**WEB APPENDICES**

When and Why Saying “Thank You” Is Better Than Saying “Sorry” in Redressing Service Failures: The Role of Self-esteem

**WEB APPENDIX A**

**Web Appendix** **Table 1: Summary of Studies**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Study**  **(type)** | **Main purpose** | | **Service failure context** | **DV(s)** | **Alternative accounts**  **examined** |
| 1A  *(lab)* | Showing the basic effect: appreciation is more effective than apology in redressing service failure | | Inferior gift | --Postrecovery satisfaction |  |
| 1B  *(lab)* | Wrong survey | --Compliance (intention to fill out the right survey) |  |
| 1C  *(field)* | Delayed compensation | --Compliance (intention to fill out a follow-up survey)  --Postrecovery satisfaction |  |
| ID  *(MTurk)* | Airline overbooking | --Postrecovery satisfaction  --Intention to file a complaint |  |
| IE  *(MTurk)* | Service provider unavailability | --Postrecovery satisfaction  --Likelihood to accept the new service arrangement  --Intention to file a complaint  --Intention to ask for compensation |  |
| 2  *(MTurk)* | Providing process evidence | Measuring self-esteem | Plumber appointment delay | --Postrecovery satisfaction  --Intention to repatronize  --Intention to recommend |  |
| follow-up  *(lab)* | Measuring self-esteem | Airline luggage delay | --Postrecovery satisfaction  --Overall satisfaction  --Intention to repatronize | --Cause attribution  --Severity inference |
| 3  *(lab)* | Manipulating self-esteem | Restaurant service delay | --Tipping likelihood  --Tip amount | --Frequency inference  --Relationship perception |
| 4  *(MTurk)* | Process-by-moderation (of narcissism) | Restaurant service delay | --Tip amount |  |
| 5  *(lab)* | Testing the moderating effect of recovery timing |  | --Tip amount | --Severity inference  --Frequency inference  --Relationship perception  --Valence |
| 6  *(lab)* | Testing the effect of combining appreciation and apology | | Online store shipping delay | --Postrecovery satisfaction | --Frequency inference  --Relationship perception |
| 7  *(MTurk)* | Testing the roles of service failure severity and utilitarian recovery | | Restaurant service delay | --Postrecovery satisfaction |  |

