Modeling perceptions of criminality and remorse from faces using a data-driven computational approach

Friederike Funk [1,2], Mirella Walker [3], & Alexander Todorov [1]

[1] Princeton University, USA
[2] University of Cologne, Germany
[3] University of Basel, Switzerland

Running header: PERCEIVED CRIMINALITY AND REMORSE
Word Count: 6322 main text

Draft version: August, 26th, 2016. This article may not exactly replicate the final version published in Cognition and Emotion. It is not the copy of record.

To be cited as: Funk, F., Walker, M., & Todorov, A. (in press). Modelling perceptions of criminality and remorse from faces using a data-driven computational approach, *Cognition and Emotion*. doi: 10.1080/02699931.2016.1227305

Friederike Funk, University of Cologne, Department of Psychology, Herbert-Lewin-Str. 10, 50931 Köln, Germany; friederike.funk@uni-koeln.de *(Corresponding author)*

Mirella Walker, University of Basel, Department of Psychology, Missionsstrasse 64A, 4055 Basel, Switzerland; mirella.walker@unibas.ch

Alexander Todorov, Princeton University, Department of Psychology, Princeton Neuroscience Institute, and Woodrow Wilson School; 323 Peretsman-Scully Hall, Princeton NJ 08540, USA; atodorov@princeton.edu

Funding acknowledgement: This research was partially supported by the Swiss National Science Foundation under grant no. 100014-135213.

Abstract

Perceptions of criminality and remorse are critical for legal decision-making. While faces perceived as criminal are more likely to be selected in police lineups and to receive guilty verdicts, faces perceived as remorseful are more likely to receive less severe punishment recommendations. To identify the information that makes a face appear criminal and/or remorseful, we successfully used two different data-driven computational approaches that led to convergent findings: one relying on the use of computer-generated faces, and the other on photographs of people. In addition to visualizing and validating the perceived looks of criminality and remorse, we report correlations with earlier face models of dominance, threat, trustworthiness, masculinity/femininity and sadness. The new face models of criminal and remorseful appearance contribute to our understanding of perceived criminality and remorse. They can be used to study the effects of perceived criminality and remorse on decision-making; research that can ultimately inform legal policies.

Keywords: social perception, faces, criminal appearance, remorse, emotion, data-driven models

Modeling perceptions of criminality and remorse from faces using a data-driven computational approach

People form impressions about others from their faces remarkably fast (Bar, Neta, & Linz, 2006; Willis & Todorov, 2006). The manifold consequences of such impressions reach from effects on electoral success to harsher criminal sentences (for reviews see, e.g., Olivola, Funk, & Todorov, 2014; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). In the present paper, we focus on two prominent facial characteristics that are especially relevant in the domain of legal decision-making: perceptions of criminality and remorse. People share stereotypes about what kinds of faces are perceived as criminal (see, e.g., Bull, 1992; MacLin & Herrera, 2006; Shoemaker, South, & Lowe, 1973). Ratings on the criminal appearance of faces show high interrater reliability, demonstrating that different people perceive the same individuals to appear criminal or not (Flowe, 2012). Criminal-looking faces are more likely to be remembered (MacLin & MacLin, 2004) and chosen more often in police lineups (Flowe & Humphries, 2011). Moreover, correlational evidence suggests that the more criminal a defendant is perceived to be (due to a facial tattoo), the more likely the defendant is to be judged guilty (Funk & Todorov, 2013). These perceptions of criminality are positively related to perceptions of dominance and negatively related to perceptions of trustworthiness (Flowe, 2012).

In addition to perceived criminality, it is well known that perceived remorse has strong effects in the context of punishment. Remorse is a troubling feeling of distress caused by a sense of guilt for past wrongs (see, e.g., Slovenko, 2006). The Oxford English Dictionary defines remorse as "deep regret or guilt for doing something morally wrong" (remorse, n., OED online, see www.oed.com). People who feel remorse admit that they have done wrong or caused harm and accept responsibility for what they did. Remorse also goes along with "a desire to atone or make

reparation by, for example, expressing remorse, making restitution to the person harmed, undergoing penance, or behaving differently in the future" (Proeve & Tudor, 2010, p. 107). The character of remorseful defendants is rated more positively (Darby & Schlenker, 1989; Robinson, Smith-Lovin, & Tsoudis, 1994; Taylor & Kleinke, 1992; Tsoudis & Smith-Lovin, 1998); and remorseful defendants are rated as less likely to recommit the offense and as deserving of less punishment (Bornstein, Rung, & Miller, 2002; Gold & Weiner, 2000; Pipes & Alessi, 1999; Rumsey, 1976). Consequently, the legal codes of many countries (e.g., Australia, Canada, England and Wales, New Zealand, Singapore, and the United States of America) include remorse as a potential mitigating factor in sentencing decisions (for a list of legal references see, e.g., Proeve & Tudor, 2010, Appendix Chapter 6). Furthermore, social psychological studies have found that the presence of remorse affects how satisfied people are after punishing transgressors (Funk, Gerlach, Walker, & Prentice, under review).

Summing up, despite the importance of both criminal and remorseful appearance in legal settings, little is known about how exactly criminality and remorse look like. What is it that makes a face appear criminal or remorseful?

Identifying the look of criminality and remorse using statistical face models

To find out the facial information that shapes first impressions from faces, different ways of visualizing this information have been introduced in recent years. Using computer-generated faces, Oosterhof and Todorov (2008) have originally modeled facial characteristics as a function of trustworthiness, dominance, and threat judgments. Todorov, Dotsch and colleagues (2013) have further validated models of attractiveness, competence, extraversion, and likeability. Using pictures of real people, Walker and Vetter (2009) have successfully modeled social skills, likeability, attractiveness, trustworthiness, risk seeking, and aggressiveness, as well as the Big

Two (agency and communion) and Big Five personality factors (neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness, see Walker & Vetter, 2016).

