ORIGINAL PAPER

A Sorry Excuse for an Apology: Examining People's Mental Representations of an Apologetic Face

Meghan George[1,2] · Joshua R. Guilfoyle[1] · Jennifer R. Steele[1] · C. W. Struthers[1]

Abstract

The goal of the current research was to gain an understanding of people's mental representations of an apologetic face. In Study 1, participants' responses were used to generate visual templates of apologetic faces through reverse correlation (Study 1a, $n = 121$), and a new set of participants (Study 1b, $n = 37$ and 1c, $n = 153$) rated that image (group-level Classification Image, CI), as well as either the inverse image (group-level anti-CI in Study 1b) or base face (in Study 1c), on apology-related characteristics. Results demonstrated that people have a mental representation of an apologetic face, and that sadness is an important feature of this template. To examine similarities between mental representations of apologetic and sad faces, participants in Study 2 generated visual templates of sad faces using reverse correlation (Study 2a, $n = 121$). New participants (Study 2b, $n = 162$) were then randomly assigned to rate the averaged face, eyes, and mouths (group-level CIs) as well as the individual visual templates (individual-level CIs) generated from both studies for either how apologetic or sad they appeared. Visual templates of apologetic and sad faces were seen as apologetic, providing evidence of the prominence of sadness in mental representations of apology.

Keywords Apology · Face perception · Reverse correlation · Nonverbal · Sad

Introduction

> You can always say sorry, but the real apology is when you hear the sadness in their voice and see the look in their eyes. And you realize that they have hurt themselves just as much.
> Kid Cudi (2012)

Apologies are interpersonal tools that can repair and improve valuable relationships. When people transgress against one another, existing social bonds become threatened, and norms dictate that transgressors should apologize if they wish to repair the damage to

✉ Meghan George
meghanlgeorge@gmail.com

1   Department of Psychology, York University, Toronto, ON, Canada

2   Department of Psychology, York University, 4700 Keele St, Toronto, ON M3J 1P3, Canada

the relationship caused by their hurtful acts (Goffman, 1955). Once an apology is offered, recipients then evaluate the apology, along with the importance of this social bond, and decide how to proceed with the relationship (Burnette et al., 2012). Because apologies are commonly used to help maintain social connections, it is important to understand what makes a person seem truly apologetic. Despite a rich literature that explores the verbal components of apologies (e.g., Bippus & Young, 2020; Kirchhoff et al., 2012; Lewicki et al., 2016), very limited research has examined the corresponding nonverbal facial cues that make a face appear apologetic. The goal of the present research was to increase our understanding of how apologies are conveyed nonverbally, by examining people's mental representations of an apologetic face. Specifically, across two studies, we used reverse correlation to examine whether people have a mental representation of an apologetic face and whether apology-related characteristics (specifically sadness) are important components of this representation.

What are the Components of an Effective Apology?

How apologies affect relationships in the wake of transgressions has long been of interest to social psychologists. Apologies allow transgressors to mitigate the loss of trust between themselves and those that they have harmed (Lewicki et al., 2016), the perceived injustice or unfairness of their behavior, and a perceived lack of respect for the target of their transgression (DeCremer & Schouten, 2008). In addition, apologizing has the potential to relieve guilt (Riek, 2010), demonstrate an understanding of their transgression, act as an outlet for empathetic responses (e.g., crying; Hornsey et al., 2019), and demonstrate sympathy for the wronged individual (Slocum et al., 2011). Although it may be difficult for offenders to admit wrongdoing (Riek, 2010; Schumann, 2014), the drawbacks of apologizing can be offset by the forgiveness that they can elicit (Riek, 2010).

Not all apologies are created equal, however, and researchers have sought to identify which elements result in the most successful outcomes (Fehr & Gelfand, 2010). For instance, researchers have found that the key verbal components of an effective apology include expressing remorse, taking responsibility, asking for forgiveness, and promising not to reoffend (Kirchhoff et al., 2012; Lewicki et al., 2016; Schlenker & Darby, 1981; Struthers et al., 2008). Yet interpersonal communication is a complex, multi-layered, and dynamic process; beyond the words that we use, nonverbal cues such as vocal tone, eye gaze, and hand gestures, are often also used to help us convey our thoughts and emotions (Poyatos, 2015). Importantly, facial expressions have been characterized as "the most significant nonverbal language" (Awasthi & Mandal, 2015, p. 1) and "the crown jewel of nonverbal communication" (Deodhare, 2015, p. 173), suggesting that apologies might also be conveyed nonverbally through our face.

Apologies as Expressions of Emotion and Motivation

Anyone who has ever been on the receiving end of an apology can likely confirm that words mean much less if they are not accompanied by appropriate nonverbal cues. But which cues are "appropriate"? Social and moral norms dictate that a transgressor should feel negative emotions such as sadness, hurt, and remorse, particularly if their actions led to the harm of others. Research into public and corporate apologies (Hornsey et al., 2019; Ten Brinke & Adams, 2015), as well as punishment decision making (Picó et al., 2020) offer examples from empirical investigations of the role of these emotions and their embodiment

in apologies. In another example using qualitative data, participants explained that, from a victim's perspective, a transgressor should focus at least some of their attention to their own feelings because expressions of negative emotions such as hurt and regret during an apology help the victim determine whether they should forgive (Slocum et al., 2011). In addition, research has found a strong link between the experience of an emotion and the outward facial expression of that emotion. For example, perceptions of sadness, anger, and happiness, have been found to correlate highly with self-reports of internal emotional experiences (see Matsumoto et al., 2008a, 2008b for a review). Therefore, it seems likely that even when a transgressor verbally communicates their emotions, nonverbal facial expressions conveying those emotions may amplify their apologetic message. Conversely, facial expressions may hinder the apology process if they are not congruent with the content of the verbal apology—something some transgressors may be either explicitly or implicitly aware of. There is evidence to suggest that when a perpetrator intentionally transgresses, they feel less guilt (Leunissen et al., 2013) and are less willing to apologize. Transgressors low in empathy—and therefore low in their ability to vicariously experience the emotions of those they have harmed—have also been shown to be less likely to apologize following wrongdoing (Howell et al., 2012; Schumann & Dragotta, 2021). In both of these examples, it seems possible that it is difficult to express emotions that are not genuinely felt, and, in the case of an apology, such disingenuous or incongruent expressions may be detrimental to the social relationship.

Extrapolating from research on verbal apologies and emotions, we posit that a key function of facial expressions associated with apologies is to communicate self-focused emotional information. However, as many of the other verbal aspects suggest, apologies also serve an other-oriented function (Schumann, 2018). Because transgressions involve both a perpetrator and a victim, an apology is socially motivated, conveying information to a victim about the importance of the relationship and a desire to repair it (Schumann, 2014). For example, the act of transgressing can damage trust between an offender and a victim, and for an apology to be perceived as genuine, the verbal apology literature suggests that trust must be re-established (Hornsey et al., 2019; Lewicki et al., 2016; Tomlinson et al., 2004). Previous research has demonstrated that trust can be signalled using nonverbal cues (Hornsey et al., 2019) and perceived in facial stimuli (see Dotsch et al., 2008). Therefore, nonverbal communication of trustworthiness may be one effective way to communicate genuine apology via facial expression.

Other socially motivated aspects of apologies that have emerged from the verbal literature include offering to compensate victims, as well as directly asking for forgiveness (Kirchhoff et al., 2012). It is unclear whether these can be communicated nonverbally; however, there is evidence to suggest that apologies from high-status individuals within a corporation are more effective relative to lower-status individuals (Ten Brinke & Adams, 2015). We argue that attempts to appear in a low-power or in a submissive position relative to the apology recipient may serve as a form of social compensation (e.g., Hornsey et al., 2019), as well as a way to surrender control and ask for forgiveness. For example, when a transgressor wants to repair his or her relationship with those they have wronged, the injured party has the power to absolve the wrongdoing and mend the relationship, or to end the relationship. Acknowledging this power could differentially serve as a form of social currency and expressing submission could be considered a nonverbal proxy for offering verbal compensation to the target of a transgression. It may also signal that the transgressor is surrendering control of the situation in asking for forgiveness. Given evidence that many of the components of an apology can be expressed nonverbally, we aimed to investigate whether people have a mental template of an apologetic face, and if so,

which apology-related characteristics are important components of this representation? To address these questions, we made use of methods and procedures from the face perception literature, specifically, reverse correlation.

