Who can be fooled?: Modeling facial impressions of gullibility

Who Can Be Fooled? Modeling Facial Impressions of Gullibility

Bastian Jaeger[1,2] and Erdem O. Meral[1]

[1]Department of Social Psychology Tilburg University

[2]Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam

Draft version: 4 May 2021

This paper is currently undergoing peer review. Comments are welcome.

Author Note

Both authors contributed equally to this manuscript.

Correspondence concerning this article should be addressed to Bastian Jaeger, Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, 1081 BT, Amsterdam, the Netherlands. E-mail: bxjaeger@gmail.com.

All data, analysis scripts, materials, and preregistration documents are available at the Open Science Framework (https://osf.io/tx6p3/).

Abstract

The success of acts of deceit and exploitation depends on how trusting and naïve (i.e., gullible) targets are. In three preregistered studies, using both theory-driven and data-driven approaches, we examine how people form impressions of gullibility based on targets' facial appearance. We find significant consensus in gullibility impressions, suggesting that people have a somewhat shared representation of what a gullible person looks like (Study 1, $n = 294$). Gullibility impressions are based on different cues than trustworthiness or dominance impressions, suggesting that they constitute dissociable facial stereotypes (Study 2, $n = 403$). Examining a wide range of facial features, we find that gullibility impressions are primarily based on resemblance to an angry facial expression. We also find that young, female, and smiling individuals are seen as more gullible (Study 3, $n = 209$). These findings suggest that gullibility impressions are based on cues linked to low levels of perceived threat.

Keywords: impression formation; face perception; gullibility; emotion resemblances, babyfacedness

Who Can Be Fooled? Modeling Facial Impressions of Gullibility

Every year, large amounts of money are lost due to e-mail scams, pyramid schemes, identity thefts, and other types of fraudulent behaviors (Witt, 2018). Aside from the financial costs, fraud victims report experiencing emotional distress including feelings of anger and stress, physical and mental health problems, and issues with their close relationships (Button et al., 2014; Modic & Anderson, 2015). While fraud is common, some groups of individuals are targeted more frequently (Burnes et al., 2017; Cohen, 2006). In fact, people differ in how gullible they are (Teunisse et al., 2019). That is, some people may be more at risk because their general tendency to trust and believe others makes them more likely to fall for a scam. How do perpetrators identify gullible targets? Previous research has shown that people spontaneously infer a wide variety of personality traits from a person's facial appearance (Todorov et al., 2015). Building on these insights, we investigate how people form facial impressions of gullibility.

Gullibility Impressions From Faces

Personality impressions from faces are formed spontaneously, quickly, and efficiently (Todorov et al., 2015). While people infer a wide variety of traits from a person's facial appearance, three core dimensions capture most variance in face judgments: trustworthiness, dominance, and attractiveness (B. C. Jones et al., 2021; Oosterhof & Todorov, 2008; Sutherland et al., 2013). These dimensions are thought to reflect the functional significance of trait impressions, with trustworthiness perceptions reflecting an evaluation of a person's intentions, dominance perceptions reflecting the evaluation of a person's ability to implement their intentions, and attractiveness perceptions reflecting the evaluation of a person's mate value (Oosterhof & Todorov, 2008; Sutherland et al., 2013). Facial impressions can have important consequences (Olivola et al., 2014). For instance, trustworthiness impressions influence criminal sentencing decisions (Wilson & Rule, 2015), personnel selection (Gomulya et al., 2017), and economic trust (Jaeger et al., 2019). In short, people spontaneously infer personality traits from facial appearance and these inferences guide a wide range of social decisions.

While a wealth of studies has examined what makes a person look trust*worthy* or dominant (Jaeger & Jones, 2020; Oosterhof & Todorov, 2008), little is known about what makes a person look trust*ing*. The goals of the current investigation are threefold.

First, we examine whether a facial gullibility stereotype exist. If people have a somewhat shared representation of what a gullible person looks like, then they should show at least some level of consensus in their gullibility judgments (Hehman et al., 2017).

Second, we examine which facial characteristics make a person look gullible. People rely on various features when forming trait impressions from faces (Todorov et al., 2015). For instance, perceptions of trustworthiness are driven by morphological characteristics such as facial width-to-height ratio (Ormiston et al., 2017), statistical characteristics such as averageness and sex-typicality (Dotsch & Todorov, 2012; Sofer et al., 2015), and resemblances between facial features and emotional expressions (Jaeger & Jones, 2020; Said et al., 2009). An extensive body of work on the antecedents and consequences of baby-ish facial features (Berry & Zebrowitz McArthur, 1985; Zebrowitz et al., 2003; Zebrowitz & Montepare, 1992) shows that babyfaced individuals are seen as naïve, trusting, and credulous (Berry & Zebrowitz McArthur, 1985, 1986; Zebrowitz & Montepare, 1992). These findings suggest that babyfacedness may be an important component of gullibility impressions. However, babyfacedness is correlated with many other facial characteristics (e.g., facial width-to-height ratio, resemblance to a happy or angry facial expression; Jaeger & Jones, 2020), which may form the basis for people's impressions. Here, we aim to provide a more comprehensive account of the basis of gullibility impressions, examining associations between a large set of facial characteristics commonly studied in the social perception literature.

Finally, we examine relationships between gullibility impressions and impressions of trustworthiness and dominance. Prior work suggests that gullible people are seen as well-intentioned and approachable, but also as relatively incompetent (Evans & van de Calseyde, 2018). This view predicts that that gullible-looking people should score high on perceived trustworthiness (a proxy for an individual's perceived intentions), but low on perceived dominance (a proxy for an individual's ability to implement their intentions; Todorov, Said, Engell, & Oosterhof, 2008). We not only examine relationships between gullibility impressions and impressions on these fundamental dimensions of social perception, we also ask whether facial gullibility is dissociable from these dimensions. That is, we test to what extent gullibility impressions are based on the same facial characteristics as impressions of trustworthiness and dominance.

Measuring Gullibility Impressions

As relatively little is known about the basis of gullibility impressions, we adopt a rather broad conceptualization of the target dimension. We aim to measure impressions that capture the perceived likelihood of targets to assume good intentions in others, which would make them more likely to become the target of lies, fraud, or other types of exploitation. In everyday language, this propensity is captured by traits such as gullibility, trust, naïveté, and credulity. Thus, although we use the label gullibility throughout the manuscript, we provide participants with a broad definition of gullibility which encompasses the extent to which they believe a person is gullible, trusting, naïve, and easily believes others (Krueger et al., 2019; Schniter & Shields, 2020).

One might argue that this conceptualization is too broad and conflates gullibility and trust, which have been treated as separate constructs in some studies (Rotter, 1967; Teunisse et al., 2019; Yamagishi et al., 1999). For example, Teunissen and colleagues (2019) argued that trust is a generalized expectation that others are trustworthy (in absence of any information about their actual trustworthiness), whereas gullibility represents a failure to detect or act upon cues that indicate a person's untrustworthiness. Although we do not disagree that there might be value in treating trust and gullibility as somewhat distinct traits, this does not mean that trust and gullibility necessarily represent dissociable facial stereotypes. For example, previous research has examined the facial features associated with impressions of competence (e.g., Oh et al., 2019). One could argue that competence can be separated into multiple conceptually distinct facets (empathy, intelligence, dexterity, etc.). Yet, this does not imply that mapping the facial features that drive general impression of competence is uninformative—in fact, these impressions have been shown to predict various social outcomes (Ling et al., 2019; Todorov et al., 2005). In a similar vein, the current work examines more general impressions of gullibility.

The Current Studies

In three preregistered studies, we model perceptions of gullibility using different sets of face stimuli and statistical techniques. In Study 1 (n = 254), we rely on reverse correlation—a data-driven technique—to create images of prototypically gullible-looking faces (Dotsch et al., 2008). We also recruit a separate sample of participants (n = 40) to test how gullibility perceptions are related to perceptions of other core dimensions of social perception.