Given the effects of criminal appearance on important legal outcomes, we aimed at identifying the facial features that determine whether faces are perceived to look criminal or not. As it is well known that a criminal stereotype exists (e.g., MacLin & Herrera, 2006) and that perceived criminality correlates with both perceived dominance and trustworthiness (Flowe, 2012) – two dimensions which have been successfully modeled in earlier research – we were confident that we would be able to create face models of criminality that identify and visualize the information people rely on when they perceive criminality in faces.

Despite the important role of remorse, surprisingly little research has dealt with the expression of remorse (cf. Keltner & Buswell, 1996). According to Ekman (1993), the expression of remorse is not unique to remorse but part of a group of emotions that share the same facial expression (the "unhappiness" emotions, p. 389). Emotions of this group (e.g., sadness, remorse, shame, and guilt) are expressed by raising the inner corners of the eyebrows, slightly raising the cheeks, and pulling the lip corners downward. Others suggest that whereas shame and embarrassment show discrete facial features, remorse does not (Keltner & Buswell, 1996). Thus, it was less clear whether it would be possible to create a statistical face model of perceived remorse. Our research is the first that uses a purely data-driven approach to empirically investigate if there is a "look of remorse" and if perceivers agree about what looks remorseful in a face and what does not.

General procedure to obtain data-driven face models

Different from earlier work that relied on a single data-driven approach to extract the information that people use when making social judgments about faces, we used two approaches – the one pioneered by Oosterhof and Todorov (2008; referred to as the Princeton approach here) and the one pioneered by Walker and Vetter (2009; referred to as the Basel approach here) – to

identify the looks of perceived criminality and remorse. The former approach has been described in detail by Todorov and colleagues (2008; 2013); and the latter approach by Walker and Vetter (2009; 2016). To summarize, in both approaches, faces are represented as points in a multidimensional, statistical face space (see Blanz & Vetter, 1999). The statistical face space is derived from the analysis of real faces and the resulting dimensions summarize the principal differences among these faces. Within this space, each face is a linear combination of the resulting dimensions. Based on judgments (here: ratings on criminal or remorseful appearance) of faces generated by the statistical face space, it is possible to identify the facial information that is perceived in a certain way (here: criminal or remorseful) and to describe them in a parametrically controlled model. This model is a new vector in the statistical face space that accounts for the maximum variance of the respective judgments.

There are two main advantages of using data-driven models to identify the facial information people use to make social judgments. First, these models are not biased in their search of the set of features that drive particular judgments (Todorov, Dotsch, Wigboldus, & Said, 2011) and can discover features that are not obvious a priori. Second, once the vector has been identified in face space, it is possible to systematically create *new* faces along this vector that appear more or less criminal-looking, for instance. This approach is completely data-driven and holistic without singling out any particular facial feature. In the present set of studies, we first identified the face models that underlie perceptions of criminality and remorse (Study 1) using both Princeton and Basel approaches. Subsequently, again using both approaches, we created and validated these models by applying them to new faces that systematically varied on the criminality or remorse vector (Study 2).

Study 1: Identifying face models of criminal and remorseful appearance

In order to identify the facial information people use to make social judgments of criminality and remorse, we had participants rate faces with known 3D-structure on their criminal or remorseful appearance (for the Princeton approach these faces are computer-generated using the software FaceGen, for the Basel approach these faces are 3-D scans of real faces, see Figure 1 for examples). As explained by Todorov and Oosterhof (Oosterhof & Todorov, 2008; Todorov & Oosterhof, 2011) and Walker and Vetter (2009; 2016), face models are derived by using the mean ratings of each face regarding a certain trait.

Interrater reliabilities served as indicators that participants perceived faces similarly. For models to be obtained with the Princeton approach, we conducted studies in which we let participants rate each face twice to calculate a participant's test-retest reliability in order to ultimately reduce error variance in participants' ratings. For the Basel face models, we used existing ratings on criminality (obtained by Walker & Vetter, 2016) and remorse (obtained by Funk et al., under review), as described below.

Method

Ratings for the Princeton approach

Eighty-three undergraduate students from the Princeton psychology subject pool were recruited for a 30-minute study and received half an hour of course credit for their participation. The study consisted of two blocks, in each of which participants, unbeknown to them, rated the same 300 computer-generated faces either regarding criminal or regarding remorseful appearance (faces available at http://tlab.princeton.edu/databases/randomfaces, see Oosterhof & Todorov, 2008; these faces were randomly generated using the FaceGen Modeller program; specifically, each face was represented as a point in a 50-dimensional shape space and a 50-dimensional

reflectance, surface texture space; the faces were randomly sampled from normal distributions centered at the average face with coordinates of 0 on all dimensions).

Criminal appearance. Forty-three of the participants (26 female, 15 male, M_{age} = 19.45, SD_{age} = 1.24) completed the criminal appearance version of the study: They were told that we were interested in the facial characteristics that influence how criminal a person looks and were asked to imagine that the person shown has committed a crime (without adding more specific information) and is now facing trial. Participants were explained that they would be asked to make the same rating for each face: "To what extent do you think this person looks criminal?" (1 = *not at all*, 7 = *very much*), and that there are no right or wrong answers, that we were interested in their spontaneous reaction, and that they should follow their gut feeling.

We calculated test-retest correlations between the first and second block for each participant before we averaged the two ratings for each face. Data from participants were discarded if their test-retest correlation of the first and second ratings for the faces was not significant ($p > .01$, $r \leq .147$). For the ratings on criminal appearance, the remaining subjects' ($N = 32$) test-retest correlations (all $p < .01$) were distributed around $M_r = .347$, $SD_r = .147$, $min_r = .148$, $max_r = .703$, a size comparable to similar studies on other kinds of face models (e.g., Toscano, Schubert, Dotsch, Falvello, & Todorov, in press). Interrater reliability was high ($IRR_{criminal} = .914$), indicating that different participants perceived faces in a similar way.