Reverse Correlation and Mental Representations

Reverse correlation is a data-driven method that has been gaining popularity in psychological research over the last several years. Originally developed for use in vision and neurophysiology research (Dotsch & Todorov, 2012), social psychologists quickly adopted the technique to identify specific features or characteristics that participants use as mental templates for various traits and domains (e.g., gender, emotion, group membership). This method allows for the analysis of small sections of the face that may differ across judgements of particular domains and generates individual- and group-level prototypes, termed *classification images* (or CIs; Dotsch & Todorov, 2012). From these prototypes, specific facial features or cues that define the appearance of certain traits or domains, such as happy or dominant, can be identified.

A great deal of research using the reverse correlation technique has focused on intergroup cognitions and attitudes (e.g., Dotsch et al., 2008; Imhoff et al., 2011). However, this technique has the potential to shed light on more abstract concepts and how they are mentally represented and communicated. For example, reverse correlation has been used to create visual templates of trustworthy faces (Dotsch & Todorov, 2012), as well as faces that differ across social class (Bjornsdottir & Rule, 2017). It can be difficult to describe or communicate certain attributes or expressions, and reverse correlation bypasses the need to physically demonstrate or verbally articulate a concept. As such, this is a particularly useful method for more complex emotions (Frijda, 1988). For the current investigation into nonverbal apologies, reverse correlation provides access to perceptual information that participants may or may not realize influences their judgments. By understanding people's mental representations of an apologetic face, we can better predict why faces are perceived and categorized as apologetic or not.

The Present Research

The goal of the present research was to determine whether people have a mental representation of an apologetic face that can be approximated using reverse correlation and whether apology-related characteristics derived from our understanding of effective verbal apologies are prominent in this visual representation. Consistent with previous reverse correlation research, each study consisted of two phases (e.g., Brooks et al., 2018; Brown-Iannuzzi et al., 2017; Lei & Bodenhausen, 2017). In Study 1a, participants were presented with 400 forced-choice trials consisting of distorted images derived from a single White male face – which was generated by creating a morphed average of faces from the Karolinska Face Database (Lundqvist et al., 1998). Each distinct distortion pattern layered over the original base face resulted in a unique-looking stimulus image (Dotsch & Todorov, 2012). Participants were asked to select which face of the pair appeared more apologetic for each trial. Their selections across each of these trials were then combined within (individual-level) and then across (group-level) all participants to create visual templates, or classification images (CI), which are the averaged noise patterns associated with each selected response (Brinkman et al., 2017). In Study 1b, a new set of participants were asked to rate the CI, as well as the reverse image (anti-CI) and the base face from which it was created (Study 1c),

on multiple apology-related emotions and characteristics that were generated from words mentioned or implied in the verbal apology literature. All measures, manipulations, and data exclusions are reported either within the methods section of each study or in the supplemental materials.

Study 1a—Image Generation

Method

Participants

One hundred and twenty-five undergraduate students participated in exchange for course credit. The sample size in previous reverse correlation research has varied widely (e.g., $N=36$, Brooks et al., 2018; $N=238$, Brown-Iannuzzi et al., 2017). As such, we decided a priori to recruit a moderately sized sample (i.e., stopping rule of 128 participants) that was the approximate midpoint of the range of samples found in the literature, and terminated data collection once we believed this sample size had been achieved. We excluded the data from one participant who did not complete the study, as well as three participants who did not meet the cut-off for correct responses recommended on the Conscientious Responders Scale (i.e., at least 3 out of 5; Marjanovic et al., 2014). The final sample consisted of 121 participants (women = 91, men = 29, other = 1, $M_{age}=19.92$ years, $SD_{age}=3.54$). Participants identified their race as South Asian ($n=37$), White ($n=33$), Middle Eastern ($n=13$), Black ($n=12$), East Asian ($n=11$), Southeast Asian ($n=9$), Latin or South American ($n=3$), Mixed Race ($n=2$), or other ($n=1$).

Measures

Reverse Correlation In the current study we used a forced-choice reverse correlation paradigm (see Brinkman et al., 2017; Todorov et al., 2011) that included 400 trials. The stimuli used in each trial were created using a grayscale image of a White male face generated by the morphed average of faces from the Karolinska Face Database (Lundqvist et al., 1998; see Fig. 1). This single image served as our base face, or the image on top of which visual distortion patterns were placed.

Visual distortions (noise patterns) were randomly generated in R using the rcicr package using open-source code (Dotsch, 2017). Each pattern was comprised of pixel regions that differed in their degree of light or dark. For each noise pattern, a complete inverse pattern was also created. All visual noise patterns were then overlaid onto the base face image. As is the nature of visual noise, this caused some areas of the underlying face to appear distinct based on how the light and dark pixel patterns interacted with the base image. The inverse pattern made areas on the stimuli that appeared light appear dark, and vice versa.

In each of the 400 trials, participants were shown two images simultaneously, each containing the same base face, one overlaid with the noise pattern and one with the inverse noise pattern (Dotsch & Todorov, 2012). Participants were asked to select which of the two images "looks more apologetic." Responses were recorded and the noise patterns associated with each of the selected stimuli were averaged using the rcicr package in R (Dotsch, 2017). This participant-specific pixel average was then once again superimposed on the base image, resulting in an image that best exemplified their internal representation of apology (individual-level CI).

Fig. 1 Base Face (AKDEF image id: MNES) used to generate stimuli for Study 1a. All random visual noise patterns were placed on top of this image

A similar process was then repeated across all participants, and an image was created for the whole sample (a group-level CI). The same process was completed using the noise patterns that were not selected by participants in each trial, resulting in individual-level and group-level *anti*-CIs.

Conscientious Responders Scale To assess whether participants were reading instructions in full and responding thoughtfully, we used the Conscientious Responders Scale (CRS; Marjanovic et al., 2014). Participants were instructed to respond in particular ways (e.g., "In response to this question, please choose option number three, *slightly disagree*") to five questions randomly distributed throughout the questionnaire portion of the study. Following recommendations outlined by Marjanovic et al. (2014), participants who correctly answered three or more questions in this scale were classified as conscientious responders and their data were retained for analyses.

Procedure

Participants completed the study individually on a desktop computer located in a laboratory at a Canadian University. A research assistant provided initial instructions and remained in a nearby room to answer any questions participants may have had. All questions and experimental stimuli were programmed using Qualtrics software. After providing informed consent, participants completed several demographic questions and then proceeded to the reverse correlation trials. After completing the reverse correlation trials, participants completed some additional exploratory scales as part of a larger study (see Supplemental Materials) and were then debriefed.

Fig. 2 Group-level Apology CI generated from Study 1a by averaging the visual noise patterns of the images selected as appearing apologetic in the forced-choice reverse correlation paradigm

Fig. 3 Group-level Apology Anti-CI generated from Study 1a by averaging the visual noise patterns of the images that were *not* selected in the forced-choice reverse correlation paradigm

Results

Reverse Correlation

The group-level CI of an apologetic face (Fig. 2), as well as the anti-CI (Fig. 3) are presented below.

Study 1b—Image Rating

To address our research questions, participants were asked to rate the group-level apologetic CI and the anti-apologetic CI (Study 1b), or the group-level apologetic CI and the base face from which it was created (Study 1c) along multiple dimensions that were generated by the authors based on their theoretical or practical significance in the verbal apology literature. For example, we collected ratings of apologetic, remorseful, regretful, and trustworthy to respectively represent the key verbal components of saying sorry, expressing remorse and regret, and promising not to reoffend (Kirchhoff et al., 2012; Lewicki et al., 2016; Schlenker & Darby, 1981; Struthers et al., 2008). Sadness and dominance were selected based on findings that public apologies that include crying (Hornsey et al., 2019) and high-level executives taking responsibility for wrongdoing (Ten Brinke & Adams, 2015) result in more positive outcomes. We also collected ratings for traits and expressions that do not have an explicit, empirically established role in nonverbal apologies, but that we felt made sense theoretically. These traits were selected following several brainstorming sessions and discussions among the authors, and included innocent, sincere, and kind. Overall, our investigation included constructs associated with emotions (e.g., sad), character traits (e.g., kind), and social categories (e.g., gender).[1]

Method

Participants

Forty-one undergraduate students completed the rating portion of the study in exchange for course credit. Our sample size was again guided by previous research (e.g., Bjornsdottir & Rule, 2017), and we decided a priori to stop data collection after 40 participants had completed the study. Participants signed up for and completed the study online. We removed data from three participants who did not complete the study and one participant who responded incorrectly to three or more of the CRS questions. After removing these participants, our final sample was 37 (women = 32, men = 5, $M_{age}=19.89$ years, $SD_{age}=3.08$). The final sample included participants who identified as South Asian ($n=13$), White ($n=9$), Middle Eastern ($n=6$), Black ($n=3$), East Asian ($n=2$), Southeast Asian ($n=1$), Mixed Race ($n=1$), Latin or South American ($n=2$). A sensitivity power analysis indicated this sample size had an ability to detect an effect size of $d=0.47$ based on 80% power and a two-tailed $\alpha=0.05$.

