2024-03

Facial first impressions following a prison sentence: Negative shift in trait ratings but the same underlying structure

Coutts, C

10.1016/j.jesp.2023.104568
Journal of Experimental Social Psychology
Elsevier

Facial first impressions following a prison sentence: Negative shift in trait ratings but the same underlying structure

Coral M. Coutts, Christopher A. Longmore, & Mila Mileva

School of Psychology, Faculty of Health: Medicine, Dentistry and Human Sciences, University of Plymouth, UK

Correspondence to:
Mila Mileva, School of Psychology, Faculty of Health: Medicine, Dentistry and Human Sciences, University of Plymouth, Drake Circus, Plymouth PL4 8AA, United Kingdom. E-mail: mila.mileva@plymouth.ac.uk

Funding: This work was supported by a British Academy Postdoctoral Fellowship (PF20\100034) awarded to Mila Mileva.

Data availability statement: The stimulus set and the data that support the findings of these studies are openly available via the Open Science Framework at https://osf.io/6jrx4/?view_only=c7c6698720164b729507a4233e3ba097

Conflict of interest disclosure: The authors declare no conflict of interest.

Ethics approval statement: The experimental procedures for all reported studies conform to relevant legislation (Declaration of Helsinki, GDPR) and were approved by the School of Psychology Ethics Committee at the University of Plymouth. Informed consent was provided by all participants prior to participation.

Abstract

The first impressions we form of unfamiliar others can often guide many important decisions such as whether someone is guilty of a crime or the severity of their sentence, even in the presence of more relevant information. While most of the current work in this context has focused on their impact during trial proceedings and sentencing, little is known about the potential impact of first impressions following a guilty sentence and the success of the subsequent reintegration into society. Here, we used a data-driven approach to address this question by first collecting unconstrained spontaneous impressions from two groups of perceivers – one group believed that the identities they were presented with had received a prison sentence, whereas the other received no additional semantic information (Study 1). This then allowed us to establish the most prevalent traits people refer to when describing their first impressions in this context and to reveal the underlying structure of these impressions using an Exploratory Factor Analysis (Study 2). We find a substantial negative shift in social evaluation following the knowledge of a prison sentence, both in terms of spontaneous descriptions and specific trait ratings. However, this additional contextual information did not affect the underlying structure of first impressions. These findings support recent social evaluation theories arguing for a more complex interplay between bottom-up visual and top-down semantic or contextual cues during the formation of facial first impressions but also reveal important constraints to the impact of such cues on the core impression formation processes.

Keywords: first impressions, social evaluation, offender perception, prison sentence

Introduction

First impressions are an inevitable part of our everyday social interactions. They are formed within a few (hundred) milliseconds with no effort whatsoever (Willis & Todorov, 2006) and have consistently been shown to reflect evaluations along two fundamental dimensions, valence (approachability) and dominance (competence), with some evidence for a third additional attractiveness dimension (Oosterhof & Todorov, 2008; Sutherland et al., 2013). While first impressions are unlikely to reflect real and stable personality characteristics (Lavan, Mileva, et al., 2021; Todorov et al., 2015), they have been consistently shown to predict important, real-world social outcomes in a range of contexts such as politics (Sussman et al., 2013; Todorov et al. 2005), employment (Fruhen et al., 2015; Linke et al., 2016) and economics (Duarte et al., 2012; Rule & Ambady, 2010, 2011). First impressions can also guide important forensic and judicial decisions even when we have access to much more relevant information (Jaeger et al., 2020; Wilson & Rule, 2015, 2016). For example, possessing facial features that are perceived to be 'criminal', relating to perceptions of social dominance and threat (Funk et al., 2017), increase the chance of being selected in a police line-up and receiving a guilty verdict, irrespective of the evidence presented (Flowe & Humphries, 2011; Funk & Todorov, 2013). Judgements of trustworthiness are also extremely important, with evidence showing that untrustworthy-looking defendants are pronounced guilty with less evidence, higher confidence rates and are also more likely to receive a death penalty sentence compared to defendants perceived as being more trustworthy (Porter et al., 2010; Wilson & Rule, 2015, 2016). Here, we explore potential top-down influences on facial first impressions and their underlying structure within a similar forensic context.