Remorseful appearance. The forty participants (25 female, 15 male, M_{age} = 19.43, SD_{age} = 1.05) who completed the remorseful appearance version of the study were told that we are interested in the facial characteristics that influence how remorseful a person looks and were asked to imagine that the person shown has committed a crime (without any specific information). It was added that there was no doubt about whether the person committed that crime or not. Participants were asked to indicate for each face "To what extent do you think the

person feels genuine remorse?" (1 = *not at all*, 7 = *very much*). After excluding participants with non-significant test-retest correlations ($p > .01$, $r \leq .127$), the remaining subjects' ($N = 36$) test-retest correlations (all $p < .01$) were distributed around $M_r = .334$, $SD_r = .097$, $min_r = .174$, $max_r = .529$, and the interrater reliability was high ($IRR_{remorse} = .927$).

[INSERT FIGURE 1 HERE]

Ratings for faces in the Basel Face Model

Criminal appearance. The data to identify criminal appearance in the Basel faces were taken from a study on personality judgments of faces conducted by Walker and Vetter (2016). Participants ($N = 1671$, 1066 female, 598 male, $M_{age} = 24.43$, $SD_{age} = 5.34$) were recruited via the SoSci Panel (Leiner, 2014) and offered the chance to take part in a lottery. They saw a random subset of three out of 153 colored 3-D scans of real faces (n for each face ~ 33) from the Basel Face Model (for details see Paysan, Knothe, Amberg, Romdhani, & Vetter, 2009). The Basel Face Model consists of persons that were conveniently sampled and instructed to show a relaxed, neutral facial expressions, as well as wear no make-up, jewelry, and facial hair during the scanning session. Participants were asked to spontaneously judge the faces on several personality traits, one of which referred to the faces' criminal appearance (item wording "The person depicted is criminal", 1 = *does not apply at all*, 5 = *fully applies*). Interrater reliability was high ($IRR_{criminal} = .821$, $N = 27$).

Remorseful appearance. Participants ($N = 542$, 218 female, 293 male Mturk workers from the U.S., $M_{age} = 32.0$, $SD_{age} = 11.9$) were recruited online to participate in an 8-minute study on "Social judgments of faces" and were paid $0.40. They saw a random subset of fifteen out of 153 3-D scans of real faces (n for each face ~ 55) from the Basel Face Model (Paysan et al., 2009).

For each face, participants were asked to imagine that this person has committed a crime and is now facing trial. In a first round, participants then indicated to what extent they thought the person feels genuine remorse. In a second round, participants saw the faces again in randomized order and also indicated to what extent the person feels guilty and regrets the crime, feels truly sorry for the victim, wants to make amends for the harm caused, and knows that the behavior was wrong (for each face $.77 \leq \alpha \leq .99$, 1 = *not at all*, 7 = *very much*, items in fixed order). These items were chosen according to the conceptual definition of remorse in order to increase the construct validity of assessed remorse. Interrater reliability of this remorse scale was high ($IRR_{\text{remorse}} = .893$, $N = 41$). In order to identify the Basel face model of remorse in a subsequent step, we used the mean ratings on the five item remorse scale for each face. Importantly, a Basel face model for the single item "remorse" looked identical to the model we derived from averaging five items that had already been used in Funk et al. (under review). In addition, across the 153 facial identities rated, the mean of the single item remorse highly correlated with the mean of the remorse scale that was ultimately used to create the Basel face model of remorse, $r = .968$. For the sake of consistency with this earlier work, we decided to use vectors based on the means of the five items.

Resulting models

For both criminal and remorseful appearance, we created statistical face models using the Princeton approach with the FaceGen Modeller program (http://facegen.com, version 3.5) as well as in the Basel face model. The resulting models are visualized in Figure 2.

[INSERT FIGURE 2 HERE]

Intercorrelations with other trait dimensions. Data-driven face models have the advantage that they are embedded in a bigger methodological framework. As described earlier, Todorov and colleagues have already collected ratings on other traits using the Princeton approach to identify corresponding face models, as have Walker and colleagues for the Basel face model. Without collecting any new data, it is possible to use these face models based on earlier ratings to calculate correlations with our new face models on criminality and remorse. In face space, face models of a certain appearance are represented as vectors. Thus, a correlation can be calculated as cosine of our newly derived vector of criminality (or remorse) and another vector of interest (e.g., threat). Face space consists of dimensions that reflect variations in shape as well as dimensions that reflect variations in reflectance (i.e. variations in pigmentation/color as well as in texture). It is possible to report these correlations separately. Correlations between criminal appearance as well as remorseful appearance with threat, dominance, and trustworthiness for both the Princeton and the Basel approach are reported in Table 1. For the Basel face model, it was additionally possible to correlate criminal and remorseful appearance with models of sadness and masculinity/femininity (with data collected by Walker & Vetter, 2016). Within the Basel model, perceptions of criminality strongly correlate with perceived masculinity, whereas perceptions of remorse strongly correlate with the perception of sadness (more precisely "feeling down"/ translation of the German term used: "niedergeschlagen"). An inspection of Figure 2 suggests that the same would hold true for the Princeton model if data on perceived sadness and masculinity existed.

Using both the Princeton and Basel approach, the criminal face models strongly positively correlate with face models of threat and dominance, as well as negatively with trust, replicating earlier findings by Flowe (2012). Regarding the face model of perceived remorse, the Basel and Princeton model show convergence regarding very minor correlations between perceived remorse

and trustworthiness. Other correlations of the remorse face model with earlier models differ between the Princeton and the Basel approach: Whereas perceived remorse is negatively correlated with threat in the Princeton model, there is no such correlation in the Basel model. Moreover, the negative correlation of perceived remorse with dominance is smaller in the Basel model than in the Princeton model.