Measures

Classification Image Rating We presented participants with the group-level CI and anti-CI produced from Phase 1 and participants rated how much each face exhibited a particular trait on a scale from 0 to 100. The images were presented individually and both the images and traits that participants were asked to judge were presented in random order. "Apologetic," "remorseful," "regretful," and "unapologetic" were rated on a 100-point scale ranging from 0 (*not at all*) to 100 (*very*). The remaining traits were rated using a semantic differential

---

[1] Participants also rated gender-related characteristics (i.e., masculinity-femininity) in Study 1b and 1c as part of a larger study examining mental representations of apology across gender.

scale, anchored with *very* at each end. These included: guilty-innocent, insincere-sincere, mean-kind, sad-happy, submissive-dominant, and untrustworthy-trustworthy.

Conscientious Responders Scale The CRS measure was identical to that in Study 1a and was used to assess participant attention (Marjanovic et al., 2014). The five questions from the CRS were presented in random order at specific intervals through the scale portion of the task.

Procedure

After consenting, participants completed the image rating task followed by several scales, which were collected as part of a larger study (see Supplementary Materials for details). Once complete, participants were thanked and debriefed.

Results

Classification Image Rating

A comparison between the apologetic CI and the anti-CI revealed that the images were rated as significantly different from each other on characteristics explicitly related to apology (i.e., apologetic and unapologetic). The means and standard deviations for ratings of both the apologetic and anti-apologetic CIs are presented in Table 1. The CI was judged to be more apologetic than the anti-CI, $t(34)=8.05$, $p<0.001$, $d=1.36$, whereas the anti-CI was rated as more unapologetic, $t(34)=-2.77$, $p=0.01$, $d=0.47$, relative to the apologetic CI. Note that in addition to the standardized d effect sizes, mean ratings provide additional information about the magnitude of perceived characteristics due to the use of a 100-point scale. That is, the apologetic face rating of 94.76 can be conceptualized as having a high degree of sadness, whereas the mean rating of 27.94 is a relatively low degree of sadness perceived in the anti-apologetic face.

We then compared the apologetic CI to the anti-apologetic CI for each of the remaining characteristics.[2] The two images were significantly different from each other on each of the characteristic ratings (all $ps<0.05$).[3] The apologetic CI received the highest rating for sad ($M=94.76$, $SD=10.24$) and, surprisingly, post-hoc analyses show that sadness was rated significantly higher than apologetic, $t(35)=7.81$, $p_{adjust}<0.001$, $d=1.30$. As can be seen in Table 1, submissive ($M=81.36$, $SD=14.09$), regret ($M=69.42$, $SD=22.72$), and remorse ($M=65.68$, $SD=23.72$) were also rated higher than apologetic.

---

[2] In instances where mean ratings of the apologetic CI fell below 50 on the semantic differential scale items, we have reverse scored them such that higher scores indicate a higher rating for the item that was on the low end of the scale. For example, the average rating of Sad-Happy was 5.24, which we present as a Sad score of 94.76.

[3] This remained true for all characteristics following Bonferroni adjustments to account for 10 comparisons, with the exception of ratings of unapologetic, $t(34)=-2.77$, $p_{adjust}=.09$, $d=.47$.

Table 1 Study 1b ratings of group-level apologetic and anti-apologetic classification images

Construct	Classification Image				t	d
	Apologetic CI		Anti-Apologetic CI			
	M	SD	M	SD		
Apologetic	56.03	27.99	10.58	13.44	8.05***	1.36
Innocent (Guilty)	60.77	26.07	18.68	16.69	8.24***	1.46
Mean (Kind)	55.00	18.76	68.78	23.73	3.36*	.60
Regret	**69.42**	**22.72**	**10.73**	**15.62**	**12.18*****	**2.03**
Remorse	**65.68**	**23.72**	**14.74**	**18.29**	**8.40*****	**1.48**
Sad (Happy)	**94.76**	**10.24**	**27.94**	**12.78**	**23.34*****	**4.06**
Sincere (Insincere)	52.85	21.62	27.26	20.10	4.41**	.79
Submissive (Dominant)	**81.36**	**14.09**	**30.52**	**15.89**	**10.75*****	**1.87**
Unapologetic	37.08	31.34	63.14	40.44	− 2.77	.47
Untrustworthy (Trustworthy)	63.25	21.90	82.22	17.54	− 3.89**	.66

Bonferroni adjustments were made to p-values to account for comparisons of the apologetic CI and anti-apologetic CI across 10 characteristics. Mean ratings for the apologetic CI that fell below 50 on semantic differential scale items were reverse scored. All ratings are therefore presented in terms of the characteristic that the apologetic CI was found to most closely represent. Characteristics in parentheses represent the alternative anchor for items measured on semantic differential scales. Bolded values represent the four most strongly identified traits

* $p < .05$, ** $p < .01$, *** $p < .001$

Study 1c—Image Rating

The Study 1b ratings provided some initial evidence that people have a mental representation of an apologetic face that could be approximated using reverse correlation. Participants rated the apologetic CI as being more apologetic, and higher in both remorse and regret than the anti-apologetic CI. In addition, this visual template of an apologetic face received the highest rating for sadness, providing some evidence that sadness is a key component of an apologetic face.

The ratings from Study 1b were limited, however, in two main ways. First, some characteristics in Study 1b were measured using a single-dimension scale (e.g., apologetic, unapologetic, remorse, regret) whereas others were rated only on a semantic differential scale (e.g., submissive to dominant). This not only made it challenging to compare across ratings, it may have led to the misinterpretation of the constructs as mutually exclusive. Second, we compared the apologetic CI to the anti-apologetic CI. Because the anti-apologetic image represents the averaged noise pattern of the unselected stimuli and not, for example, an *unapologetic* face, these ratings may lack face validity and could inflate our results. To address these issues, in Study 1c, we reduced the number of items rated in Study 1a to the ones that had received the highest ratings and/or we deemed to be the most central to our research questions, and each of these seven apology-related characteristics were then rated using a single-dimension scale (i.e., from 0, *not at all,* to 100, *very*). These items included: apologetic, regretful, remorseful, sad, submissive, trustworthy, and dominant. In addition, participants rated the apologetic-CI and the base face.

Method

Participants

A total of 153 undergraduates from the Psychology Department's undergraduate participant pool (women = 104, men = 47, prefer not say = 2, M_{age} = 19.90, SD_{age} = 4.32) participated in Study 1c in exchange for course credit. Participants identified as White (n = 40), South Asian (n = 38), Middle Eastern (n = 24), Black (n = 18), Southeast Asian (n = 10), Mixed Race (n = 7), East Asian (n = 6), Other (n = 4), Latin American (n = 3), and South American (n = 3). For this study, we conducted an a priori power analysis which indicated that a sample size of 62 participants was required to detect an effect size of f = 0.15, which was based on findings from previous reverse correlation research (Lloyd et al., 2017) at 80% power. The final sample was larger than the initial goal because it was run in combination with another study, which required a larger sample size. A sensitivity analysis indicated that data from 153 participants was enough to detect a minimum effect of f = 0.09 at 80% power, α = 0.05.

Procedure

The procedure was similar to Study 1b, with four key differences. First, participants rated the apologetic CI as well as the base face from which it was created.[4] Second, to allow for more appropriate comparisons, each image was rated on 7 apology-related characteristics using a single-dimension scale from 0 (*not at all*) to 100 (*very*).[5] Third, the images were presented one at a time and the viewing order was randomized between participants such that participants rated a specific image type (e.g., base face, apologetic face) on each randomized characteristic before rating another image type. In each trial, a single question appeared at the top, asking the participant the extent to which they thought the face exhibited that particular characteristic. A response was required to move to the next trial. Finally, this study was completed after participants completed a brief unrelated study which had been combined in the interest of time.