**WEB APPENDIX B**

**Web Appendix Table 2: Summary of Gender Effects Across All Studies[[1]](#footnote-1)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Studies | DV | Gender as a covariate | | Gender as a moderator |
| Main or interaction effect | The effect of gender | Interaction effect |
| Study 1A | Postrecovery satisfaction | F(2, 135) = 15.97,  *p* < .01 | F(1, 135) = .80,  *p* = .37 | F(2, 133) = .33,  *p* = .72 |
| Study 1D | Postrecovery satisfaction | F(1, 113) = 3.73,  *p* = .06 | F(1, 113) = .24,  *p* = .62 | F(1, 112) = 1.72,  *p* =.19 |
| Intention to file a complaint | F(1, 113) = 3.71,  *p* = .07 | F(1, 113) = .07,  *p* = .79 | F(1, 112) = .34,  *p* = .56 |
| Study 1E | Postrecovery satisfaction | F(1, 195) = 10.18,  *p* < .01 | F(1, 195) = 3.96,  *p* = .05 | F(1, 194) = 3.04[[2]](#footnote-2),  *p* = .08 |
| Intention to file a complaint | F(1, 195) = 6.14,  *p* = .01 | F(1, 195) = 3.34,  *p* = .07 | F(1, 194) = 1.71,  *p* = .19 |
| Likelihood to accept the new service arrangement | F(1, 195) = 4.86,  *p* = .03 | F(1, 195) = .49,  *p* = .48 | F(1, 194) = .15,  *p* = .70 |
| Intention to ask for compensation | F(1, 195) = 5.37,  *p* = .02 | F(1, 195) = .003,  *p* = .96 | F(1, 194) = 1.71,  *p* = .19 |
| Study 2 | Postrecovery satisfaction | F(2, 190) = 33.30,  *p* <.01 | F(1,190) = .07,  *p* = .79 | F(2, 188) = .41,  *p* = .66 |
| Intention to repatronize | F(2, 190) = 9.83,  *p* <.01 | F(1, 190) = .00,  *p* = .99 | F(2, 188) = .84,  *p* = .43 |
| Intention to recommend | F(2, 190) = 13.61,  *p* <.01 | F(1,190) = .04,  *p* = .84 | F(2,188) = 1.16,  *p* = .32 |
| Self-esteem | F(2, 190) = 21.71,  *p* < .01 | F(1,190) = 1.67,  *p* = .20 | F(2, 188) = .24,  *p* = .79 |
| Study 2 follow-up | Postrecovery satisfaction | F(1, 131) = 8.00,  *p* < .01 | F(1,131) = 1.57,  *p* = .21 | F(1, 130) = .70,  *p* = .40 |
| Overall satisfaction | F(1, 131) = 10.52,  *p* < .01 | F(1,131) = 1.33,  *p* = .25 | F(1,131) = 1.54,  *p* =. 22 |
| Postrecovery satisfaction | F(1,131) = 10.87,  *p* < .01 | F(1,131) = 4.03,  *p* < .05 | F(1,130) = 3.68[[3]](#footnote-3),  *p* = .06 |
| Self-esteem | F(1,131) = 10.30,  *p* < .01 | F(1,131) = .06,  *p* = .80 | F(1, 130) = 5.16[[4]](#footnote-4),  *p* =.03 |
| Study 3 | Tipping likelihood | Interaction beween prerecovery self-esteem and recovery strategy:  F(1, 254) = 6.74,  *p* = .01 | F(1, 254) = .03,  *p* = .86 | Three-way:  F(1, 251) = .92,  *p* =.34 |
|  | Tip amount | Interaction beween prerecovery self-esteem and recovery strategy:  F(1, 254) = 3.86,  *p* = .05 | F(1, 254) = 2.38,  *p* = .12 | Three-way:  F(1, 251) = .33,  *p* = .57 |
| Study 4 | Tip amount | Interaction between narcissism and recovery strategy: t(248) = 1.77, *p* = .08 | t(248) = .74,  *p* = .46 | Three-way:  t(245) = .003,  *p* = 1.00 |
| Study 5 | Tip amount | Interaction between recovery timing and symbolic recovery:  F(1, 165) = 3.92,  *p* = .05 | F(1, 165) = .13,  *p* = .72 | Three-way:  F(1, 162) = 1.39,  *p* = .24 |
| Study 6 | Postrecovery satisfaction | F(2,203) = 3.35,  *p* = .04 | F(1, 203) = 0,  *p* = 1.00 | F(2, 201) = .58,  *p* = .56 |
|  | Overall satisfaction | F(2, 203) = 3.65,  *p* = .03 | F(1, 203) = .84,  *p* = .36 | F(2, 201) = 1.88,  *p* =.16 |
| Study 7 | Postrecovery satisfaction | Interaction between symbolic recovery, utilitarian recovery and failure severity:  F(1, 530) = 8.91,  *p* <.01 | F(1, 530) = .20,  *p* = .66 | Four-way:  F(1, 523) = .37,  *p* =.55 |

**WEB APPENDIX C**

**Study 2 Follow-up**

This follow-up study has four objectives. First, we further examine the proposed underlying mechanism that elevated self-esteem mediates the advantageous effects of appreciation. Second, it attempts to improve the manipulation of the recovery strategies by keeping the wording as similar as possible (“*Thank you for the wait. We appreciate it!*” vs. “*Sorry about the wait. We apologize!*”). Third, it examines cause attribution (i.e., apology vs. appreciation might imply that the cause of the service failure is internal to the company, or that the failure is stable and under the control of the company) and severity inference (i.e., apology vs. appreciation might imply that the service failure is more severe) as alternative accounts. Finally, to further assess the robustness of our effect, we use a different, airline service failure situation.

***Method***

This study employed a two-cell, one-way (recovery strategy: appreciation vs. apology) between-subjects design. One hundred thirty-four undergraduate students (47.01% female; Mage = 21.54 years, SD =1.64) from a large Midwestern university in the United States completed the study for course credit and were randomly assigned to one of the two conditions.

Participants were asked to imagine the following service encounter: “You went on a trip and your flight arrived at your destination on time. However, you were told that your luggage had been misplaced and would be delayed for about 2 hours. The airline delivered your luggage to the hotel. When the customer representative delivered the luggage, he said: “*Thank you for the wait. We appreciate it!*” (“*Sorry about the wait. We apologize*!”)

Afterward, participants indicated their satisfaction with the way the airline handled the luggage delay using the same measure as used previously (α = .98; averaged to form a postrecovery satisfaction index), reported their overall satisfaction with the airline (“The service of this airline is very thoughtful,” “I am very satisfied with this airline,” “The service of this airline is very good”; 1 = “strongly disagree”, 7 = “strongly agree”; α = .90; averaged to form an overall satisfaction index). and indicated their repatronizing intention (willingness to fly with the airline again; 1 = “not at all,” 7 = “very much”).