In Studies 2 and 3, we measure perceptions of gullibility for two sets of real, rather than computer-generated faces. In both studies, we test whether there is a shared gullibility stereotype by estimating the level of consensus in participants' judgments. In Study 2 (n = 403) we measure gullibility perceptions of 183 emotionally neutral U. S. American targets. We examine relationships between gullibility impressions and a wide range of facial characteristics that are commonly studied in the social perception literature. Many facial characteristics are correlated, making it difficult to disentangle their unique effects. To identify the most important predictors of gullibility impressions and build a parsimonious model of facial gullibility, we rely on statistical methods from machine learning (i.e., cross-validation and elastic net regression Hastie et al., 2009; Yarkoni & Westfall, 2017).

In Study 3 (n = 209) we measure gullibility perceptions of 171 German targets. Going beyond emotionally neutral faces, we examine whether smiles, which are typically seen as a sign of positive intentions (Martin et al., 2017; Mehu et al., 2007), increase perceived gullibility. Moreover, previous work on the demographic characteristics of fraud victims suggest that young and old (vs. middle-aged) people and women are more likely to be targets of fraudulent behavior (Jagatic et al., 2007; Sheng et al., 2010). We therefore test whether young, old, and female indivduals are perceived as more gullible.

All data, analysis scripts, materials, and preregistration documents are available at the Open Science Framework (https://osf.io/tx6p3/). We report how our sample sizes were determined and all data exclusions and measures for each study.

Study 1

The objectives of Study 1 were to (a) investigate the facial features that differentiate between persons perceived as low or high on gullibility and (b) test how perceptions of gullibility are related to perceptions of other core trait dimensions (Oosterhof & Todorov, 2008; Sutherland et al., 2013). To this end, participants completed a reverse-correlation image classification task (Dotsch & Todorov, 2012). Individuals were presented with two facial photographs, both depicting the same face with average features. One image was superimposed with a random noise pattern and the other image was superimposed with the reverse of the random noise pattern. This creates slight and completely random differences in facial appearance. Across a large number of trials in which noise patterns were varied, participants were asked to choose the more gullible-looking face. Afterwards, all noise patterns of the

selected images were averaged. That is, the noise patterns that were judged to produce a more gullible-looking facial appearance were combined and superimposed on the base image, producing an approximation of a prototypically gullible-looking version of the base image. The noise patterns that were not selected were also combined to produce an anti-gullible face. This approach has been successfully used to visualize a variety of facial stereotypes for traits such as trustworthiness, dominance, warmth, and competence (Dotsch & Todorov, 2012; Imhoff et al., 2013).

In the second phase, which was preregistered, an independent sample of participants rated the prototype images. We attempted to validate the results of the reverse correlation task by testing whether the gullibility prototype is indeed rated as more gullible than the anti-gullibility prototype. We also obtained ratings of trustworthiness, dominance, and attractiveness. Based on the results of prior work (Evans & van de Calseyde, 2018), we predicted that the gullibility prototype would be seen as more trustworthy, but less dominant than the anti-gullibility prototype. We did not have strong expectations regarding differences in attractiveness. However, as attractiveness ratings are correlated with trustworthiness ratings (and other evaluations of a person's intentions) in computer-generated faces (Oosterhof & Todorov, 2008), we predicted higher attractiveness for the gullibility prototype.

Methods

Image construction. In the image construction phase, we recruited 255 students who completed the study for partial course credit. One participant (0.39%) was excluded due to missing data, leaving a final sample of 254 participants (M_{age} = 19.62, SD_{age} = 1.91, 89.76% female). Note that our sample size was considerably larger than typical sample sizes for reverse correlation tasks (e.g., Imhoff et al., 2013; Oliveira et al., 2019). Participants completed a two images forced choice variant of the reverse correlation task (Dotsch et al., 2008; Dotsch & Todorov, 2012). On every trial, two images of the same base face (a morphed image of 50 male faces from the Karolinska Directed Emotional Faces database Lundqvist, Flykt, & Öhman, 1998) were displayed. One image was superimposed with a random noise pattern, whereas the other image was superimposed with the negative of the random noise pattern. The position of the two images and the order in which the different noise patterns were displayed were randomized.

Across 300 trials, participants selected the face that most resembled a gullible person. Participants were told that a person who scores high on gullibility "*is very trusting, naïve, and*

easily believes others" and that a person who scores low on gullibility "*is very distrusting, skeptical, and easily suspicious of others.*"

We created a group-level classification image (CI) by averaging the noise patterns on the images that participants had chosen as more gullible. We also created a group-level anti-CI by averaging the noise patterns of the images that participants did not choose. In this case, the anti-CI is supposed to represent a person who is not gullible.

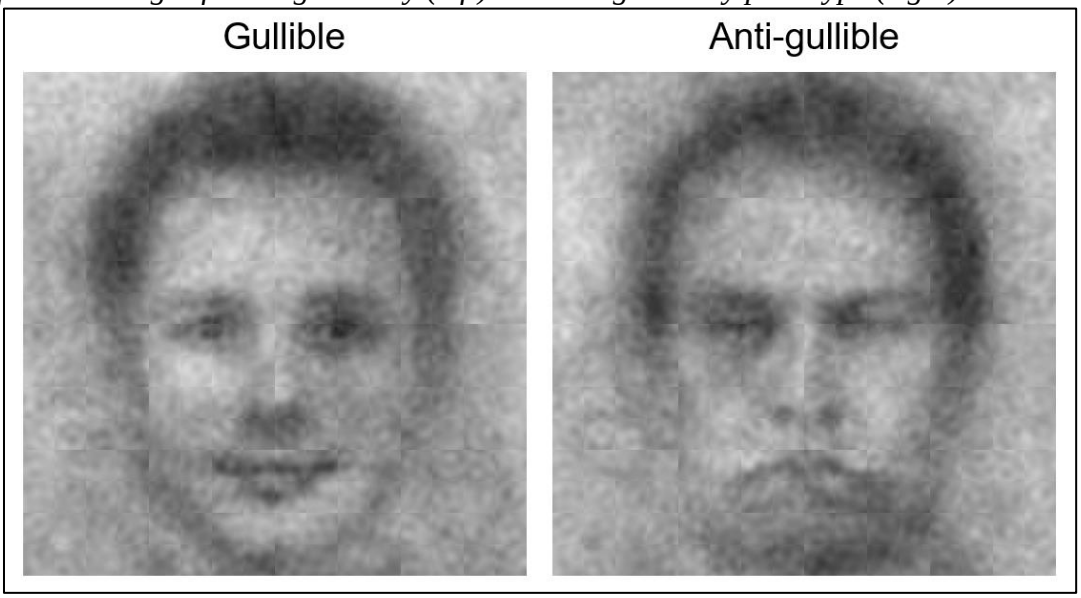
Image rating. Simulation results show that data from 40 independent raters yield stable average ratings even when consensus is relatively low (Hehman, Xie, Ofosu, & Nespoli, 2018). Therefore, in the image rating phase, we recruited 40 American Mechanical Turk workers (37.50% female, M_{age} = 33.90, SD_{age} = 7.54) to rate the gullibility and anti-gullibility prototypes on four trait dimension: gullibility, trustworthiness, dominance, and attractiveness. All participants first indicated their gullibility ratings and then their ratings for the other three dimensions (in a randomized order). We displayed one image at a time and the position of the two face prototypes was counterbalanced. Participants indicated their responses on a scale that ranged from 1 (*not at all* [trait]) to 9 (*extremely* [trait]).

Results

The face prototypes resulting from the reverse correlation task are displayed in Figure 1. The gullibility prototype is shown on the left and the anti-gullibility prototype is shown on the right. A paired-samples *t*-test showed that the gullibility prototype was indeed seen as more gullible (M = 6.30, SD = 1.88) than the anti-gullibility prototype (M = 4.55, SD = 2.28), $t(39)$ = 3.21, p = .003, Hedge's g = 0.50.