---

[4] The anti-CI was also included to maintain consistency with Study 1b. However, inclusion of the anti-characteristic image is not necessary for all research questions, as has been established in the reverse correlation literature (e.g., Brown-Iannuzzi et al., 2017; Lloyd et al., 2017). As such, in Study 1c we focused only on the comparisons between the CI and the base face. Comparisons with the anti-CI were similar to Study 1b and these comparisons are provided in the Supplemental Materials.

[5] To remain consistent with Study 1b, and for exploratory purposes unrelated to the current research, we also retained the three semantic differential scales: submissive-dominant, male–female, and masculine-feminine that were rated after all of the single-dimension ratings were completed. We also collected ratings of gender (male and female) and gender-related characteristics (masculine and feminine) using single-dimension scales as part of a larger study.

Table 2 Study 1c ratings and comparisons of the group-level apologetic classification image and base face

Construct	Classification image				t	d
	Apologetic face		Base Face			
	M	*SD*	*M*	*SD*		
Apologetic	**73.14**	**25.53**	**35.26**	**23.94**	**14.18*****	**1.15**
Regret	**75.25**	**23.24**	**34.12**	**22.96**	**16.51*****	**1.33**
Remorse	**70.19**	**24.57**	**35.03**	**24.24**	**12.59*****	**1.02**
Sad	**86.36**	**19.24**	**38.94**	**24.71**	**20.13*****	**1.63**
Submissive	59.90	25.55	5.31	22.02	5.72***	.46
Trustworthy	37.74	22.26	43.56	22.22	− 2.43	.20
Dominant	24.34	22.77	53.85	22.18	− 11.68***	.94

Bonferroni adjustments were made for *p*-values within the apology-related characteristics category, and accounted for eight comparisons in total (i.e., comparisons of the apologetic and base faces for seven characteristics, as well as the comparison between Apologetic and Sad which is the highest rated characteristic). Bolded values represent the four most strongly identified traits

*$p < .05$, **$p < .01$, ***$p < .001$

Results

Classification Image Rating: Apology

As predicted, the apologetic face ($M = 73.14$, $SD = 25.53$) generated using reverse correlation was perceived as significantly more apologetic than the base face ($M = 35.26$, $SD = 23.94$),[6] $t(152) = 14.18$, $p < 0.001$, $d = 1.15$. This provides additional evidence that participants have a mental representation of an apologetic face and that a visual template of an apologetic face can be approximated using reverse correlation.

In addition, as can be seen in Table 2, results from multiple paired-sample t-tests indicated that ratings for regretful, remorseful, sad, and submissive were all significantly higher for the apologetic face than for the base face, whereas dominance was found to be lower for the apologetic face than for the base face and ratings for trustworthy did not differ from the base face. As in Study 1b, ratings of sadness ($M = 86.36$, $SD = 19.24$) were again higher than ratings of apologetic, providing additional evidence that the apologetic CI was perceived as significantly more sad than apologetic, $t(152) = 6.87$, $p < 0.001$, $d = 0.56$.

Discussion

In Study 1, we found evidence that people have a mental representation of an apologetic face and that a visual template of this face could be generated using reverse correlation. Specifically, participants in Study 1 perceived apology to a significantly greater extent in the apologetic CI relative to both the anti-CI (Study 1b) and the neutral White male base

---

[6] Due to non-normal distributions of both variables, Wilcoxon signed-rank tests were also conducted. Results from these robust tests are similar to those presented here and can be found in Supplemental Materials.

face from which the apologetic face was created (Study 1c). Similarly, the ratings of characteristics typically associated with being apologetic, including regret and remorse, were higher for the apologetic CI relative to both the anti-CI and base face.

Somewhat surprisingly, across both Study 1b and 1c, the CI—which was created specifically to represent an apologetic face—was rated most highly on sadness relative to all other characteristics, including apologetic. This provides some initial evidence that sadness is an important component of an apologetic face. One related explanation for this finding is that basic emotions such as sadness may be more easily perceived relative to secondary or social emotions, such as those expressed during an apology, including being "apologetic." However, given the high ratings that this apologetic CI received, it may also be the case that sadness is the *only* facial expression needed to express apology nonverbally, and perhaps all other components are expected to be communicated verbally. That is, making a sad face might be sufficient to make a face appear apologetic. We examined this possibility in Study 2.

Study 2

The goal of Study 2 was to further examine the relationship between apologetic and sad faces. To examine this question, we compared the visual templates of an apologetic face and a sad face. Specifically, in Study 2a, we had a new set of participants generate visual templates of a sad face using the same procedure and images from Study 1, with the exception that participants were now asked to select the image that looked sadder. In Study 2b, participants were shown the apologetic and sad CIs and were randomly assigned to rate all faces as either apologetic or sad. This resulted in a 2 (Classification Image: Apologetic or Sad) × 2 (Rating: Apologetic or Sad) mixed design with the first factor within subjects (blinded) and the second between. This allowed us to examine whether a CI generated to be sad was perceived differently from a CI generated to be apologetic.

In addition, in Study 2 we obtained ratings not only of the group-level CIs, but also of the individual-level CIs. Recent research suggests that the two-phase structure of the reverse correlation procedure can result in inflated Type I error, particularly when group-level CIs alone are evaluated (Cone et al., 2020). One way to control for these error rates in the image rating phase is by obtaining ratings of individual-level CIs (i.e., those faces generated by averaging the noise patterns of each individual generator's response in the generation phase), in addition to the group-level average. As such, in Study 2b, we also had participants provide ratings for each individual-level CI, as well as the group-level CIs.

Given that apologetic faces in Study 1 received high ratings of sadness, we expected that sad faces would be rated equally (if not more) sad. What was less clear, however, was whether the sad faces would be judged as comparably apologetic. One additional question for Study 2 was whether differences in the sad and apologetic CIs would emerge for specific facial regions. As such, participants in Study 2b were also asked to rate group-level CIs that were cropped to display only the eye region or only the mouth region (see Fig. 4). Similar occluded or cropped stimuli have been used in previous research (e.g., Hutchings et al., 2021; Rule et al., 2008) and we used a comparable method to examine whether
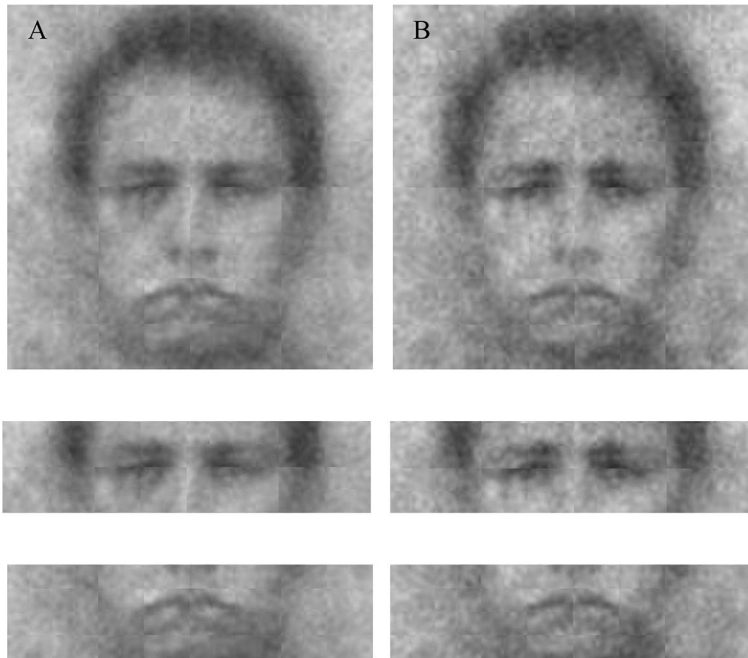
Fig. 4 Images in Column A are the Apologetic Group-level CI generated in Study 1a in full (top), cropped at the eyes (middle) and cropped at the mouth (bottom). Column B shows the Sad Group-level CI generated in Study 2a in full (top), cropped at the eyes (middle) and cropped at the mouth (bottom)

differences would emerge in the ratings for apology and sad for the eyes and/or mouth regions from the apologetic and sad CIs.

Study 2a—Image Generation

The goal of Study 2a was to generate a visual template of a sad face using the same approach as was used to generate the apologetic template in Study 1a.

