WHO'S SORRY NOW? AN INVESTIGATION OF HOW GENDER SHAPES THE APPEARANCE AND JUDGMENT OF APOLOGETIC FACES

MEGHAN GEORGE

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PSYCHOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

October 2020

Abstract

Research suggests that successful apologies include key verbal components such as admitting responsibility and asking for forgiveness. However, there is limited research examining the nonverbal aspects of apology and specifically whether people have a mental representation of apologetic faces. In three studies, reverse correlation was used to examine mental representations and judgments of apologetic faces that differed by gender. In each study, a visual template of an apologetic face was created using the responses of participants who completed a perceptual judgement task designed to estimate people's mental representations. In a second phase, a separate group of participants rated the apologetic face as well as the base face from which it was created on various apology- and gender-related characteristics. In each study, the generated apologetic face was consistently rated as being more apologetic, regretful, and remorseful than the base face, suggesting that people's mental representation of an apologetic face can be approximated using reverse correlation. Sadness was the highest rated characteristic for each face and ratings of sadness significantly predicted ratings of apology for three of the four visual templates created, suggesting that sadness is an important nonverbal aspect of an apologetic face. Submissiveness also emerged as a significant predictor of apology for three of the four faces; by contrast, trustworthiness, was not found to be a consistent characteristic seen in these apologetic templates. Male and female perceivers did not differ significantly in their ratings, and this was true regardless of the gender of the generator and target face. However, women generated an apologetic face from a female base face that was later judged to be significantly more apologetic than the apologetic face generated by men. These results suggest that men and women agree on their evaluations of apologetic faces, despite differences in their mental representations. This work is the first to demonstrate that

people hold mental representations of an apologetic face, and that sadness is a key characteristic perceived in faces generated to appear apologetic. Coupled with the literature on verbal apologies, the current research contributes to our practical and theoretical understanding of apology.

Keywords: apologies, gender, face perception, reverse correlation

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Who's sorry now? An investigation of how gender shapes the appearance and judgment of apologetic faces

Apologies are interpersonal tools that can repair valuable social bonds and improve relationships (Bippus & Young, 2019; Schumann, 2018). When individuals harm one another, social norms dictate that transgressors should apologize if they wish to fix damage to the relationship (Goffman, 1955). Victims then evaluate the apology on various characteristics, such as perceived sincerity, and respond with forgiving or unforgiving behaviour (Burnette et al., 2012). Because apologies are meant to repair and improve damaged relationships, researchers have attempted to identify which components of an apology are most likely to lead to positive relationship outcomes (Lazare, 2004; Tavuchis, 1991). Despite a rich literature that has examined the verbal components of successful apologies (e.g., Bippus & Young, 2019; Kirchhoff et al., 2012; Lewicki et al., 2016), little is known about the corresponding nonverbal facial cues that make one *appear* apologetic (cf., George et al., 2020).

The goal of this dissertation research was to examine people's visual representation of an apologetic face using the reverse correlation procedure. Across three studies, apologetic faces were generated by participants based on male (Studies 1 & 3), male-female morphed (Study 2), or female (Study 3) base faces to determine (a) whether people have a mental representation of an apologetic face that can be approximated using the reverse correlation procedure, (b) what apology-related characteristics are most prominent and most predictive of apology ratings in visual templates of an apologetic face, (c) whether such visual representations of an apologetic face will consistently be perceived as more masculine or feminine than the base face, and (d) whether perceived apology is moderated by the gender of the generator, perceiver, or target.

Using Reverse Correlation to Determine What Apologetic Faces Look Like

Social scientists have long been interested in how apologies affect relationships in the wake of a transgression (Lazare, 2004, Tavuchis, 1991). Transgressions can damage trust between the person committing the offence and the victim (Lewicki et al., 2016), can represent injustice or unfairness, and can even signify a lack of respect for victims (DeCremer & Schouten, 2008). To mitigate the negative influence of transgressions on relationship maintenance, transgressors may wish to apologize for their behaviour, as this has the potential to relieve guilt (Riek, 2010), demonstrate that they understand they have done something wrong, and show sympathy for the victim (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 elicit (Riek, 2010). Not all apologies are created equal, however, and researchers have sought to identify which elements most successfully predict forgiveness (Fehr & Gelfand, 2010). The findings of this research suggest that the key verbal components of a successful apology include taking responsibility, expressing remorse, asking for forgiveness, and promising not to reoffend (Kirchhoff et al., 2012; Lewicki et al., 2016; Schlenker & Darby, 1981; Struthers et al., 2008).

Although what a person says can be an essential part of an effective apology, there are other factors that can contribute to the complex, multi-layered, and dynamic process of interpersonal communication. For instance, the body language of both communicators and perceivers, as well as the use of silence or pauses, convey information beyond what is expressed verbally (Poyatos, 2015). Nonverbal cues such as vocal tone, eye gaze, and hand gestures, are also important communication tools, as are embodied emotions, such as crying to express remorse (Hornsey et al., 2019). Although each of these cues contribute to our understanding of the nonverbal aspects of apology, facial expressions in particular have been described as the

"most significant" form of nonverbal communication (Awasthi & Mandal, 2015, p. 1) and may therefore be relied on to a greater extent than other cues when evaluating apologies. However, there is currently a paucity of existing empirical evidence to guide the conceptualization of apologetic facial cues (cf., George et al., 2020; Hornsey et al., 2019). It is therefore unclear whether apologies are conceptualized differently when thinking about them in an abstract sense (e.g., describing what an apology is and what qualities it entails) versus thinking about them in terms of visual aspects of the social actors who deliver such apologies.

To address this gap in the literature, I previously tested whether an apologetic visual template could be generated using a reverse correlation technique (George et al., 2020). 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 (see Dotsch & Todorov, 2012), social psychologists have adopted the technique to examine how people mentally represent faces that communicate traits such as trustworthiness and dominance (Dotsch & Todorov, 2012), as well as broad social information such as status (Bjornsdottir & Rule, 2017), friendship networks (Kunst et al., 2019), and potential mate attractiveness (Karremans et al., 2011). For example, in a recent study, Lei and Bodenhausen (2017) investigated the role of racial bias in how one thinks about individuals from different social classes. Using a Black/White bi-racial base face, the experimenters asked participants who varied in their level of social class prejudice to generate images of low-, middle-, and high-income individuals. Results indicated that those high in prejudice generated low-income faces that were subsequently rated as appearing Black and high-income faces that were rated as appearing White.

Reverse correlation has also been used to examine ambiguous social categories, such as

sexual and political orientation (Tskhay & Rule, 2015). Across a series of studies, researchers demonstrated that, when constructing mental representations of individuals from groups that lack specific visual cues used to categorize them, such as gay and straight men, participants relied on other characteristics associated with each group. That is, faces generated to represent gay males seemed to appear happy and straight male faces seemed to appear angry to outside observers. Results also pointed to an association between happy and liberal, as well as angry and conservative. More recently, this research has been incorporated into a theoretical body of work that suggests accurate categorization of social targets is a function of one's experience with a specific group (Rule & Sutherland, 2017). For example, despite a lack of explicit cues identifying a face as gay or straight, individuals who have extensive experience interacting with gay men may call upon memories of past interactions to inform their categorization of a novel face. Reverse correlation therefore has the ability to clarify and concretize abstract constructs, such as how sexual orientation may manifest in physical facial features, by bypassing the need to verbally articulate them. Participants are able to generate images without necessarily having conscious knowledge of factors (e.g., race or affect) that might influence their mental representations of specific categories (e.g., social class or sexual orientation). Due to the large number of trials needed for this method, differences among each stimulus image are difficult, if not impossible, to detect, and it is only the resultant images that appear to demonstrate any recognizable features or expressions. As such, we are able to study how multiple constructs interact without concern for evoking demand characteristics.

In an initial study that made use of this method to examine people's mental representation of an apologetic face (George et al., 2020), I asked participants to completed a reverse correlation task in which they viewed 400 visually 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). These visually distorted images were then viewed two at a time and participants were asked to select which face among each pair appeared more apologetic across hundreds of trials. The distortion patterns of the selected images were then averaged, resulting in a final image meant to approximate people's mental representation of an apologetic face (Brinkman et al., 2017).

To determine whether the generated face was perceived as apologetic, a new sample of participants, naïve to what the image was intended to represent, were asked to rate the face on how apologetic they thought it appeared. In addition, the face was rated on various other characteristics that have been directly or indirectly associated with apology in the literature. These ratings were intended to help identify possible characteristics that make apologetic faces distinct and recognizable as expressing apology; because there was no previous research examining this question, this was largely exploratory. Some attributes were derived from the literature on the key verbal components of an apology and included remorse and regret (Lewicki et al., 2016; Scher & Darley, 1997; Struthers et al., 2008), guilt (Riek, 2010), submissiveness (a proxy for surrendering power and control as a form of social compensation; e.g., Hornsey et al., 2019), and trustworthiness (Hornsey et al., 2019; Lewicki et al., 2016; Tomlinson et al., 2004). Other attributes were selected from a list of apology descriptors derived through informal discussions among the research team. Overall, we investigated constructs associated with emotions and expressions (i.e., apologetic, guilty, regretful, remorseful, sad, and un-apologetic), character traits (i.e., kind, sincere, and trustworthy), and traits associated with social categories (i.e., male and female, masculinity and femininity).

In line with expectations, the apologetic face generated from this procedure was rated as more apologetic than the anti-apologetic face (generated by averaging the inverse or unselected noise patterns). Both images were rated as similarly masculine, and this is likely because they were derived from the same male face. Somewhat surprisingly, however, the apologetic face was rated as particularly sad and submissive relative to the other attributes including apologetic and trustworthy, providing some initial evidence that appearing sad and submissive (but not necessarily trustworthy) are key components of people's mental representation of an apologetic face.

Stereotypes and Biases Influence How We Think About Apologies

These initial findings suggest that people's mental representation of an apologetic face can be visualized using the reverse correlation procedure and that appearing sad and submissive are key components of an apologetic face. Notably, however, the apologetic face, as well as all evaluations, were based on a male base image. As such, it is unclear whether similar findings would emerge for other social groups. One social group that might be particularly likely to affect perceptions of apology is gender. Gender norms and stereotypes exist within, and can be applied to, the domain of apologies, and which stereotypes one relies upon has the potential to shape how both apologies and apologizers are perceived.

There is substantial evidence to suggest that the world is perceived through a gendered lens (e.g., Brewer, 1988; Cloutier et al., 2014). As humans evolved, the ability to distinguish individuals based on their biological sex was necessary for reproduction (Martin & Slepian, 2020). Gender roles are said to have developed out of a need to survive and maintain a genetic lineage, with men going to hunt for food if the women were unable to do so because of the need to care for and breastfeed small children. This long association between biological sex and

gender roles/behaviour undoubtedly shaped the way certain roles and behaviours are described today. Specifically, the terms masculinity and femininity reflect a constellation of traits that apply to certain behaviours or social roles rather than gender. It is through gender norms and stereotypes, however, that the terms masculinity and femininity have become so strongly linked to gender. Indeed, this link between concepts can function as a cognitive tool to help process information about new people quickly and efficiently. It is much easier to process congruent information about one's gender and their gender characteristics (e.g., a strong man, a dainty woman), than it is incongruent information (e.g., a timid man, an assertive woman; Martin & Slepian, 2020).

As gender roles evolve and stereotypes become less relevant, there have been attempts to move away from using traits that are highly tied to biological sex. For example, social psychologists often use terms such as competence and agency as synonymous of masculine traits, and warmth and communality to describe feminine traits. Despite this shift in academic terminology, the terms masculine and feminine are still very much a part of our vernacular. So much so, in fact, that masculine and feminine are used to describe more than just social targets. Wilkie and Bodenhausen (2015) demonstrated that numbers are ascribed gendered traits, with odd numbers between one and ten being described as masculine and even numbers as feminine. Given that gender can apply to social and non-social targets, it is possible that both the characteristics of apologizers and apologies may be associated with specific gender classifications.

There is some evidence to suggest that apologies may be gendered and that when imagining an apologetic face – or any face, for that matter – a male face may automatically come to mind. Gender norms have contributed to the historical and persistent power and privilege that

have been afforded to men and this norm has become so deeply ingrained in the collective unconscious that men often represent the standard against which all other social categories are compared (Bailey et al., 2020; Hamilton, 1991; Hegarty & Buechel, 2006; Hegarty & Pratto, 2004; Merritt & Harrison, 2006). An important consequence of gender norms is that they lead to repeated exposure of men and women in specific roles and engaging in specific behaviours. Without sufficient examples to counter this exposure, gendered associations are formed, become automatic, and influence attitudes (Finnegan et al., 2015; Olson & Fazio, 2001). The implicit assumption that men are the default form is so robust that even the use of gender-neutral language has not eliminated this pro-male bias (Lassonde & O'Brien, 2013). Men continue to be perceived as more agentic and less communal than women, a finding that is reflected in gender distributions across stereotypically male (e.g., CEO, scientist) and female (e.g., kindergarten teacher, nurse) occupations (Ceci et al., 2009; Eagly & Steffen, 1984).

Men occupy a large proportion of high-status positions including politicians and CEOs (Wei & Ran, 2019). This is particularly relevant to the current research because when individuals in highly visible careers transgress, it is common for them to apologize publicly, and such public displays often suggest that a severe transgression has occurred. The apology is thought to vary as a function of the severity of the offense (Wohl et al., 2012), and so admitting wrongdoing on a large scale and accepting the potentially harsh judgments and consequences from victims and non-victims alike would imply a serious offense. A public apology is also likely to improve the chances of being forgiven, as willingly lowering oneself to a position of low power, dominance, and/or social status has been shown to lead to successful apologies (MacLachlan, 2013). In fact, contrast effects suggest that it may actually be easier to recognize submission in a high-status individual because of the large discrepancy between it and their typically highly dominant

demeanour (Thayer, 1980). In other words, it may be more difficult to detect when a submissive individual is attempting to lower their status within a relationship, as the magnitude of apparent change from their typical submissive state is small.

Despite support for the hypothesis that men automatically come to mind when thinking about an apologetic face, it is possible that presenting an image of a White male primes individuals to think about apologies in the context of a male transgressor. Indeed, participants in my previous study (George et al., 2020) typically gave faces a rating of approximately 78 on a scale from 0 (*completely feminine*) to 100 (*completely masculine*), indicating high masculinity relative to femininity. Because of the strong association between male and masculine (and female and feminine), this is not surprising. Evaluations of the male apologetic face as submissive and sad, however, were not consistent with typical gender stereotypes, which depict men as dominant relative to women across domains such as social status, personality, and even body posture (de Lemus et al., 2012). These stereotypes are supported by evidence from the face perception literature, which demonstrates that as a neutral face is morphed to appear more dominant, it is rated as more masculine, whereas morphing to appear less dominant results in higher ratings of femininity (Oosterhof & Todorov, 2008). In addition, emotional expressions such as sadness are often associated with women because of the stereotype that women are particularly emotionally expressive (Adams et al., 2015) and emoting functions as a socially sanctioned way for women – but not men – to signal their affiliative tendencies (Fischer & LaFrance, 2015). In fact, because an apology is a reconciliatory behaviour, it can be classified as promoting a communal (versus agentic) and warm (versus competent) mindset, both of which overlap significantly with the concept of femininity (Martin & Slepian, 2020).

Given the stereotypes that women are more submissive and less dominant than men

(Adams et al., 2015), that women tend to be more emotionally expressive than men (e.g., Deng et al., 2016; Fischer & LaFrance, 2015), and that sadness and low dominance have been found to co-occur in face perception (Knuston, 1996), one might predict that associations between apologies, their attributes, and the gender stereotypes evoked by those attributes, would bias conceptualization of an apologetic face toward a *female/feminine* appearance in the absence of an unambiguous prime such as the male base image used in reverse correlation. Indeed, work by Schumann and Ross (2010) found that women do indeed apologize more than men (though the severity of transgressions differ across gender). As was mentioned above, repeated exposure to a stimulus helps to establish an implicit association, and if we witness women apologizing frequently, gender and apologizing behaviour can become inextricably linked. If a single apologetic face prototype does exist, it could arguably be a consolidation of faces from experience (see Minda, & Smith, 2011 for a summary of prototype categorization models).