Lastly, in face space, perceptions of criminality and remorse are negatively correlated for the Princeton models: the correlation for shape is r = -.54, and r = -.59 for reflectance, respectively. For the Basel face model the correlation between criminal and remorseful perception is not significant. r = -.06 (for shape), and r = .02 (for reflectance).

[INSERT TABLE 1 HERE]

Discussion

In both the Princeton and Basel models, perceptions of criminality and remorse showed high interrater reliability and led to convergent models of criminal and remorseful appearance. As can be seen in Figure 2, for criminality and remorse both the Princeton and Basel approaches led to similar models. Informal visual inspection of different versions of the same faces with low and high levels of criminality suggests that stronger perceptions of criminality go along with more masculine faces that have more prominent chins, smaller eyes, lowered eyebrows, and darker pigmentation. Importantly, these detailed descriptions only exemplify the differences between the faces. The resulting face models capture holistic changes and visualize all changes in appearance that matter for the respective traits, they are not reducible to single facial features. Replicating earlier work, and quantified through correlations with earlier face models, perceptions of criminality are related to perceptions of dominance and untrustworthiness, and are very similar to

perceptions of threat. Contributing to the existing research on perceived criminal appearance, our models are the first that are completely data-driven and were derived without any prior assumptions.

Studying the perception of remorse was also successful: in both the Princeton and Basel models, raters showed high interrater agreement. Consistent with earlier work (Ekman, 1993; Keltner & Buswell, 1996), an informal comparison of faces generated by both the Princeton and Basel face models illustrates that perceived remorse is characterized by elevated inner corners of the eyebrows, as well as by lip corners pulled downward and lighter pigmentation. Correspondingly, the Basel face model of remorse correlates with perceived sadness, mirroring the link between expressed remorse and expressed sadness (Ekman, 1993).

The resulting face models in Figure 2 highlight similarities as well as dissimilarities between the two approaches. If the resulting faces are used as stimuli in experimental research, there are different advantages for one model over the other depending on the context in which the facial stimuli are used. Whereas the Princeton approach allows for more powerful manipulations because the face does not need to fit into a specific surrounding, the pictures created with the Basel approach look more realistic like real individuals and not like computer-generated faces or faces that are digitally altered. One downside of this naturalistic approach where derived faces look like photographs, though, is that changes in appearance are more subtle. This difference in subtleness might explain why the Princeton and Basel approach do not show identical correlations between criminal and remorseful appearance with other traits. Importantly, though, the direction of the correlations was identical for all traits but threat, and facial markers of trustworthiness did not significantly correlate with remorseful appearance in both Princeton and Basel models. All in all, these findings demonstrate strong convergence between the Princeton

and Basel approach and emphasize the internal validity of the resulting face models of criminality and remorse.

Study 2: Validation of the face models of criminal and remorseful appearance

In order to validate the derived face models of criminality and remorse empirically (i.e., to test if the face space vector we used to create faces indeed evokes variations in perceptions of criminality and remorse, respectively), we created faces varying on the new vectors for five different identities both in the Princeton and the Basel face model. For the Princeton model, we used five facial identities that have been used before in empirical studies (Todorov et al., 2013). The faces were maximally different from each other to increase generalizability of the effects found. For the Basel face model, we randomly selected portraits of five male identities with relaxed, neutral facial expression from the Basel Face Database (Walker, Schönborn, Greifeneder, & Vetter: The Basel Face Database, see www.mirellawalker.com/face-database/). For each of the five facial identities used, we created five different versions with varying manipulation strength (of remorse or criminality, respectively) resulting in twenty-five facial stimuli per set. The manipulation strength we chose ranged from very low, low, and neutral to high and very high. For the Basel faces the manipulation strength refers to -6 *SD*, -3 *SD*, 0, +3 *SD*, and +6 *SD*, and for the Princeton faces the manipulation strength refers to -3 *SD*, -1.5 *SD*, 0, +1.5 *SD*, and +3 *SD*, respectively. These standard deviations refer to the underlying dimensions in a particular face space (extracted in Study 1) and cannot be compared between the Princeton and the Basel face models (i.e., 6 *SD* in the Basel model is not twice as much as 3 *SD* in the Princeton model, it is only twice as much as 3 *SD* in the Basel model). We chose these particular steps in manipulation strength to obtain a broad range of stimuli while at the same time avoiding too extreme and potentially weird-looking faces (e.g., +/- 8 *SD* in the Basel face model or +/- 4 *SD* in the Princeton model would look more extreme, but could also make the face appear

unrealistic). Using repeated within-subject designs, we let participants rate a particular set of twenty-five faces twice (resulting in fifty facial stimuli for each participant), either regarding criminality or remorse.

Method

One hundred fifty-nine participants (72 female, 87 male; M_{age} = 34.8, SD_{age} = 11.2; 130 identified as White, 13 as Black, and 20 as Arab, East Asian, Latino, Native, South Asian or Other) were recruited on Mturk for a 5-minute survey on face evaluation and were randomly assigned to participate in one of the four different versions of the study: Princeton faces criminal, Princeton faces remorse, Basel faces criminal, or Basel faces remorse. Participant gender, age or ethnicity did not affect any of the findings reported below. Originally, we collected data from 160 participants who were randomly assigned to the four versions of the study. Data for the Princeton remorse version (originally N = 41, IRR = .980) had to be collected again because of an error in the data collection software: instead of showing a low remorse face (-3 *SD*) for the fifth facial identity, the computer program showed a face low in criminality (-3 *SD*). Importantly, the pattern of the findings for the other facial stimuli that were used correctly was identical to the findings reported below for the new and final Princeton remorse version (referred to throughout the manuscript, N = 40). Results for the original Princeton remorse version (N = 41) are available in the supplementary online material.