Method

Participants

A total of 131 participants were recruited from an undergraduate research participant pool and received course credit in exchange for their involvement. As in Study 1 and based on previous research, we decided a priori to stop recruitment at 125 participants. However, because we intended participants in Study 2b to rate equal numbers of individual-level CIs for both the apologetic and sad faces, we subsequently oversampled by six participants to account for additional exclusions. Following the exclusion of three participants who failed to complete the study and seven participants who answered fewer than three questions correctly on the Conscientious Responders Scale (Marjanovic et al., 2014), the final

sample size consisted of 121 participants (women = 89, men = 32, M_{age} = 19.36 years, SD_{age} = 2.48). The racial composition of the sample consisted of South Asian (n = 34), White (n = 22), Black (n = 20), Middle Eastern (n = 13), Southeast Asian (n = 10), Mixed Race (n = 7), Latin American (n = 6), East Asian (n = 5), and Other (n = 4).

Procedure

The procedure was similar to Study 1a, with a few exceptions. First, data collection took place during the COVID-19 pandemic and as such, participants completed the study online using the Qualtrics platform. Participants were not required to complete the study on a desktop computer; however, the software data suggest that only three participants completed the study on a mobile device. Second, instead of asking participants to select which of two faces in each of the 400 trials looked more apologetic, the instructions now asked, "Which face looks sadder?".

Results

Reverse Correlation

CIs were created using the rcicr package in R (Dotsch, 2017). As in Study 1, a group-level CI was generated using the noise patterns from the images in each forced-choice pair that were selected (see Fig. 4). In addition, CIs were generated at the individual-level, such that one CI per each of the 121 participants was created (see Supplemental Materials for examples).

Study 2b—Image Rating

Method

Participants

Data were collected from 207 undergraduate participants in exchange for course credit. Two participants did not consent to participate in the study, 31 failed to complete the study in its entirety, and 12 were removed after being identified as numerical outliers.[7] Following these exclusions, the final sample consisted of 162 participants (women = 125, men = 37, M_{age} = 20.79, SD_{age} = 4.84). Participants identified as South Asian (n = 45), Middle Eastern (n = 29), Black (n = 26), White (n = 18), Southeast Asian (n = 14), East Asian (n = 11), Other (n = 8), Mixed Race (n = 5), Latin American (n = 4), and South American (n = 2). A sensitivity power analysis reveals an ability to detect a between-subjects effect size of f = 0.22 based on 80% power and a two-tailed alpha level of p = 0.05.

---

[7] Please see Supplemental Materials for further information about how and why outliers were removed.
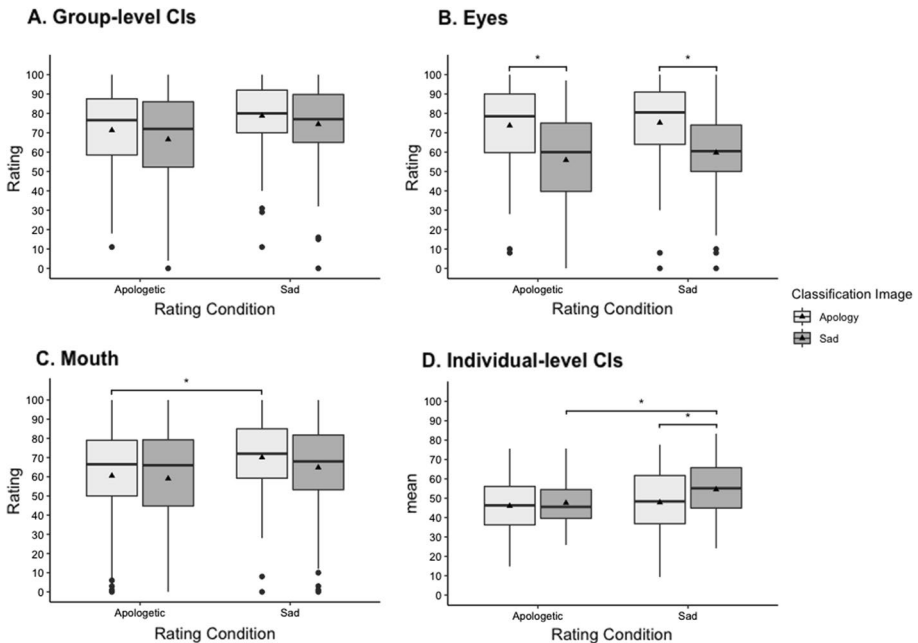
Fig. 5 Boxplots comparing Study 2b ratings of Apologetic and Sad for Apologetic and Sad CIs at the group-level (**A**), cropped at the eyes (**B**), cropped at the mouth (**C**), and at the individual-level (**D**). Shaded areas within each box indicate the range of scores between the 1st and 3rd quartile. Vertical whiskers represent the range of ratings, with dots indicating outliers. Medians are represented by the horizontal bars within each box; mean ratings are represented by triangles.* $p < .05$

Measures

Classification Image Rating Participants were randomly assigned to one of two rating conditions: Apologetic or Sad. Those in the Apologetic condition were instructed to provide ratings of how apologetic each face appeared, while those in the Sad condition were instructed to rate the faces for how sad each appeared. Ratings were measured on a scale from 0 (*not at all*) to 100 (*very*). Regardless of condition, all participants were asked to judge a total of 251 full or partial faces. Images included the 121 individual-level apologetic CIs generated in Study 1, and the 121 individual-level sad CIs generated in Study 2 (see Supplemental Materials for additional information). In addition, participants rated the base face, the group-level CI and group-level anti-CI for each face type (apologetic and sad), as well as cropped versions of both the eyes and the mouth for each face type (see Fig. 4). All images were presented in random order.

Procedure

Because data collection took place during the COVID-19 pandemic, participants were not able to complete the study on campus. Instead, participants accessed the study online at their own convenience. Restrictions were put in place such that the survey had to be completed on a desktop or laptop computer. If participants attempted to start the survey on a

mobile device, they received a message instructing them to log on again on another device. Participants completed an informed consent form, followed by a brief demographic questionnaire, the image rating task with a total of 251 rating trials and 5 CRS trials, and a debriefing.

Results

Four separate 2 (Condition: Rate Apologetic, Rate Sad)×2 (Classification Image: Apologetic, Sad) ANOVAs were conducted to compare ratings of the group-level CIs, the group-level CIs cropped at the eyes, the group-level CIs cropped at the mouth and the averaged individual-level CI ratings. Results of each analysis are presented in Fig. 5.

A significant main effect of condition was found when comparing the group-level CIs, $F(1, 320)=10.37$, $p=0.001$. Across both images, ratings of sad ($M=76.58$, $SD=18.85$) were higher than ratings of apologetic ($M=68.93$, $SD=23.90$). Pairwise comparisons revealed that this was the case for the apologetic CI, $t(143.1)=-2.34$, $p=0.02$, $d=0.37$, as well as the sad CI, $t(141.19)=-2.17$, $p=0.03$, $d=0.34$. However, these comparisons no longer reached significance following Bonferroni adjustments ($p_{adjust}=0.08$ and $p_{adjust}=0.13$ respectively). Neither the main effect of Classification Image, $F(1, 320)=3.63$, $p=0.06$, nor the interaction, $F(1, 320)=0.01$, $p=0.93$, were significant.

Next, we compared the ratings of the CIs cropped to include only the eyes. The 2 (Condition: Rate Apologetic, Rate Sad)×2 (Classification Image: Apologetic, Sad) ANOVA revealed a significant main effect of Classification Image, $F(1, 320)=45.76$, $p<0.001$, such that the apologetic CI ($M=74.48$, $SD=20.60$) received significantly higher ratings than the sad CI ($M=57.90$, $SD=23.40$). Pairwise comparisons indicated that this was true for participants rating apologetic, $t(75)=5.85$, $p_{adjust}<0.001$, $d=0.67$, as well as sad, $t(85)=5.40$, $p_{adjust}<0.001$, $d=0.58$. The main effect of Condition was not significant, $F(1, 320)=1.16$, $p=0.28$, and again the interaction did not reach significance, $F(1, 320)=0.24$, $p=0.63$.