By definition, first impressions are quick, zero-acquaintance judgements. However, we are often presented with bits of information about a person before meeting them and this semantic knowledge has the scope to influence our evaluations. Traditionally, theories and models of facial first impressions have taken a more feed-forward approach. This is reflected in the overgeneralisation processes that form the basis of face-based first impressions (Zebrowitz & Montepare, 2008). For example, neutral faces resembling subtle characteristics of positive affect are generally evaluated more favourably (Said et al., 2009) and adult faces with features resembling those of infant faces are attributed related

qualities such as being submissive, naïve and less competent (Zebrowitz & Montepare, 1992). Thus, research in this domain has generally focused on identifying the bottom-up impact of different features or patterns within the human face on social evaluation, leaving any potential top-down effects coming from the perceiver or the context relatively unexplored.

Outside of face-based impressions, the opposite pattern has emerged in social psychological research, where the main focus has been on the perceiver or the interaction between the perceiver and their target. Here, there are many accounts of implicitly or explicitly activated affect, personality traits or stereotypes guiding social evaluation processes (Martin et al., 1990; Newman et al., 1996; Newman & Uleman, 1990). For example, priming the trait reckless might lead to a more negative attitude towards a potentially dangerous behaviour (and the person displaying this behaviour) compared to priming the trait brave (Higgins et al., 1977). Information inconsistent with an existing stereotype can also play an inhibitory role and produce weaker relevant spontaneous trait inferences (Wigboldus et al., 2003) and these biases do not necessarily rely on chronic stereotypes but can even be produced with arbitrarily assigned groups (Otten & Moskowitz, 2000).

Similar contextual top-down influences have also been shown by making the concept of aggression more accessible via subtle priming which is particularly relevant to the context of the present work. In one of the first such studies, Srull and Wyer (1979) use the ‘Donald paradigm’ where perceivers were first asked to unscramble a number of sentences, some of which described aggressive behaviour. Then, in a seemingly unrelated study, they were presented with a short description of a person named Donald that was ambiguous in terms of aggressive behaviour. Perceivers who were primed with aggressive content sentences then perceived Donald’s behaviour as more aggressive and hostile. This aggression priming can further interact with group stereotypes, where even out-group members associated with no explicit aggressiveness stereotypes can be perceived more negatively following priming (Otten et al., 2007).

Although such spontaneous trait inferences and facial first impressions have been studied somewhat independently of one another, it is likely that they rely on similar mechanisms. There is already evidence that the same two-dimensional structure reported for facial impressions can be seen in more general social cognition models of warmth and

competence that capture our evaluation of stereotypes, events, and even objects (Fiske et al., 2007; Wiggins, 1979; Wojciszke, 1994). It is therefore reasonable to assume that similar top-down influences can also guide facial first impressions. In fact, there is already some work showing that a perceiver's level of prejudice can affect their mental representation of the physical appearance of outgroup members. Dotsch et al. (2008), for instance, compared classification images created by Dutch perceivers with high, moderate and low levels of prejudice against Moroccan outgroup target identities and found that the images of the high prejudice group were perceived as significantly more criminal and less trustworthy than the images created by the moderate or the low groups. The perception of facial cues can also be affected by social context and the assumed relationship between a target and an observer in particular (Tuk et al., 2009).

Recent developments and theories in the field are beginning to acknowledge these processes in facial impressions, with recent research (e.g., Freeman et al., 2020; Oh et al., 2021) presenting evidence for a more complex interplay between bottom-up visual and contextual (semantic or affective) top-down cues that guide our perception (Wildman & Ramsey, 2021). For example, the relationship between the two underlying dimensions in first impressions, trustworthiness and dominance, could be transformed with access to demographic cues such as age and gender (Hehman et al., 2014; Mileva et al., 2019; Sutherland et al., 2014). It is, therefore possible, that access to further semantic knowledge might produce a similar re-structuring or prioritising of the underlying first impression dimensions. So far, however, these effects have only been explored in certain independent traits. Here, we take a more holistic approach by exploring the effect of the interplay between bottom-up visual cues extracted from the human face and existing top-down semantic information that guides our social judgements on the underlying structure of facial first impressions.