Participants were explained that they will see 50 pictures of faces and that some of the faces might look more or less similar to each other. In the two remorse versions, participants read "Your task is to rate how remorseful you think a face looks. (Remorse is a feeling of being sorry for doing something bad or wrong in the past.)" The wording for the two criminal versions was respectively "Your task is to rate how criminal you think a face looks." Subsequently, all of the

participants read that we are interested in their spontaneous judgment about each face and that there are no right or wrong answers.

In the first half of the survey, participants saw 25 different faces in randomized order (consisting of five different facial identities each with five varying degrees of appearance manipulation) and rated for each face how remorseful (in the two remorse versions) or how criminal (in the two criminal versions) the person looks (1 = *not remorseful/criminal at all*; 7 = *very remorseful/criminal*). In the second half of the survey, unbeknown to them, participants saw the exact same 25 faces again, again in randomized order, and made the same judgments. At the end of the survey, participants were asked to provide some demographic information (age, gender, political orientation, and ethnic background). They were thanked and paid $0.50.

Results

Preliminary analyses

For each participant, a test-retest correlation was computed for the ratings of the first and second half of the study. Overall, the test-retest correlations for each participant were distributed around $r \sim .8$ (see Figure 3), $M_r = .627$, $SD_r = .284$, $Med_r = .711$, $min_r = -.177$, $max_r = .958$, indicating that participants showed high consistency in the relative order of their ratings for the first and second trial. Naturally, test-retest correlations in Study 2 were higher than in Study 1. One explanation for the difference is that the faces in Study 2 had been manipulated to induce specific impressions, leading to higher intrapersonal agreement between the two trials resulting in higher test-retest correlations, whereas the neutral faces in Study 1 were more homogeneous in appearance, making the rating task more difficult and leading to relatively lower test-retest correlations (for a discussion about how the level of homogeneity in a set of stimuli affects the correlation between ratings, see Hönekopp, 2006). Another explanation for the different sizes of

correlations is that participants rated fewer faces in Study 2, which probably also increased participants' levels of consistency between the first and second rating.

[INSERT FIGURE 3 HERE]

Note that a low test-retest correlation can either mean that participants did not form a coherent impression of a face and therefore show low consistency in the relative order of their ratings for the first and second trial, or that participants did not pay attention to the survey or were unmotivated and just "clicked through" the faces. Either way, although it leads to more noise, we decided to keep participants with a low test-retest correlation in the validation sample to employ the most conservative way to validate the face models.

For the main analyses, the two ratings for each face were averaged. Across the four versions, those averaged ratings had high interrater reliability, *IRR* (*N* = 159) = .993, indicating that different participants rated the faces similarly.

[INSERT FIGURE 4 HERE]

[INSERT FIGURE 5 HERE]

Main analyses

Data were analyzed using repeated measures ANOVAs with two within-subjects factors: *manipulation strength* ranging from very low (--), low (-), neutral (-/+), to high (+) and very high (++) and *facial identity* consisting of the five different facial identities. For each of the models, we tested the linear and quadratic trends of manipulation strength, predicting that a linear trend would explain more variance than a quadratic trend.

For both Princeton and Basel models, remorse and criminality ratings differed significantly between the different levels of manipulation strength, see Figures 4 and 5 for mean ratings. Linear trends were all significant and explained more variance than quadratic trends, see Table 2. For some of the analyses, facial identity was also a significant factor, indicating that different facial identities differ in their baseline appearance of criminality or remorse. The face models were sometimes more powerful for some identities than for others, indicated by interaction effects.

Yet, regardless of the faces' different baseline appearances, in both Princeton and Basel models, manipulation strength was always perceived as intended, and linear trends explained a sizable amount of rating variance. The same findings emerged when only participants with a positive test-retest correlation $r > 0$ ($N = 148$) were included in the final analyses, see supplementary online material.

[INSERT TABLE 2 HERE]

Discussion

In sum, all four face models could be successfully validated. With both the Princeton and the Basel approaches, we identified and validated face models of perceived criminality and remorse. Our face models are the first that can show the facial information people rely on when making inferences about criminality or remorse. As such, they visualize shared stereotypes about "the criminal look" independent of external facial features (like hairstyle, facial tattoos, or facial hair, see MacLin & Herrera, 2006), as well as about "the look of remorse" independent of any verbal or nonverbal features (see, e.g., tenBrinke, MacDonald, Porter, & O'Connor, 2012), and add to the scientific understanding of these stereotypes.

These models will be useful for researchers who study the effects of perceived criminality or remorse, as both face models can be used for research programs that advance psychological theory (for instance to study the effect of perceived remorse on justice-related satisfaction, see Funk et al., under review) or to study applied questions (for instance the extent to which criminal appearance affects sentencing decisions, for the general idea see Funk & Todorov, 2013).

Which of the two methods (Basel or Princeton) a researcher should decide to use depends on the scope of the research, as both methods have particular advantages and disadvantages. With the Princeton faces, it is possible to flexibly vary the social perception of a face, even if it goes along with a change in the facial identity or the gender category (see Figure 2). With FaceGen one can randomly create as many new faces of any kind as one needs; the possibilities are limitless. Faces from the Princeton model are always computer faces, however, whereas the Basel face model allows for a more natural look. This natural look is the biggest advantage of the Basel face model. However, using faces from this model requires the researcher to make a trade-off between a natural look and a strong manipulation, because the manipulated facial identity always remains the same and the manipulated face needs to fit into the surroundings, such as hairstyle or clothes. As a consequence, faces derived with the Princeton models can easily be used for large scale studies that require many facial stimuli, for instance, because the goal is to detect certain psychological mechanisms. The Basel models, on the other hand, have the advantage of manipulating realistic variations of real photographs, which enables researchers to investigate effects concerning particular facial identities of choice, for instance, and without participants realizing that the presented faces are manipulated in regards to a certain trait.

