was 100, by 99 participants (24.8%). The next most frequent was one box, by 27 participants (6.8%). Participants tended to check multiples of 10 more often than other numbers.

Table 1. Range, sample size and median number of boxes checked in each rank category.

Rank	Range	*n*	Number of boxes checked (Median)
1	0	5	0
2	1–10	82	3
3	12–20	28	20
4	22–30	18	30
5	33–40	15	40
6	46–50	29	50
7	55–60	22	60
8	65–70	22	70
9	74–80	35	80
10	83–90	29	90
11	95–100	115	100

The Spearman rank coefficients between each pair of factors indicate that instrumentality was most highly correlated with rank (Table 2). The null hypothesis, that the degree of decrease of deviance from inputting the predictor variables to the model was zero, was rejected (χ^2 = 97.13, df = 7, p < .001). Gender, closeness, relationship period, instrumentality, and scenario significantly contributed to the rank (pseudo R^2 = .214; Table 3). The highest odds ratio was for instrumentality.

Because the number of participants who checked all 100 boxes was extremely high, we did a post-hoc, dichotomous analysis of instrumentality. The instrumentality score of the best friend of the 99 participants who checked all 100 boxes (0.65 ± 0.59) was significantly higher than that of the other participants (0.37 ± 0.61; t = −4.02, df = 169.88, p < .001, 95% CI = [−0.415, −0.141], Cohen's d = 0.46)

Discussion

We measured willingness to apologize for an imagined interpersonal transgression against a participant's real friend by the number of boxes they checked. Among the factors expected to affect that number, the primary factor was the instrumentality of the friend, which replicated the results of Ohtsubo and Yagi (2015). Moreover, the cost of the apology was larger in the stained book scenario than in the no-show scenario, which was also consistent with Ohtsubo and Yagi (2015). The participants of the current study were recruited via a survey agency and had a wider age range (20s to 60s) compared to Ohtsubo and Yagi (2015), whose participants were undergraduates in their late teens to early 20s. These results show that the valuable relationship hypothesis is supported by the newly developed method and with a more varied sample.

It is interesting to note that in Ohtsubo and Yagi (2015) the participants answered about the degree of a fictitious cost they would pay for an apology for a fictitious mistake, whereas in the present study, the participants paid a real cost of effort and time for the same fictitious mistake. This means that people pay real costs by just imagining a situation for which they should apologize. If a costly apology were to be decided on a rational basis, it is unlikely that anyone would actually pay a cost, albeit a small one such as checking these boxes, for an imaginary mistake. That is, these results suggest that the costly apology is triggered quickly and intuitively by a heuristic.

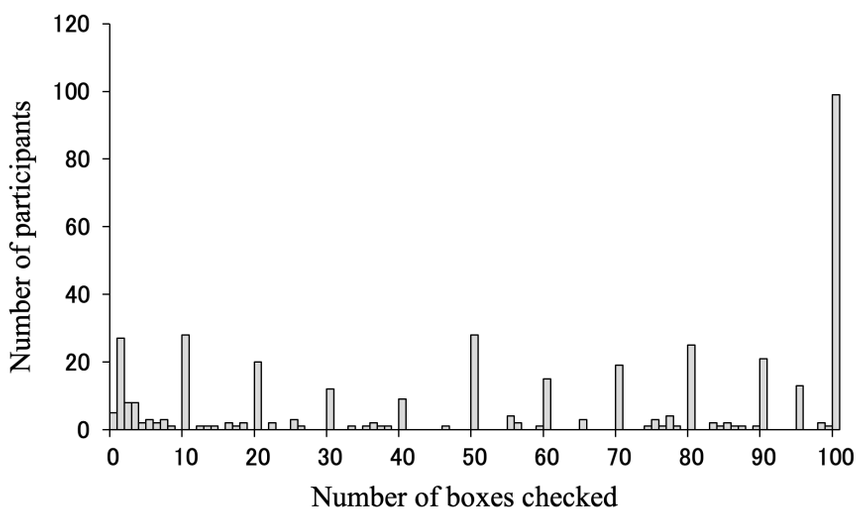

Figure 2. Distribution of the number of boxes checked.

Table 2. Spearman rank coefficients between each factor.

Factor	median	1	2	3	4	5	6	7
1. check (rank)	8							
2. gender (1 = male, 0 = female)		−.169**						
3. Age (year)	45	.213**	.029					
4. Closeness	6	.257**	−.124*	.037				
5. Relationship period (month)	242.5	.255**	.054	.622**	.209**			
6. Instrumentality	0.38	.278**	.006	.064	.333**	.033		
7. Expected forgiveness	3	−.035	.032	.022	.067	.072	.013	
8. Scenario (1 = stained book)		.153**	.000	−.003	.014	.036	.031	−.176**

* $p < .05$, ** $p < .01$

Table 3. Odds ratio of each factor on the number of boxes checked.