To do that, we focus on a more forensically-relevant context and consider the effect of ex-offender status awareness on facial first impressions which could then have serious implications for offender rehabilitation and reintegration into society. While previous studies have already established how facial first impressions might guide decision-making within the criminal justice system, the way in which offenders are perceived outside of this context remains unclear (Hirschfield & Piquero, 2010). Given the vast evidence for the impact of first impressions in pre-sentencing procedures, it is likely that these same

perceptions could affect the reintegration of offender's post-release as well and currently, we know very little about how the knowledge of a criminal conviction might influence the first impressions attributed to ex-offenders. As a result, the potential impact of these impressions on the success of their reintegration has also remained largely unexplored (Austin & Hardyman, 2004).

Criminal conviction history is an extremely important factor within the forensic psychology literature that has been shown to affect many different aspects of court trial proceedings and sentencing decisions – from the probability of receiving a guilty verdict to pre-trial jury selection (Atkin & Cramer, 2012). For example, there is evidence for higher conviction rates when jurors are aware of previous criminal convictions (Eisenberg & Hans, 2009). These are further amplified by weaker prosecuting evidence, juror instructions that conviction history can serve to judge present guilt and by the degree of similarity between any previous crimes and the present one/s (Greene & Dodge, 1995; Wissler & Saks, 1985). Such studies demonstrate that the knowledge of a conviction can lead to the formation of damaging impressions of ex-offenders based on their criminal history, which can negatively impact them in subsequent trials.

Some propose that this ex-offender bias could partly be minimised through the process of juror selection. In fact, there is already a considerable amount of work that focusses on identifying such stigmatising attitudes in juror candidates who can then be de-selected from the jury in order to ensure a fairer and more impartial trial (e.g., Atkin & Cramer, 2012). However, it is likely that adverse ex-offender biases are not limited to the courtroom but also permeate society with scarce (if any) opportunities to control for them. For instance, the stigmatisation they face could reduce their chances of obtaining housing or legitimate employment which could be further destabilising and lead to repetitive criminogenic attitudes and behaviours (Levenson & Cotter, 2005; Pritikin, 2008; Wormith et al., 2007).

What is more, previous criminal convictions are often disclosed. Within the UK, those convicted of child sexual offences are put on a national database accessible to parents, guardians, and carers of children through the child sex offender disclosure scheme (Home Office, 2013). The police may also decide to disclose a conviction to new and existing partners or people within the same household, if the household is shared (Nacro, 2022).

Offenders must often disclose criminal convictions to employers and many convictions are flagged up through Disclosure and Barring Service screenings (DBS) which thus causes them to fail and notifies the employer (GOV.UK, 2022). This increases the likelihood of any ex-offender biases found in jury decision-making to extend to post-release reintegration attempts, highlighting the need to establish how criminal conviction history can influence our behaviours and attitudes in everyday life.

The present set of studies aims to establish the extent of such biases making use of the substantial literature on facial first impressions (see Sutherland & Young, 2022 and Todorov et al., 2015 for reviews). Given how much we already know about the effects of criminal conviction history in more forensic and applied contexts, it is surprising that little is known about its effects on first impressions and, more theoretically important, on their underlying structure. Here, we take a data-driven approach to establish how knowledge of prior criminal convictions might change the overall underlying structure of first impressions as well as affect some of the most prevalent first impression traits. In Study 1, we collect spontaneous unconstrained descriptors attributed to a set of unfamiliar faces in order to establish the most commonly used and referred to first impression traits. Critically, before providing these descriptors, some participants received the additional information that the images they will be presented with will show people who are currently serving time in prison, whereas other participants did not receive any additional instructions. This allowed us to explore any differences in the valence of first impressions driven purely by this contextual knowledge and not by the physical properties of face images. In Study 2, we collected ratings of the key first impression traits identified in Study 1 and used Exploratory Factor Analysis (EFA) to identify any differences in the underlying structure of first impressions brought about by being aware of a previous prison sentence.