General Discussion

Our goal was to identify the facial information people use when making inferences about criminality and remorse by visualizing this information in computational face models. We used a completely data-driven statistical face approach without any a priori assumptions about the underlying facial features or structures. Our results demonstrated that people show consensus about the trait perception of criminality, as well as about the state perception of remorse. The resulting face models visualize the facial information that people perceive as criminal- or remorseful-looking.

As for criminal appearance, our face models replicate earlier work that has linked perceptions of criminality to perceptions of dominance and untrustworthiness (Flowe, 2012). These models capture holistic changes and are not reducible to single facial features. Yet, informal visual comparisons of different versions of the same faces with low and high levels of criminality suggest that faces perceived to look criminal have more prominent chins, smaller eyes, lowered eyebrows, and darker pigmentation. Within the face models, perceived criminality positively correlated with earlier face models of dominance and threat, and negatively correlated with trustworthiness. Criminal appearance has been found to affect decisions related to the criminal justice system (see for instance Flowe & Humphries, 2011; Funk & Todorov, 2013). For researchers, our findings present two tools to create faces that systematically vary along the perceived criminality dimension in order to advance our knowledge on the effect of appearance on legal outcomes (see for instance recent findings by Wilson & Rule, 2015, on the link between untrustworthy appearance and extreme sentencing outcomes).

As for remorseful appearance, the derived face models add to the scientific understanding of perceived remorse. Although remorse is used in the criminal justice system and plays an influential role (Bandes, 2016), little is known about the information people rely on when they

infer remorse. First, the ease with which we could identify face models demonstrates that people show consensus about what is perceived to look remorseful in a face and what is not. Second, our face models can visualize the facial information people rely on when they decide if a person shows genuine remorse or not (see Figure 2). Although the face models are holistic and do not refer to single facial features, informally comparing versions of the same faces with low and high levels of remorse suggests that perceived remorse goes along with lighter pigmentation, raised inner corners of the eyebrows, and lip corners that are pulled downward. Within the face models, perceived remorse positively correlated with earlier face models of sadness and negatively correlated with earlier models of dominance. Although expressed remorse might lead a perceiver to trust the remorseful individual because of more positive character ratings (see e.g., Darby & Schlenker, 1989), on the level of facial features the facial markers of remorse were not correlated with facial markers of trustworthiness. This divergence between the perceivers' reactions and the facial features of a perceived trait is interesting, yet not necessarily surprising, because facial markers of trustworthiness have been shown to resemble emotional happiness expressions (Todorov et al., 2013), whereas remorse belongs to the "unhappiness" emotions instead (Ekman, 1993).

All in all, our models of perceived remorse and criminality are in line with many earlier findings on impression formation from faces based on emotion overgeneralization (see, e.g., Said, Sebe, & Todorov, 2009; Zebrowitz, 2004; for a review, see Todorov et al., 2015). These prior studies have shown that resemblance to emotional expressions influences trait impressions from faces. Untrustworthy-looking faces, for example, and to a smaller extent dominant-looking faces resemble angry faces (Oosterhof & Todorov, 2008). Given the very high correlations of the model of criminal appearance with untrustworthiness and dominance, it is not surprising that traces of anger can be seen in the most criminal looking faces.

Although remorse can be differentiated from related emotion concepts theoretically, it was not the goal of the current work to empirically distinguish perceived remorse from perceived guilt or perceived regret. Our goal was to study if people agree about the facial features that constitute the "look of remorse". Conceptually, experienced remorse differs from regret and guilt (see e.g., Proeve & Tudor, 2010; as well as Taylor, 1996). Guilt focuses more on the person who committed the act, for instance, whereas remorse is more related to the committed act itself and evokes action tendencies aiming at making up for the previous wrong. Remorse can also be differentiated from regret in that remorse refers to events for which a person feels responsible, whereas regret can be felt about any kind of event. Some scholars suggest that despite conceptual differences, empirically, experienced remorse and guilt might be difficult to disentangle (Proeve & Tudor, 2010). Similarly, it is possible that corresponding face models of perceived remorse and perceived guilt might look very similar, but this is an empirical question left for future studies to investigate.

Future studies could look at the spontaneous inference of the presence or absence of remorse and the conditions of its detection. In our design, we provided participants with contextual knowledge when they were confronted with the state of remorse: participants were instructed to imagine that the individuals shown had committed a crime. It is possible, for instance, that the absence of remorse can only be inferred if a context is provided (see also Ekman, 1993) and that – without this contextual information – a presumably remorseless-looking face is possibly perceived to look happy.

In addition, it would be an interesting avenue for future research to develop face models that allow for variations in eye gaze. In real-world settings, sincere remorse can also be expressed by covering one's head or by avoiding eye contact (Corwin, Cramer, Griffin, & Brodsky, 2012).

These expressions are likely to be important for the perception of remorse, yet at this point our face models are not able to detect this aspect of remorseful appearance.

Conclusion

Despite the importance of perceived criminality and remorse in the context of legal-decision making, so far researchers knew little about how exactly perceived criminality and remorse look like. Despite differences in the faces used and in the participants' samples (German-speaking or the U.S.), the two approaches showed convergent results. With the new Basel and Princeton face models, we now better understand which features people rely on when they infer criminality or remorse from other people's faces, and we can visualize perceived criminality and remorse with any kind of face. In addition, researchers now have the possibility to use these powerful tools in empirical research designs to create empirically valid stimuli that vary in perceived criminality or remorse. With the help of these facial stimuli, it will be possible to study the effects of perceived criminality and remorse on legal decision-making and to advance research that can ultimately inform legal policies.