However, it is also possible that gender stereotypes, as well as expectations surrounding apologies, are incorporated into *perceptions and judgments* of apologetic faces. Research in person perception has demonstrated that stereotypes bias how individuals are perceived and categorized. For example, Freeman and colleagues (2011) conducted a study in which they presented participants with faces that varied by race on a continuum from Black to White. Target images also differed as a function of status, operationalized by the clothing the target was wearing. Results indicated that participants were more likely to categorize racially ambiguous targets as Black when they were seen wearing a janitor uniform (low status), and White when they were seen wearing a business suit (high status) (Freeman et al., 2011). The authors concluded that existing status stereotypes (i.e., that White individuals hold high status jobs and Black individuals hold low status jobs) were competing with the incoming racial information and

influencing the way target faces were categorized. Stereotypes may function in a similar way for apologies, with stereotypes about apologies (e.g., that they demonstrate gender-related characteristics such as submissiveness) competing with gender information, and/or stereotypes about gender (e.g., women apologize more than men) competing with information about the magnitude of the apology being offered.

More recent evidence suggests that the perception of a single category can activate associated stereotypes, which in turn can shape the perception of a separate and distinct category. For example, Stolier and Freeman (2016) showed participants a series of faces that varied across three categories: race, sex, and emotional expression. Participants completed a mouse-tracking task wherein they categorized each face on a single dimension (e.g., sex). Their results suggest that when faces share overlapping associations or stereotypes (i.e., Black and Male, Happy and Female), stereotypes seem to interfere with the categorization process. When viewing a smiling Black female, participants were slower to categorize the face as female. The authors posit that this is due to the cyclical process of perceiving and attributing stereotypes to faces. The face is initially perceived as Black, which is highly associated with the male gender category. This strong association then inhibits the identification of the face as female, even in the presence of another, albeit weaker, association between happy and female (Stolier & Freeman, 2016). If similar stereotypes exist about who apologizes or about what characteristics are attributed to an apologizer, one would predict that these stereotypes work simultaneously to shape the perception and evaluation of apologetic faces. For example, if one holds a stereotype that apologizers are typically feminine and submissive, a man may be perceived as more feminine when offering an apology compared to when he is not. Endorsing such a stereotype may also affect how apologetic an individual is perceived to be, with people more closely matching the stereotype (i.e., women

in this example) rated as more apologetic than those who do not match the stereotype (i.e., men).

In line with this work, there is evidence to suggest that face perception can also be influenced by group-level factors associated with both the perceiver and target (Hehman et al., 2017; Hehman et al., 2019; Xie et al., 2019). Specifically, an interaction between perceiver and target gender has been found to influence impression formation such that ingroup member smiles are judged to be more benevolent than outgroup smiles (Paulus & Wentura, 2014; Xie et al., 2019). Indeed, this ingroup bias has been applied to the study of apologies, with those offered by ingroup members leading to more forgiveness than those from an outgroup (Wohl et al., 2012). In addition, apologies from men have been found to be more effective than apologies from woman, however this finding is only significant for female participants who read a scenario about a man offering them an apology (Walfisch et al., 2013). Research from the emotion literature suggests that this ingroup-outgroup interaction may be less important than the gender of perceivers alone. Women have been found to perceive faces as more expressive than men and are more accurate in identifying mixed emotional expressions (Hall & Matsumoto, 2004). Although women and men are equally skilled in identifying faces exhibiting individual emotions, women seem to be more attuned to subtle differences within a single emotion category (Hall & Matsumoto, 2004). In sum, research suggests that the social category of targets and perceivers can influence perceptions, and in the current research I aimed to examine these potential moderators in the creation and perception of visual templates of apology based on male, female, and gender-morphed base faces.

Present Research

The main purpose of the present research was to create visual templates of apologetic faces. Despite having some initial evidence from my previous study to suggest that an apologetic

face can be generated using reverse correlation, in the current studies I sought to replicate and extend this finding. Specifically, the first goal of this research was to confirm that visual representations of apologetic faces can be generated using the reverse correlation procedure. Across three studies, I presented participants with stimuli derived from a male (Studies 1, 3), gender-morphed (Study 2), and female (Study 3) base face to develop apologetic visual templates that were subsequently rated by an independent sample of participants on apology and apology-related characteristics. I predicted that the apologetic images created from these base faces would be perceived as significantly more apologetic than their base face. Such findings would add to the literature by suggesting that people have a mental representation of an apologetic face that can be approximated using the reverse correlation procedure.

My second goal was to investigate whether faces judged to be apologetic are also perceived as exhibiting other apology-related characteristics. The results of my initial study based on a male base face suggested that people's visual template of an apologetic face was rated as being particularly sad and submissive. In the current research I sought to replicate and extend this finding by determining whether these two characteristics would be rated highly regardless of the gender of the base face from which the apologetic face was created. If appearing sad and submissive are key components of an apologetic face, I would predict that they would emerge regardless of the gender of the base face. By contrast, if seeing a face exhibiting counter-stereotypical characteristics, such as a sad and submissive male face, results in higher ratings of the characteristics because of the contrast between the apologetic face and base face, I would expect that these findings would only replicate when a male base face was used.

The third goal of this research was to examine the relationship between gender, gender-related characteristics, and apologies. I set out to determine whether visual representations of an

apologetic face would consistently be perceived as more masculine or feminine than the base face from which it was created. Because an argument could be made that apologies are typically associated with women due to the frequency with which women apologize, and also that they are typically associated with men due to memorable examples of high-powered individuals offering apologies, I tested competing hypotheses. Specifically, I tested whether 1) apologetic classification images would be rated as *less* masculine and more feminine relative to the base face due to associations between apologies and stereotypically feminine characteristics, or 2) the apologetic face would be rated as *more* masculine relative to the base face given that gender norms place men as the default human form.

A final goal was to examine whether ratings of apology were moderated by the gender of the image generator (Studies 2 & 3), gender of the image rater (Studies 2 & 3), or gender of the target face (Study 3). Although the analyses were largely exploratory, I tentatively predicted a few key patterns. Given that women have been found to perceive faces as being more expressive and are more accurate in identifying mixed emotional expressions relative to men (Hall & Matsumoto, 2004), I hypothesized that an apologetic face generated by women would reflect their experience perceiving faces with heightened emotional acuity. Such perceptions may contribute to mental representations that express a greater degree of emotion, which in turn may be rated as more apologetic than an apologetic face generated by men. This possibility was tested in Studies 2 and 3. Given that judgments can also be biased by group membership, it also seemed possible that there would be a moderating effect of participant gender on face generation and ratings, with perceptions of gender in-group faces differing to those of gender out-group faces. This possibility was tested in Study 3. I predicted that participants would demonstrate an in-group bias such that faces generated by in-group members would appear more apologetic, and

that participants would rate in-group faces as more apologetic that out-group faces.

All materials, data, and syntax for data analysis in R (R Core Team, 2020) for each of these three studies can be found on the Open Science Framework (https://osf.io/5t9n7/?view_only=409bb3bc6eb84d6894df5620e095d47b).

Study 1

The main purpose of Study 1 was to replicate my previous finding (George et al., 2020) that reverse correlation can be used to generate an apologetic face. I predicted that the apologetic image generated in my previous study would be rated by a new sample of participants as significantly more apologetic than the base face from which it was created. If supported, these findings would provide additional evidence that people have a mental representation of an apologetic face that can be approximated using reverse correlation.

The second goal of the current research was to increase our understanding of what differentiates an apologetic face by determining which apology-related characteristics received the highest ratings and which were most predictive of ratings of apology. Building on my previous findings, I was particularly interested in determining whether the generated visual template of an apologetic face would again receive high ratings of sad and submissive, and whether these would be significant predictors of apology. Due to the exploratory nature of my original research, some characteristics were initially measured using a semantic differential scale (e.g., submissive to dominant, feminine to masculine) whereas others were measured using a single-dimension scale (e.g., remorse, regret), making it challenging to compare across characteristics and/or to test for predictors of apology. To extend these initial findings, I reduced the list of characteristics used in the original study by removing those that were only weakly detected in the apologetic face, while retaining characteristics associated with key verbal

components of apology. In the current study, seven apology-related characteristics, as well as gender (male and female) and gender-related characteristics (masculine and feminine) were each measured using a single-dimension scale (i.e., from 0 = *not at all* to 100 = *very*).

The final goal of Study 1 was to examine how gender would be rated for an apologetic face. Because the base face was created from exclusively male faces, it was expected that participants would identify the apologetic face as male relative to female, and masculine relative to feminine. However, what was less clear was whether those ratings would differ across the apologetic and base faces. I predicted that if people relied on gender-stereotypes associating women with being sad and submissive, the apologetic face would be perceived as less masculine and more feminine than the base face; by contrast, if people relied on a prototype of a male apologizer, the apologetic face would be perceived as more masculine (and potentially less feminine) than the base face.

Method

Participants and Procedure

A total of 153 undergraduates (M_{age} = 19.90, *SD* = 4.32, female = 104) participated in Study 1. The sample was racially diverse, and included White (26%), South Asian (25%), Middle Eastern (15.5%), Black (12%), South East Asian (6.5%), Mixed (4.5%), East Asian (4%), Latin American (4%), and Not Listed (2.5%) participants. A priori power calculations indicated that a sample size of 62 participants was required to detect an effect size of f = .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 =.09 at 80% power, p = .05.

All participants arrived at the lab and provided their consent to participate in two unrelated computer tasks. After completing a brief unrelated study, participants were directed to a link where they provided study-specific consent, and then completed the study. Participants were shown a grey-scale image of a White male face that was generated to be a visual template of an apologetic person. This stimulus picture was the result of my previous reverse correlation study (George et al., 2020; see Figure 1) and participation in the current study was restricted to those who had not previously completed a reverse correlation study in our laboratory. In addition to the apologetic face, they were also asked to rate the original base face from which the apologetic image was created.[1] All 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 11 randomized characteristics before rating another image type. A single question appeared at the top of each trial, asking each participant the extent to which they thought the face exhibited a particular characteristic. All images were given a rating from 0 (*not at all*) to 100 (*extremely*) on the following characteristics: apologetic, regretful, remorseful, sad, submissive, dominant, trustworthy, male, female, masculine, and feminine. This resulted in a total of 11 ratings for the apologetic face and 11 ratings for the base face. A response was required in order to move to the next trial. After each face had been rated on all 11 characteristics, a final block of trials instructed participants to again rate the faces, however this block consisted of ratings which combined two characteristics into a single scale. Specifically, each face was rated on three 100-point scales, viewed in random order, that ranged from: extremely submissive to extremely dominant, extremely female to extremely male, and

[1] Ratings for an addition image known as the anti-apologetic image were also collected. Because the anti-apologetic image represents the averaged noise pattern of the unselected stimuli and not, for example, an *unapologetic* face, it was determined that ratings of this face would not provide meaningful data or comparisons, and they were therefore not included in the analyses. 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).

extremely feminine to extremely masculine.

Figure 1

Male Base Face and Apologetic Face

Results

Ratings of Apology

In order to test my first hypothesis that a visual template of an apologetic face can be successfully constructed using reverse correlation, I compared ratings of apologetic for the apologetic face (M = 73.14, SD = 25.53) to those for the base face (M = 35.26, SD = 23.94). The distribution of each variable was non-normal and attempts to correct this using data transformations were unsuccessful. I therefore performed a Wilcoxon signed-rank test on the apology ratings to account for the data distribution.[2] Due to the nature of the analyses, results are reported as median scores and effect sizes were calculated using the sample size and a standardized statistic. Means are reported in tables for ease of interpretation. As predicted, the

[2] The majority of variables across all three studies were non-normally distributed. Transformations including square root, reciprocal, and log transformations, were conducted on all datasets and failed to consistently correct the distribution of all variables. Nonparametric tests were therefore used in Studies 1-3. It should be noted that parametric tests were also conducted, and results are consistent with those obtained from the nonparametric analyses.

apologetic face (*Mdn* = 80.00) generated using reverse correlation was perceived as significantly more apologetic than the base face (*Mdn* = 33.00), $p < .001$, $r_{effect\ size} = .77$. Both the apologetic and base faces were found to be significantly different from the midpoint of the scale, $p < .001$, $r_{effect\ size} = .66$, and $p < .001$, $r_{effect\ size} = .52$, respectively, with the apologetic classification image being rated as more apologetic than the midpoint, and the base face being rated as less than the midpoint. In addition to the standardized *r* effect sizes, the ratings on a 100-point scale provide additional information about the magnitude of perceived apology. That is, the mean rating of 35.26 can be conceptualized as a relatively low degree of apology perceived in the base face, whereas the apologetic face rating of 73.14 is perceived as having a high degree of apology.

Apology-Related Characteristics

The second goal was to increase our understanding of which apology-related characteristics are present in an apologetic face. In order to address this goal, I first compared the apology-related characteristic ratings for the apologetic image with those for the base face. Results from multiple Wilcoxon signed-rank tests indicated that ratings for sad, regretful, remorseful, 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 (see Table 1 for rank ordered mean ratings of the apologetic face). Consistent with this pattern, on the semantic differential scale that ranged from *submissive* to *dominant*, the apologetic face was rated as more submissive (*Mdn* = 29.00) than the base face (*Mdn* = 54.00).

In addition, I examined the rank order of the mean ratings of the apologetic face. As can be seen from the mean values listed in Table 1, sad was perceived to a greater extent on the apologetic face than any other characteristic, and direct comparisons confirmed that it was rated

significantly higher than all other characteristics, $p < .001$, including apologetic. Regretful, remorseful, and apologetic, which received the next highest ratings, were not different from each other, ps > .17.

Table 1

Mean Ratings and Mean Comparisons Across Faces for Study 1

Characteristic	Apologetic Face *M (SD)*	Base Face *M (SD)*	*p*	$r_{effect\ size}$
Apologetic	73.14 (25.53)	35.26 (23.94)	< .001	.77
Sad	86.36 (19.24)	38.94 (24.71)	< .001	.84
Regretful	75.25 (23.24)	35.03 (24.24)	< .001	.81
Remorseful	70.19 (24.57)	34.12 (22.96)	< .001	.74
Submissive	59.90 (25.55)	45.31 (22.02)	< .001	.45
Trustworthy	37.74 (22.26)	43.56 (22.22)	.24	.18
Dominant	24.34 (22.77)	53.85 (22.18)	< .001	.71
Male	70.77 (22.75)	81.28 (18.32)	< .001	.44
Masculine	57.54 (26.33)	72.73 (20.55)	< .001	.56
Feminine	28.59 (23.42)	23.88 (22.36)	.02	.25
Female	26.76 (24.17)	21.74 (23.00)	.14	.18

Note. Significance levels were calculated using Wilcoxon signed-rank tests due to the non-normal distributions of each variable. Bonferroni adjustments were made for *p*-values within the apology-related characteristics category, and accounted for 11 comparisons in total (i.e., comparisons of the apologetic and base faces for six characteristics, as well as comparisons across the highly rated characteristics for the apologetic face). Within each grouping above, characteristics are listed from highest to lowest scores of the apologetic face.

Next, I calculated the mean difference between ratings of the apologetic and base face (see Table 2 for rank ordered mean difference scores). By examining a mean difference score, I was able to account for the fact that each of the characteristics was perceived to some extent in

the base face. For example, if a characteristic is rated highly in the base face, ratings of that same characteristic become less meaningful in the apologetic face because it no longer distinguishes one face from another.

Table 2

Mean Difference in Ratings Across Faces for Study 1

Characteristic	Mean Difference (Apologetic Face – Base Face) *M (SD)*
Apologetic	37.88 (33.05)
Sad	47.42 (29.14)
Regretful	41.13 (30.82)
Remorseful	35.16 (34.54)
Dominant	-29.51 (31.24)
Submissive	14.59 (31.52)
Trustworthy	-5.82 (29.64)
Masculine	-15.20 (24.83)
Male	-10.51 (23.63)
Female	5.02 (26.84)
Feminine	4.71 (23.68)

Note: Within each grouping, characteristics are listed in rank order from highest to lowest absolute value of the mean difference scores.