Based on evidence for ex-offender stigma, we expect that criminal offending will lead to a more negative social evaluation based both on spontaneous descriptors (Study 1) and specific trait ratings (Study 2). Given the prominent role and outcomes of trustworthiness and attractiveness judgements in related forensic contexts such as eyewitness testimony and court sentencing decisions (Efran, 1984; Porter et al., 2010; Sigall & Ostrove, 1974; Wilson & Rule, 2015), it is also expected that knowledge of criminal offending will results in significant decreases in ratings of these two traits. As first impressions have been shown to have a wide range of social consequences (Milazzo & Mattes, 2016; Olivola & Todorov,

2010; Rule & Ambady, 2011; Zebrowitz & McDonald, 1991), any evidence for a negative shift in the evaluation of (thought to be) criminal offenders might reflect the way they are treated by society upon their reintegration in many key areas of 'normal' life and would likely have an impact on the success of their rehabilitation. We report all measures, manipulations, and exclusions in these studies. Sample size was determined before any data analysis was conducted for Study 1, however ratings of a larger set of images were collected for Study 2 after the data from a smaller set were analysed (see Participants section in Study 2 for further details). Sample size was based on the number of free descriptors used by Oosterhof & Todorov (2008) for Study 1 and Hehman et al. (2018) who suggest that between 20-30 raters are needed in order to achieve a stable mean trait impression rating from faces, with 95% confidence at a corridor of stability of +/- 0.50 on a 1-7 Likert scale for each trait for Study 2.

Study 1

In this study, participants were presented with a number of unfamiliar face images and were asked to freely describe their initial spontaneous impressions of the presented identities. Some participants were instructed that the identities they will encounter are currently serving time in prison while other participants did not receive any additional instructions. This allowed us to establish how this key piece of inferential information might affect any subsequent judgements made about the person irrespective of any differences within the physical structure of the faces. These free descriptors informed the initial step in our data-driven approach that aimed to establish the most common words and traits people use to describe their first impressions in everyday life.

Method

Participants

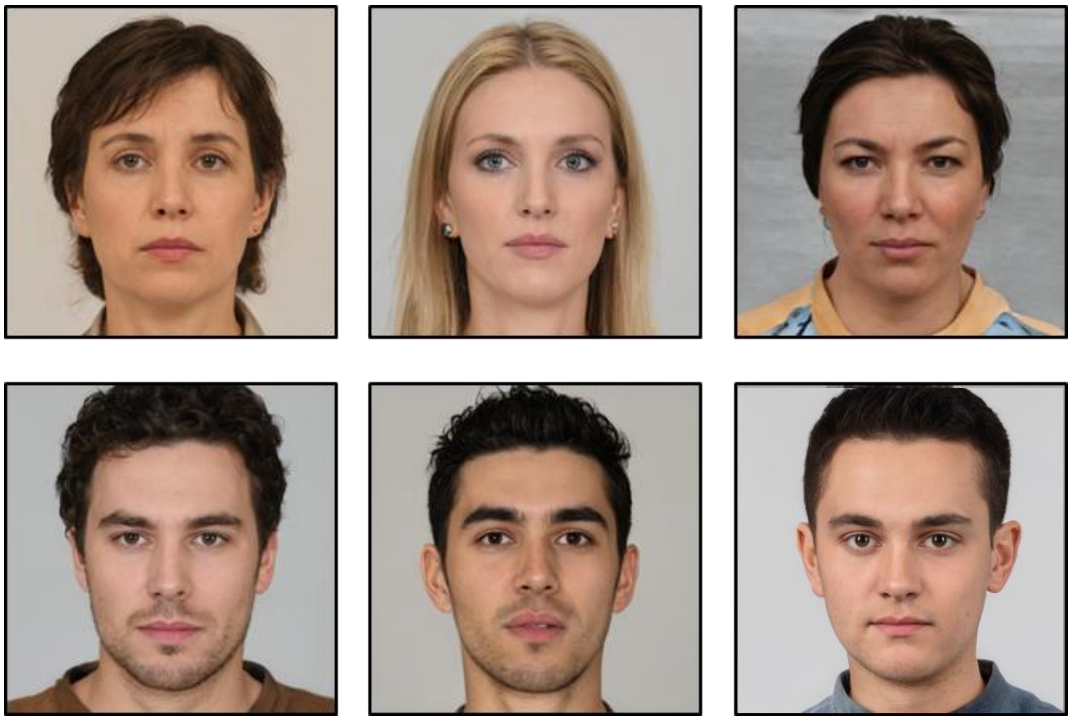
A total of 20 participants (19 female, *M* = 21, age range = 18-40) from the University of X were recruited for the study. All were over the age of 18 years old and had normal or corrected-to-normal vision. Participants received course credit for their participation. Experimental procedures were approved by the School of Psychology Ethics Committee at the University of X and informed consent was provided prior to participation.