Acknowledgments: We thank Virginia Falvello, Lauren Feldman, and Matthias Keller for excellent research assistance.

References

Bandes, S. A. (2016). Remorse and Criminal Justice. *Emotion Review*, *8*(1), 14–19. http://doi.org/10.1177/1754073915601222

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*, 269–278. http://doi.org/10.1037/1528-3542.6.2.269

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, *19*(7), 187–194. http://doi.org/10.1145/311535.311556

Bornstein, B. H., Rung, L. M., & Miller, M. K. (2002). The effects of defendant remorse on mock juror decisions in a malpractice case. *Behavioral Sciences & the Law*, *20*, 393–409. http://doi.org/10.1002/bsl.496

Bull, R. (1992). Physical appearance and criminality. *Current Psychological Reviews*, *2*, 269–281. http://doi.org/10.1007/bf02684461

Corwin, E. P., Cramer, R. J., Griffin, D. A., & Brodsky, S. L. (2012). Defendant remorse, need for affect, and juror sentencing decisions. *Journal of the American Academy of Psychiatry and the Law Online*, *40*, 41–49.

Darby, B. W., & Schlenker, B. R. (1989). Children's reactions to transgressions: Effects of the actor's apology, reputation and remorse. *British Journal of Social Psychology*, *28*, 353–364. http://doi.org/10.1111/j.2044-8309.1989.tb00879.x

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*, 384–392.

Flowe, H. D. (2012). Do characteristics of faces that convey trustworthiness and dominance underlie perceptions of criminality? *PLoS ONE*, *7*(6), e37253. http://doi.org/10.1371/journal.pone.0037253

Flowe, H. D., & Humphries, J. E. (2011). An examination of criminal face bias in a random sample of police lineups. *Applied Cognitive Psychology*, *25*, 265–273. http://doi.org/10.1002/acp.1673

Funk, F., Gerlach, T. M., Walker, M., & Prentice, D. A. (under review). Beyond retribution: Wronged people have transformative justice motives toward transgressors.

Funk, F., & Todorov, A. (2013). Criminal stereotypes in the courtroom: Facial tattoos affect guilt and punishment differently. *Psychology, Public Policy, and Law*, *19*, 466–478. http://doi.org/10.1037/a0034736

Gold, G. J., & Weiner, B. (2000). Remorse, confession, group identity, and expectancies about repeating a transgression. *Basic and Applied Social Psychology*, *22*, 291–300. http://doi.org/10.1207/15324830051035992

Hönekopp, J. (2006). Once more: is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 199–209. http://doi.org/10.1037/0096-1523.32.2.199

Keltner, D., & Buswell, B. N. (1996). Evidence for the distinctiveness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition and Emotion*, *10*, 155–171.

Leiner, D. J. (2014). Convenience Samples from Online Respondent Pools: A case study of the SoSci Panel. Working Paper. Retrieved from https://www.researchgate.net/publication/259669050

MacLin, M. K., & Herrera, V. (2006). The criminal stereotype. *North American Journal of Psychology*, *8*, 197–207.

MacLin, O. H., & MacLin, M. K. (2004). The effect of criminality on face attractiveness, typicality, memorability and recognition. *North American Journal of Psychology*, *6*, 145–154.

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, *18*, 566–570.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the USA*, *105*, 11087–11092.

Paysan, P., Knothe, R., Amberg, B., Romdhani, S., & Vetter, T. (2009). A 3D face model for pose and illumination invariant face recognition. *Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) for Security, Safety and Monitoring in Smart Environments*.

Pipes, R. B., & Alessi, M. (1999). Remorse and a previosuly punished offense in assignment of punishment and estimated likelihood of a repeated offense. *Psychological Reports*, *85*, 246–248. http://doi.org/10.2466/pr0.1999.85.1.246

Proeve, M., & Tudor, S. (2010). *Remorse - Psychological and jurisprudential perspectives*. Burington, VT: Ashgate.

Robinson, D. T., Smith-Lovin, L., & Tsoudis, O. (1994). Heinous crime or unfortunate accident? The effects of remorse on responses to mock criminal confessions. *Social Forces*, *73*, 175–190. http://doi.org/10.2307/2579922

Rumsey, M. C. (1976). Effects of defendant background and remorse on sentencing judgments. *Journal of Applied Social Psychology*, *6*, 64–68. http://doi.org/10.1111/j.1559-1816.1976.tb01312.x

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, *9*(2), 260–264. http://doi.org/10.1037/a0014681

Shoemaker, D. J., South, D. R., & Lowe, J. (1973). Facial stereotypes of deviants and judgments of guilt or innocence. *Social Forces*, *51*, 427–433.

Taylor, C., & Kleinke, C. L. (1992). Effects of severity of accident, history of drunk driving, intent, and remorse on judgments of a drunk driver. *Journal of Applied Social Psychology*, *22*, 1641–1655. http://doi.org/10.1111/j.1559-1816.1992.tb00966.x

Taylor, G. (1996). Guilt and remorse. In Rom Harré & W. Gerrod Parrott (Eds.), *The Emotions: Social, Cultural and Biological Dimensions* (pp. 57–73). London, UK: SAGE. Retrieved from http://sk.sagepub.com/books/the-emotions/n5.xml

tenBrinke, L., MacDonald, S., Porter, S., & O'Connor, B. (2012). Crocodile tears: Facial, verbal and body language behaviours associated with genuine and fabricated remorse. *Law and Human Behavior*, *36*, 51–59. http://doi.org/10.1037/h0093950

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion (Washington, D.C.)*, *13*(4), 724–738. http://doi.org/10.1037/a0032335

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

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*(5), 19–45.