The 2 (Condition: Rate Apologetic, Rate Sad)×2 (Classification Image: Apologetic, Sad) ANOVA was again used to compare ratings of the CIs cropped to include only the mouth. This analysis resulted in a main effect of Condition, $F(1, 320)=7.93$, $p=0.005$, with ratings of sadness ($M=67.42$, $SD=22.52$) again being higher than ratings of apologetic ($M=59.79$, $SD=26.24$). The main effect of Classification Image, $F(1, 320)=1.64$, $p=0.20$, was not significant, nor was the interaction effect, $F(1, 320)=0.49$, $p=0.49$.[8] Despite the lack of interaction, post-hoc analyses suggest that the apologetic face was rated significantly higher by participants evaluating sadness than by those evaluating apologetic, $t(142.65)=-2.61$, $p_{adjust}=0.04$, $d=0.41$. Ratings of the sad face did not differ significantly by condition, $t(152.15)=-1.41$, $p_{adjust}=0.64$, $d=0.22$.

Due to rising concerns about the error rates associated with the two-phase reverse correlation procedure (Cone et al., 2020), we also conducted the 2 (Condition: Rate Apologetic, Rate Sad)×2 (Classification Image: Apologetic, Sad) ANOVA using the averaged ratings of each individual-level CI. We found a significant main effect of Condition, $F(1,$

---

[8] Due to the non-normal distributions and mixed design, the same 2 (Condition: Rate Apologetic, Rate Sad)×2 (Classification Image: Apologetic, Sad) robust ANOVAs were also conducted using the WRS2 package in R (Mair & Wilcox, 2020). The pattern of results, which can be found in the Supplemental Materials, was the same.

320) = 13.37, $p < 0.001$, and a main effect of Classification Image, $F(1, 320) = 11.30$, $p < 0.001$, which were qualified by a significant interaction effect, $F(1, 320) = 4.34$, $p = 0.04$. Post-hoc between-subjects comparisons revealed that participants did not differ in their ratings of apologetic for the apologetic CI ($M = 46.27$, $SD = 11.48$) or the sad CI ($M = 47.63$, $SD = 13.00$), $t(75) = -2.49$, $p_{adjust} = 0.06$, $d = 0.29$. However, when ratings were averaged across all individual CIs, the sad CI ($M = 54.44$, $SD = 9.32$) was rated as significantly more sad than the apologetic CI ($M = 48.14$, $SD = 8.67$), $t(85) = -14.86$, $p_{adjust} < 0.001$, $d = 1.60$. In addition, sad CIs were rated as significantly higher in sadness relative to participants evaluating the same faces for apologetic, $t(134.18) = -3.79$, $p_{adjust} < 0.001$, $d = 0.60$. The direct comparison between ratings of apologetic and sad for the apologetic CI was not significant, $t(138.53) = -1.16$, $p_{adjust} > 0.99$, $d = 0.18$.

Discussion

The current results replicate and extend the findings from Study 1. The group-level apologetic CI generated in Study 1 was again rated as apologetic ($M = 75.25$ on a 100-point scale from *not at all* to *very*), providing further evidence that people have a mental representation of an apologetic face that can be approximated using reverse correlation. Interestingly, as in Study 1, this visual template of an apologetic face again received higher ratings of sadness relative to apologetic, suggesting that a main component in people's mental representation of an apologetic face is sadness. In addition, in Study 2, the visual template of an apologetic face and the new visual template of a sad face, were not perceived as differing in how apologetic they appeared. That is, an image generated to be apologetic was *not* perceived to be more apologetic than a face generated to be sad. This was also true when ratings of apologetic were compared across the apologetic and sad individual-level CIs. Taken together, these findings provide additional evidence of the importance of sadness in people's representation of an apologetic face. These findings are also consistent with the emotion literature, which suggests that basic or primary emotions are easier to identify (or at least identified to a greater extent) than social or secondary emotions (Ortony et al., 1990)—possibly because experiences of those easily-categorizable emotions are familiar (Barrett, 2006). Most importantly, these findings also suggest that one way to appear apologetic is to look sad.

In order to gain further insight into potential similarities and differences in how apology and sadness are conveyed by different facial regions, we also had participants rate the cropped eyes and mouth of the group-level apologetic and sad faces. When considering only the eyes, the apologetic CI was rated as significantly more apologetic and significantly sadder than the sad CI. When considering only the mouths, on the other hand, apologetic and sad CIs were rated comparably, but the mouths were rated higher in sadness than in apology; despite a lack of interaction, post-hoc analyses suggested this was driven by the apologetic CI. It is unclear exactly why this is the case, but these findings are consistent with the possibility that the eyes can convey more complex emotions such as feeling apologetic, whereas basic emotions such as sadness are more easily identified by the mouth. Finally, it is worth noting that when the individual-level apologetic and sad CIs were compared, the averaged ratings of sadness were higher for the sad CIs than for the apologetic CIs. This finding is consistent with the possibility that the visual template of a sad face is not identical to an apologetic face, rather that sadness might be an important and central feature of an apologetic face.

General Discussion

In the current research we found evidence that people have a mental representation of an apologetic face that can be approximated using reverse correlation and that sadness is a key component of this representation. This research contributes to a growing body of evidence that the reverse correlation method can be used to generate visual images that reflect people's mental representations. Specifically, in Study 1 we generated a classification image (CI) of an apologetic face using this method and naïve participants in both Studies 1b and 1c subsequently rated this image as appearing apologetic. We also examined which of several characteristics, derived from the verbal apology literature, would be perceived in this apologetic CI. Sadness consistently emerged as the highest rated characteristic, receiving even higher ratings than apologetic in both studies. In addition, as expected, regret, remorse, and submissive consistently received high ratings in Studies 1b and 1c, suggesting that these are also perceived components of an apologetic face. However, not all characteristics received high ratings. For example, ratings of trustworthiness were relatively low and did not differ significantly from the base face. Although the verbal literature suggests that being trustworthy, and specifically conveying the intention not to reoffend, is an important component of an effective verbal apology (Kirchhoff et al., 2012; Lewicki et al., 2016; Schlenker & Darby, 1981; Struthers et al., 2008), trustworthiness did not emerge as a key nonverbal feature in an apologetic face.

Because the apologetic CI from Study 1 received such high ratings of sadness, we compared apologetic and sad faces in Study 2 to determine whether nonverbal expressions of apology were systematically different from expressions of sadness. We found no differences in people's ratings of how apologetic the apologetic and sad CIs appeared, and this was true regardless of whether the group-level CIs or individual-level CIs were assessed. This was also despite the fact that, at least for the individual-level CIs, the sad CIs were perceived as being sadder than the apologetic CIs, suggesting that these visual templates were not identical. In addition, differences were detected when comparing ratings of the eyes of the apologetic versus sad group-level CIs, with the eyes of the apologetic CI being rated as significantly more apologetic *and* sadder than the eyes of the sad CI. Although no significant differences were found when comparing the within-subjects ratings of the mouths of the CIs as either apologetic or sad, individuals judging sadness provided higher ratings of the apologetic CI's mouth relative to those judging apologetic.

There are several possibilities for why the apologetic CI was rated significantly more sad than apologetic. First, sadness may indeed be the characteristic driving the perception of nonverbal apology. Our findings demonstrate that sad faces are judged similarly to apologetic faces, suggesting that when offering an apology, a transgressor may only need to appear sad to produce a successful apology. This may be an easy heuristic for transgressors because it is a basic emotion and potentially easier to conceptualize and express (perhaps especially when it is not genuine) than apology. A second possibility for our findings is that the concept of "apologetic" is not a singular thing, but rather an amalgamation of multiple different characteristics and expressions. Although ratings of the CIs were significantly different from each other only when considering the eyes, the mean ratings of both apologetic and sad were higher for the apologetic face in all but one instance.[9] We believe that these trends may be a result of the degree of expression associated with an apologetic

---

[9] This was only the case for the individual-level CIs.

face, and in particular, apologetic eyes. That is, participants can identify both apologetic and sad expressions to a greater degree in apologetic faces compared to sad faces. Interestingly, however, the data also suggests that participants tasked with rating both CIs in terms of sadness consistently provided higher ratings on average compared to participants rating apologetic. This could imply that we are confident in our ability to identify sadness, possibly because it is a common and basic emotion, whereas we are slightly more reserved in our judgment of apologetic due to the additional cognitive processing required to identify multiple characteristics in a single face (Olszanowski et al., 2018). Taken together, these findings suggest that sadness may be easy to identify, however apology may be easy to express. Future research that asks participants to directly compare apologetic CIs to sad CIs in a forced-choice trial may be a first step to investigating whether the faces, as a whole, can be differentiated, which could pave the way for research on genuine apologies. Perhaps seeing the two CIs next to each other will show that "apologetic" is not a singular construct, but rather an amalgamation of multiple different characteristics and expressions.