Materials

Stimuli consisted of 30 face images (15 female) created using StyleGAN2 technology[1] (Karras et al. 2020). All images were resized to 400 x 400 pixels and were presented in colour. All face images were front-facing with a neutral emotional expression and all identities had minimal make-up and accessories (see Figure 1 for examples). All 30 faces appeared Caucasian to minimise racial biases associated with criminality perception (Eberhardt et al., 2004) as well as social evaluation (Zebrowitz et al., 1993).

Figure 1

Examples of the 30 StyleGAN2 Photo Stimuli Used In Studies 1 and 2.



Participants were also asked to complete an 11-item questionnaire designed to assess their personal beliefs, attitudes, and opinions on three core aspects of prisoner rehabilitation: personal relationships, employment, and potential future sentencing. The questionnaire comprised of five main questions: 1) How likely would you be to employ someone with a criminal conviction? 2) How comfortable would you feel working alongside someone who had a criminal conviction? 3) How willing would you be to have any kind of relationship (i.e., friend, romantic partner, close acquaintance) with someone who held a criminal conviction? 4) How much more likely would you be to believe someone is guilty of a crime they were accused of if they had already been convicted of a crime in the past? 5) How much do you believe those with a criminal conviction can be rehabilitated back into

[1] All images were downloaded from the https://thispersondoesnotexist.com/ website.

society? Participants were asked to respond to all questions using a 7-point scale and they also had the opportunity to elaborate on their ratings using open text entry boxes.

Procedure

The experiment was created and hosted on the online testing platform, Qualtrics (Provo, UT, USA). In the beginning of each experimental session, participants were randomly assigned to one of two conditions: the prison sentence awareness condition, whereby they were instructed that they would be presented with images of people currently serving time in prison or the control condition, whereby they were not provided with any additional information about the identities depicted in the face images. Critically, participants in both conditions were presented with the exact same face images. For the free descriptor task, each participant was sequentially presented with 30 face images and was asked to freely describe their first impression of each person using a textbox located under the image. Face presentation order was randomised individually for each participant. The task was not timed and participants were encouraged to provide as many descriptors as possible. Following the free descriptor task, all participants had an unlimited time to complete the 11-item questionnaire. The entire study took approximately 30 minutes to be completed.