Todorov, A., & Oosterhof, N. N. (2011). Modeling social perception of faces. *Signal Processing Magazine, IEEE*, *28*, 117–122. http://doi.org/10.1109/MSP.2010.940006

Toscano, H., Schubert, T. W., Dotsch, R., Falvello, V., & Todorov, A. (In press). Physical strength as a cue to dominance: A data-driven approach. *Personality and Social Psychology Bulletin.*

Tsoudis, O., & Smith-Lovin, L. (1998). How bad was it? The effects of victim and perpetrator emotion on responses to criminal court vignettes. *Social Forces*, *77*, 695–722. http://doi.org/10.2307/3005544

Walker, M., Schönborn, S., Greifeneder, R., & Vetter, T. (under review). The Basel Face Database. Development and validation of a stimulus set systematically modeled regarding the Big Two and the Big Five personality diensions.

Walker, M., & Vetter, T. (2009). Portraits made to measure: Manipulating social judgments about individuals with a statistical face model. *Journal of Vision*, *9*, 1–13. http://doi.org/10.1167/9.11.12

Walker, M., & Vetter, T. (2016). Changing the personality of a face: Perceived Big Two and Big Five Personality Factors modeled in real photographs. *Journal of Personality & Social Psychology, 110*(4), 609-624. http://dx.doi.org/10.1037/pspp0000064

Willis, J., & Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*, 592–598. http://doi.org/10.1111/j.1467-9280.2006.01750.x

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, *26*(8), 1325–1331. http://doi.org/10.1177/0956797615590992

Zebrowitz, L. A. (2004). The origins of first impressions. *Journal of Cultural and Evolutionary Psychology*, *2*(1–2), 93–108. http://doi.org/10.1556/JCEP.2.2004.1-2.6

Tables

Table 1. *Intercorrelations of criminal and remorseful appearance in the Princeton model and Basel model with other perceived traits in the respective face space*

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

	Princeton model *N* = 50 components				**Basel face model** *N* = 198 components			
	Criminal appearance		Remorseful appearance		Criminal appearance		Remorseful appearance	
	Shape	*Reflectance*	*Shape*	*Reflectance*	*Shape*	*Reflectance*	*Shape*	*Reflectance*
Threat	.97***	.98***	-.47***	-.54***	.93***	.94***	-.05	.03
Dominance	.91***	.93***	-.58***	-.64***	.62***	.62***	-.22**	-.25***
Trustworthiness	-.78***	-.80***	.09	-.17	-.90***	-.90***	-.01	-.13
Sadness					.14*	.18*	.58***	.62***
Masculinity/Femininity					-.68***	-.67***	.03	.07

Table 2. *Tests of the face models: Linear trends indicate that derived face models are valid face models of criminality and remorse*

		p	η_p^2
Princeton faces criminal			
manipulation strength	$F(1.53, 56.62) = 140.64$	***	.792
Linear trend	$F(1, 37) = 174.87$	***	**.825**
Quadratic trend	$F(1, 37) = 70.78$	***	.657
facial id	$F(3.03, 112.01) = 7.31$	***	.165
interaction term	$F(8.59, 317.76) = 1.12$	n.s.	
Princeton faces remorse			
manipulation strength	$F(1.52, 59.33) = 157.71$	***	.802
Linear trend	$F(1, 39) = 195.28$	***	**.834**
Quadratic trend	$F(1, 39) = 59.86$	***	.605
facial id	$F(4, 156) = 22.43$	***	.466
interaction term	$F(10.09, 393.68) = 6.57$	***	.144
Basel faces criminal			
manipulation strength	$F(1.37, 53.40) = 68.63$	***	.638
Linear trend	$F(1, 39) = 82.62$	***	**.679**
Quadratic trend	$F(1, 39) = 17.13$	***	.305
facial id	$F(3.08, 120.03) = 27.83$	***	.416
interaction term	$F(9.58, 373.48) = 4.61$	***	.106
Basel faces remorse			
manipulation strength	$F(1.60, 63.98) = 88.91$	***	.690
Linear trend	$F(1, 40) = 114.53$	***	**.741**
Quadratic trend	$F(1, 40) = 15.71$	***	.282
facial id	$F(4, 160) = 14.86$	***	.271
interaction term	$F(9.87, 394.80) = 4.05$	***	.092

Note: *** $p < .001$; $N = 159$; non-integer *F*-values indicate that degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

Figure captions

Figure 1. Examples of the neutral faces participants rated in Study 1 to obtain face models of criminal and remorseful appearance (A) Examples of the 300 randomly generated facial identities in FaceGen (Todorov et al., 2008); (B) Examples of the 153 3-D scans of real faces from the Basel Face Model (Paysan et al., 2009), copyright by Thomas Vetter.

Figure 2. The resulting face models (A) Princeton approach criminal, (B) Basel face model criminal, (C) Princeton approach remorse, (D) Basel face model remorse, each ranging from strongly reduced salience (--), slightly reduced salience (-), original (+/-), to slightly enhanced salience (+), and strongly enhanced salience (++). These pictures were validated in Study 2 along with pictures from four other facial identities for each face model. For b) and d) see the Basel Face Database (Walker, Schönborn, Greifeneder, & Vetter, under review), copyright by Mirella Walker.

Figure 3. Distribution of the subjects' (N = 159) test-retest correlations between the first and second trial collapsed across the different face models (each consisting of ratings on the same 25 faces).

Figure 4. Mean ratings for criminal and remorseful faces with manipulation strength varying from -3 *SD*, -1.5 *SD*, 0, +1.5 *SD,* to +3 *SD* (generated with the Princeton model). Bars represent standard errors of the mean.

Figure 5. Mean ratings for criminal and remorseful faces with manipulation strength varying from -6 *SD*, -3 *SD*, 0, +3 *SD*, to +6 *SD* (generated with the Basel face model). Bars represent standard errors of the mean.