We then assessed participants’ state self-esteem with the same measure as used in Study 2 (α = .94; averaged to form a state self-esteem index). In addition, we measured cause attribution of the delay (four dimensions: locus of causality, external control, stability, and personal control; McAuley, Duncan, and Russell 2007) and severity inference (How do you feel about the delay?; 1 = “short/not acceptable/not reasonable”, 7 = “long/acceptable/reasonable (reverse coded)”; α = .72; averaged to form a severity inference index). Finally, participants provided basic demographic information.

***Results***

***Postrecovery responses.***ANOVAs conducted on the postrecovery satisfaction index, overall satisfaction index, and repatronizing intention consistently revealed a significant effect of recovery strategy (postrecovery satisfaction: F(1, 132) = 8.21, *p* = .005, ηp2 = .06; overall satisfaction: F(1,132) = 10.75, *p* = .001, ηp2 = .08; repatronizing intention F(1,132) = 11.07, *p* = .001, ηp2 = .08). Planned contrasts showed that appreciation led to greater postrecovery satisfaction (M = 4.66, SD = 1.38 vs. M = 3.88, SD = 1.78), overall satisfaction (M = 4.23, SD = 1.36 vs. M = 3.45, SD = 1.37), and re-patronizing intention than apology (M = 4.36, SD = 1.48 vs. M = 3.53, SD = 1.38) (see Web Appendix Figure 1).

***Self-esteem.***The same ANOVA performed on the state self-esteem index again showed a significant effect of recovery strategy (F(1,132) = 10.34, *p* = .002, ηp2 = .07). Participants in the appreciation condition (M = 4.38, SD = 1.60) reported higher self-esteem than those in the apology condition (M = 3.52, SD = 1.47) (see Web Appendix Figure 1).

***Alternative accounts*.** No significant difference was found between appreciation and apology across all four dimensions of cause attribution (locus of causality, external control, stability, and personal control, *ps* > .56). Similarly, no difference was found in severity inference (*p* >.71) (see Web Appendix D for more details).

**Figure 1**

***Mediation analysis*.** To demonstrate that self-esteem mediates the effect of recovery strategy on postrecovery satisfaction, overall satisfaction, and repatronizing intention, we performed three separate mediation analyses (PROCESS Model 4; Hayes 2013) with 5000 bootstrapping iterations. The results showed that the indirect effects of recovery strategy through self-esteem were significant (for postrecovery satisfaction: b = .53, SE = .18, 95% CI [.20, .92]; for overall satisfaction: b = .56, SE = .19, 95% CI [.22, .94]; for re-patronizing intention: b = .45, SE = .15, 95% CI [.17, .76]).

***Discussion***

This study further confirmed the role of self-esteem in mediating the advantageous effect of appreciation versus apology in redressing service failures. Moreover, the study casted doubts on two “phrasing-as-information” accounts by showing that the two recovery strategies did not affect participants’ attribution of the cause of the service failure and their severity inference.

**WEB APPENDIX D**

**Alternative Accounts Test in All Studies**

Across Studies 2 -6, we examined five alternative accounts related to “phrasing as information” (i.e., consumers may draw inferences from the wording of the recovery messages, which might account for the superior effect of appreciation): (1) cause attribution, where apology (vs. appreciation) might imply that the cause of the service failure is internal to the company, or that the failure is stable and under the control of the company; (2) severity inference, where apology (vs. appreciation) might imply that the service failure is more severe; (3) frequency inference, where apology (vs. appreciation) might imply that the service failure occurs on a more regular basis; (4) relationship perception, where appreciation (vs. apology) might lead consumers to perceive their relationship with the service provider as more communal oriented (vs. exchange oriented); (5) valence of the recovery message, where appreciation (vs. apology) might set a positive tone and imply the service provider’s expectations of the customers’ patience, inducing them to behave consistently with their expected “merits.” To examine these accounts, we first examine whether these factors are influenced by the recovery strategy. In addition, we include these factors as covariates in the analyses to examine whether our key effect still holds after statistically controlling for these factors. Below, we report these analyses in detail.

***Study 2 Follow-up:***

***Cause attribution.*** Following McAuley, Duncan and Russel (1992), we derived seperate indices for the four dimensions of causal attribution: locus of causality (α = .68), external control (α = .68), stability (α = .49), and personal control (α = .71). Four separate one-way (recovery strategy) ANOVAs were performed on these four dimensions, respectively. The results showed that recovery strategy did not affect these cause attributions (all (F(1, 132) < .98, all *p*s > .56). In addition, including these four dimensions as covariates in the one-way (recovery strategy) ANOVAs performed on the postrecovery responses did not change the significance of the effects of recovery strategy (postrecovery satisfaction: F(1, 128) = 8.71, *p* < .01, ηp2 = .06; overall satisfaction: F(1, 128) = 11.36, *p* < .01, ηp2 = .08; repatronizing intentions: F(1, 128) = 13.28, *p* < .01, ηp2 = .09).