Results & Discussion

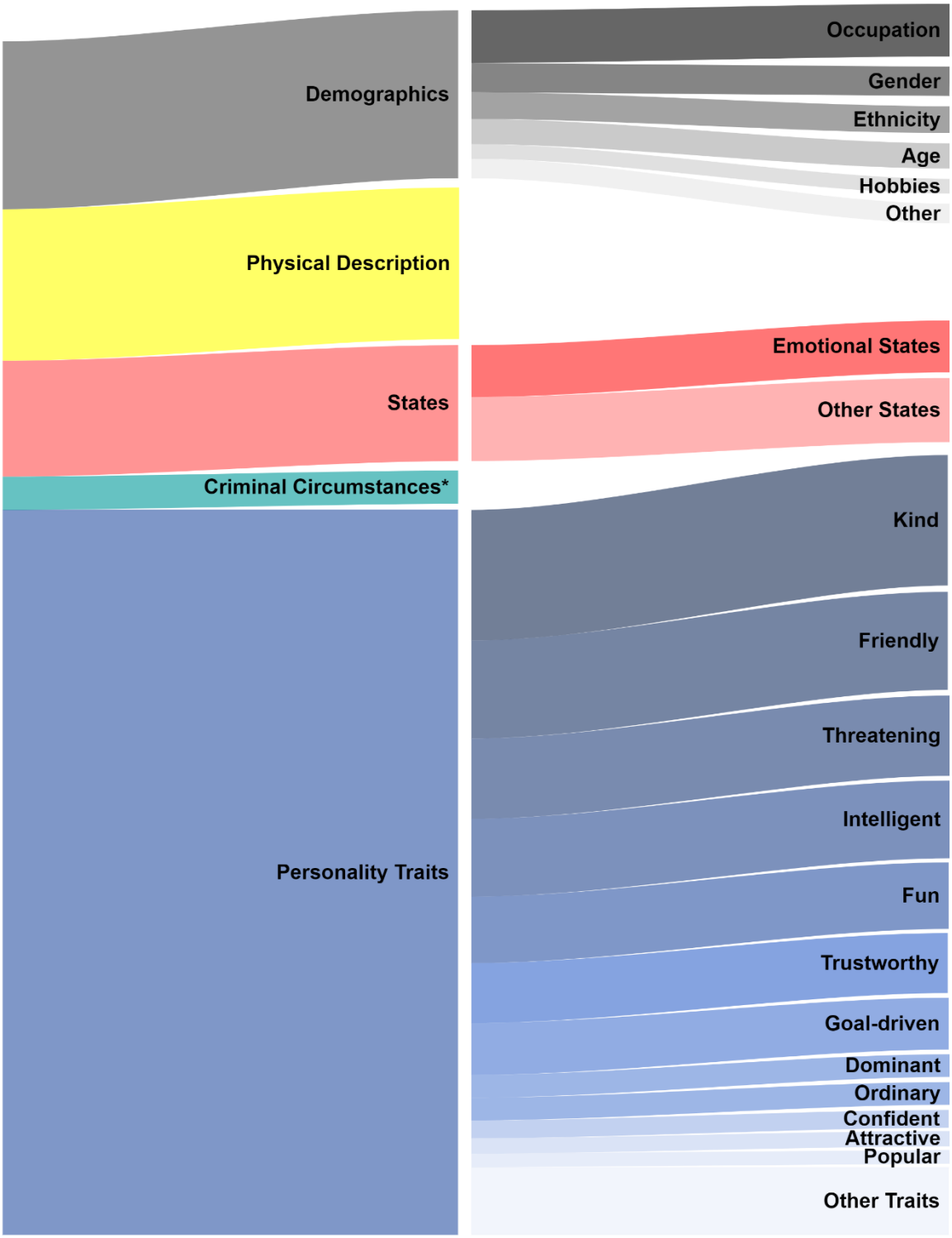
Overall, participants provided 1486 spontaneous descriptors of their first impressions. Each participant used 6.44 words on average to describe their impression of each face, demonstrating that they were engaged in the task and were able to form rich first impressions. Figure 2 shows the broad structure of the spontaneous descriptors, which followed five main categories. These included demographics (13.9%), with specific references to occupation, gender, ethnicity, age, hobbies and others (e.g., ‘university teacher’, ‘construction worker’, ‘young man’, ‘middle class’, ‘enjoys gardening’), descriptors of physical appearance (12.6%, e.g., ‘dark hair’, ‘green eyes’, ‘needs a haircut’), emotional and other states (9.6%, e.g., ‘angry’, ‘moody’, ‘tired’, ‘upset’), specific references to criminal circumstances which were exclusive to the descriptors provided by participants in the prison sentence condition (2.8%, e.g., ‘accidentally ended up in prison’, ‘drug related crime’, ‘good person who has taken a wrong decision’) and finally, descriptors relating to first impression traits (60.2%) such as kind, friendly, intelligent, etc. A very limited number of spontaneous descriptors were deemed unclassifiable (N = 14, 0.95% of descriptors).

Most Common Traits Selection

A total of 78 unique first impression characteristics were identified across the two conditions. These characteristics were then independently classified into broad categories by two researchers using the online blackboard platform Miro (www.miro.com) with any disagreements resolved collaboratively following independent classification. For example, the characteristics ‘clever’, ‘smart’, ‘educated’, ‘knowledgeable’, etc. were included in a broad category labelled ‘intelligence’. Based on the conceptual distinctiveness of the categories and the number of references in each category, 10 traits were selected to best represent the free descriptor dataset. These traits accounted for 76.8% of all first impression descriptors and included: kind (161 units, 18%), friendly (121 units, 13.5%), threatening (99 units, 11.1%), intelligent (96 units, 10.7%), trustworthy (74 units, 8.3%), goal-driven (45 units, 5%), dominant (28 units, 3.1%), confident (22 units, 2.5%), successful (22 units, 2.5%) and attractive (19 units, 2.1%). In addition, previous literature has identified shyness and warmth as being important social traits regarding offender perception (Edens, 2009; Henderson et al., 2014). Therefore, they were also included, resulting in a final set of 12 main traits to be used in Study 2. Sampling the naturalistic and spontaneous traits and words people use to describe their facial first impressions is therefore not only a more objective approach, but it also more accurately approximates how facial first impressions are formed in everyday life.