Figure 1

Classification images for the gullibility (left) and anti-gullibility prototype (right)



Next, we examined ratings of trustworthiness, dominance, and attractiveness. In line with our predictions, the gullible face was seen as more trustworthy (M = 7.00, SD = 1.22) than the anti-gullible face (M = 3.98, SD = 2.21), $t(39) = 7.18$, $p < .001$, Hedge's $g = 1.12$. The gullible face was also seen as less dominant (M = 4.65, SD = 1.99) than the anti-gullible face (M = 6.63, SD = 1.63), $t(39) = 4.53$, $p < .001$, Hedge's $g = 0.71$. Finally, the gullible face was seen as more attractive (M = 6.55, SD = 1.47) than the anti-gullible face (M = 3.68, SD = 2.13), $t(39) = 7.83$, $p < .001$, Hedge's $g = 1.44$.

Discussion

Study 1 provided initial insights into what people think a gullible person looks like by identifying some of the facial features that differentiate persons perceived as scoring high or low on gullibility. A visual inspection of the face prototype suggests that a gullible-looking person has baby-faced features, with a smiling mouth, and widely opened eyes. Ratings by an independent sample of participants confirmed that the gullibility prototype was indeed seen as more gullible than the anti-gullibility prototype, suggesting that people somewhat agree on which facial features make a person look gullible. Moreover, the gullible face was seen as more trustworthy and less dominant than the anti-gullible face. These findings are in line with the notion that gullible people are seen as well-intentioned (i.e., high on trustworthiness) and harmless (i.e., low on dominance) individuals (Evans & van de Calseyde, 2018).

Study 2

In Study 2, we used naturalistic, rather than computer-generated face images. This allowed us to test how naturally occurring variations in facial features influence gullibility perceptions. The goals of Study 2 were three-fold. First, we again tested whether a shared facial gullibility stereotype exists, by examining consensus in participants' perceptions. Second, we aimed to investigate the basis of gullibility perceptions. We examined associations between a large set of facial characteristics gullibility perceptions. Specifically, we focused on various characteristics have been studied extensively in the social perception literature (e.g., resemblances to emotional expressions, attractiveness, babyfacedness), as well as a broad set of demographic (e.g., gender and age) and morphological characteristics (e.g., facial width-to-height ratio). Finally, results of Study 1 suggested that perceptions of gullibility are related to perceptions of trustworthiness and dominance, which represent core dimensions on which faces are evaluated (Oosterhof & Todorov, 2008). This raises the question of whether facial stereotypes of gullibility can be dissociated from facial stereotypes of trustworthiness and dominance. We examined this question by testing whether facial characteristics differently predicted judgments of gullibility, trustworthiness, and dominance. To answer these questions, we leveraged the extensive norming data of the Chicago Face Database (Ma et al., 2015). The database contains photographs of a large and diverse set of targets, and measurements of a wide range of facial characteristics.

Methods

This study was preregistered and we explicitly mention if a decision or analysis was not specified a priori.[1]

Participants. We again aimed to recruit 40 raters per image set (i.e., a total of 120 participants; Hehman et al., 2018) with the final sample size being determined by how many participants completed the study within two weeks. In total, we recruited 512 first-year psychology students from a Dutch university who completed the study in return for partial course credit. In line with our preregistration, data from 102 participants (19.92%) who indicated poor

[1] Following the suggestions of two anonymous reviewers, we revised our analytical approach for Study 2, which now substantially deviates from our preregistered analysis plan. The results of all preregistered analyses are reported in the Supplemental Materials.

or basic English proficiency, from 3 participants (0.01%) who always indicated the same rating across all trials, and from 4 participants (0.01%) whose response time was shorter than 100 milliseconds on at least 10% of all trials were excluded from analyses, leaving a final sample of 403 participants (M_{age} = 20.00, SD_{age} = 2.12; 80.15% female).

Stimuli. We selected all 183 images of Caucasian adults (93 male and 90 female) from the Chicago Face Database (Ma et al., 2015). All targets displayed a neutral facial expression and were photographed from a fixed distance against a uniform background. The estimated age of targets ranged from 17 to 50 (M_{age} = 27.87, SD_{age} = 5.84). We created three sets of stimuli. Each stimulus set contained 61 images (30 female).

Procedure. Participants were randomly assigned to one stimulus set. They were asked to rate how gullible they think the person in the photo is on a scale from 1 (*not at all*) to 7 (*extremely*). We used the same gullibility description as in Study 1. We created average gullibility ratings per image. These averages were based on ratings by a minimum of 129 participants.

Facial characteristics. For each target, the Chicago Face Database contains data on an extensive set of facial characteristics. Here, we focused on four sets of facial characteristics: demographic variables (gender and age), resemblances to emotional expressions (the extent to which facial features resemble an expression of fear, anger, disgust, happiness, sadness, and surprise), psychological traits (perceived attractiveness, femininity, babyfacedness, prototypicality, and unusualness), and morphological traits (luminance, fWHR, lip fullness, eye size, and nose shape). Data on gender and race were provided by the photographed individuals. Data on emotional resemblances and psychological traits represent subjective ratings provided by participants, with each characteristic representing the mean rating from an average of 44 independent raters (*Min* = 21, *Max* = 131). Morphological features were measured in Adobe Photoshop. A more detailed description of the variables and how they were measured is provided by Ma and colleagues (2015).

Results

All analyses were performed in R (R Core Team, 2021).

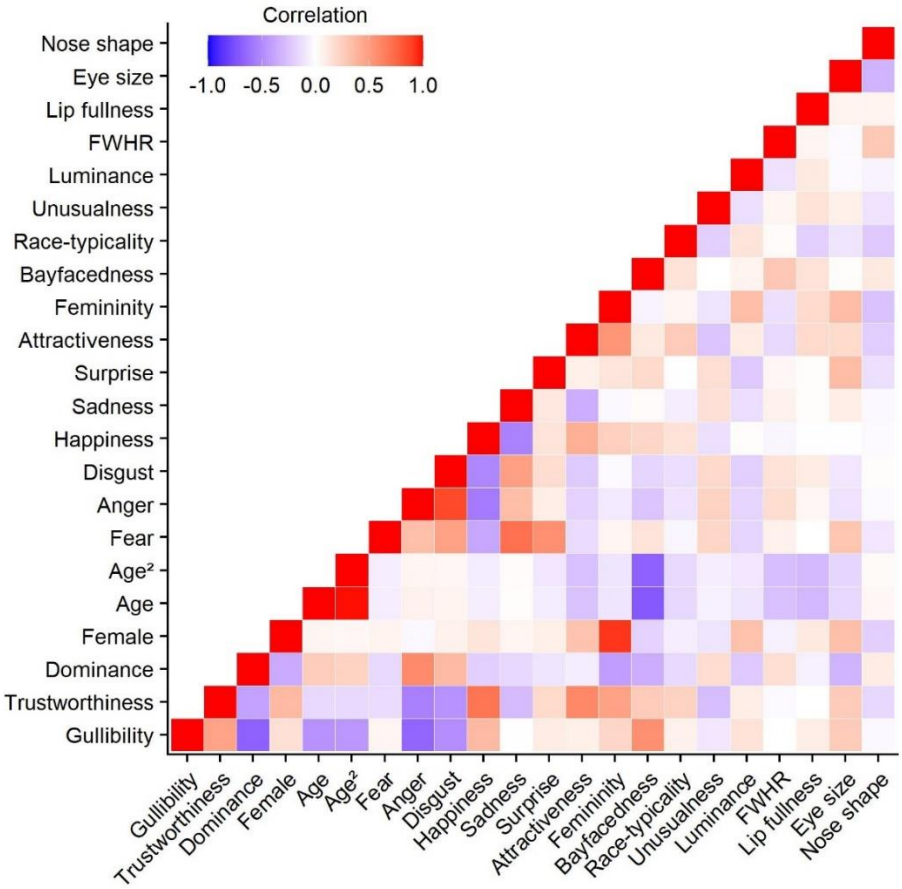
Consensus. First, we examined consensus in gullibility ratings. We computed intraclass correlation coefficients (*ICC*s) using the psych package (Revelle, 2011). Participants showed relatively low, but significant consensus in their judgments across the three sets, *ICC* = .078, $p <$

.001 (set 1: *ICC* = .062, $p < .001$, 95% CI [.043, .092]; set 2: *ICC* = .095, $p < .001$, 95% CI [.068, .138]; set 3: *ICC* = .077, $p < .001$, 95% CI [.055, .113]).