A similar, though not identical, pattern was also found when comparing *mean difference* scores across each characteristic. The distribution of the mean difference scores did not meet the assumption of normality and a Friedman's ANOVA was therefore used to determine whether ratings differed significantly across characteristics. The ANOVA was significant, $\chi^2(6) = 421.76$, $p < .001$. Follow up multiple comparisons were conducted to compare the top-ranked characteristics. Wilcoxon signed-rank tests indicated that the magnitude of change in sad ratings

from the base to apologetic faces once again emerged as the largest value (Mdn_{diff} = 49.00), however this did not differ significantly from other large mean differences including regretful, *difference* = 53.50, and apologetic, *difference* = 91.50, based on a critical difference of 114.81.[3] Although the comparison between relative ratings of sad and remorseful was identified as significant from the omnibus test, *difference* = 122.00, it was no longer significant following a Bonferroni adjustment accounting for comparisons of the top three characteristics, $p = .07$. The magnitude of change in sad ratings was, however, significantly different from the magnitude of change for all other characteristics, all ps < .01.

Next, in order to determine the relative contribution of each characteristic in predicting ratings of apology, I conducted two multiple linear regression analyses. Due to the somewhat exploratory nature of the study, I used an all-subsets approach in which I entered all apology-related (i.e., non-gender) characteristics into a linear regression. Results indicated that the model with the best fit included submissive ($\beta = 0.33$, $p < .001$), regretful ($\beta = .31$, $p < .001$), sad ($\beta = 0.16$, $p = .03$), and remorseful ($\beta = 0.13$, $p = .05$), $F(5, 147) = 30.12$, $p < .001$, $\Delta R^2 = .49$. Dominance was included in this model, however it was not a significant predictor, $p = .10$. Although both models have the same ΔR^2 values and are not significantly different from each other, $F(1, 146) = 0.62$, $p = .43$, the reduced model has the advantage of being more parsimonious due to the fact that it has fewer predictors (Vandekerckhove et al., 2015). See Tables 3 and 4 for characteristic correlations and multiple regression analyses.

[3] Rank data was used to compare ratings across characteristics. To calculate the present results, each participant's responses were ranked from their lowest- to highest-rated characteristic. The rank values were then summed for each characteristic separately and the differences reported represent the difference between these sums for each comparison. A two-group comparison was identified as significant if the difference value was larger than the critical difference, which was calculated using a z value for seven measurements at $\alpha = .05$ (Eisinga et al., 2017; Field et al., 2012).

Table 3

Apology-Related Characteristic Correlations for Study 1

Characteristic	1	2	3	4	5	6	7
1. Apologetic	--						
2. Sad	.46***	--					
3. Regret	.57***	.58***	--				
4. Remorse	.42***	.29***	.35***	--			
5. Submissive	.52***	.20**	.31***	.30***	--		
6. Trustworthy	.13	.05	.05	.20*	.02	--	
7. Dominant	-.35***	-.25**	.31**	-.33***	-.27***	-.03	--

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

Table 4

Multiple Regression Results for Apology for Study 1

Characteristic	B	95% CI for B		SE B	β	ΔR^2	VIF[a]
		LL	*UL*				
Base Model						.49	
Constant	2.36	-15.13	19.84	8.85			
Regret	0.34***	0.18	0.50	0.08	0.31***		
Submissive	0.33***	0.20	0.45	0.06	0.33***		
Sad	0.21*	0.02	0.40	0.10	0.16*		
Remorse	0.12	-0.01	0.26	0.07	0.12		
Dominant	-0.12	-0.26	0.02	0.07	-0.11		
Trustworthy	0.05	-0.08	0.19	0.07	0.05		
Reduced model						.49	
Constant	3.68	-13.47	20.82	8.68			
Regret	0.34***	0.18	0.50	0.08	0.31***		1.67
Submissive	0.33***	0.20	0.45	0.06	0.33***		1.19
Sad	0.21*	0.02	0.40	0.10	0.16*		1.56
Remorse	0.13	-0.00	0.27	0.07	0.13		1.28
Dominant	-0.12	-0.26	0.02	0.07	-0.10		1.19

Note. CI = confidence interval; VIF = Variance Inflation Factor

[a] VIF > 5 indicates critical multicollinearity, VIF > 1 indicates moderate multicollinearity

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

Gender and Gender-Related Characteristics

My final aim was to examine perceptions of gender. I first used ratings of perceived gender as a manipulation check to ensure that both the apologetic and base faces were identified as male. Because the base face, and therefore the apologetic face that was produced from it, were composed exclusively of images of men, it was expected that participants would rate both faces as looking more male than female. Once again, the assumption of normality was not met, and so a 2 (Image Type: Apologetic, Base) x 2 (Perceived Gender: Male, Female) Friedman's ANOVA was conducted. A significant effect revealed that ratings of gender differed across images, χ^2 (3) = 243.67, $p < .001$. Post hoc comparisons were adjusted using Bonferroni correction to account for eight comparisons, and results indicated that the apologetic face was rated as significantly more male (*Mdn* = 75.00) than female (*Mdn* = 20.00)*, difference* = 188.50*, p* < .01, as was the base face (Mdn_{male} = 83.00; Mdn_{female} = 15.00), *difference* = 288.50, $p < .01$. The apologetic face was rated as significantly less male than the base face, *difference* = 74.50, $p < .001$, $r_{effect\ size}$ = .44. The apologetic face was not rated as significantly different from the base face in terms of how female it appeared, *difference* = 25.50, $p = .23$, $r_{effect\ size}$ = .18. The ratings of male for the apologetic face were significantly different from the midpoint of the scale, $p < .001$, $r_{effect\ size}$ = .04, as were the ratings of male for the base face, $p < .001$, $r_{effect\ size}$ = .83. This suggests that they were both rated as being significantly more male than not, however effect sizes indicate that this difference was larger for the base face. Ratings of female were also significantly different from the midpoint of the scale, with the apologetic*,* $p < .001$, $r_{effect\ size}$ = .69, and base face, $p < .001$, $r_{effect\ size}$ = .76, found to be less female than the midpoint. In addition, although ratings of gender on a scale from female to male were not included in the ANOVA, the scale means are in line

with ratings of male, although again the base face (*Mdn* = 80.00) was rated as more male (relative to female) than the apologetic face (*Mdn* = 65.00), $p < .001$, $r_{effect\ size} = .46$, see Table 1.

To investigate whether an apologetic face was perceived to be more masculine or feminine than the base face, I conducted a 2 (Image Type: Apologetic, Base) x 2 (Gender-Related Characteristic Rating: Masculine, Feminine) Friedman's ANOVA. Ratings of gender-related characteristics differed significantly across images, χ^2 (3) = 188.60, $p < .001$. Post hoc analyses were Bonferroni adjusted to account for eight comparisons. Results indicated that both faces were rated as significantly more masculine than feminine ($difference_{base\ face} = 277.00$, $p < .001$; $difference_{apologetic\ face} = 121.00$, $p < .001$). However, the apologetic face was rated as less masculine (*Mdn* = 61.00) than the base face (*Mdn* = 75.00), *difference* = 110.00, $p < .001$. The faces differed significantly in their ratings of femininity as well (apologetic face: *Mdn* = 27.00; base face: *Mdn* = 20.00), *difference* = 46.00, $p = .03$. Each difference value was deemed significant if it exceeded a critical difference of 59.58. As with ratings from female to male, the ratings on the scale from feminine to masculine were consistent with ratings of masculinity alone (apologetic face: *Mdn* = 58.00; base face: *Mdn* = 70.00). Again, ratings for both the apologetic face ($p < .001$, $r_{effect\ size} = .56$) and base face ($p < .001$, $r_{effect\ size} = .76$) were significantly higher than the midpoint of the scale, demonstrating that they were both perceived as masculine relative to feminine, but with greater masculinity perceived in the base face relative to the apologetic face. Comparisons to the midpoint confirm that although the apologetic face was perceived as masculine, it was rated as less masculine and more feminine than the base face.

Discussion

Replicating previous findings, participants in Study 1 perceived apology to a significantly greater extent in the apologetic image relative to the neutral White male base face from which

the apologetic face was created. This difference in ratings of apology across faces provides additional evidence that reverse correlation can be used to create a visual template of an apologetic face. Ratings of additional apology-related characteristics provide insight into potential factors that contribute to people's visual template of an apologetic face. The apologetic face was rated as more sad, regretful, remorseful, and submissive than the base face, as well as less dominant. Although these results are consistent with my previous findings, the low ratings of trustworthiness, and lack of difference from the base face, contradict what one might predict based on the verbal apology literature. Some research suggests that a successful verbal apology should include a promise to the victim that the transgression will not reoccur in the future (Kirchhoff et al., 2012; Lewicki et al., 2016), which might suggest the need to convey trustworthiness. Despite an ability to perceive trustworthiness in facial stimuli (Dotsch & Todorov, 2012), the current findings provide evidence that signals of trustworthiness may not be a key component in perceptions of an apologetic face.

When one considers the absolute magnitude of the mean difference of each rating across the apologetic and base faces, it appears as though the largest difference is in ratings of sadness. This would suggest that sadness plays a particularly influential role when communicating apology nonverbally. In addition, consistent with my previous research (George et al., 2020), multiple linear regression analyses support the possibility that both sadness and submissiveness are important; perceptions of submissiveness were the largest predictor of apology, followed by perceptions of sadness, and, as might be expected, regret and remorse.

Ratings pertaining to the gender of an apologetic face offer some initial information about the biases we hold about apologizers. Regardless of the stereotypes that might exist about who should or who frequently apologizes, participants identified the apologetic face as matching

the gender of the image used to create the target faces (i.e., male). However, both ratings of gender category (i.e., male or female) and gender-related characteristics (i.e., masculine or feminine), provided evidence that an apologetic face created from a male base face is seen as less masculine relative to the male base face from which it was created.

Study 2

Study 1 provided some initial insight into how apologetic faces are conceptualized and perceived, however given that a male base face was used, it is possible that these results would not generalize beyond a single social target (i.e., White males). In a pre-registered Study 2, my first aim was therefore to replicate and extend these findings by using reverse correlation to generate an apologetic face using a novel base image. A gender-morphed base face, created by combining 8 male and 8 female faces, was used in an initial image generation phase, and the apologetic face that resulted from Phase 1 was then shown to participants in Phase 2 for rating. As in Study 1, I hypothesized that the apologetic face would once again be rated as more apologetic than the base face.

The second goal was to examine whether the apology-related characteristics that were found in Study 1 would also be rated highly on the new apologetic face and whether their relative ability to predict ratings of apology would be similar. I anticipated that sadness would be highly rated, and that sadness and submissiveness would again emerge as significant predictors of apology. Such findings would provide additional evidence for key characteristics identified in visual templates of apology, as well as provide some initial evidence that this may be the case regardless of the social category of the target.

The third goal of Study 2 was to further examine how masculine and feminine this apologetic visual template would be rated relative to the base face from which it was created.

Gender ratings from Study 1 suggest that an apologetic face created from a male base face is recognized as male, yet it is perceived as less male (*Mdn* = 75.00) and less masculine (*Mdn* = 61.00) than the base face (Mdn_{male} = 83.00, and $Mdn_{masculine}$ = 75.00, respectively). Given the use of a gender-morphed face in Study 2, I was interested to see if the new apologetic face would be rated as male, and whether ratings of masculinity would again be lower for the apologetic face relative to the base face. Together with the results of Study 1, such a finding would provide converging evidence that apologies are associated with men but that apologetic faces are also perceived to be lower in masculinity.

Finally, in the current study, I examined whether ratings of apology would be moderated by the gender of the image generator and/or the gender of the image rater (i.e., perceiver). In order to address this goal, an equal number of male and female participants were recruited for both the image generation and image rating phases of Study 2. Because women have been found to perceive faces as being more expressive and are more accurate in identifying mixed emotional expressions relative to men (Hall & Matsumoto, 2004), I expected women's heightened emotion perception to manifest in their mental representations as an apologetic classification image that appears and is rated as more apologetic than an apologetic image generated by men. I further expected that this would be particularly true for female (versus male) raters, owing again to their attention to emotional expressions.

Method

Phase 1: Image generation

Participants and Procedure

An initial sample of 73 participants was recruited from the Department of Psychology undergraduate participant pool in exchange for course credit. Data from three participants was

removed because they did not complete the study, and data from one additional participant was removed because they answered the majority of questions on the Conscientious Responders Scale (CRS; Marjanovic et al., 2014) incorrectly (i.e., a total score of less than 3 out of 5). After exclusions, the final sample consisted of 69 participants (M_{age} = 19.10, *SD* = 1.89, female = 34). Participants within the sample identified as South Asian (29%), Black (17%), White (17%), Middle Eastern (12%), South East Asian (9%), Mixed (7%), East Asian (3%), Latin American (4.5%), and Aboriginal (1.5%). Participant recruitment was limited to individuals who identified as either male or female.[4] The purpose of this exclusion was to ensure that participants identify with only one gender, allowing the comparison of ratings across two distinct perceiver gender categories.[5] This sample size was determined to be adequate based on current norms within the reverse correlation literature, which suggest that as few as 30 participants per condition are needed to produce high quality resulting images (e.g., Bjornsdottir et al., 2019; Lick et al., 2013).

Participants completed the study on desktop computers located in a computer laboratory on campus. As many as 20 participants completed the study during a single testing session, under the supervision of a research assistant who ensured that participants worked quietly and independently. Each participant was provided with the URL that took them to the website where the survey was housed. After providing informed consent and basic demographic information, participants proceeded to the reverse correlation trials. After completing the study, participants were thanked and debriefed.

---

[4] Respondents were asked to self-identify their gender, however their biological sex at birth is unknown.

[5] The automatic processing of gender information develops early, as gender is an important dimension by which we categorize individuals (Dunham et al., 2016). Biological sex and gender are historically conflated terms and this relationship has undoubtedly led to a binary definition of gender as either male or female (Martin & Slepian, 2020). I do not adhere to, nor do I endorse this definition of sex or gender. Both of these terms have been highly influenced by social constructionism and many scholars agree that sex and gender exist outside of this binary definition (for a review, see Hyde et al., 2019). That said, male and female remain the primary labels used to describe others and, coupled with the association between gender and gender-related characteristic (i.e., stereotypes or gender roles), the use of binary gender categories was the most appropriate way to address my research questions.

Stimuli

Reverse Correlation

The current study used a forced-choice version of the reverse correlation paradigm (see Brinkman et al., 2017; Todorov et al., 2011). Unlike the base face used in typical reverse correlation studies, including the image used in Study 1, which consist of a grayscale image of a Caucasian male face generated by the morphed average of faces from the Karolinska Face Database (Lundqvist et al., 1998; see Figure 1), the base image for Study 2 was created by morphing White male and female faces together (Brinkman et al., 2019). The individual faces were once again grayscale images selected from the Karolinska Face Database (Lundqvist et al., 1998), and were combined using WebMorph (DeBruine, 2017), an online face transformation program (see Figure 2). This single image served as the base face on top of which the visual distortion patterns appeared. A sample trial from the image generation phase of Study 2 can be seen in Figure 3.

Figure 2

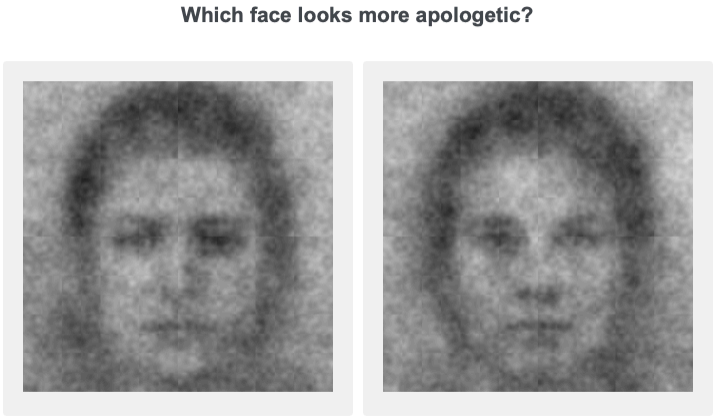
Gender-Morphed Base Face

Visual distortion or noise patterns were randomly generated in R using the rcicr package and 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 of the visual noise patterns were then overlaid onto the base face image. As is the nature of visual noise, it 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 caused areas on the stimuli that appeared light appear dark, and vice versa. A total of 400 trials were presented individually and consisted of two images presented side-by-side on the screen. Both images in a single trial contained the same base face with inverse noise patterns on each (Dotsch & Todorov, 2012). Instructions at the top of each screen asked participants to select which of the two images looked "more apologetic."