Figure 2

*A Sankey Diagram Showing the Broad Categories and Sub-Categories of the Spontaneous Descriptors Collected in Study 1. * Descriptors in This Category Were Exclusively from Participants Assigned to the Prison Sentence Condition.*



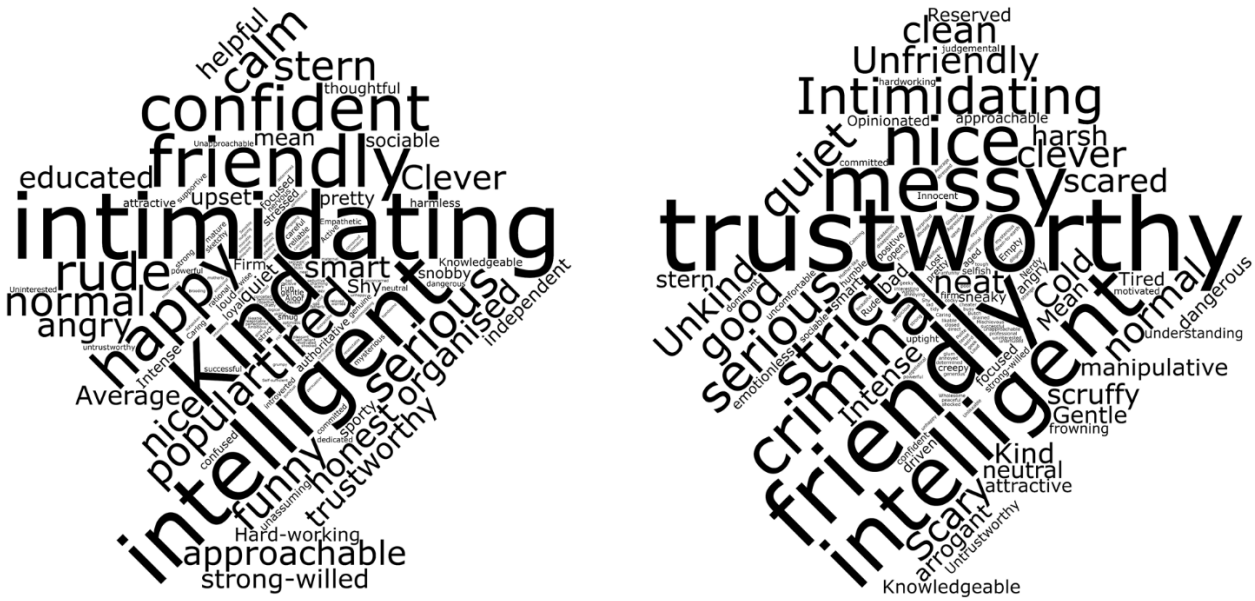
Thematic Word Clouds

As a preliminary inspection of the differences in the spontaneous descriptors provided by participants in the two conditions, a separate word cloud was created using the first impression descriptors provided by participants instructed that they will be presented with images of people serving time in prison and participants who did not receive any additional

information about the people they were asked to describe (see Figure 3). Each word cloud depicts high frequency descriptors with a larger font.

Figure 3

Word Clouds Depicting Spontaneous First Impressions of People Thought to be Serving Time in Prison (Right) and People for Whom No Additional Semantic Information Was Available (Left). Larger Font Size Represents More Frequent Descriptors.