Factor	Odds ratio	95% CI	VIF
Gender	0.496**	0.339–0.728	1.029
Age	1.011	0.993–1.031	1.709
Closeness	1.256*	1.043–1.512	1.233
Relationship period	1.002**	1.001–1.004	1.808
Instrumentality	2.207**	1.513–3.221	1.149
Expected forgiveness	0.958	0.734–1.251	1.042
Scenario (1 = stained book)	1.724**	1.187–2.503	1.029

* $p < .05$, ** $p < .01$

Kiyonari et al. (2000) proposed the "social exchange heuristic", which helps to achieve mutual cooperation. They suggested that cooperation in the one-shot prisoner's dilemma could be explained by a heuristic that encourages participants to perceive the situation as an assurance game, in which cooperation is a better choice than defection insofar as the partner is cooperating as well. The costly apology may be caused by the same kind of social heuristic, to prevent defection due to some error and thus maintain mutual cooperation. However, since the degree of cost is affected by the instrumentality of the victim, there might be a function in the mind that calculates the degree of cost from the instrumentality. In future studies, it might be possible to clarify these relationships using the checkbox method, which has made quantitative cost evaluation possible.

There were a few our findings that did not replicate Ohtsubo and Yagi (2015). While they reported that the higher the expectation of forgiveness, the less cost the participant paid, expected forgiveness had no significant effect in our study. On the other hand, the significant positive effect of closeness we found was not reported in Ohtsubo and Yagi (2015), which might be due to the fact that the "Inclusion of the Other in the Self" scale was used to measure closeness in Ohtsubo and Yagi (2015), while a 7-point Likert scale was used in our study. Moreover, although Ohtsubo and Yagi (2015) did not examine the effect of relationship length (probably due to the fact that their participants were undergraduates), in our study the relationship period positively affected the number of checked boxes. It is possible that differences in methods from previous studies, such as the age range of the participants and the measurement methods, might have contributed to these differences.

The distribution of the number of boxes checked shows that about a quarter of the participants checked all 100 boxes. Although we measured the time and effort required to check the boxes as the willingness to incur a cost, checking all 100 boxes might be too easy compared to the intent to apologize. Because this might cause a ceiling effect, we could consider increasing the number of boxes in further studies. Another tendency of the distribution was that the number of boxes checked was concentrated in multiples of 10. Since there were 10 checkboxes in each row, participants might tend to stop checking the boxes at a pleasing, round number. One possible countermeasure would be to eliminate this checkstop cue. For example, there might be a way to arrange checkboxes in a spiral pattern or to present a randomized pattern of arrays. In human social life, possible topics relevant to costly signaling are sincerity, insistence, and strength of commitment to another person. The checkbox method we examined in this study has the merits that it can be performed easily within the format of traditional web survey services and makes participants pay a real cost of effort and time. In addition, cross-cultural comparisons are much easier than with typical Likert scales because there is no need to translate subtle nuances. It has the potential to allow us to gather a lot of quantitative data for willingness to incur a cost from wider samples in a way that is more useful than a simple Likert scale.

Acknowledgement

We appreciate Dr. Yohsuke Ohtsubo for permitting to use the questionnaire and an anonymous reviewer for helpful comments. This work was supported by JSPS KAKENHI Grant Number 20H01755.

Supplementary Material

Data are available online.

References

Cheney, D. L., Seyfarth, R. M., & Silk, J. B. (1995). The role of grunts in reconciling opponents and facilitating interactions among adult female baboons. *Animal Behaviour*, *50*(1), 249–257. https://doi.org/10.1006/anbe.1995.0237

Cheon, B. K., Melani, I., & Hong, Y.-Y. (2020). How USA-centric is psychology? An archival study of implicit assumptions of generalizability of findings to human nature based on origins of study samples. *Social*

Psychological and Personality Science, *11*(7), 928-937. https://doi.org/10.1177/1948550620927269

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

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, *144*(4), 517–546. https://doi.org/10.1016/S0022-5193(05)80088-8

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*, 29. https://doi.org/10.1038/466029a

Kiyonari, T., Tanida, S., & Yamagishi, T. (2000). Social exchange and reciprocity: Confusion or a heuristic? *Evolution and Human Behavior*, *21*(6), 411–427. https://doi.org/10.1016/S1090-5138(00)00055-6

Laidre, M. E., & Johnstone, R. A. (2013). Animal signals. *Current Biology*, *23*(18), R829-R833. https://doi.org/10.1016/j.cub.2013.07.070

Ohtsubo, Y. (2015). Understanding reconciliation from an evolutionary social psychological perspective: The valuable relationships hypothesis and a costly apology model. *Japanese Journal of Social Psychology*, *30*(3), 191–212. https://doi.org/10.14966/jssp.30.3_191 (in Japanese with English abstract)

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

Shimizu, H. (2016). An introduction to the statistical free software HAD: Suggestions to improve teaching, learning and practice data analysis. *Journal of Media, Information and Communication*, *1*, 59–73. https://ci.nii.ac.jp/naid/120005744983 (in Japanese with English abstract)

Smith, A., McCauley, T. G., Yagi, A., Yamaura, K., Shimizu, H., McCullough, M. E., & Ohtsubo, Y. (2020). Perceived goal instrumentality is associated with forgiveness: A test of the valuable relationships hypothesis. *Evolution and Human Behavior*, *41*(1), 58-68. https://doi.org/10.1016/j.evolhumbehav.2019.09.003

Zahavi, A. & Zahavi, A. (1997). The handicap principle: A missing piece of Darwin's puzzle. Oxford University Press.

Zechmeister, J. S., Garcia, S., Romero, C., & Vas, S. N. (2004). Don't apologize unless you mean it: A laboratory investigation of forgiveness and retaliation. *Journal of Social and Clinical Psychology*, *23*(4), 532–564. https://doi.org/10.1521/jscp.23.4.532.40309