Figure 3

Sample Image Generation Trial



Conscientious responding

The Conscientious Responders Scale (CRS; Marjanovic et al., 2014) instructs participants to respond in particular ways to five questions randomly distributed throughout the study. For example, one question reads "In response to this question, please choose option number three,

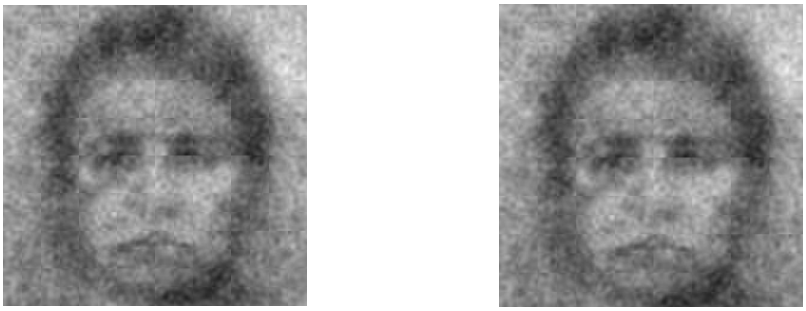
slightly disagree." If participants responded incorrectly to three or more questions, they were classified as a random responder and were excluded from the final data set. Participants who correctly answered three or more questions on this scale were classified as conscientious responders and their data was retained for analyses.

Results

Using the rcicr package (Dotsch, 2017) in R (R Core Team, 2020), participant responses were grouped by gender and averaged (see Figure 4) to generate two group-level classification images of an apologetic face.

Figure 4

Male- and Female-Generated Apologetic Faces



Note. Apologetic faces were created by averaging responses of male (left) and female (right) participants at the group-level.

Method

Phase 2: Image Rating

Participants and Procedure

Two hundred and sixteen undergraduates participated in the rating phase of Study 2. Based on the same criteria used in Study 1, five participants were excluded because they did not

provide consent (n = 2) or they did not provide their anonymous identification number (n = 3) which meant I was unable to confirm that their entry was unique (i.e., not a duplicate). In addition, 23 participants were removed because they did not complete the study, and an additional 37 participants were removed because they answered the majority of questions on the Conscientious Responders Scale (CRS; Marjanovic et al., 2014) incorrectly (i.e., a total score of less than 3 out of 5). The CRS was adapted for the current study to include a face and asked participants to provide a specific rating. To match the format of the rest of the study, these ratings were made on a sliding scale. Because it could be difficult to select an exact number on the sliding scale, any response within +/- 3 of the required number was considered a correct answer. After exclusions, the final sample consisted of 151 participants (M_{age} = 20.38, SD = 4.21, female = 78). Participants identified as South Asian (25%), White (23%), East Asian (15%), Middle Eastern (14%), South East Asian (9%), Black (6%), Mixed (4%), Not Listed (2%), and Latin American (2%). Because different statistical analyses were used to address each research question individually, the sample size and power calculations reported here were based on the highest order analytic design. Based on a 2 (Rater Gender: Male, Female) x 2 (Generator Gender: Male, Female) mixed factorial design, it was determined using G*Power (Faul et al., 2007) that a sample size of 98 participants was required to detect a medium effect size of f^2 = .25 with 80% power. A sensitivity power analysis revealed that the final sample of 151 was able to detect a minimum effect of f = .10 based on 80% power and a two-tailed alpha level of p = .05.

Participants completed the study online and were instructed to use a desktop computer. After providing their consent, participants completed the image rating task. The image rating phase of Study 2 was identical to that of Study 1, except that in this study, no semantic

differential scales were included, and participants were shown three grey-scale images of a White face: the gender-morphed base face, the apologetic face image generated in Phase 1 by male generators, and the apologetic face image generated in Phase 1 by female generators. All 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 11 randomized characteristics before rating another image type. A single question appeared at the top of each trial, asking participants the extent to which they thought the face exhibited a particular characteristic. All images were given a rating from 0 (*not at all*) to 100 (*very*) on the following characteristics: apologetic, regretful, remorseful, sad, trustworthy, dominant, submissive, male, female, masculine, and feminine. This resulted in a total of 11 ratings for each of the apologetic and base faces. Responses were required in order to move to the next trial. To assess that participants were paying attention to the task, five questions adapted from the CRS (Marjanovic et al., 2014) were randomly presented during the survey. Once the task was completed, participants were thanked in writing and debriefed.

Results

Ratings of Apology

In order to address my first two research goals, I averaged the ratings of the male- and female-generated faces on all characteristics. By averaging these scores, I was able to use a simplified statistical analysis to address the current research question.[6] After collapsing ratings across the two apologetic faces, the rating of apology (*Mdn* = 65.00) was then compared to the base face (*Mdn* = 28.00) using a Wilcoxon signed-rank test. Consistent with my hypotheses, and with the results of Study 1, the faces were rated as significantly different from each other, ($p <$

[6] A two-sample *t*-test comparing ratings for male- and female-generated apologetic faces revealed no significant difference in how apologetic the faces were perceived.

.001, $r_{effect\ size}$ = .70), with the averaged ratings of the apologetic face being significantly more apologetic than the midpoint, $p < .001$, $r_{effect\ size}$ = .37, and the base face being rated as significantly less apologetic than the midpoint, $p < .001$, $r_{effect\ size}$ = .71.[7] The *p* values reported were Bonferroni adjusted to account for three comparisons. Considering the ratings as unstandardized effect sizes, the mean base face rating of 28.41 on a 100-point scale of apologetic was comparable to the base face in Study 1 (M = 35.26). However, the effect size for the current apologetic face was not quite as large as previously found (M = 73.14), with a mean apology rating of only 59.28.

Apology-Related Characteristics

To examine which apology-related characteristics differentiate an apologetic face from a base face when a gender-morphed image is used, the ratings of each apology-related characteristic were again averaged across the male- and female- generated apologetic faces, and compared to ratings of the base face (see Table 5 for rank ordered mean ratings of the apologetic face). Multiple Wilcoxon signed-rank tests indicated that, similar to Study 1, the apologetic face was rated as significantly more sad, regretful, remorseful, and submissive than the base face and significantly less dominant (all *p*s < .001, see Table 5). There was no difference in ratings of trustworthiness. Direct comparisons of the characteristics for the apologetic face alone confirmed that ratings of sadness were significantly higher than all other characteristics, including apologetic, all *p*s < .001 following Bonferroni adjustments. Regret ratings were also significantly higher than ratings of apologetic, p = .02.

[7] Because the Wilcoxon signed-rank test compares samples based on rank data, the test statistic that is generated is a sum of rankings and therefore does not represent a meaningful, interpretable value. For this reason, and in line with reporting standards, test statistics are not presented (Field et al., 2012).

Table 5

Mean Characteristic Ratings and Comparisons Across Faces for Study 2

Characteristic	Apologetic Face *M (SD)*	Base Face *M (SD)*	*p*	$r_{effect\ size}$
Apologetic	59.28 (24.28)	28.41 (22.19)	< .001	.70
Sad	73.84 (20.47)	35.26 (23.58)	< .001	.79
Regretful	64.22 (22.28)	30.87 (23.38)	< .001	.77
Remorseful	60.10 (22.14)	29.31 (20.57)	< .001	.74
Submissive	54.16 (24.07)	44.07 (22.91)	< .01	.30
Trustworthy	38.79 (21.14)	42.56 (22.98)	.91	.12
Dominant	29.85 (20.98)	47.52 (23.45)	< .001	.59
Characteristic	Apologetic Face *M (SD)*	Base Face *M (SD)*	*p*	*r*
Male	50.13 (25.22)	29.25 (21.63)	< .001	.57
Masculine	47.76 (23.74)	32.25 (21.89)	< .001	.47
Female	46.86 (24.95)	74.03 (19.79)	< .001	.70
Feminine	46.07 (23.68)	73.35 (18.31)	< .001	.70

Note. Significance levels were calculated using Wilcoxon signed-rank tests due to the non-normal distributions of each variable. As in Study 1, Bonferroni adjustments were made for *p*-values within the apology-related characteristics category, and accounted for 11 comparisons in total (i.e., comparisons of the apologetic and base faces for six characteristics, as well as comparisons across the four highest-rated characteristics for the apologetic face). Also, within each grouping, characteristics are listed from highest to lowest scores of the apologetic face.

As in Study 1, the difference between mean ratings of the apologetic and base faces were calculated to demonstrate the magnitude of difference in scores for each characteristic. These differences are listed in rank order in Table 6. The rank order of apology-related characteristics was identical to Study 1, despite a different base face being used. As in Study 1, and in the raw ratings from both Studies 1 and 2, sadness again emerged as the top-rated characteristic. A post hoc direct comparison between the difference scores for sad and apologetic revealed that the change in perceived sadness was larger than the change in perceived apology from base to

apologetic face, *difference* = 324.50, $p < .01$, $r_{effect\ size} = .27$, and this exceeded the critical difference of mean rank sums of 114.05, confirming statistical significance.

Table 6

Mean Difference in Ratings Across Faces for Study 2

Characteristic	Mean Difference (Apologetic Face – Base Face) *M (SD)*
Apologetic	30.87 (32.76)
Sad	38.58 (31.14)
Regret	33.34 (29.52)
Remorse	30.78 (29.74)
Dominant	-17.67 (26.42)
Submissive	10.09 (31.26)
Trustworthy	-3.77 (25.93)
Feminine	-27.28 (28.78)
Female	-27.17 (30.58)
Male	20.88 (31.66)
Masculine	15.51 (29.73)

Note: Within each grouping, characteristics are listed in rank order from highest to lowest absolute value of the mean difference scores.

To address the question of how apology-related characteristics contribute to visual perceptions of apology, I once again ran a multiple regression analysis. The best model for the current data was one in which all variables were included, and this accounted for 69.60% of the variance in apology ratings, $F(6, 144) = 58.09$, $p < .001$, $\Delta R^2 = .70$. Unlike Study 1, remorseful accounted for the largest proportion of variance ($\beta = 0.341$ $p < .001$), followed by trustworthy ($\beta = 0.23$, $p < .001$), submissive ($\beta = 0.23$, $p < .001$), and sad ($\beta = 0.17$, $p = .01$), see Table 8. Regret and dominance were not identified as significant predictors, $\beta = 0.16$, $p = .05$ and $\beta = 0.08$, $p = .08$, respectively. This is somewhat surprising for regret, in particular, as it was found

to be highly correlated with other predictive variables such as remorse (see Table 7 for correlation table).

Table 7

Apology-Related Characteristic Correlations for Study 2

Characteristic	1	2	3	4	5	6	7
1. Apologetic	--						
2. Sad	.62***	--					
3. Regret	.71***	.71***	--				
4. Remorse	.74***	.61***	.76***	--			
5. Submissive	.61***	.41***	.55***	.50***	--		
6. Trustworthy	.44***	.17*	.26**	.31***	.20*	--	
7. Dominant	-.21**	-.18*	- .15	- .17*	- .16*	.10	--

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

Table 8

Multiple Regression Results for Apology for Study 2

Characteristic	B	95% CI for B		SE B	β	ΔR^2	VIF[a]
		LL	*UL*				
Full Model						0.70	
Constant	-7.22	-16.99	2.55	4.94			
Remorse	0. 34***	0.19	0.50	0.08	0.31***		2.57
Trustworthy	0.27***	0.16	0.38	0.06	0.23***		1.14
Submissive	0.24***	0.13	0.35	0.01	0.23***		1.49
Sadness	0.19*	0.04	0.35	0.08	0.16*		2.07
Regret	0.17	-0.00	0.35	0.09	0.16		3.28
Dominant	-0.10	-0.20	0.01	0.05	-0.08		1.08

Note. CI = confidence interval; VIF = Variance Inflation Factor

[a] VIF > 5 indicates critical multicollinearity, VIF > 1 indicates moderate multicollinearity

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

Gender and Gender-Related Characteristics

To test whether the apologetic face generated from a gender-morphed base would be rated as more male or more female than the base image, a 2 (Image Type: Apologetic, Base) x 2 (Perceived Gender: Male, Female) Friedman's ANOVA was conducted. Ratings of gender differed across images, χ^2 (3) = 154.61, $p < .001$. Bonferroni-adjusted post hoc analyses for four comparisons indicated that, unexpectedly, the base face was rated as significantly less male (*Mdn* = 28.00) than female (*Mdn* = 76.00), *difference* = 273.50, $p < .001$, $r_{effect\ size} = .07$, suggesting that for participants, this gender-morphed face did not look gender ambiguous but rather seemed more female than male. By contrast, the ratings for the apologetic face did appear gender-ambiguous and did not differ by ratings of gender ($Mdn_{male} = 50.00$, $Mdn_{female} = 46.86$), *difference* = 23.50, $p > .99$, $r_{effect\ size} = .07$. In addition, unlike Study 1, the apologetic face was rated as more male and less female than the base face (*difference* = 157.00, $p < .001$, $r_{effect\ size} = .57$; *difference* = 140.00, $p < .001$, $r_{effect\ size} = .68$, respectively). The critical difference for each comparison was 59.19.

An additional 2 (Image Type: Apologetic, Base) x 2 (Perceived Gender Characteristics: Masculine, Feminine) Friedman's ANOVA was conducted to investigate perceived masculinity and femininity associated with apologies. Gender characteristic ratings differed significantly across images, χ^2 (3) = 154.32, $p < .001$. Follow up post hoc analyses with Bonferroni adjustments for four comparisons were consistent with ratings of gender, such that the base face was rated as significantly less masculine (*Mdn* = 30.00) than feminine (*Mdn* = 76.00), *difference* = 264.00, $p < .001$, $r_{effect\ size} = .04$, whereas the ratings of gender characteristics did not differ significantly for the apologetic face ($Mdn_{masculine} = 50.00$, $Mdn_{feminine} = 50.00$), *difference* = 16.00, $p > .99$, $r_{effect\ size} = .04$. Comparisons across faces indicated that the apologetic face was both significantly more masculine than the base face, *difference* = 88.00, $p < .001$, $r_{effect\ size} = .47$, and

less feminine than the base face, *difference* = 192.00, $p < .001$, $r_{effect\ size} = .70$. The critical difference for each comparison was 59.19.

Moderation of Apology Ratings

Finally, I assessed the possible moderation of generator and/or rater gender on ratings of apology. Due to non-normal distributions and the mixed design, an ANOVA-type statistic test was conducted (Brunner et al., 1997; Noguchi et al., 2012). Results of the 2 (Rater Gender: Male, Female) x 2 (Generator Gender: Male, Female) between-within analysis yielded no significant main effects or interactions, all *p*s > .36.

Discussion

Study 2 provided additional evidence for my first hypothesis that people have a mental representation of an apologetic face and that reverse correlation can be used to produce a visual template of apology that is seen as more apologetic than a base face. This was true when people saw an apologetic face based on a male base face in Study 1, and when they saw an apologetic face based on a gender-morphed face in Study 2. This provides additional evidence that people have a mental representation of an apologetic face that can be approximated using reverse correlation, and this is true for a gender-morphed face that resulted in a gender-ambiguous apologetic face.