Facial characteristics and gullibility perceptions. Next, we examined which facial characteristics were associated with perceptions of gullibility. A correlation matrix with all variables is displayed in Figure 2. The strongest correlations were observed for resemblance to an angry facial expression, $r(181) = -.66$, $p < .001$, babyfacedness, $r(181) = .57$, $p < .001$, resemblance to a disgusted facial expression, $r(181) = -.49$, $p < .001$, age, $r(181) = -.47$, $p < .001$, and resemblance to a happy facial expression, $r(181) = .36$, $p < .001$. Thus, individuals with babyfaced features and individuals who appeared happy were perceived as more gullible, whereas older individuals and individuals who appeared angry or disgusted were perceived as less gullible.

Figure 2

Correlations between facial characteristics and perceptions of gullibility, trustworthiness, and dominance (Study 2)



Differences with trustworthiness and dominance perceptions. Next, we examined whether there were differences between perceptions of gullibility, trustworthiness, and dominance. Replicating findings of Study 1, we found that perceptions of gullibility were positively related to perceptions of trustworthiness, $r(181) = .47$, $p < .001$, and negatively related to perceptions of dominance, $r(181) = -.68$, $p < .001$. This raises the question of how much facial stereotypes for gullibility, trustworthiness, and submissiveness (i.e., low dominance) overlap. To address this question, we examined to what extent perceptions of gullibility, trustworthiness, and submissiveness are based on the same facial features. For each of the 18 facial characteristics, we estimated a multilevel regression model with random intercepts per target. We regressed trait ratings on the facial characteristic in question (e.g., babyfacedness), a dummy variable indicating which trait was judged (gullibility vs. trustworthiness vs. submissiveness), and their interaction. The interaction effect shows whether there was a significant difference in the association between a facial characteristic and perceptions on the three trait dimensions. For example, the interaction effect for babyfacedness would show whether there was a significant difference in how strongly babyfacedness was associated with (a) perceptions of gullibility vs. trustworthiness and (b) perceptions of gullibility vs. submissiveness.

For the sake of brevity, we only show the interaction effects of all models in Table 1. Results showed that many facial characteristics (10 out of the 18 examined here) were differently related to perceptions of gullibility and trustworthiness (see Table 1, column 1). For example, babyfaced targets were seen as more gullible and more trustworthy, but babyfacedness was stronger predictor of gullibility perceptions than trustworthiness perceptions. In a similar vein, attractiveness was more strongly related to perceptions of trustworthiness than to perceptions of gullibility.

A similar pattern was observed when comparing perceptions of gullibility and submissiveness. Many characteristics (7 out of the 18 examined here) were differently related to perceptions of gullibility and submissiveness (see Table 1, column 2). For example, babyfaced targets were seen as more gullible and more submissive, but babyfacedness was a stronger predictor of gullibility perceptions than submissiveness perceptions. Similarly, feminine-looking targets were seen as more gullible and more submissive, but femininity was a stronger predictor of submissiveness perceptions than gullibility perceptions. Together, these results show that

many facial characteristics influence perceptions of gullibility, trustworthiness, and dominance to different degrees.

Table 1
Differences in the relationships between various facial characteristics and perceptions of gullibility, trustworthiness, and submissiveness (i.e. reverse-scored dominance).

	Gullibility vs. Trustworthiness	Gullibility vs. Submissiveness
Female	0.407 **	0.410 **
Age	0.308 ***	0.213 **
Fear	-0.194 **	0.109
Anger	0.113	0.073
Disgust	0.023	0.133
Happiness	0.328 ***	-0.161 *
Sadness	-0.287 ***	0.174 *
Surprise	0.083	0.006
Attractiveness	0.524 ***	0.003
Femininity	0.269 ***	0.212 **
Babyfacedness	-0.300 ***	-0.221 **
Race-typicality	0.159 *	0.093
Unusualness	-0.184 *	-0.088
Luminance	-0.064	0.080
fWHR	-0.018	-0.169 *
Lip fullness	-0.091	-0.034
Eye size	0.0001	0.046
Nose shape	-0.141	-0.067

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

Elastic net regression. Finally, we examined which facial characteristics are most predictive of gullibility perceptions. Many of the facial characteristics we examined here were correlated (see Figure 2). Examining zero-order correlations provides limited insight into their

unique relationships with gullibility perceptions. To identify the most important predictors of gullibility perceptions—that is, to build a parsimonious model of facial gullibility—we simultaneously entered all facial characteristics (except gender and age) in the same regression model. When the number of predictors is high and (some) correlations between predictors are large, ordinary least squares regression often produces biased estimates and overfitted models (Hastie et al., 2009; Tibshirani, 1996). We therefore relied on elastic net regression (Hastie et al., 2009; Zou & Hastie, 2005), which shrinks predictors to reduce overfitting and performs variable selection by setting the coefficients of uninformative predictors to zero (for similar approaches, see Hehman et al., 2020; Jaeger & Jones, 2020). To avoid overfitting, we also performed nested cross-validation (Yarkoni & Westfall, 2017) using the caret package (Kuhn, 2008). In this procedure, the data is first split into a training set and a test set. The training set is used to estimate the model. Then, the model's predictive fit (i.e., the root-mean-square error, RMSE) is assessed for the test set. The RMSE indicates how close, on average, the model's predicted values are to the actually observed values. This process is then repeated many times with different random splits of the data. The procedure prevents overfitting, as the model's performance is always tested on new data (Yarkoni & Westfall, 2017).