***Severity inference***. A one-way (recovery strategy) ANOVA performed on severity inference showed that recovery strategy did not influence participants’ perceived severity of the service failure (F(1, 132) = .13, *p* = .72). In addition, including severity inference as a covariate in the one-way (recovery strategy) ANOVAs performed on the postrecovery responses did not change the significance of the effects of recovery strategy (postrecovery satisfaction: F(1, 131) = 8.21, *p* < .01, ηp2 = .06; overall satisfaction (F(1, 131) = 11.17, *p* < .01, ηp2 = .08); repatronizing intentions: F(1, 131) = 12.36, *p* < .01, ηp2 = .09).

***Study 3:***

***Frequence Inference.*** A 2 (prerecovery self-esteem) × 2 (recovery strategy) ANOVA performed on frequence inference revealed only a marginally significant main effect of recovery strategy (F(1,255) = 3.13, *p* = .08, ηp2 = .01). However, participants in the appreciation condition believed that delays occur more frequently in the restaurant (M = 4.66, SD = 1.19 ) than those in the apology condition (M = 4.39, SD = 1.26). Therefore, this effect could not account for the results we found previously and rendered our hypothesis test even more conservative. Furthermore, including frequency inference as a covariate in the 2 (prerecovery self-esteem) × 2 (recovery strategy) ANOVAs performed on postrecovery responses did not change the significance of the interaction (for tipping likelihood:F(1, 254) = 6.74, *p* = .010, ηp2 = .03; for tip amount: F(1, 254) = 3.61, *p* = .06, ηp2 = .01).

***Relationship Perception***. The same ANOVA was conducted on participants’ perceived relationship with the server. Neither the main effects nor the interaction effect approached significance (*p*s >.12), suggesting that the self-esteem manipulation and, more important, the recovery strategy manipulation did not change participants’ relationship perception. The grand mean (M = 5.32, SD = 1.30) was significantly higher than the scale midpoint of 4 (t(258) = 16.46, *p* < .01), suggesting that participants generally considered the relationship a business relationship (vs. friendship). In addition, including relationship perception as a covariate in the 2 (prerecovery self-esteem) × 2 (recovery strategy) ANOVAs performed on postrecovery responses did not change the significance of the interaction (for tipping likelihood:F(1, 254) = 6.67, *p* = .01, ηp2 = .03; for tip amount: F(1, 254) = 3.57, *p* = .06, ηp2 = .01).

***Study 5:***

***Severity Inference.*** The 2 (recovery timing) × 2 (recovery strategy) ANOVA conducted on severity inference revealed only a significant main effect of recovery timing (F(1, 166) = 18.01, *p* < .01, ηp2 = .10). Specifically, participants viewed the service failure to be more severe when the recovery was communicated before (M = 3.95, SD = 1.09) than after (M = 3.23, SD = 1.15) the service failure, which suggests that precursory recovery effort might backfire. In addition, including severity inference as a covariate in the 2 (recovery timing) × 2 (recovery strategy) ANOVA performed on the tip amount did not change the significance of the interaction (F(1, 165) = 3.38, *p* = .07, ηp2 = .02).

***Frequency inference.***The 2 × 2 ANOVA conducted on frequency inference revealed no significant effects (*p*s >.18). In addition, after including frequency inference as a covariate in the 2 × 2 ANOVA performed on the tip amount, the interaction between recovery timing and recovery strategy remained significant (F(1, 165) = 3.91, *p* = .05, ηp2 = .02).

***Relationship perception.***Similarly, the 2 × 2 ANOVA conducted on relationship perception revealed only a significant main effect of recovery timing (F(1, 166) = 4.16, *p* = .04, ηp2 = .02). Specifically, postfailure recovery led participants to perceive the relationship to be more like business relationship (M = 5.39, SD = 1.27) than prefailure recovery (M = 4.99, SD = 1.30). In addition, after including relationship perception as a covariate in 2 × 2 ANOVA performed on the tip amount, the interaction between recovery timing and recovery strategy remained significant (F(1, 165) = 3.87, *p* = .05, ηp2 = .02).