Ratings of apology-related characteristics in the current study were also consistent with the results of Study 1, as ratings of the apologetic face in both studies received higher ratings for apologetic, sad, regretful, remorseful, and submissive than the base face, and lower ratings of dominance. Again, sadness received the highest ratings, with ratings of apology falling significantly below sadness and regret. Although sadness remained a significant predictor of ratings of apology, in the current study it accounted for less variability relative to Study 1, as

well as relative to ratings of remorseful and trustworthy, suggesting that for this gender-morphed face, perceived sadness might have had less predictive ability for ratings of how apologetic the face appeared.

The findings regarding the gender and gender-related characteristics begin to shed light on potential stereotypes and biases associated with apologies. Surprisingly, the gender-morphed base face was perceived as female relative to male, yet the apologetic face received similar ratings for both male and female, suggesting that the apologetic face was perceived as gender ambiguous. When the difference in ratings between the faces is considered, it appears as though faces originally identified as looking female are rated as looking less so when altered to represent nonverbal apology, with ratings of female and feminine similar in magnitude to ratings of male and masculine for the apologetic face. By contrast, although ratings of male and masculine in Study 1 were lower in the apologetic face relative to the base face, ratings of female and feminine did not differ. This lends support to the theory that when we think of expressions of apology, we default to applying them to men. Such an explanation would suggest that apologies have been influenced by social construction, not unlike applying assumptions of maleness to gender-neutral terms such as doctor or professor. In Study 3, both a male and female base face were used to gain further insight into how ratings of gender and gender-related characteristics differ among base and apologetic faces, and to determine whether female apologetic faces would be rated as more female or feminine relative to male or masculine.

My hypothesis regarding the moderation of ratings by rater gender was not supported in the present study. One explanation for this result is that no such relationship exists. Alternatively, the rating/perception of apology-related characteristics may rely on the target's group membership. Similar to nonverbal "accents" (Marsh et al., 2003), women may be more attuned

to pick up on nonverbal expressions on in-group member faces. If this is the case, it is not surprising that ratings did not differ for men or women, as the apologetic face was not identified as distinctly belonging to either the male or female gender. The role of social-category membership was explored further in Study 3.

Study 3

The first goal of Study 3 was to replicate and extend the results of Studies 1 and 2 by examining whether an apologetic face could be created using reverse correlation from both a male and a female base face. To do so, I used the male base face from Study 1 as well as a female base face created to be comparable (Lundqvist et al., 1998). These images were then used to produce a new male and a female apologetic face using reverse correlation. I predicted that both of these faces would be rated as more apologetic than their respective base faces.

The second goal was to further understand the specific apology-related characteristics that contribute to making an apologetic face distinct from a neutral base face. In the previous two studies, sadness received the highest raw and relative ratings when people rated the apologetic faces. I examined whether this would again be the case for a new apologetic face created from a male base and whether this would extend to an apologetic face created from a female base face. In addition, in the current study, I hypothesized that each of the apologetic faces would again be rated high in regret, remorse, and submissiveness, and that these variables would be highly correlated and predictive of apology, as was found previously. Given the inconsistency of trustworthiness ratings in Studies 1 and 2, it was unclear how this characteristic would be perceived in each of the stimulus faces. As for ratings of dominance, there is evidence to suggest that females and males are judged differently on this characteristic outside of an apology context

(Oosterhof & Todorov, 2008; Sutherland et al., 2015), and so it was unclear as to whether or not gender would affect ratings of dominance for faces expressing apology.

The third goal was to again measure ratings of perceived gender and gender-related characteristics in visual templates of apologetic faces and see if similar patterns emerged for male and female targets. The results of Study 1 suggest that an apologetic face might be perceived to be less masculine than a base face, however that base image was comprised entirely of male faces. The base face used in Study 2 was comprised of an equal number of male and female faces and is in line with what some researchers have used as a gender-ambiguous stimulus (e.g., Brinkman et al., 2019). Despite the equally male/female composition of the gender-morphed base face in Study 2, it was rated as highly female-looking, with the apologetic face perceived as less feminine and more masculine in comparison. Using separate male and female base faces in Study 3, I predicted that the gender of the apologetic faces would be perceived as matching the gender of the base faces from which the images were created (i.e., males would be rated as male and females as female). Given the different pattern of results from Studies 1 and 2, it was unclear whether the gender-related characteristics of masculine and feminine would be rated higher or lower for the apologetic faces relative to the base faces and whether this would depend on the gender of the base face. It was also unclear if the apologetic faces would be perceived as more male/masculine than female/feminine (as in Study 1), or if these ratings would be similar (as in Study 2). Each of these possibilities was examined.

In addition, in Study 3, I examined three potential moderators of apology ratings: the gender of the image generator, the gender of the target image, and the gender of the raters. Specifically, I examined the possibility that the gender of the raters would influence perceptions of apology such that female raters would perceive a greater degree of all characteristics relative

to male raters, who may not be as aware of subtle differences in expression due to gender differences in identifying complex mixed emotions (Hall & Matsumoto, 2004). I also predicted that I would find a main effect of target gender, however it was unclear how such an effect would manifest. It was possible that the apologetic male faces would be rated as more apologetic than the apologetic female faces because they match a prototype developed from repeated exposure to high-powered men offering memorable public apologies (MacLachlan, 2013; Wei & Ran, 2019; Wohl et al., 2012). Alternatively, it was possible that apologetic female faces would be rated as more apologetic than apologetic male faces because they match a prototype developed from extensive and repeated exposure to women apologizing for even minor transgressions (Schumann & Ross, 2010).

The interactions among the factors were also of interest. I examined whether participants would generate outgroup faces that would later be rated as more apologetic than the faces of the generator's ingroup. Intergroup research has consistently shown that people hold ingroup biases and tend to attribute negative attributes to outgroup relative to ingroup members (e.g., Dotsch et al., 2008; Hewstone et al., 2002). In the current context, a person who transgresses may be implicitly and/or explicitly assumed to possess negative attributes, which could result in biased depictions of transgressors. Finally, I predicted that the gender of the rater would influence ratings of target faces that differed by gender. Such a finding would be attributable to similar in- and out-group biases.

Method

Phase 1: Image generation

Participants and Procedure

An initial sample of 140 undergraduate students participated in the image generation

phase of Study 3. Three participants were removed because they were identified as having responded randomly (n = 1), having not completed the study (n = 1), and as choosing not to identify their gender identity. Therefore 137 undergraduates (M_{age} = 19.82, SD = 3.89, female = 70) were included in the final dataset. Participants identified as White (25%), South Asian (18%), Black (15%), Middle Eastern (14%), East Asian (10%), South East Asian (8%), Mixed (5%), Not Listed (3%), and Latin American (2%). Because image generation was divided into four conditions (men generating male faces, men generating female faces, women generating male faces, and women generating female faces), I aimed to recruit a total sample size of at least 120 participants based on norms within the reverse correlation literature (e.g., Bjornsdottir et al., 2019; Lick et al., 2013). The procedure was identical to the image generation phase in Study 2. For Study 3, however, participants were randomly assigned to one of two conditions (female or male base face). They arrived at a computer lab on campus, where they were directed to a condition-specific online survey. Participants provided consent, completed the 400 forced-choice trials in which they selected "Which face looks more apologetic?", were thanked and debriefed.

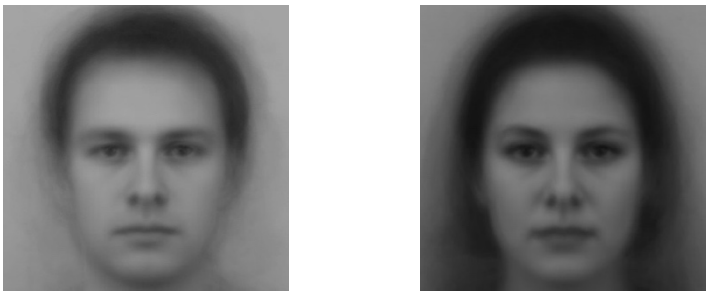
Stimuli

Reverse Correlation

The stimuli used in the current reverse correlation study were generated from White male and female faces. The male base face was identical to that used in Study 1. The female face was created on WebMorph (DeBruine, 2017) by averaging eight randomly selected female faces from the Karolinska Face Database (Lundqvist et al., 1998; see Figure 5). All images were grayscale, and visual distortion patterns were generated in R (R Core Team, 2020) using the rcicr package (Dotsch, 2017) and placed on top of the male and female base faces separately.

Figure 5

Male and Female Base Faces



Note. The male base face was identical to the one used in Study 1. All random visual noise patterns were placed on top of these images.

Conscientious responding

The Conscientious Responders Scale (CRS; Marjanovic et al., 2014) for this phase was identical to Phase 1 of Study 2.

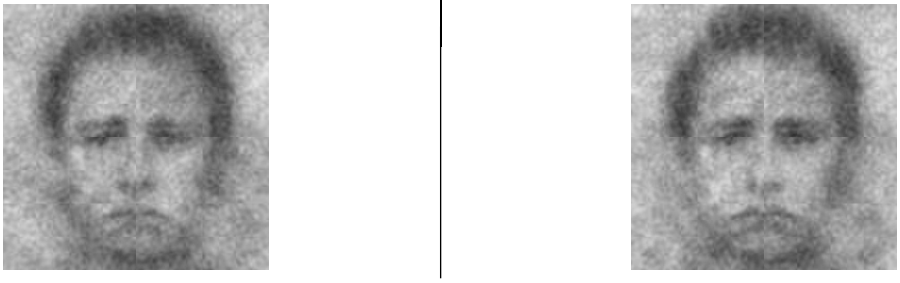
Results

The 'rcicr' package in R (Dotsch, 2017), was used to create an average face for each participant based on their responses. These individual averages were then combined with same-gender participant averages within each condition. This resulted in four classification images: a male face generated by men, a male face generated by women, a female face generated by men, and a female face generated by women (see Figure 6).

Figure 6

Male and Female Apologetic Faces Separated by Generator Gender

Male-Generated Apologetic Faces | Female-Generated Apologetic Faces



Male-Generated Apologetic Faces

Female-Generated Apologetic Faces

Method

Phase 2: Image Rating

Participants and Procedure

Two-hundred and ninety-five surveys were completed by undergraduate students for Phase 2 of Study 3. Participants were removed from the data because they did not consent (n = 1), did not finish (n = 31), failed the CRS (n = 60), completed the study more than once (n = 2), indicated that they did not identify with a binary definition of gender (n = 1), and because they did not provide their anonymous identification number, which meant I could not verify that their data represented a unique entry (n = 1). After exclusions, the total sample size consisted of 199 undergraduates (M_{age} = 21.62, SD = 5.19, female = 99), with participants identifying as South Asian (26%), White (25%), Middle Eastern (13%), Black (10%), East Asian (9.5%), Mixed (7.5%), South East Asian (5.5%), Latin American (2%), Not Listed (1%), and Polynesian (.5%), Based on the same exclusion criteria as Study 2, prior to data collection, a sample of 82 participants was set as the minimum sample size required to detect a medium effect of f^2 = .25 at 80% power. This calculation applies to the main effect of Rater Gender in the moderation analysis, as it required the largest sample size relative to all other analyses. A sensitivity power

analysis reveals that a sample of 199 was able to detect a minimum effect of $f = .16$ based on 80% power and a two-tailed alpha level of $p = .05$.

The procedure was identical to that of Study 2, with the exception that participants in the current study were asked to rate six faces: the female and male base faces, the female apologetic classification image generated in Phase 1 by female and male generators, and the male apologetic classification image generated in Phase 1 by female and male generators (see Figures 5 and 6). Each face was rated on all characteristics before moving on to the next face, and the order in which the faces were presented was randomized. Participants were sent a link to the study upon signing up and asked to complete it online on a desktop computer. Five adapted CRS questions (Marjanovic et al., 2014) similar to those used in Study 2 were used to assess attention, however the images associated with each question was updated with stimuli from Study 3. Once again, these questions were randomly selected and presented among the other trials at specific intervals. The online study consisted of an informed consent, image rating trials, and a debriefing.

Results

Ratings of Apology

To examine whether reverse correlation could be used to create apologetic faces from both a male and a female base face, I first compared ratings of apologetic and base images for male and female faces separately, collapsing across the gender of the generators. Consistent with my hypothesis, as well as with findings from Studies 1 and 2, both the male (*Mdn* = 72.50) and female (*Mdn* = 65.00) apologetic faces were significantly more apologetic than the male (*Mdn* = 23.00) and female (*Mdn* = 28.00) base faces, $p < .001$, $r_{effect\ size} = .79$, and $p < .001$, $r_{effect\ size} = .79$, respectively. As well, all faces were significantly different from the midpoint of the apology

rating scale. Specifically, the male (*Mdn* = 72.50, $p < .001$, $r_{effect\ size} = .65$) and female (*Mdn* = 72.50, $p < .001$, $r_{effect\ size} = .61$) apologetic faces were rated significantly higher than the midpoint of the scale. By contrast, the male (*Mdn* = 23.00, $p < .001$, $r_{effect\ size} = .67$) and female (*Mdn* = 28.00, $p < .001$, $r_{effect\ size} = .68$) base faces were rated as significantly less apologetic than the midpoint. Both standardized (*r*) and unstandardized (*M*) effect sizes suggest that base faces were perceived as very low in apology, and apologetic faces were seen as relatively high. The male and female apologetic faces ($p = .06$, $r_{effect\ size} = .19$), as well as the male and female base faces ($p > .99$, $r_{effect\ size} = .04$), did not differ in their ratings of how apologetic they appeared.

Apology-Related Characteristics

To test whether the same apology-related characteristics found in apologetic faces in my previous studies extended to stimuli from a distinct social category (i.e., gender), I compared apology-related characteristic ratings of male and female apologetic and base faces separately. Multiple paired-sample Wilcoxon signed-rank tests confirmed once again that apologetic faces were perceived as significantly more sad, regretful, remorseful, and submissive, as well as less dominant, than their respective base face, regardless of the gender of the target face being rated (all ps < .001, see Table 9). For the first time, ratings of trustworthy were found to differ significantly, with the female apologetic face (*Mdn* = 43.00) rated as *less* trustworthy than the base face (*Mdn* = 50.00), $p < .001$, $r_{effect\ size} = .36$. Consistent with the previous studies, there was no difference in trustworthy ratings for the male faces.

After ranking these characteristics for the male and female faces separately, sadness was again found to have the highest rating for both male (*Mdn* = 82.50) and female (*Mdn*= 79.50) faces, and this characteristic was rated significantly higher than the next highest ranked characteristic: remorse for male faces (*Mdn* = 72.00), $p < .001$, $r_{effect\ size} = .64$, regret for female

faces (*Mdn* = 67.00), $p < .001$, $r_{effect\ size} = .63$, see Table 9.

Table 9

Mean Characteristic Ratings and Comparisons Across Faces, Study 3

	Apologetic Male Face	Male Base Face	*p*	$r_{effect\ size}$
Characteristic	*M (SD)*	*M (SD)*		
Apologetic	67.56 (20.77)	28.49 (23.70)	< .001	.79
Sad	78.77 (17.48)	31.68 (24.35)	< .001	.85
Regret	68.46 (18.16)	30.36 (23.99)	< .001	.81
Remorse	68.34 (18.77)	28.54 (25.17)	< .001	.83
Submissive	57.00 (21.58)	36.75 (24.35)	< .001	.55
Trustworthy	39.87 (20.39)	38.93 (22.73)	> .99	.02
Dominant	31.27 (18.04)	57.86 (24.76)	< .001	.70
Male	77.48 (17.78)	86.82 (17.61)	< .001	.56
Masculine	66.37 (21.13)	77.22 (22.29)	< .001	.45
Feminine	25.44 (20.91)	20.33 (21.43)	< .01	.26
Female	22.68 (20.21)	14.83 (18.64)	< .001	.41
	Apologetic Female Face	Female Base Face	*p*	$r_{effect\ size}$
Characteristic	*M (SD)*	*M (SD)*		
Apologetic	63.66 (18.39)	29.25 (22.91)	< .001	.79
Sad	75.67 (18.73)	28.65 (24.31)	< .001	.84
Regret	64.93 (19.44)	26.46 (21.96)	< .001	.81
Remorse	64.36 (19.23)	28.65 (22.84)	< .001	.82
Submissive	57.97 (21.20)	39.57 (22.58)	< .001	.54
Trustworthy	40.44 (20.59)	50.13 (22.72)	<.001	.36
Dominant	28.02 (18.42)	55.60 (23.60)	< .001	.73
Male	19.19 (16.25)	12.24 (15.98)	< .001	.43
Masculine	23.32 (17.60)	16.94 (19.60)	< .001	.33
Feminine	75.51 (16.77)	84.92 (17.21)	< .001	.53
Female	79.53 (17.85)	89.10 (13.77)	< .001	.54

Note. Wilcoxon signed-rank tests were used for each comparison due to the non-normal distributions of each variable. All *p*-values within the apology-related characteristics category were Bonferroni-adjusted for 11 comparisons (i.e., comparisons of the apologetic and base faces for six characteristics, and comparisons across the highest-rated characteristics for the apologetic face). Characteristics in each grouping are listed from highest to lowest scores of the apologetic face.