Limitations and Future Directions

The generalizability of the current results are limited by the fact that we used a White male base face, and our data may therefore only speak to nonverbal apologies as they apply to White men. This is an important consideration given what we know about the expression and interpretation of facial cues across groups. For example, evidence suggests that racial biases influence how emotional expressions are identified and categorized, with White individuals high in implicit racial bias faster to identify anger in Black (relative to White) faces (Hugenberg & Bodenhausen, 2003). As well, women tend to be more expressive than men (Adams et al., 2015; Fischer & LaFrance, 2015). However, it is unclear if and how differences in emotional expression and perception might influence the apology process. Our understanding of this topic would benefit greatly from using faces that vary across gender and race. In addition, cultural differences in the norms associated with displaying emotions (Matsumoto et al., 2008a, 2008b), may impact expectations and perceptions of nonverbal apologies. This is a particularly important avenue for future research given our increasing globalization and the fact that, in some cases, the implications of misinterpreting an apology could even have dire consequences. For example, an apology that is offered in a judicial setting (e.g., sentencing) and is perceived as insincere could lead to harsher sentencing; similarly, cultural differences in the need for and displays of apology could result in international tensions between leaders and key decision makers. Extending the present research to also examine groups that face different stereotypes will expand our understanding of the prejudices that may bias how we repair and maintain our personal relationships. In addition to using only White faces, all stimuli in the current studies relied on base faces that were digitally manipulated (morphed averages), which may limit the generalizability of results. Although the base faces were generated by morphing naturalistic photographs of faces together, the resulting face represents an 'average' face (Langlois et al., 1994) and does not represent a real individual. Such 'average' faces are often rated as more attractive than the component faces (Langlois & Roggman, 1990), and despite appearing altered by the visual noise in the reverse correlation paradigm, it is possible that attractiveness could affect ratings of apologetic and sad. It is unclear whether similar results would be obtained if base faces instead consisted of single naturalistic photographs of faces.

Continuing to understand how apologies manifest on real faces and how natural expressions contribute to successful apologies would undoubtedly lead to myriad implications.

For instance, knowing how to emulate an apologetic expression could allow transgressors to demonstrate empirically tested characteristics while offering apologies, whether or not they are sincere. Similarly, however, it may be possible—and even necessary—to train recipients of apologies, such as judges, to accurately detect apologetic characteristics and assess whether a transgressor's messaging is genuine. Combining verbal and nonverbal apology cues and measuring subsequent apologetic ratings will also allow for an investigation of the relative contribution of what is said during an apology and what is expressed facially. It may be the case that transgressors are only perceived as genuinely apologetic once a certain threshold of congruence between verbal and nonverbal messaging is achieved.

Another limitation of the current research is that we asked people to generate and rate apology in the absence of any contextual information. Including information about the severity of a transgression may result in varying intensities of mental representations of apologetic expressions, and nonverbal cues needed to convey apology that differ depending on the severity of the transgression. Schlenker and Darby (1981) argue that more severe transgressions require the inclusion of more apology components in order to be successful. For example, simply stating "I'm sorry" may be sufficient when apologizing for arriving late to a meeting, whereas more of an explanation for one's behavior would be expected when apologizing for committing corporate fraud. Consistent with this notion, it seems possible that apologies that follow severe transgressions might require more intense or prolonged facial expressions than those following milder transgressions in order for the face to be judged as apologetic. In the absence of context, raters may not readily engage in thinking about abstract constructs such as nonverbal apology, but instead perceive more basic or concrete concepts associated with nonverbal cues, such as emotions (i.e., sadness).

Including contextual information in both the generation and rating phases in future studies may help to disentangle sad faces in response to a major life event (e.g., the loss of a loved one) relative to a small emotion-provoking event (e.g., watching an emotional commercial on TV) and help elucidate whether subtypes of sad and apologetic facial expressions exist. In the current studies, it is unclear whether participants thought of any specific event either when selecting the more apologetic or sad images across multiple trials in each study's image generation and/or rating phases, and this could be an important avenue for future research. For example, future studies might provide participants with a vignette describing a transgression or a sad event and manipulate the severity of the event between conditions at the generation and/or rating phases of the reverse correlation procedure. Alternatively, context could be established by asking participants to recall an event from their past prior to completing the study. Results obtained from studies such as these would contribute to an understanding of the role of context in how we think about apologies. It seems possible that, if sadness is a key feature of an apologetic face, similar results would emerge, regardless of the context, and this is a worthwhile avenue for future research.

Finally, future research would benefit from determining whether it is possible to fake an apologetic appearance, whether perceivers can differentiate between genuine and simulated apologies, whether either of these scenarios have implications for forgiveness outcomes, and whether or to what degree the context surrounding a transgression affects these factors. Measuring forgiveness as an outcome variable after exposing individuals to various apologetic faces will be an important step toward determining the ecological validity of the current findings. If apologies and their actual and perceived sincerity influenced forgiveness, this could have important implications in contexts such as sentencing procedures in the criminal justice system. Similarly, research that explores different combinations of verbal and nonverbal apologies may also be useful in furthering our understanding of apologies.

One way to test which facial cues most successfully communicate an apology would be to manipulate the number of apology cues one receives from a transgressor, and determine the likelihood of forgiveness. For example, does an apology need to contain verbal indicators of sadness, submissiveness, and nonverbal apology, or are facial cues enough to satisfy the recipient's need for the transgressor to express remorse? It seems possible that apologies are an example of nonverbal communication that relies on combinations of cues, including facial expressions as well as context (Poyatos, 2015).

Conclusion

In conclusion, the current findings suggest that people have a mental representation of an apologetic face, and that sadness is an important component of this representation. The findings presented in Studies 1 and 2 have important implications in our social lives, as all relationships will likely require us to both offer and evaluate an apology at some point. Garnering a deeper understanding of our expectations when it comes to apologetic faces—as well as a better understanding of differences between apologetic and sad expressions—may help smooth communication following transgressions and ultimately improve our interpersonal relationships.

Declarations

References

Adams, R. B., Jr., Hess, U., & Kleck, R. E. (2015). The intersection of gender-related facial appearance and facial displays of emotion. *Emotion Review, 7*(1), 5–13. https://doi.org/10.1177/1754073914544407

Awasthi, A., & Mandal, M. K. (2015). Facial expressions of emotions: Research perspectives. In M. Mandal & A. Awasthi (Eds.), *Understanding facial expressions in communication* (pp. 1–18). Springer, New Delhi. https://doi.org/10.1007/978-81-322-1934-7_1

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Bippus, A. M., & Young, S. L. (2020). How to say "I'm Sorry:" Ideal apology elements for common interpersonal transgressions. *Western Journal of Communication, 84*(1), 43–57. https://doi.org/10.1080/10570314.2019.1610787

Bjornsdottir, R. T., & Rule, N. O. (2017). The visibility of social class from facial cues. *Journal of Personality and Social Psychology, 113*(4), 530. https://doi.org/10.1037/pspa0000091

Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology, 28*(1), 333–361. https://doi.org/10.1080/10463283.2017.1381469

Brooks, J. A., Stolier, R. M., & Freeman, J. B. (2018). Stereotypes bias visual prototypes for sex and emotion categories. *Social Cognition, 36*(5), 481–493. https://doi.org/10.1521/soco.2018.36.5.481

Brown-Iannuzzi, J. L., Dotsch, R., Cooley, E., & Payne, B. K. (2017). The relationship between mental representations of welfare recipients and attitudes toward welfare. *Psychological Science, 28*(1), 92–103. https://doi.org/10.1177/0956797616674999

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

Cone, J., Brown-Iannuzzi, J. L., Lei, R., & Dotsch, R. (2020). Type I error is inflated in the two-phase reverse correlation procedure. *Social Psychological and Personality Science*. https://doi.org/10.1177/1948550620938616

De Cremer, D., & Schouten, B. C. (2008). When apologies for injustice matter: The role of respect. *European Psychologist, 13*(4), 239–247. https://doi.org/10.1027/1016-9040.13.4.239

Deodhare, D. (2015). Facial expressions to emotions: A study of computational paradigms for facial emotion recognition. In M. K. Mandal & A. Awasthi (Eds.), *Understanding facial expressions in communication* (pp. 173–198). Springer.