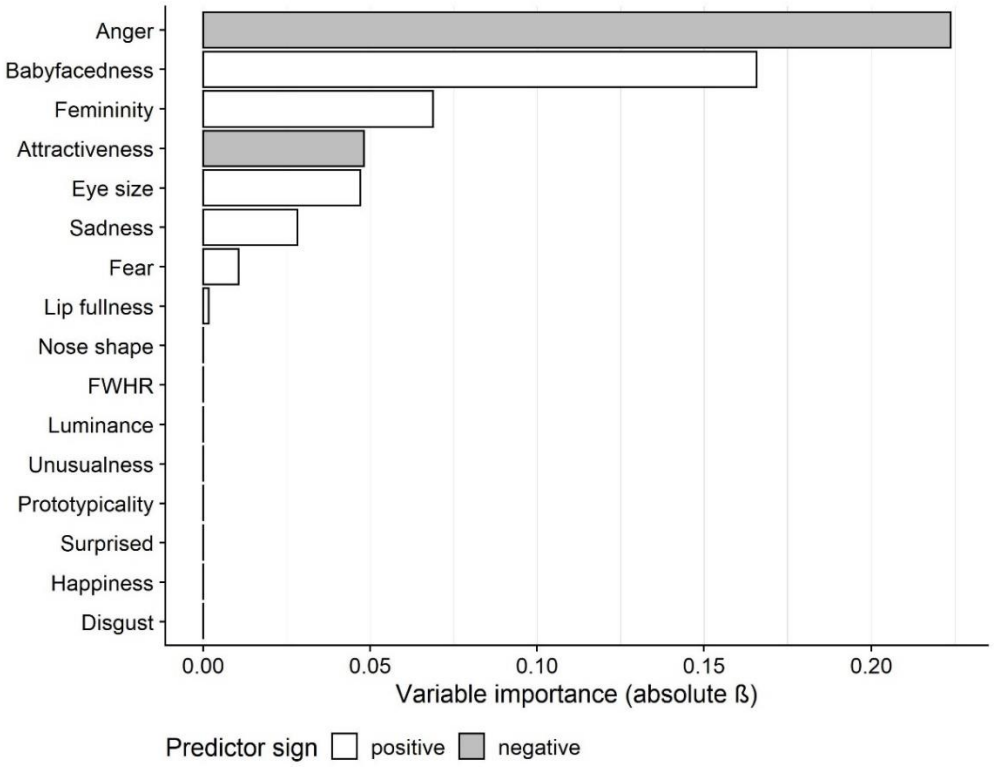
Elastic net models have two hyperparameters that require tuning: α, which controls the degree to which the model shrinks coefficients, and λ, which determines how aggressively coefficients are set to zero. To identify which combination of α and λ maximized the fit of our model, we performed nested cross-validation. We split our dataset into ten folds. For each split of the data, a ten-fold grid search was carried out to derive the best hyperparameters before predicting the held out tenth fold. We repeated this process 100 times. This allowed us to identify at which levels of α and λ our model's predictive fit was maximized (i.e., RMSE was minimized). We then implemented our final model with the optimal α and λ parameters, again relying on 10-fold cross-validation with 100 repeats.[2]

[2] We do not report any inferential statistics for results of these analyses. Indicators of statistical significance, such as *p*-values, depend on the size of a sample. To test the predictive power of our models, we examined their average performance across many different splits of the data. This means that our analyses were based on sample sizes that can be arbitrarily increased. This

Our model predicted perceptions of gullibility to within 0.232 points on our 7-point scale (M_{RMSE} = 0.232, SD_{RMSE} = 0.035). The model explained 66% of the variance in gullibility perceptions (M_{R^2} = 0.663, SD_{R^2} = 0.118). We examined which facial features contributed most to the predictive accuracy of the model (see Figure 3). Resemblance to an angry facial expression ($\bar{\beta}$ = -0.224) was the strongest predictor of gullibility perceptions, followed by babyfacedness ($\bar{\beta}$ = 0.167). To a lesser extent, facial femininity ($\bar{\beta}$ = 0.069), attractiveness ($\bar{\beta}$ = -0.048), and eye size ($\bar{\beta}$ = 0.047) also showed unique predictive power. All other facial characteristics were less important predictors of gullibility perceptions. Thus, these results suggest that gullibility perceptions are mostly based on resemblance to an angry facial expressions and babyfacedness.

Figure 3

Relationship between different facial characteristics and gullibility perceptions



Note. Coefficients were derived from an elastic net model with 10-fold cross-validation and 100 repeats.

way, any coefficient that is not exactly zero would ultimately be significant, making *p*-values uninformative (Troitzsch, 2014; see also Yarkoni & Westfall, 2017).

Discussion

Study 2 provided additional insights into which facial characteristics contribute to perceptions of gullibility. First, participants showed low, but significant agreement in their gullibility impressions, suggesting that there is a somewhat shared facial gullibility stereotype.

We also examined associations between gullibility impressions and a large set of facial characteristics that have previously been shown to form the basis for other trait impressions (Said et al., 2009; Stirrat & Perrett, 2010; Todorov et al., 2008). The strongest zero-order correlations were observed for resemblance to an angry facial expression, babyfacedness, resemblance to a disgusted facial expression, age, and resemblance to a happy facial expression. Using methods from machine learning (i.e., cross-validation, regularization), we also examined the unique predictive power of all facial characteristics in order to test which characteristics are the most important predictors of gullibility impressions. Our elastic net model explained 66% of the variance in gullibility impressions and predicted impressions to within 0.232 points on a 7-point scale. Resemblance to an angry facial expressions and babyfacedness contributed most to the predictive accuracy of the model. Thus, our results suggest that gullibility impressions were primarily based on the extent to which emotionally neutral faces resemble a facial expression of anger and, to a lesser extent, babyfacedness. These results are in line with the idea that gullibility impressions reflect perceptions of a low threat level of a target.

Finally, our results also showed that perceptions of gullibility are somewhat distinct from perceptions of trustworthiness and dominance, with many facial characteristics differently influencing judgments of the three traits. This suggests that the facial gullibility stereotype is dissociable from facial trustworthiness and dominance stereotypes.

Study 3

Study 3 extended our previous analyses in three ways. First, Study 1 and Study 2 focused on perceptions of gullibility in neutral faces. However, facial expressions—and smiles in particular—are common in everyday life (Martin et al., 2017), especially in situations in which people might be motivated to evaluate a person's gullibility (e.g., social networking or dating websites). We therefore examined the effect of smiling on gullibility perceptions. Smiles are usually seen as a signal of affiliation and positive intentions (Martin et al., 2017; Mehu et al., 2007). We therefore predicted that smiling individuals would be seen as more gullible than emotionally neutral individuals.

Second, previous studies suggest that young and old (vs. middle-aged) individuals are particularly likely to be targets of fraudulent behavior (Cohen, 2006; Jagatic et al., 2007; Sheng et al., 2010). In Study 2, we found a negative correlation between age and perceived gullibility. That is, young, but not old individuals were seen as more gullible. However, the oldest individual in the image set was only 50 years old. In Study 3, we therefore examined perceptions of gullibility across a wider age range. We tested whether young adults (ca. 20 to 30 years old) and old adults (ca. 70 to 80 years old) would be seen as more gullible than middle-aged adults (ca. 35 to 45 years old).

Third, we again examined gender differences in perceived gullibility. Study 2 showed that women were seen as more gullible than men when using a stimulus set of U.S. American individuals. Here, we tested whether this finding replicates in a stimulus set of German individuals who varied more in their age.

Methods

This study was preregistered and we explicitly mention if a decision or analysis was not specified a priori.

Participants. We again aimed to recruit at least 40 independent raters per image set (i.e., a total of 200 participants, Hehman, Xie, Ofosu, & Nespoli, 2018), with the final sample size being determined by how many participants completed the study within two weeks. In total, we recruited 212 first-year psychology students from a Dutch university who completed the study in return for partial course credit. In line with our preregistration, data from one participant (0.47%) who always indicated the same rating across all trials, and from two participants (0.95%) whose response time was shorter than 100 milliseconds on at least 10% of all trials were excluded from analyses, leaving a final sample of 209 participants (M_{age} = 20.19, SD_{age} = 2.39; 72.73% female, 26.79% male, 0.48% other). We also preregistered to exclude participants who indicate poor or basic English proficiency. However, due to a technical error, data on English proficiency was not assessed.

Stimuli. We selected 171 images of Caucasian adults from the FACES Database (Ebner et al., 2010). We selected images of 58 young targets (M_{age} = 24.2, SD_{age} = 3.4, 29 female), 56 middle-aged targets (M_{age} = 49.0, SD_{age} = 3.9, 27 female), and 57 old targets (M_{age} = 73.2, SD_{age} = 2.8, 29 female). All targets were photographed from a fixed distance against a uniform background. For each target, we selected an image in which they were displaying a neutral facial

expression and an image in which they were displaying a happy facial expression. Thus, the total number of images was 342 with gender (male vs. female) and age group (young vs. middle-aged vs. old) varying between targets and facial expression (neutral vs. smiling) varying within targets. We created five sets of stimuli. Each stimulus set contained either 63 or 72 images with approximately equal numbers of male and female targets and young, middle-aged, and old targets.