See Appendix for ratings of male and female apologetic faces separated by Generator Gender.

The mean differences between ratings of the apologetic and base faces for both male and female targets are listed in Table 10. The rank order of the characteristics was generally consistent with that found in Studies 1 and 2, with sad, regretful, apologetic, and remorseful perceived highly for both target faces. Post hoc direct comparisons between the difference scores for sad and apologetic revealed that the change in perceived sadness was larger than the change in perceived apology from base to apologetic face, and this was true of the male face, *difference* = 137.50, $p < .001$, $r_{effect\ size} = .26$, and the female faces, *difference* = 147.50, $p < .001$, $r_{effect\ size} = .32$ (the critical difference after Bonferroni adjustments was 130.90).

Table 10

Mean Difference in Ratings Across Faces, Study 3

	Mean Difference (Apologetic Male Face – Male Base Face)	Mean Difference (Apologetic Female Face – Female Base Face)
Characteristic	*M (SD)*	*M (SD)*
Apologetic	39.07 (31.41)	34.41 (28.39)
Sad	47.09 (27.94)	45.11 (29.32)
Regret	39.80 (29.11)	38.47 (28.20)
Remorse	38.10 (29.03)	35.71 (28.48)
Dominant	-26.59 (28.44)	- 27.58 (28.00)
Submissive	20.25 (31.56)	18.41 (30.04)
Trustworthy	0.94 (27.18)	-9.69 (26.59)
Masculine	-10.85 (23.57)	5.38 (20.80)
Male	-9.34 (19.70)	6.96 (18.20)
Female	7.85 (21.51)	-9.57 (17.78)
Feminine	5.12 (22.34)	-9.41 (19.12)

Note: Within each grouping, characteristics are listed in rank order from highest to lowest absolute value of the mean difference scores.

Multiple linear regression analysis was again used to investigate the relative predictive

ability of apology-related characteristics. Analyses were conducted on male and female target faces separately to determine whether similar characteristics are predictive of apology for these different social targets. Regardless of target gender, both models were significant and predicted similar amounts of variance overall (male target: $F(4, 194) = 91.76$, $p < .001$, $\Delta R^2 = .65$; female target: $F(4, 194) = 77.74$, $p < .001$, $\Delta R^2 = .62$). Remorse and regret accounted for the largest variance for male (remorse: $\beta = 0.39$, $p < .001$; regret: $\beta = 0.27$, $p < .001$) and female (remorse: $\beta = 0.38$, $p < .001$; regret: $\beta = 0.39$, $p < .001$) faces. For apologetic male faces, sadness and submissiveness were also significant predictors of apology ($\beta = 0.16$, $p = .01$; $\beta = 0.14$, $p < .01$, respectively). This was not the case for apologetic female faces, however. Unlike the model for male faces, trustworthiness was found to significantly predict apology for apologetic female faces ($\beta = 0.11$, $p = .02$), but both sadness and submissiveness were not. Results from both regression analyses, as well as the correlations among variables, can be found in Tables 11 and 12. For each of the apologetic male and female faces, models containing all predictors did not differ significantly from the reduced models with fewer predictors. Despite this, the reduced models may be considered to be superior because they are more parsimonious (Vandekerckhove et al., 2015).

Table 11

Apology-Related Characteristic Correlations, Study 3

Male Face

Characteristic	1	2	3	4	5	6	7
1. Apologetic	--						
2. Sad	.64***	--					
3. Regret	.72***	.71***	--				
4. Remorse	.74***	.63***	.71***	--			
5. Submissive	.49***	.32***	.45***	.46***	--		
6. Trustworthy	.19**	.13	.10	.19**	.21**	--	
7. Dominant	-.29***	-.24***	-.34***	-.27***	-.24***	.18**	--

Female Face

Characteristic	1	2	3	4	5	6	7
1. Apologetic	--						
2. Sad	.58***	--					
3. Regret	.72***	.68***	--				
4. Remorse	.71***	.63***	.72***	--			
5. Submissive	.40***	.28***	.40***	.35***	--		
6. Trustworthy	.30***	.22**	.25***	.21**	.18**	--	
7. Dominant	-.23**	-.33***	-.29***	-.24***	-.23**	.08	--

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

Table 12

Multiple Regression Results for Apology, Study 3

Characteristic	B	95% CI for B		SE B	β	ΔR^2	VIF[a]
		LL	*UL*				
Male Face						.65	
Base Model							
Constant	-4.55	-14.73	5.64	5.16			
Remorse	0.44***	0.29	0.58	0.07	0.38***		
Regret	0.29***	0.14	0.44	0.08	0.36***		
Sadness	0.18*	0.04	0.33	0.07	0.16*		
Submissive	0.12*	0.03	0.22	0.05	0.12*		
Trustworthy	0.05	-0.04	0.15	0.05	0.05		
Dominant	-0.04	-0.15	0.06	0.05	-0.04		
Male Face						.65	
Reduced model							
Constant	-5.85	-14.24	2.54	4.25			
Remorse	0.45***	0.31	0.59	.07	0.39***		2.26
Regret	0.29***	0.14	0.45	.08	0.27 ***		2.71
Sadness	0.19*	0.04	0.34	.07	0.16*		2.19
Submissive	0.14**	0.04	0.23	.05	0.14**		1.32
Female Face							
Base Model						.61	
Constant	6.97	-2.12	16.06	4.61			
Remorse	0.35***	0.22	0.48	.06	0.37***		
Regret	0.33***	0.20	0.47	.07	0.35***		
Trustworthy	0.10*	0.02	0.18	.04	0.11*		
Submissive	0.08	-0.01	0.16	.04	0.09		
Sadness	0.05	-0.07	0.18	.06	0.06		
Dominant	0.01	-0.10	0.09	.05	0.01		

Characteristic	B	95% CI for B		SE B	β	ΔR^2	VIF[a]
		LL	*UL*				
Female Face							
Reduced model						.61	
Constant	8.22	1.59	14.85	3.36			
Remorse	0.37***	0.25	0.49	0.06	0.38***		2.01
Regret	0.36***	0.24	0.48	0.06	0.38***		2.21
Trustworthy	0.10*	0.02	0.18	0.04	0.11*		1.08
Submissive	0.08	-0.01	0.16	0.04	0.09		1.22

Note. CI = confidence interval; VIF = Variance Inflation Factor

[a] VIF > 5 indicates critical multicollinearity, VIF > 1 indicates moderate multicollinearity

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

Gender and Gender-Related Characteristics

In order to address my third research question, a 2 (Image Type: Apologetic, Base) x 2 (Perceived Gender: Male, Female) within-subjects Friedman's ANOVA was conducted for male and female target faces separately. Significant differences were found between the apologetic and base faces for male, χ^2 (3) = 446.42, $p < .001$, and female, χ^2 (3) = 474.11, $p < .001$, targets. As was found in Study 1, the apologetic male face was rated as less male (*Mdn* = 80.50) and more female (*Mdn* = 17.50) than the base face (Mdn_{male} = 92.00, Mdn_{female} = 10.00), *difference* = 120.50, $p < .001$, $r_{effect\ size}$ = .56 and *difference* = 82.50, $p < .001$, $r_{effect\ size}$ = .41, respectively. The apologetic female face, by contrast, was rated as more male (*Mdn* = 16.50) and less female (*Mdn* = 82.50) than the base face (Mdn_{male} = 7.00, Mdn_{female} = 94.00), *difference* = 91.00, $p < .01$, $r_{effect\ size}$ = .43 and *difference* = 98.00, $p < .001$, $r_{effect\ size}$ = .43, respectively. Comparisons within each image type confirmed that the male face was rated as looking more male than female for the apologetic (*difference* = 263.00, $p < .001$, $r_{effect\ size}$ = .82) and base face (*difference* = 466.00, $p <$

.001, $r_{effect\ size}$ = .86). The opposite pattern was found for the female face (*difference* = 284.50, p < .001, $r_{effect\ size}$ = .85; *difference* = 473.50, p < .001, $r_{effect\ size}$ = .87).

Similar 2 (Image Type: Apologetic, Base) x 2 (Gender-Related Characteristics: Masculine, Feminine) analyses were conducted to investigate gender-related characteristics. Once again, differences were found between the apologetic and base faces for male, χ^2 (3) = 338.59, p < .001, and female faces, χ^2 (3) = 440.52, p < .001. The male apologetic face was rated as less masculine (*Mdn* = 70.00) than the male base face (*Mdn* = 80.00), *difference* = 95.00, p < .001, $r_{effect\ size}$ = .45, where p was Bonferroni-adjusted to account for four comparisons. The male apologetic face was also rated as more feminine (*Mdn* = 21.50) than the base face (*Mdn* = 14.00), *difference* = 49.00, however this fell below the critical difference of 67.95 and was therefore not statistically significant. The reverse pattern was found for the female face. The apologetic female face received higher ratings of masculinity (*Mdn* = 20.00*)* than the base face (*Mdn* = 10.00), *difference* = 65.50, although this did not exceed the critical difference of 67.95. The female face was also rated lower on femininity (*Mdn* = 76.00) than the base (*Mdn* = 90.00), *difference* = 114.50, p < .001, $r_{effect\ size}$ = .53.

The ratings of masculinity and femininity provided additional evidence of the strong association between gender and gender characteristics. That is, the apologetic male faces were rated as significantly more masculine relative to feminine ($difference_{apologetic}$ = 251.00, p < .001, $r_{effect\ size}$ = .75; $difference_{base\ face}$ = 395.00, p < .001, $r_{effect\ size}$ = .82), and the female apologetic faces were rated as significantly more feminine than masculine ($difference_{apologetic}$ = 277.50, p < .001, $r_{effect\ size}$ = .86; $difference_{base\ face}$ =457.50, p < .001, $r_{effect\ size}$ = .86).

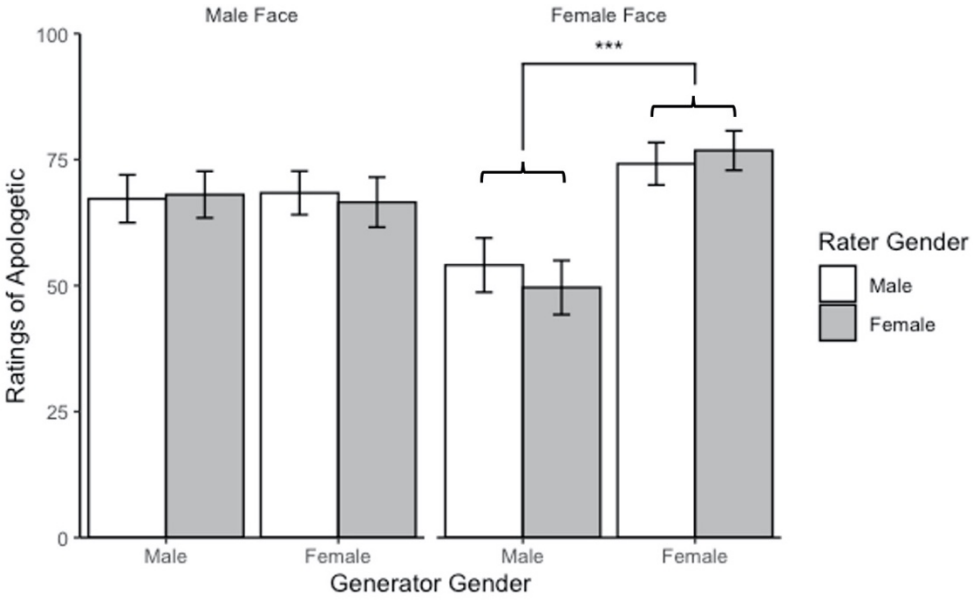
Moderation of Apology Ratings

Finally, I examined whether and how gender affected ratings of apology. An ANOVA-

type statistic test (Brunner et al., 1997; Noguchi et al., 2012) was used due to the non-normal distributions and the mixed design. A 2 (Rater Gender: Male, Female) x 2 (Generator Gender: Male, Female) x 2 (Target Gender: Male, Female) ANOVA-type statistic test revealed a significant main effect of Generator Gender ($p < .001$) and a marginally significant main effect of Target Gender ($p = .05$) after conducting Bonferroni adjustments. These main effects were qualified by a significant Generator Gender by Target Gender interaction, $\chi^2(3) = 2208.9$, $p < .001$ (see Figure 7). Follow up analyses revealed that men generated an apologetic face from the female base face (*Mdn* = 59.00) that was subsequently rated as *less* apologetic than any of the other apologetic faces (male-generated male face: *Mdn* = 71.00, *difference* = 146. 50, $p < .001$, $r_{effect\ size} = .10$; female-generated male face: *Mdn* = 71.00, *difference* = 163.50, $p < .001$, $r_{effect\ size} = .51$; female-generated female face: *Mdn* = 79.00, *difference* = 252.00, $p < .001$, $r_{effect\ size} = .38$). By contrast, women generated an apologetic face from the female base face that was subsequently rated as *more* apologetic than any of the other apologetic faces (male-generated male face: *difference* = 105.50, $p < .001$, $r_{effect\ size} = .31$; female-generated female face: *difference* = 88.50, $p < .001$, $r_{effect\ size} = .10$). The critical difference for all direct comparisons was 67.95 and all *p* values were Bonferroni adjusted for nine comparisons. No other effects were significant, *p*s > .87.

Figure 7

Generator Gender by Target Gender Interaction for Male and Female Target Faces



Note. Mean values have been used to represent ratings of apologetic.

***$p < .001$.

Discussion

As in Studies 1 and 2, the apologetic faces generated in the current study were rated as significantly more apologetic than the base faces from which they were created. This provides additional evidence that people have a mental representation of an apologetic face and that reverse correlation can be used to create a visual estimate of this representation. In addition, both the male and female apologetic faces showed comparable patterns in terms of the ranking of the apology-related characteristics that were perceived. Specifically, participants rated both apologetic faces as highly sad, regretful, and remorseful. Although this pattern of rankings was similar to Studies 1 and 2, Study 3 differed in terms of the ability of the apology-related characteristics to predict apology. For the first time, ratings for trustworthiness were found to differ significantly across the base and apologetic faces, however this was only true for female

faces. The female apologetic face was rated as *less* trustworthy than the base face, whereas this difference was not significant for male faces. Trustworthy ratings for female, but not male, faces also significantly predicted apology ratings, and this replicated what was found in Study 2 for the gender-morphed face. As ratings of trustworthy, along with regretful and remorseful, increased, perceived apology increased as well. Ratings of apology for apologetic male faces were also predicted by higher levels of regretful and remorseful, however sad and submissive are also significant predictors. The results of these multiple regression analyses suggest that, although overall apology ratings may be similar for male and female faces, it is possible that the criteria used to predict these judgments may depend on who is apologizing.

Unexpectedly, apologetic faces were not systematically seen as more male or more female. Instead, in Study 3, the male apologetic face was rated as less male and masculine and more female and feminine relative to the male base face, whereas the opposite pattern emerged for the female faces, with the apologetic face being rated as less female and feminine and more male and masculine than the female base face. It is unclear why this is the case; however, these findings are consistent with the possibility that apologizing leads one to be rated as less gender-prototypical.