***Study 6:***

***Frequency inference.***A one-way (recovery strategy) ANOVA performed on frequency inference showed that recovery strategy did not influence participants’ frequency inference (F(2, 204) = .43, *p* = .65, ηp2 < .01). Entering frequency inference as a covariate in the one-way (recovery strategy) ANOVA performed on postrecovery satisfaction did not change the effect of recovery strategy (F(2, 203) = 3.67, *p* = .03, ηp2 = .04).

***Relationship perception****.* The same ANOVA conducted on relationship perception showed that recovery strategy did not change participants’ relationship perception (F(2, 204) = .78, *p* = .46, ηp2 < .01). In addition, including relationship perception as a covariate in the one-way ANOVA (recovery strategy) performed on postrecovery satisfaction did not change the effect of recovery strategy (F(2, 203) = 3.56, *p* = .03, ηp2 = .03).

**WEB APPENDIX E**

**Post-test of Self-esteem Manipulation in Study 3**

***Method***

This study employed a two-cell, one-way (performance feedback: positive/90th percentile vs. negative/10th percentile) between-subjects design. One hundred eighteen Mturk workers (66.10% female; Mage = 39.12 years, SD =12.66) completed the study for monetary compensation and were randomly assigned to one of the two conditions.

Similar to the procedure used in the main study, participants first completed the Word Generation Task and received the positive (90th percentile) or negative (10th percentile) performance feedback. Afterward, they read the introduction of our self-esteem measure: “We all have our way of thinking about ourselves. How we think about ourselves, for example, or how confident we feel, can change from time to time or from situation to situation. Please think about the feedback you just received and then indicate how you felt when you received this feedback.” Participants then responded to the three-item measure of state self-esteem used in Study 2: (1) At this moment, I take a positive attitude toward myself; (2) At this moment, I feel that I am a person of worth; (3) At this moment, I believe I am a valuable person (1 = strongly disagree, 7 = Strongly agree; α = .96; averaged to form a state self-esteem index).

In addition to this postfeedback self-esteem measure, we used another, “postfeedback versus prefeedback” measure to help participants more accurately report how their state self-esteem had changed as a function of our manipulation (positive vs. negative performance feedback) compared to their prefeedback state. In other words, this measure explicitly used the prefeedback state as the baseline, against which the postfeedback state was gauged. Specifically, participants read: “Now, let’s say your level of self-esteem before conducting the Word Generation Task was 0, what is the number that best describes your level of self-esteem after you received feedback about your performance on the Word Generation Task?” They then responded to the question: “If your self-esteem has increased after feedback, please pick a positive (+) number. If your self-esteem has decreased after the feedback, pick a negative (-) number” (-3 = decreased a lot, 0 = stayed the same, +3 = increased a lot). Finally, participants reported their demographic information.

***Results***

The ANOVA conducted on the state self-esteem index showed that participants reported a higher level of self-esteem after receiving the positive (M = 5.88, SD = 1.13) versus negative (M = 4.52, SD = 1.99; F(1,116) = 20.17, *p* < .01, ηp2 = .32) feedback. In addition, the measure assessing the change in participants’ state self-esteem showed that the positive feedback significantly increased participants’ self-esteem (M = 1.14, SD = 1.20; t-tested against midpoint (0 = stay the same): t(58) = 7.30, *p* < .01), while the negative feedback significantly decreased self-esteem (M = -.68, SD = 1.29; t-tested against midpoint (0 = stay the same): t(58) = 4.03, *p* < .01). These results together suggest that our manipulation successfully increased or decreased participants’ self-esteem as intended.

**WEB APPENDIX REFERENCES**

McAuley, E., Duncan, T., and Russell, D. (1992), “Measuring Causal Attributes: The Revised Causal Dimension Scale (CDSII), *Personality and Social Psychology Bulletin*, 18, 566-573.

DeWall, C. Nathan, Roy F. Baumeister, Nicole L. Mead, and Kathleen D. Vohs (2011), “How Leaders Self-Regulate Their Task Performance: Evidence That Power Promotes Diligence, Depletion, and Disdain,” *Journal of Personality and Social Psychology*, 100 (1), 47–65.

1. We examine whether gender interfered with our proposed effect across all studies except for Study 1b, in which gender was not recorded for participants who scanned the QR code, and Study 1c, in which gender was not recorded. [↑](#footnote-ref-1)
2. The main effect of recovery strategy was still significant (F(1, 194) = 11.42, *p* < .01) [↑](#footnote-ref-2)
3. The main effect of recovery strategy was still significant (F(1, 130) = 10.27, *p* < .01). [↑](#footnote-ref-3)
4. The main effect of recovery strategy was still significant (F(1, 130) = 9.68, *p* <.01). [↑](#footnote-ref-4)