Procedure. Participants were randomly assigned to one stimulus set. They were asked to rate how gullible they think the person in the photo is on a scale from 1 (*not at all*) to 7 (*extremely*). We again used the same gullibility description as in Studies 1 and 2.

Results

Consensus. First, we computed intraclass correlation coefficients (*ICC*s) to estimate consensus in gullibility ratings across participants. Participants showed significant consensus in their judgments across the five sets of images, $ICC = 0.278$, $p < .001$ (set 1, $ICC = .320$, $p < .001$, 95% CI [.249, .414], set 2, $ICC = .210$, $p < .001$, 95% CI [.158, .282], set 3, $ICC = .327$, $p < .001$, 95% CI [.260, .415], set 4, $ICC = .254$, $p < .001$, 95% CI [.197, .333], and set 5, $ICC = .170$, $p < .001$, 95% CI [.123, .239]).

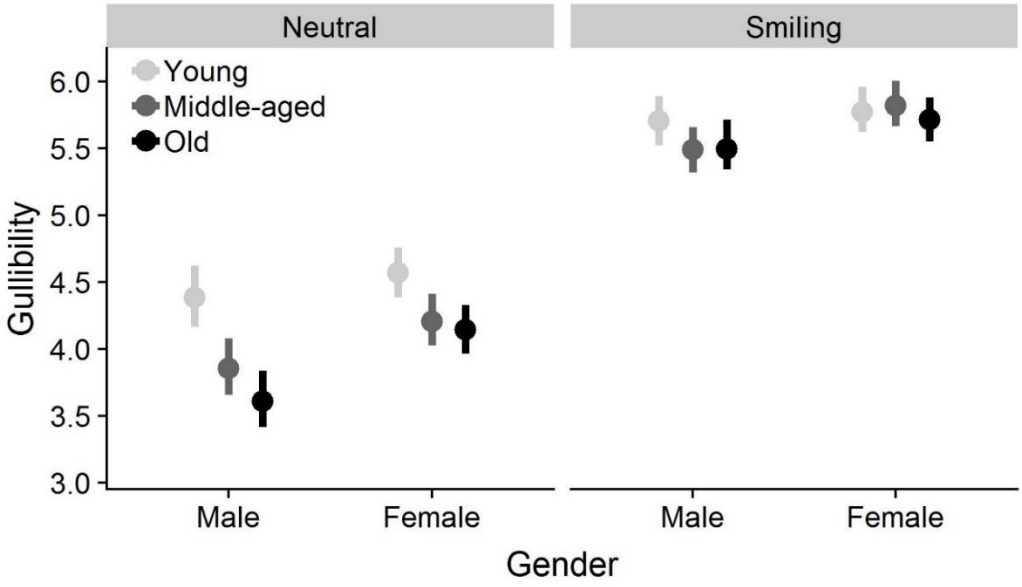
Gullibility perceptions. Next, we estimated a multilevel regression model in which we regressed gullibility ratings on facial expression (neutral vs. smiling), gender (male vs. female), and age (young vs. middle-aged vs. old, with middle-aged as the reference category; see Figure 4). We included random intercepts for participants and targets (to account for differences in average gullibility ratings across participants and targets) and random slopes for all predictors (to account for variance in effects across participants and targets). In line with our predictions, we found that smiling individuals were perceived as more gullible than neutral individuals, $\beta = 1.546$, $SE = 0.096$, $t(252.3) = 16.10$, $p < .001$, and women were seen as more gullible than men, $\beta = 0.218$, $SE = 0.049$, $t(200.8) = 4.47$, $p < .001$. We also found that young individuals were seen as more gullible than middle-aged individuals, $\beta = 0.183$, $SE = 0.060$, $t(162.0) = 3.04$, $p = .003$. However, contrary to our expectations, we found no significant difference in perceived gullibility between older and middle-aged individuals, $\beta = -0.017$, $SE = 0.062$, $t(140.9) = 0.28$, $p = .78$.

We also explored whether there were any interaction effects between facial expression, gender, and age. We again estimated a multilevel regression model with random intercepts per participant and target. This time, we included all two-way and three-way interaction effects

between the three variables. There were significant interaction effects between facial expression and age. To probe these interaction effects, we examined the effect of facial expression for young, middle-aged, and old targets. Smiling (vs. neutral) individuals were seen as more gullible when they were young, β = 1.264, *SE* = 0.100, $t(197.3) = 12.63$, $p < .001$, middle-aged, β = 1.626, *SE* = 0.110, $t(187.2) = 14.77$, $p < .001$, and old, β = 1.753, *SE* = 0.115, $t(182.9) = 15.23$, $p < .001$. However, the effect of a happy facial expression on gullibility perceptions was significantly smaller for young (vs. middle-aged) individuals, β = -0.321, *SE* = 0.123, $t(156.9) = 2.62$, $p = .010$, and significantly larger for old (vs. middle-aged) individuals, β = 0.269, *SE* = 0.124, $t(157.1) = 2.18$, $p = .031$. No other two-way or three-way interactions were significant.

Figure 4

The influence of gender, age, and facial expression on perceptions of gullibility



Note. Dots denote predicted gullibility ratings derived from a multilevel regression model. Error bars denote bootstrapped 95% confidence intervals.

Discussion

In the current study, we examined how gender, age, and facial expression (i.e., smiling vs. neutral) contribute to perceptions of gullibility. We found that women were seen as more gullible than men and that young individuals (ca. 20 to 30 years old) were seen as more gullible than middle-aged individuals (ca. 35 to 45 years old). Contrary to our expectations, we did not find that old individuals (ca. 70 to 80 years old) were seen as more gullible than middle-aged

individuals. Finally, results showed that smiling increased targets' perceived gullibility. In fact, smiling individuals were seen as more gullible irrespective of their gender or age.

General Discussion

What makes a person look gullible? Across three preregistered studies using both data-driven and theory-driven methods, we examined how people form impressions of gullibility based on others' facial appearance. We investigated (a) whether people share stereotypes regarding what a gullible person looks like, (b) how gullibility impressions are related to impressions of core dimensions of social perception (i.e., trustworthiness and dominance), and (c) which facial characteristics people rely on to form gullibility impressions.

We consistently find that people somewhat agree on what a gullible person looks like. In the reverse correlation task in Study 1, people repeatedly selected which random noise pattern increased perceived gullibility when superimposed on a face. Averaging all selected and unselected noise patterns clearly produced faces with different appearances (see Figure 1). This suggests the presence of a shared understanding of what a gullible person looks like because the absence of a common stereotype would lead to a random or highly idiosyncratic selection of noise patterns, which, when averaged, would result in two identical faces. The idea that people have a somewhat shared representation of a gullible face was also corroborated by analyzing gullibility impressions of naturalistic faces. Results showed significant levels of consensus in gullibility judgments (*ICC* = .078, when judging emotionally neutral faces in Study 2, *ICC* = .278, when judging neutral and smiling faces in Study 3). Hehman and colleagues (2017) found ICCs of .234 for trustworthiness and .131 for dominance in their analysis of 700,000 trait impressions. Thus, the current results suggest that consensus in gullibility impressions is low to average compared to other personality impressions. Taken together, we find that people show some consensus in what a gullible person looks like and their level of consensus is comparable to the consensus observed in judgments of other personality dimensions.

While people can infer a variety of traits from faces (Oosterhof & Todorov, 2008), not every trait is characterized by a unique facial appearance. In fact, previous studies have shown that faces are primarily judged along three dimensions: trustworthiness (representing judgments of a target's intentions), dominance (representing judgments of a target's abilities), and attractiveness (representing judgments of a target's mate valu;e Oosterhof & Todorov, 2008; Sutherland et al., 2013). How were perceptions of gullibility related to these core dimensions of

social perception? Gullibility is conceptually associated with approachability and positive intentions, but also with low competence and harmlessness (Evans & van de Calseyde, 2018). In line with this view, perceptions of gullibility were positively related to perceptions of trustworthiness, but negatively to perceptions of dominance (Studies 1 and 2).