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science, 3*(5), 562–571. https://doi.org/10.1177/1948550611430272

Dotsch, R., Wigboldus, D. H., Langner, O., & van Knippenberg, A. (2008). Ethnic outgroup faces are biased in the prejudiced mind. *Psychological Science, 19*(10), 978–980. https://doi.org/10.1111/j.1467-9280.2008.02186.x

Dotsch, R. (2017). rcicr: Reverse correlation image classification toolbox. R package version 0.4.0. https://CRAN.R-project.org/package=rcicr

Fehr, R., & Gelfand, M. J. (2010). When apologies work: How matching apology components to victims' self-construals facilitates forgiveness. *Organizational Behavior and Human Decision Processes, 113*(1), 37–50. https://doi.org/10.1016/j.obhdp.2010.04.002

Fischer, A., & LaFrance, M. (2015). What drives the smile and the tear: Why women are more emotionally expressive than men. *Emotion Review, 7*(1), 22–29. https://doi.org/10.1177/1754073914544406

Frijda, N. H. (1988). The laws of emotion. *American Psychologist, 43*(5), 349–358. https://doi.org/10.1037/0003-066X.43.5.349

Goffman, E. (1955). On face-work: An analysis of ritual elements in social interaction. *Psychiatry, 18*(3), 213–231. https://doi.org/10.1080/00332747.1955.11023008

Hornsey, M. J., Wohl, M. J., Harris, E. A., Okimoto, T. G., Thai, M., & Wenzel, M. (2019). Embodied remorse: Physical displays of remorse increase positive responses to public apologies, but have negligible effects on forgiveness. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000208

Howell, A. J., Turowski, J. B., & Buro, K. (2012). Guilt, empathy, and apology. *Personality and Individual Differences, 53*(7), 917–922. https://doi.org/10.1016/j.paid.2012.06.021

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science, 14*(6), 640–643. https://doi.org/10.1046/j.0956-7976.2003.psci_1478.x

Hutchings, R. J., Simpson, A. J., Sherman, J. W., & Todd, A. R. (2021). Perspective taking reduces intergroup bias in visual representations of faces. *Cognition, 214*, 104808. https://doi.org/10.1016/j.cognition.2021.104808

Imhoff, R., Dotsch, R., Bianchi, M., Banse, R., & Wigboldus, D. H. (2011). Facing Europe: Visualizing spontaneous in-group projection. *Psychological Science, 22*(12), 1583–1590. https://doi.org/10.1177/0956797611419675

Kid Cudi. (2012). https://quotefancy.com/quote/1152059/Kid-Cudi-You-can-always-say-sorry-but-the-real-apology-is-when-you-hearthe-sadness-in

Kirchhoff, J., Wagner, U., & Strack, M. (2012). Apologies: Words of magic? The role of verbal components, anger reduction, and offence severity. *Peace and Conflict: Journal of Peace Psychology, 18*(2), 109. https://doi.org/10.1037/a0028092

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1*(2), 115–121. https://doi.org/10.1111/j.1467-9280.1990.tb00079.x

Langlois, J. H., Roggman, L. A., & Musselman, L. (1994). What is average and what is not average about attractive faces? *Psychological Science, 5*(4), 214–220. https://doi.org/10.1111/j.1467-9280.1994.tb00503.x

Lei, R. F., & Bodenhausen, G. V. (2017). Racial assumptions color the mental representation of social class. *Frontiers in Psychology, 8*, 519. https://doi.org/10.3389/fpsyg.2017.00519

Leunissen, J. M., De Cremer, D., Folmer, C. P. R., & Van Dijke, M. (2013). The apology mismatch: Asymmetries between victim's need for apologies and perpetrator's willingness to apologize. *Journal of Experimental Social Psychology, 49*(3), 315–324. https://doi.org/10.1016/j.jesp.2012.12.005

Lewicki, R. J., Polin, B., & Lount, R. B., Jr. (2016). An exploration of the structure of effective apologies. *Negotiation and Conflict Management Research, 9*(2), 177–196. https://doi.org/10.1111/ncmr.12073

Lloyd, E. P., Kunstman, J. W., Tuscherer, T., & Bernstein, M. J. (2017). The face of suspicion: Suspicion of Whites' motives moderates mental representations of Whites. *Social Psychological and Personality Science, 8*(8), 953–960. https://doi.org/10.1177/1948550617699251

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology Section Karolinska Institutet, 91*, 630.

Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods, 52*(2), 464–488. https://doi.org/10.3758/s13428-019-01246-w

Marjanovic, Z., Struthers, C. W., Cribbie, R., & Greenglass, E. R. (2014). The conscientious responders scale: A new tool for discriminating between conscientious and random responders. *SAGE Open, 4*(3), 2158244014545964.

Matsumoto, D., Keltner, D., Shiota, M. N., O'Sullivan, M., & Frank, M. (2008a). Facial expressions of emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 211–234). The Guilford Press.

Matsumoto, D., Yoo, S. H., & Fontaine, J. (2008b). Mapping expressive differences around the world: The relationship between emotional display rules and individualism versus collectivism. *Journal of Cross-Cultural Psychology, 39*(1), 55–74. https://doi.org/10.1177/0022022107311854

Olszanowski, M., Kaminska, O. K., & Winkielman, P. (2018). Mixed matters: Fluency impacts trust ratings when faces range on valence but not on motivational implications. *Cognition and Emotion, 32*(5), 1032–1051. https://doi.org/10.1080/02699931.2017.1386622

Ortony, A., Clore, G. L., & Collins, A. (1990). *The cognitive structure of emotions*. Cambridge University Press.

Picó, A., Gračanin, A., Gadea, M., Boeren, A., Aliño, M., & Vingerhoets, A. (2020). How visible tears affect observers' judgements and behavioral intentions: Sincerity, remorse, and punishment. *Journal of Nonverbal Behavior, 44*(2), 215–232. https://doi.org/10.1007/s10919-019-00328-9

Poyatos, F. (2015). Nonverbal communication in interaction: Psychology and literature. In Aleksandra Kostić & Derek Chadee (Eds.), *The social psychology of nonverbal communication* (pp. 248–271). Palgrave Macmillan.

Riek, B. M. (2010). Transgressions, guilt, and forgiveness: A model of seeking forgiveness. *Journal of Psychology and Theology, 38*(4), 246–254. https://doi.org/10.1177/009164711003800402

Rule, N. O., Ambady, N., Adams, R. B., Jr., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of male sexual orientation. *Journal of Personality and Social Psychology, 95*(5), 1019. https://doi.org/10.1037/a0013194

Schlenker, B. R., & Darby, B. W. (1981). The use of apologies in social predicaments. *Social Psychology Quarterly*. https://doi.org/10.2307/3033840

Schumann, K. (2014). An affirmed self and a better apology: The effect of self-affirmation on transgressors' responses to victims. *Journal of Experimental Social Psychology, 54*, 89–96. https://doi.org/10.1016/j.jesp.2014.04.013

Schumann, K. (2018). The psychology of offering an apology: Understanding the barriers to apologizing and how to overcome them. *Current Directions in Psychological Science, 27*(2), 74–78. https://doi.org/10.1177/0963721417741709

Schumann, K., & Dragotta, A. (2021). Empathy as a predictor of high-quality interpersonal apologies. *European Journal of Social Psychology*. https://doi.org/10.1002/ejsp.2786

Slocum, D., Allan, A., & Allan, M. M. (2011). An emerging theory of apology. *Australian Journal of Psychology, 63*(2), 83–92. https://doi.org/10.1111/j.1742-9536.2011.00013.x

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

Ten Brinke, L., & Adams, G. S. (2015). Saving face? When emotion displays during public apologies mitigate damage to organizational performance. *Organizational Behavior and Human Decision Processes, 130*, 1–12. https://doi.org/10.1016/j.obhdp.2015.05.003

Todorov, A., Dotsch, R., Wigboldus, D. H., & Said, C. P. (2011). Data-driven methods for modeling social perception. *Social and Personality Psychology Compass, 5*(10), 775–791. https://doi.org/10.1111/j.1751-9004.2011.00389.x

Tomlinson, E. C., Dineen, B. R., & Lewicki, R. J. (2004). The road to reconciliation: Antecedents of victim willingness to reconcile following a broken promise. *Journal of Management, 30*(2), 165–187. https://doi.org/10.1016/j.jm.2003.01.003