Finally, although gender did moderate ratings of apology, the effect was not consistent with what was originally predicted. The gender of the generator was the only significant main effect to emerge, and this was qualified by a generator gender by target gender interaction. Having the same gender identity as the target did not influence ratings of apology, with men and women identifying similar levels of apology for each face. This finding does not replicate previous research demonstrating that women are able to differentiate subtle or complex emotions to a greater extent than men (Hall & Matsumoto, 2004), and would therefore benefit from

additional investigation. Despite similarities in how men and women *perceive* apologetic faces, the findings do seem to support the theory that women are particularly *expressive* (Adams et al., 2015). This can be seen in the interaction between generator and target gender, with women generating female faces that were later seen as more apologetic compared to all other faces. However, this expressiveness may be limited to interactions with in-group members. Men generated female faces that were rated as significantly less apologetic than all other faces, suggesting that perhaps women express apologies differently when they are directed at men relative to women, and experiences interacting with apologetic women then shape men's and women's mental representations differently.

General Discussion

In the present research I set out to contribute to the apology literature by examining whether people have a mental representation of an apologetic face that could be estimated using reverse correlation. Across three studies, apologetic faces were created from male (Studies 1 and 3), gender-morphed (Study 2), and female (Study 3) base faces. As predicted, in each study, the apologetic faces were subsequently rated by an independent group of participants to be significantly more apologetic than their respective base faces. In addition, the median rating of apology was consistently higher than the midpoint of the scale for all apologetic faces, and the base faces were consistently lower than the midpoint, regardless of the target gender. These findings provide evidence that the reverse correlation procedure was successful at creating visual templates and that these templates were perceived as consistent with people's mental representation of an apologetic face.

A second goal of this research was to determine what apology-related characteristics were most prominent in this visual template. One of the most interesting and important findings

to emerge from this research was that, across each of the studies, sadness was consistently given the highest ratings for each of the apologetic faces. In fact, in each study, sadness was perceived to a significantly greater extent than apology itself, despite the fact that these images were created to be visual templates of apologetic – and not sad – faces. Taken together, the findings suggest that expressing sadness may be the key to being perceived as apologetic, however it is unclear why this may be the case. One possibility is that basic emotions may be more easily perceived relative to secondary or compound emotions. Although speculative, it may also be the case that sadness is the *only* facial expression we associate with nonverbal apologies, and all other components are expected to be communicated verbally. Given the robustness of the finding that sadness is rated highly across each of the images, the results suggest that sadness is important regardless of the gender of the transgressor or the gender of the victim, as men and women are perceived similarly. The one caveat to this is in the ability for ratings of sadness to predict ratings of apology. For most of the apologetic images, sadness emerged as a significant predictor of ratings of apology; however, for the apologetic face created from the female base face, this was not the case, a finding that is worthy of additional investigation. Nonetheless, it was highly correlated with ratings of apologetic (r = .58) suggesting that the more participants saw sadness, the more apologetic they perceived this female target face to be. Overall, this is consistent with the possibility that displaying sadness is extremely important when apologizing.

Another key apology-related characteristic that I examined in the present research was trustworthiness. In the verbal literature, promises not to reoffend are mentioned as important aspects of a successful apology (Kirchhoff et al., 2012; Lewicki et al., 2016; Schlenker & Darby, 1981; Struthers et al., 2008), and this might be conveyed by appearing trustworthy. However, consistent with my previous findings (George et al., 2020), ratings of trustworthiness did not

appear to vary consistently across apologetic and base faces. The female, but not male, face in Study 3 was the only instance in which ratings of trustworthy emerged as significantly different from the base face, and surprisingly, the female apologetic face was rated as *less* trustworthy than the base face. An examination of the means suggests that this was because the female base face was perceived as more trustworthy than other faces and not because the female apologetic face (whose mean rating was in line with ratings of trustworthy for male faces) was perceived to be less trustworthy than other faces. Although speculative, one possible reason for this finding could be that women are generally perceived as more trustworthy than men, however seeing an apologetic female face may bring to mind the frequency with which women apologize (Schumann & Ross, 2010). This, in turn, may lead to beliefs that women transgress – and therefore betray trust – more than men. If this is the case, being sensitive to situations that may harm a relationship and offering an apology in atonement may ironically be a detriment to building and maintaining trust. Discrepancies in trustworthiness ratings may also result from how participants were thinking about the act of apologizing. An individual thinking about how *they* would signal trustworthiness following a transgression may perceive an apologetic face as more trustworthy than someone who is imagining they are being apologized to by someone who has recently betrayed trust. It is also important to note that trustworthy emerged as a significant positive predictor of apologetic for the female and gender-morphed apologetic faces. Although apologetic faces were generally rated relatively low on trustworthiness, these ratings predicted perceptions of apology, and this was true when other apology-related characteristics were included in the model.

Across all three studies, faces also received comparable ratings of apologetic, regretful, and remorseful. Both Pearson correlations and regression analyses confirmed that these

characteristics were consistently related. They share similar levels of abstractness when it comes to nonverbal expression and function as social (Hareli & Eisikovits, 2006) or secondary emotions – which differ from primary emotions such as sadness (Wohl et al., 2012). It is possible, therefore, that the nonverbal characteristics that were labelled apologetic, regretful, and remorseful in the current research represent a single latent construct that, in addition to sad, accurately predicts an expression associated with apologizing. The word "apology" might therefore be semantically different when describing a facial expression versus a behaviour. Follow up factor analyses are needed to provide further information about the relationships between variables and any latent constructs they might represent. Nonetheless, given that regret and remorse are feelings that are often associated with a transgression, the high ratings, as well as the consistent relationship that emerged between ratings of these characteristics and ratings of apology, provide additional evidence that these faces provided a good visualization of people's mental representation of apologetic faces.

Submissiveness, as well as a lack of dominance, were also examined as potentially key characteristics of an apologetic face. Submissive was not rated particularly high across each of the studies, with mean ratings ranging from 54 to 60 on a 100-point scale that ranged from 0 (*not at all*) to 100 (*very submissive*). However, submissive did emerge as a significant predictor of apology for all but the female face, albeit typically accounting for less variance than sad, regretful, and remorseful. This variable was included in my original research (George et al., 2020) as an outcome variable in its own right, however given the association between submissive and feminine (and by extension, female), I anticipated that it would remain a highly rated characteristic when judged in male, gender-morphed, and particularly female faces. In both Studies 1 and 3, apologetic male faces were judged to be more submissive and less masculine

relative to their base faces, supporting the predicted association between constructs and suggesting that apologies may be thought of as submissive and/or feminine behaviours. However, submissive was not a significant predictor of apology for female faces, and in fact, the apologetic female face was rated as less feminine than the female base face. One possible explanation for these findings is that the characteristics in question operate differently when predicting apology for male and female target faces, such that submissive is an important component for male apologizers only. Alternatively, the association between gender-characteristics and stereotypically gendered apology-related characteristics may simply be weaker than originally expected. Future research aimed at parsing out the relationship between the constructs in an apology context as well as in other domains would prove beneficial. For example, do perceptions of an individual as submissive automatically lead to perceptions of femininity (and/or vice versa), and are such perception stable across scenarios? As well, measuring the degree to which participants rate themselves on gender-related characteristics and submissiveness may provide additional data points that could be used to scale ratings of target faces to better inform this relationship (Wood & Eagly, 2015). Such self-report ratings could be included in mediation and moderation models to examine how one's perception of themselves might shape how they perceive others.

A third goal of the current research was to examine whether apologetic faces would be rated differently from the base face on gender and gender-related characteristics. Surprisingly, the findings seem to consistently suggest that apologetic faces are rated as less gender-prototypic than base faces, and it is unclear why this was the case. Although this could be interpreted as detrimental to some individuals – for example, those who wish to express and maintain a specific gender identity or appear to possess characteristics stereotypically associated with masculinity or

femininity – this does not seem to reflect a belief that apologizing is a gendered behaviour. Apologetic male faces were rated as less masculine and more feminine than male base faces, and apologetic female faces were rated as less feminine and more masculine than their base faces. Importantly, however, the pattern of perceived gender-related characteristics never resulted in a complete shift toward male faces appearing more feminine than masculine or female faces appearing more masculine than feminine. Apologies may lead to temporary changes in how masculine or feminine one is perceived, but they do not appear to threaten one's identity to the point where an individual that is more masculine relative to feminine is seen as more feminine than masculine (or vice versa) while apologizing.

Finally, the gender moderation analyses in Studies 2 and 3 provided only partial support for my hypotheses. The mental representations of men and women differed for female faces such that women generated a visual template of a female apologetic face that was rated as more apologetic than that of men. This is consistent with research showing that women are more expressive than men (Adams et al., 2015; Fischer & LaFrance, 2015), however it is possible that such expressiveness is specific to interactions with other women. Female faces generated by men were rated as significantly less apologetic than those generated by women. Given the findings in the current research that men and women perceive similar levels of characteristics from faces, it seems possible that women express differing levels of nonverbal apologies when addressing men versus women.

Even though women have a visual representation of an apologetic female face that is more expressive than men, men and women judged apologetic faces similarly regardless of who the face belongs to. These findings have widespread implications for interpersonal relationships of all kinds. Individuals involved in romantic relationships, for example, often encounter

instances where an apology is necessary. Le and colleagues (2020) recently found that, even outside of an apology context, accurate perception of a romantic partner's appeasement expressions (e.g., embarrassment, shame, submissiveness, guilt) were linked to greater relationship quality. The appeasement expressions map onto our understanding of verbal apologies, and the findings from my current research complement this by demonstrating that perceivers can identify such characteristics on novel faces. Within the apology domain, it seems possible that one's ability to identify apology-related characteristics can lead to higher quality romantic relationships.

Limitations and Future Directions

As a preliminary investigation into apologetic faces, this work provided visual templates of what one might expect a transgressor to express during an apology. Because it was unclear as to whether apologies even have associated facial expressions, an initial goal of this work was simply to produce a consistent visual template. The limitations of the current studies are therefore due, in part, to the somewhat exploratory nature of the research goals. For example, previous reverse correlation research has compared mental representations of social groups along specific stereotyped dimensions associated with said groups (e.g., gender and emotion, Brooks, Stolier, & Freeman, 2018; social class and race, Lei & Bodenhausen, 2017). Because no such comparison group was explicitly outlined in the current dissertation, a comparison was instead made to the neutral base image to demonstrate that apology could be perceived in an apologetic face above and beyond anything perceived in the base face. The two images differ in their resolution and clarity, however, potentially obscuring any meaningfully significant results and compromising the validity of the results.

Several approaches could be used to address such a limitation in future work to help

mitigate issues of image quality. For example, Walker and Keller have developed a reverse correlation technique that uses face-space modelling to generate stimuli by applying randomly varying vectors onto base images (2019). Manipulating image vectors does not result in visual noise, and stimuli generated using this method therefore match base faces in terms of picture clarity. Alternatively, applying the *infoVal* metric (Brinkman et al., 2019) could help to validate the current findings by eliminating the need to compare apologetic faces to a base. The *infoVal* metric is produced by randomly resampling the data multiple times and is intended to identify aspects of an image that are meaningful visual signals relative to false positives. The use of resampling in this method allows for comparisons across different groupings of real data and eliminates the need to compare images of different resolution or clarity.

An additional benefit of using the above-mentioned methods is that they are likely to decrease potential visual differences that might be perceived among the classification images themselves. If the image clarity appears subjectively different to individual raters, this could have consequences for subsequent ratings. For example, the male-generated female face in Study 3 may appear to some perceivers to be less clear or more pixilated than the other apologetic faces, and this could potentially explain why it received lower apology ratings in comparison. This is an empirical question that could be addressed by the alternative approaches such as *infoVal* analyses and face-space modelling in future research. Analyzing data using the *infoVal* metric has also been recommended as a way to reduce the inflated Type I error rates that have recently been tied to the reverse correlation procedure (Cone et al., 2020).

Researchers have also suggested that high error rates can be partially controlled by using participant-level apologetic faces (i.e., those faces generated by averaging the noise patterns of each individual generator's response in Phase 1), rather than the group-level average, as was

done in the present studies. Using individual-level averaged faces would address this limitation, however it would require a larger number of trials to generate each apologetic face. As well, having raters view individual faces from the generation phase and provide ratings of multiple characteristics for each would require a very large sample size. The feasibility of such a design becomes even more unwieldy if one opts for a between-subjects design. Such a design may limit participant fatigue and potential order effects associated with the presentation of stimuli, however it would again require a larger sample size. Such concerns regarding the feasibility of using participant-level faces may explain why reverse correlation researchers have yet to adopt this method of analysis broadly and/or why examples of this approach have yet to be published in the major journals within the field. Finally, another option would be to have participants generate a neutral face from the base face that might appear to have a more similar visual quality relative to the apologetic face.

One additional limitation to the current research is that participants were constrained to a specific list of the characteristics that they were asked to rate. Although the characteristics used in the present research were derived from a larger list used in my previous work (George et al., 2020), it is interesting to consider whether similar results would emerge if a larger and/or different pool of characteristics was used. Conducting more qualitative, data-driven research in which participants are asked to describe characteristics that they would anticipate in an apologetic face, or to generate lists of typical characteristics rather than limiting them to a pre-selected list of attributes, may be one way to address this limitation.

Another potential limitation of the present research, and an avenue for future research, is that there was no context provided during either of the generation or rating phases. As such, it is not clear whether people hold different mental representations of an apologetic face in the

context of a major transgression (e.g., infidelity, act of violence in a moment of anger) relative to a minor transgression (e.g., an accidental bump, or eating the last piece of cake). Similarly, it is unclear whether participants thought of any specific transgression either when selecting the more apologetic images across multiple trials in Phase 1 or when rating the apologetic images in Phase 2. As Schlenker and Darby (1981) argue, 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 behaviour may 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). Future work that provides more information to raters will help to speak to this issue and will provide new insights into whether the nonverbal cues that convey apology are consistent across a variety of contexts. An interesting future direction might therefore be to provide participants with a vignette describing a transgression and manipulate the severity of transgression between conditions at Phase 1 and/or Phase 2 of the reverse correlation procedure. Alternatively, context could be established by asking participants to recall major or minor transgressions 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.

The current findings suggest that gender may play a role in how one thinks about and

perceives apologetic faces. This was particularly evident in apologetic ratings for female faces, which differed based on the gender of the generator, as well as the gender of the rater. Ratings for the male faces were much more consistent across generator and participant gender. Participants were limited to judging faces that varied systematically across gender, however other social categories experience stereotypes that may interfere with the apology process as well. For example, when Brown-Iannuzzi and colleagues asked participants to generate a visual representation of a welfare recipient, the resulting face was rated as Black, female, unlikable, incompetent and lazy relative to the generated image of non-welfare recipients (Brown-Iannuzzi et al., 2017). If Black women are associated with this idea of a "welfare queen," it seems reasonable to predict that their apologies may be perceived as insincere and therefore less apologetic than the apologies of other social group members. Extending the present research to examine groups that face different stereotypes will provide a greater understanding of the prejudices that may be biasing how we repair and maintain personal relationships. Accounting for participant race may also help elucidate any biases present in the apology process. Although all three of the current studies involved racially diverse samples, no single group was large enough to justify analyzing the data based on participant race. Expanding participant recruitment may lead to samples with large enough racial subgroups to allow for a preliminary investigation into the role of generator and rater race. In addition to using only White faces, all of the 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. It is therefore unclear whether similar results would be obtained if base faces instead consisted of single naturalistic

photographs of a male or female face.

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 transgressor to demonstrate empirically tested characteristics while offering apologies, whether or not they are sincere. Similarly, however, it may be possible – and even necessary – to also train recipients of apologies to accurately detect apologetic characteristics and judge 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. Of course, such interventions make several assumptions that have yet to be tested. 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. One context in which answers to these questions would be highly relevant is in the criminal justice system. If apologies and their actual and perceived sincerity influenced forgiveness, this could have important implications in sentencing.