We found relatively strong zero-order correlations between perceptions of gullibility and trustworthiness (r = .47) and between perceptions of gullibility and dominance (r = -.68). Yet, we also found that many facial features were differently related to perceptions of gullibility, trustworthiness, and dominance. For example, the extent to which neutral faces resembled an expression of anger influenced perceptions of gullibility (but not trustworthiness), whereas resemblance to a happy facial expression influenced perceptions of trustworthiness (but not gullibility). Similarly diverging patterns were observed when comparing gullibility and dominance perceptions. These results suggest that perceptions of gullibility, trustworthiness, and dominance represent dissociable facial stereotypes.

Which facial cues do people rely on to form gullibility judgments? Both data-driven (Study 1) and theory-driven approaches (Study 2) yielded converging results. In Study 1, we relied on reverse-correlation to visualize the stereotypical facial appearance of gullible individuals. A visual inspection of the facial features that distinguished between high and low gullibility showed that smiling, baby-faced individuals were seen as gullible. This was confirmed in Study 2 where we examined correlations between impressions of gullibility and a large set of facial characteristics. Individuals with babyfaced features and individuals who appeared happy were perceived as more gullible, whereas older individuals and individuals who appeared angry or disgusted were perceived as less gullible.

Many facial characteristics are correlated, which makes it difficult to disentangle their unique associations with impressions (Jaeger & Jones, 2020; A. L. Jones, 2019). To understand which facial characteristics are most important for explaining gullibility impressions, we simultaneously modeled relationships with 16 facial characteristics. We relied on cross-validated elastic net regression models (Zou & Hastie, 2005), which shrinks uninformative predictors to create parsimonious models and avoid overfitting. The model was able to predict gullibility impressions to within 0.23 points on a 7-point scale and explained 66% of the variance. This performance is similar to the performance of other models that were trained on a similar set of facial characteristics to predict trustworthiness impressions (Jaeger & Jones, 2020). Results

showed that resemblance to an angry facial expressions and babyfacedness were the most informative predictors. Targets whose emotionally neutral faces appeared angry were seen as less gullible, whereas targets with a babyfaced appearance were seen as more gullible. Thus, even though we found correlations between many facial characteristics and gullibility impressions, resemblance to an angry facial expressions and babyfacedness were by far the most predictive characteristics.

A host of prior work has examined the antecedents and consequences of baby-ish facial features (Berry & Zebrowitz McArthur, 1985; Zebrowitz & McDonald, 1991; Zebrowitz & Montepare, 1992). However, our studies go beyond this work by examining associations between gullibility impressions and a host of other facial characteristics (including babyfacedness). Our results can be seen as a strong test of the importance of babyfacedness (relative to other facial characteristics) for predicting gullibility impressions. In fact, even though we do find that babyfacedness was one of the most informative predictors of gullibility impressions, it was not the most informative one. Thus, in line with prior work (Berry & Zebrowitz McArthur, 1985, 1986; Zebrowitz & Montepare, 1992), we find that babyfaced individuals are seen as more gullible, but we also find that resemblance to an angry facial expression is an even more important facial characteristic when forming impressions of gullibility.

Finally, going beyond emotionally neutral faces, Study 3 investigated the influence of smiles on gullibility perceptions. Smiles are typically interpreted as signals of affiliation and positive intentions (Martin et al., 2017; Mehu et al., 2007) and should therefore have a positive influence on perceived gullibility. Indeed, we found that irrespective of their gender or age, smiling individuals were seen as more gullible that emotionally neutral individuals. Taken together, findings from all three studies suggest that when forming impressions of gullibility, people rely on a suite of facial cues that are seen as indicating high levels of approachability and low levels of threat.

Limitations and Future Directions

The present studies showed that people rely on various facial features to form judgments of gullibility. Do people actually rely on these cues, for example, when deciding whom to target for financial scams? Previous work on the demographic characteristics of fraud victims suggest that this might indeed be the case. For instance, women and young people are more likely to fall

victim to scams (Jagatic et al., 2007; Sheng et al., 2010). However, the influence of gullibility perceptions probably extends beyond situations in which one party is attempting to exploit another. In many mixed-motive situations—such as the prisoner's dilemma (Camerer, 2003; Rapoport & Chammah, 1970)—mutual gains are only realized if both parties trust the other to cooperate. No matter whether a person wants to establish a cooperative relationship or exploit the other's trust, a necessary requirement for both goals is that the other party cooperates. Additional studies are needed to test whether people with a gullible facial appearance are indeed more likely to be sought out in such situations.

Future studies should also examine the accuracy of gullibility perceptions. It is possible that people agree on what a gullible person looks like and that they rely on their gullibility inferences when making decisions even though these judgments are not correlated with a person's actual gullibility. In fact, many studies have shown that people rely on impressions of trustworthiness, competence, or other personality traits, even though the accuracy of these impressions is very low (Bonnefon et al., 2017; Jaeger et al., 2020). More work is needed to test whether gullibility perceptions contain a kernel of truth.

In all three studies, participants indicated their gullibility perceptions by rating to what extent they believe that a target is very trusting, naïve, and easily believes others (high gullibility) or very distrusting, skeptical, and easily suspicious of others (low gullibility). Although this broad definition conflates gullibility and trust, which some view as distinct concepts (Rotter, 1967; Teunisse et al., 2019; Yamagishi et al., 1999), it should be noted that even if trust and gullibility represent different personality characteristics, this does not necessarily mean that they represent different facial stereotypes. Additional studies are needed to test whether people associate traits such as gullibility, trustfulness, and naïveté with different facial appearances or whether judgments of these traits are driven by the same facial features.

Prior work on scamming suggests that the elderly are at higher risk of falling victim of scams (Burnes et al., 2017; Cohen, 2006). In contrast, we did not find that older adults were seen as more gullible than younger or middle-aged adults (Study 3). Thus, additional work is needed to understand the relationship between age and (perceived) gullibility. Finally, future studies should also focus on the relationship between emotional expressions and perceptions of gullibility. We examined the effect of smiling because smiles are one of the most common facial expressions in everyday life (Martin et al., 2017). However, in Study 2 we found that gullibility

impressions were best predicted by resemblance to an angry facial expression (in emotionally neutral faces). Thus, angry facial expressions may lead to lower perceptions of gullibility and this effect may be even stronger than the effect of smiling.

Conclusion

Across three studies, using both theory-driven and data-driven methods, we examined how people form gullibility impressions based on facial appearance. Results showed that (a) people show some agreement in what a gullible person looks like, (b) gullibility judgments are based on different facial features than judgments of trustworthiness and dominance, suggesting that they represent dissociable facial stereotypes, and (c) people rely on facial cues that are seen as indicators of positive intentions (e.g., smiles) and low levels of threat (e.g., babyfacedness) when forming gullibility impressions. These findings are in line with the view that gullible people are perceived as approachable and harmless.