Conclusion

The need for social connection is powerful, but relationships are rarely without conflict. Apologies allow us to maintain our relationships and fulfill the need to belong following transgressions. Researchers have identified a number of key components of an apology that most

effectively repair relationships. In the present research I set out to contribute to the apology literature by determining if apologies could be expressed by way of facial expressions, and what that would look like. The findings of the current research suggest that people have a mental representation of an apologetic face. Apologetic visual templates were consistently rated as appearing sad, suggesting that this is an important feature in the nonverbal expression of apology, and that this is true regardless of whether the apologetic face is male, female, or gender-morphed. Although some gender differences emerged, the results suggest that people's mental representation of an apologetic face may be a great equalizer with regard to gender; male and female apologetic faces created by and evaluated by men and women seemed to have a similar profile. These findings provide some initial insight into people's mental representation of an apologetic face. In order to further understand the generalizability of these effects, it will be important to conduct additional research to examine whether similar effects emerge across contexts, and for targets from different social groups, such as race and socioeconomic status, to ensure that – when it comes to repairing relationships – everyone is on the same page.

References

Adams Jr, R. B., 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 *Understanding Facial Expressions in Communication* (pp. 1-18). Springer, New Delhi. https://doi.org/10.1007/978-81-322-1934-7_1

Bailey, A. H., LaFrance, M., & Dovidio, J. F. (2020). Implicit androcentrism: Men are human, women are gendered. *Journal of Experimental Social Psychology*, *89*, 103980. https://doi.org/10.1016/j.jesp.2020.103980

Bippus, A. M., & Young, S. L. (2019). How to say "I'm sorry:" Ideal apology elements for common interpersonal transgressions. *Western Journal of Communication*, 1-15. 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

Bjornsdottir, R. T., Yeretsian, S., West, G. L., & Rule, N. O. (2019). When words become borders: Ingroup favoritism in perceptions and mental representations of Anglo-Canadian and Franco-Canadian faces. Group Processes & Intergroup Relations. https://doi.org/10.1177/1368430218789495

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in Social Cognition* (Vol. 1, pp. 1–36). Psychology Press.

Brinkman, L., Goffin, S., van de Schoot, R., van Haren, N. E., Dotsch, R., & Aarts, H. (2019). Quantifying the informational value of classification images. *Behavior Research Methods*, *51*(5), 2059-2073. https://doi.org/10.3758/s13428-019-01232-2

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

Brunner, E., Dette, H., & Munk, A. (1997). Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association*, *92*(440), 1494-1502. https://doi.org/10.1080/01621459.1997.10473671

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

Campbell, J. I., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, 44(4), 1255-1265. https://doi.org/10.3758/s13428-012-0186-0

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: sociocultural and biological considerations. *Psychological bulletin*, *135*(2), 218. https://doi.org/10.1037/a0014412

Cloutier, J., Freeman, J. B., & Ambady, N. (2014). Investigating the early stages of person perception: The asymmetry of social categorization by sex vs. age. *PloS one*, *9*(1), e84677. https://doi.org/10.1371/journal.pone.0084677

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

DeBruine, L. (2017). Webmorph (Version v0.0.0.9001). *Zenodo*. doi:10.5281/zenodo.1073696

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

De Lemus, S., Spears, R., & Moya, M. (2012). The power of a smile to move you: Complementary submissiveness in women's posture as a function of gender salience and facial expression. *Personality and Social Psychology Bulletin*, *38*(11), 1480-1494. https://doi.org/10.1177/0146167212454178

Deng, Y., Chang, L., Yang, M., Huo, M., & Zhou, R. (2016). Gender differences in emotional response: Inconsistency between experience and expressivity. *PloS one*, *11*(6). https://doi.org/10.1371/journal.pone.0158666

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

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

Dunham, Y., Baron, A. S., & Banaji, M. R. (2016). The development of implicit gender attitudes. *Developmental Science*, *19*(5), 781-789. https://doi.org/10.1111/desc.12321

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of personality and social psychology*, *46*(4), 735. https://doi.org/10.1037/0022-3514.46.4.735

Eisinga, R., Heskes, T., Pelzer, B., & Te Grotenhuis, M. (2017). Exact p-values for pairwise comparison of Friedman rank sums, with application to comparing classifiers. *BMC Bioinformatics*, *18*(1), 68. https://doi.org/10.1186/s12859-017-1486-2

Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological Science, 3*(1), 34–38. https://doi.org/10.1111/j.1467-9280.1992.tb00253.x

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/BF03193146

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

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage Publications.

Finnegan, E., Oakhill, J., & Garnham, A. (2015). Counter-stereotypical pictures as a strategy for

overcoming spontaneous gender stereotypes. Frontiers in Psychology, 6, 1291. https://doi.org/10.3389/fpsyg.2015.01291

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

Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: Social status cues shape race perception. *PLoS One, 6*(9), e25107. https://doi.org/10.1371/journal.pone.0025107

George, M., Guilfoyle, J. R., van Monsjou, E., & Struthers, C. W. (2020). *Facial indicators of nonverbal apology and how they are perceived*. Manuscript in preparation.

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

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, *4*(2), 201. https://doi.org/10.1037/1528-3542.4.2.201

Hamilton, M. C. (1991). Masculine bias in the attribution of personhood: People= male, male= people. *Psychology of Women Quarterly*, *15*(3), 393-402. https://doi.org/10.1111/j.1471-6402.1991.tb00415.x

Hareli, S., & Eisikovits, Z. (2006). The role of communicating social emotions accompanying apologies in forgiveness. *Motivation and Emotion, 30*(3), 189-197. https://doi.org/10.1007/s11031-006-9025-x

Hegarty, P., & Buechel, C. (2006). Androcentric reporting of gender differences in APA journals: 1965–2004. *Review of General Psychology*, *10*(4), 377-389. https://doi.org/10.1037/1089-2680.10.4.377

Hegarty, P., & Pratto, F. (2004). The differences that norms make: Empiricism, social constructionism, and the interpretation of group differences. *Sex Roles*, *50*(7-8), 445-453. https://doi.org/10.1023/B:SERS.0000023065.56633.cb

Hehman, E., Stolier, R. M., Freeman, J. B., Flake, J. K., & Xie, S. Y. (2019). Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Social and Personality Psychology Compass*, *13*(2), e12431. https://doi.org/10.1111/spc3.12431

Hehman, E., Sutherland, C. A., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, *113*(4), 513. https://doi.org/10.1037/pspa0000090

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*(1), 575-604. https://doi.org/10.1146/annurev.psych.53.100901.135109

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. J*ournal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000208

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*, *74*(2), 171. http://dx.doi.org/10.1037/amp0000307

Karremans, J. C., Dotsch, R., & Corneille, O. (2011). Romantic relationship status biases memory of faces of attractive opposite-sex others: Evidence from a reverse-correlation paradigm. *Cognition*, *121*(3), 422-426. https://doi.org/10.1016/j.cognition.2011.07.008

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

Kunst, J. R., Onyeador, I. N., & Dovidio, J. F. (2019). *Knowledge about individuals' interracial friendships is systematically associated with mental representations of race, traits, and group solidarity*. PsyArXiv. https://psyarxiv.com/hgwr3/

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, *20*(3), 165-182. https://doi.org/10.1007/BF02281954

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

Lassonde, K. A., & O'Brien, E. J. (2013). Occupational stereotypes: activation of male bias in a gender-neutral world. *Journal of Applied Social Psychology*, *43*(2), 387-396. https://doi.org/10.1111/j.1559-1816.2013.01008.x

Lazare, Aaron (2004). *On apology*. New York: Oxford University Press.

Le, B. M., Côté, S., Stellar, J., & Impett, E. A. (2020). The distinct effects of empathic accuracy for a romantic partner's appeasement and dominance emotions. *Psychological Science*, *31*(6), 607-622. https://doi.org/10.1177/0956797620904975

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

Lewicki, R. J., Polin, B., & Lount Jr, R. B. (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

Lick, D. J., Carpinella, C. M., Preciado, M. A., Spunt, R. P., & Johnson,K. L. (2013). Reverse-

correlating mental representations of sex-typedbodies: The effect of number of trials on image quality. *Frontiers in Perception Science, 4*, Article 476. https://doi.org/10.3389/fpsyg.2013.00476

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. https://doi.org/10.1037/t27732-000

MacLachlan, A. (2013). Gender and Public Apology. *Transitional Justice Review*, *1*(2), 6. https://doi.org/10.5206/tjr.2013.1.2.6

Marjanoiic, 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). https://doi.org/10.1177/2158244014545964

Marsh, A. A., Elfenbein, H. A., & Ambady, N. (2003). Nonverbal "accents" cultural differences in facial expressions of emotion. *Psychological Science*, 14(4), 373-376. https://doi.org/10.1111/1467-9280.24461

Martin, A. E., & Slepian, M. L. (2020). The primacy of gender: Gendered cognition underlies the big two dimensions of social cognition. *Perspectives on Psychological Science*, https://doi.org/10.1177/1745691620904961

Merritt, R. D., & Harrison, T. W. (2006). Gender and ethnicity attributions to a gender-and ethnicity-unspecified individual: Is there a people= white male bias?. *Sex roles*, *54*(11-

12), 787-797. https://doi.org/10.1007/s11199-006-9046-7

Minda, J. P., & Smith, J. D. (2011). Prototype models of categorization: Basic formulation, predictions, and limitations. *Formal approaches in categorization*, 40-64. https://doi.org/10.1017/CBO9780511921322.003

Noguchi, K., Gel, Y. R., Brunner, E., & Konietschke, F. (2012). nparLD: An R software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software, 50*(12), 1-23. https://doi.org/10.18637/jss.v050.i12

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*, *12*(5), 413-417. https://doi.org/10.1111/1467-9280.00376

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087-11092. https://doi.org/10.1073/pnas.0805664105

Paulus, A., & Wentura, D. (2014). Threatening joy: Approach and avoidance reactions to emotions are influenced by the group membership of the expresser. *Cognition and Emotion, 28,* 656–677. https://doi.org/10.1080/02699931.2013.849659

Poyatos, F. (2015). Nonverbal Communication in Interaction: Psychology and Literature. In *The Social Psychology of Nonverbal Communication* (pp. 248-271). Palgrave Macmillan, London. https://doi.org/10.1057/9781137345868_12

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

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/10.1016/j.paid.2012.06.021

Rule, N. O., & Sutherland, S. L. (2017). Social categorization from faces: Evidence from obvious and ambiguous groups. *Current Directions in Psychological Science*, *26*(3), 231-236. https://doi.org/10.1177/0963721417697970

Scher, S. J., & Darley, J. M. (1997). How effective are the things people say to apologize? Effects of the realization of the apology speech act. *Journal of Psycholinguistic Research*, *26*(1), 127-140. https://doi.org/10.1023/A:1025068306386

Schlenker, B. R., & Darby, B. W. (1981). The use of apologies in social predicaments. *Social Psychology Quarterly*, 271-278. 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., & Ross, M. (2010). Why women apologize more than men: Gender differences in thresholds for perceiving offensive behavior. *Psychological Science*, *21*(11), 1649-1655. https://doi.org/10.1177/0956797610384150

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

Stolier, R. M., & Freeman, J. B. (2016). Neural pattern similarity reveals the inherent intersection of social categories. *Nature neuroscience*, *19*(6), 795-797. https://doi.org/10.1038/nn.4296

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

Sutherland, C. A., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology*, *106*(2), 186-208. https://doi.org/10.1111/bjop.12085

Tavuchis, N. (1991). *Mea culpa: A sociology of apology and reconciliation*. Stanford University Press.

Thayer, S. (1980). The effect of expression sequence and expressor identity on judgments of the intensity of facial expression. *Journal of Nonverbal Behavior*, *5*(2), 71-79. https://doi.org/10.1007/BF00986510

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

Tskhay, K. O., & Rule, N. O. (2015). Emotions facilitate the communication of ambiguous group memberships. *Emotion*, *15*(6), 812. https://doi.org/10.1037/emo0000077

Vandekerckhove, J., Matzke, D., Wagenmakers, E.-J. (2015). Model comparison and the principle of parsimony. In Busemeyer, J.R., Wang, Z., Townsend, J.T., Eidels, A. (Eds.)

Oxford handbook of computational and mathematical psychology (pp. 300–317). New York: Oxford University Press.

Walfisch, T., Van Dijk, D., & Kark, R. (2013). Do you really expect me to apologize? The impact of status and gender on the effectiveness of an apology in the workplace. *Journal of Applied Social Psychology*, *43*(7), 1446-1458. https://doi.org/10.1111/jasp.12101

Walker, M., & Keller, M. (2019). Beyond attractiveness: A multimethod approach to study enhancement in self-recognition on the Big Two personality dimensions. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspa0000157

Wilkie, J. E., & Bodenhausen, G. V. (2015). The numerology of gender: Gendered perceptions of even and odd numbers. *Frontiers in Psychology*, *6*, 810. https://doi.org/10.3389/fpsyg.2015.00810

Wei, H., & Ran, Y. (2019). Male versus female: How the gender of apologizers influences consumer forgiveness. *Journal of Business Ethics*, *154*(2), 371-387. https://doi.org/10.1007/s10551-017-3440-7

Wohl, M. J., Hornsey, M. J., & Bennett, S. H. (2012). Why group apologies succeed and fail: Intergroup forgiveness and the role of primary and secondary emotions. *Journal of Personality and Social Psychology*, *102*(2), 306. https://doi.org/10.1037/a0024838

Wood, W., & Eagly, A. H. (2015). Two traditions of research on gender identity. *Sex Roles*, *73*(11-12), 461-473. https://doi.org/10.1007/s11199-015-0480-2

Xie, S. Y., Flake, J. K., & Hehman, E. (2019). Perceiver and Target Characteristics Contribute to Impression Formation Differently Across Race and Gender. *Journal of Personality and Social Psychology*, *117*(2), 364. https://doi.org/10.1037/pspi0000160

APPENDIX

Mean Characteristic Ratings and Comparisons Across Apologetic Faces, Study 3

Male Generators

	Apologetic Male Face	Apologetic Female Face	*p*	$r_{effect\ size}$
Characteristic	*M (SD)*	*M (SD)*		
Apologetic	67.64 (23.53)	51.85 (27.06)	< .001	.48
Sad	78.42 (19.92)	67.14 (26.17)	< .001	.40
Regret	69.68 (22.60)	54.99 (27.58)	< .001	.44
Remorse	68.96 (21.91)	53.49 (26.97)	< .001	.52
Submissive	55.54 (24.63)	50.90 (26.53)	.12	.18
Trustworthy	37.95 (22.98)	37.97 (24.33)	> .99	.01
Dominant	32.49 (21.40)	34.2 (25.07)	> .99	.01
Male	78.86 (19.69)	21.14 (20.93)	< .001	.86
Masculine	69.21 (22.51)	25.54 (22.27)	< .001	.82
Feminine	21.94 (21.82)	73.84 (20.74)	< .01	.26
Female	20.48 (20.99)	78.13 (20.26)	< .001	.85

Female Generators

	Apologetic Male Face	Apologetic Female Face	*p*	$r_{effect\ size}$
Characteristic	*M (SD)*	*M (SD)*		
Apologetic	67.47 (23.33)	75.47 (20.45)	< .001	.33
Sad	79.13 (20.39)	84.21 (18.81)	.002	.26
Regret	67.01 (24.20)	74.86 (21.25)	.001	.28
Remorse	67.96 (21.87)	75.23 (21.58)	.001	.28
Submissive	58.47 (25.55)	65.05 (24.88)	.01	.23
Trustworthy	41.79 (23.40)	42.90 (23.79)	> .99	.03
Dominant	30.06 (21.36)	21.84 (19.39)	< .001	.38
Male	76.10 (21.53)	17.25 (18.63)	< .001	.85
Masculine	63.53 (25.34)	19.11 (20.03)	< .001	.82
Feminine	28.94 (24.72)	77.18 (19.89)	< .001	.82
Female	24.88 (24.68)	80.93 (20.03)	< .001	.86

Note. Wilcoxon signed-rank tests were used for each comparison due to non-normal variable distributions. All *p*-values were Bonferroni-adjusted for 11 comparisons. Characteristics are listed in the same order as in Table 9.