References

Berry, D. S., & Zebrowitz McArthur, L. (1985). Some components and consequences of a babyface. *Journal of Personality and Social Psychology*, *48*(2), 312–323. https://doi.org/10.1037/0022-3514.48.2.312

Berry, D. S., & Zebrowitz McArthur, L. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin*, *100*(1), 3–18. https://doi.org/10.1037/0033-2909.100.1.3

Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2017). Can we detect cooperators by looking at their face? *Current Directions in Psychological Science*, *26*(3), 276–281. https://doi.org/10.1177/0963721417693352

Burnes, D., Henderson Jr, C. R., Sheppard, C., Zhao, R., Pillemer, K., & Lachs, M. S. (2017). Prevalence of financial fraud and scams among older adults in the United States: A systematic review and meta-analysis. *American Journal of Public Health*, *107*(8), E13–E21. https://doi.org/http://dx.doi.org/10.2105/AJPH.2017.303821

Button, M., Lewis, C., & Tapley, J. (2014). Not a victimless crime: The impact of fraud on individual victims and their families. *Security Journal*, *27*(1), 36–54. https://doi.org/10.1057/sj.2012.11

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Cohen, C. A. (2006). Consumer Fraud and the Elderly. *Journal of Gerontological Social Work*, *46*(3–4), 137–144. https://doi.org/10.1300/j083v46n03_08

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. J., Langner, O., & Van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, *19*(10), 978–980. https://doi.org/10.1111/j.1467-9280.2008.02186.x

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, *42*(1), 351–362. https://doi.org/10.3758/BRM.42.1.351

Evans, A. M., & van de Calseyde, P. P. F. M. (2018). The reputational consequences of generalized trust. *Personality and Social Psychology Bulletin*, *44*(4), 492–507. https://doi.org/10.1177/0146167217742886

Gomulya, D., Wong, E. M., Ormiston, M. E., & Boeker, W. (2017). The role of facial appearance on CEO selection after firm misconduct. *Journal of Applied Psychology*, *102*(4), 617–635. https://doi.org/http://dx.doi.org/10.1037/apl0000172

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.

Hehman, E., Ofosu, E. K., & Calanchini, J. (2020). Using environmental features to maximize prediction of regional intergroup bias. *Social Psychological and Personality Science*, 1–9. https://doi.org/10.1177/1948550620909775

Hehman, E., Sutherland, C. A. M., 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), 513–529. https://doi.org/10.1037/pspa0000090

Hehman, E., Xie, S. Y., Ofosu, E. K., & Nespoli, G. A. (2018). *Assessing the point at which averages are stable: A tool illustrated in the context of person perception*. https://psyarxiv.com/2n6jq/

Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in Psychology*, *4*, 1–8. https://doi.org/10.3389/fpsyg.2013.00386

Jaeger, B., & Jones, A. L. (2020). *Which facial features are central in impression formation?* https://doi.org/10.31234/osf.io/9c57t

Jaeger, B., Oud, B., Williams, T., Krumhuber, E. G., Fehr, E., & Engelmann, J. B. (2020). *Can people detect the trustworthiness of strangers based on their facial appearance?* https://doi.org/10.31234/osf.io/ayqeh

Jaeger, B., Sleegers, W. W. A., Evans, A. M., Stel, M., & van Beest, I. (2019). The effects of facial attractiveness and trustworthiness in online peer-to-peer markets. *Journal of Economic Psychology*, *75 Part A*, 102125. https://doi.org/10.1016/j.joep.2018.11.004

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM*, *50*(10), 94–100. https://doi.org/10.1145/1290958.1290968

Jones, A. L. (2019). *Beyond average: Using face regression to study social perception*. https://doi.org/10.31234/osf.io/dpmzq

Jones, B. C., DeBruine, L. M., Flake, J. K., Aczel, B., Adamkovic, M., Alaei, R., Alper, S., Álvarez Solas, S., Andreychik, M. R., Ansari, D., Arnal, J. D., Babincák, P., Balas, B., Baník, G., Barzykowski, K., Baskin, E., Batres, C., Beaudry, J. L., Blake, K. R., … Chartier, C. R. (2021). To which world regions does the valence-dominance model of social perception apply? *Nature Human Behaviour*. https://psyarxiv.com/n26dy/

Krueger, J. I., Vogrincic-Haselbacher, C., & Evans, A. M. (2019). We need a credible theory of gullibility. In J. P. Forgas & R. F. Baumeister (Eds.), *Homo Credulus: The social psychology of gullibility*. Taylor & Francis.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 159–160. https://doi.org/10.1053/j.sodo.2009.03.002

Ling, L., Luo, D., & She, G. (2019). Judging a book by its cover: The influence of physical attractiveness on the promotion of regional leaders. *Journal of Economic Behavior and Organization, 158*, 1–14. https://doi.org/10.1016/j.jebo.2019.01.005

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces.*

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Martin, J., Rychlowska, M., Wood, A., & Niedenthal, P. (2017). Smiles as multipurpose social signals. *Trends in Cognitive Sciences, 21*(11), 864–877. https://doi.org/10.1016/j.tics.2017.08.007

Mehu, M., Grammer, K., & Dunbar, R. I. M. (2007). Smiles when sharing. *Evolution and Human Behavior, 28*(6), 415–422. https://doi.org/10.1016/j.evolhumbehav.2007.05.010

Modic, D., & Anderson, R. (2015). It's all over but the crying: The emotional and financial impact of internet Fraud. *IEEE Security & Privacy, 13*(5), 99–103. https://doi.org/10.1109/MSP.2015.107

Oh, D., Buck, E. A., & Todorov, A. (2019). Revealing hidden gender biases in competence impressions of faces. *Psychological Science, 30*(1), 65–79. https://doi.org/10.1177/0956797618813092

Oliveira, M. J. B., Garcia-Marques, T., Dotsch, R., & Garcia-Marques, L. (2019). Dominance and competence face to face: A reverse-correlation approach. *European Journal of Social Psychology*. https://doi.org/10.1002/ejsp.2569

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, *18*(11), 566–570. https://doi.org/10.1016/j.tics.2014.09.007

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

Ormiston, M. E., Wong, E. M., & Haselhuhn, M. P. (2017). Facial-width-to-height ratio predicts perceptions of integrity in males. *Personality and Individual Differences*, *105*, 40–42. https://doi.org/10.1016/j.paid.2016.09.017

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Rapoport, A., & Chammah, A. M. (1970). *Prisoner's Dilemma - A Study in Conflict and Cooperation*. The University of Michigan Press.

Revelle, M. W. (2011). Package ' psych. *October*, 1–250.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*, 651–665. https://doi.org/10.1111/j.1467-6494.1967.tb01454.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. https://doi.org/10.1037/a0014681

Schniter, E., & Shields, T. W. (2020). Gender, stereotypes, and trust in communication. *Human Nature*, *31*(3), 296–321. https://doi.org/10.1007/s12110-020-09376-3

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010). Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*, 373–382. https://doi.org/10.1145/1753326.1753383

Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What is typical is good: the influence of face typicality on perceived trustworthiness. *Psychological Science*, *26*(1), 39–47. https://doi.org/10.1177/0956797614554955

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, *21*(3), 349–354. https://doi.org/10.1177/0956797610362647

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118. https://doi.org/10.1016/j.cognition.2012.12.001

Teunisse, A. K., Case, T. I., Fitness, J., & Sweller, N. (2019). I should have known better: Development of a self-report measure of gullibility. *Personality and Social Psychology Bulletin*, 014616721985864. https://doi.org/10.1177/0146167219858641

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623–1626. https://doi.org/10.1126/science.1110589

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*(1), 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Troitzsch, K. (2014). Analysing simulation results statistically: Does significance matter? In D. Francisca, G. Pereira Dimuro, & H. Coelho (Eds.), *Interdisciplinary Applications of Agent-Based Social Simulation and Modeling* (pp. 88–105). IGI Global.

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

Witt, P. (2018). *The top frauds of 2018*. Federal Trade Commission.

Yamagishi, T., Kikucki, M., & Kosugi, M. (1999). Trust, gullibility, and social intelligence. *Asian Journal of Social Psychology*, *2*(1), 145–161. https://doi.org/10.1111/1467-839X.00030

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Zebrowitz, L. A., Fellous, J. M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review*, *7*(3), 194–215. https://doi.org/10.1207/S15327957PSPR0703_01

Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior*, *15*(6), 603–623. https://doi.org/10.1007/BF01065855

Zebrowitz, L. A., & Montepare, J. M. (1992). Impressions of babyfaced males and females across the lifespan. *Developmental Psychology*, *28*(6), 1143–1152.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *67*(5), 768. https://doi.org/10.1111/j.1467-9868.2005.00527.x