At first glance, the pattern revealed through the prison sentence awareness and the control group word clouds seems surprising as the most frequent descriptor used in the prison sentence condition was a positive one - ‘trustworthy’, whereas the most frequently used descriptor in the control condition was negative – ‘intimidating’. This might be driven by the fact that all images were specifically chosen to show a neutral emotional expression, however, participants in both groups were presented with the exact same set of images. Nevertheless, exploring the word clouds further shows that overall the prison sentence word cloud contained more negative high frequency descriptors such as ‘intimidating’, ‘strict’, ‘unfriendly’, ‘unkind’, ‘messy’, ‘criminal’ and ‘scary’ whereas there were overall more positive descriptors in the control word cloud including ‘kind’, ‘friendly’, ‘intelligent’, ‘popular’, ‘confident’, ‘happy’ and ‘approachable’.

Quantitative Content Analysis

In order to substantiate our qualitative observations from the word clouds as well as to provide a more formal analysis of the positivity of spontaneous descriptors attributed to people believed to be serving time in prison, we carried out a quantitative content analysis following Sutherland et al. (2014). This was a by-item analysis and a sensitivity analysis in GPower (Erdfelder et al., 1996) indicated that with the present sample (of items, not participants given the by-item analysis), alpha of .05 and 80% power, the minimum detectable effect is *dz* = 0.583. See Supplementary Figure 1 for specific details of the sensitivity analysis calculation.

First, all descriptors were blind coded by two judges as either conventionally positive or negative (e.g., “intelligent”, “snobby”), if they referred to generally positive or negative habits and dispositions (e.g., “active and sporty”, “could be quite sneaky especially with women”) or triggered positive or negative emotions (e.g., ‘mysterious but in a scary way’). Judge coding agreement was high (Kappa = .74, $p < .05$) and all disagreements were resolved before analysis. We used a by-items analysis, where an index of overall valence was calculated separately for each face image by dividing the number of positive words by the number of all negative words provided by all participants, separately for those in the prison sentence and in the control condition. This was possible since participants assigned to both conditions were presented with the exact same face images. Similarly to Sutherland et al. (2014), we added the constant 0.5 to all data cells to allow for division without error (Gart & Zweifel, 1967). One of the face images received no negative descriptors across all participants in the control condition which resulted in an exceedingly high valence index for that particular image. Thus, it was removed from the data set as an extreme outlier[2].

The valence data were not normally distributed in both the prison sentence and the control conditions (W (29) = .77, $p < .001$ for the prison sentence condition and W (29) = .84, $p < .001$ for the control condition). Therefore, a Wilcoxon signed-rank test was used to compare the spontaneous descriptor positivity in the two conditions. Faces in the prison sentence condition (M = 1.15, SD = 1.09) received significantly fewer positive descriptors than the same faces in the control condition (M = 3.41, SD = 2.77), (Z = 4.34, $p < .001$, dz = 0.805).These results show that people thought to be serving time in prison received

[2] Note that analysing the full data set, including this outlier, produced the same result (Z = 4.43, $p < .001$, dz = 0.810).

significantly less positive spontaneous descriptors compared to descriptors attributed to the same face images in the absence of any additional semantic information. Moreover, participants did not receive any additional information about the specific crimes committed by the presented identities which has been previously found to impact offender perception (Tan et al., 2016). This suggests that it was the knowledge of the identity's criminality alone, which had a detrimental effect on perceivers' first impressions.

Consistent with the existing literature (Funk et al., 2017), many participants used stereotypes of criminality to determine whether the faces appeared to 'look' criminal. This could suggest that these stereotypes may have also impacted the way these faces were perceived, which resulted in a significantly more negative overall evaluation. Our results, therefore, extend this literature by demonstrating that, although faces which appear to look 'criminal' are subjected to harsher first impressions and treatment within the criminal justice system (Flowe & Humphries, 2011), the knowledge of a person's criminal conviction could be enough to subject them to less positive first impression judgments, irrespective of their facial features. Such findings could have serious implications for offenders' successful reintegration into society.

Study 2

Having identified the main categories of spontaneous descriptors people use to reveal their first impressions, in Study 2 we continue our data-driven approach (cf Oosterhof & Todorov, 2008) by collecting trait ratings of the identified 12 traits (attractive, trustworthy, dominant, warm, intelligent, confident, friendly, kind, successful, shy, driven, and threatening) using a larger image set of 100 faces. Again, participants were either instructed that the presented identities are serving time in prison or they did not receive any additional information. This approach allowed us to establish any specific differences in the way people with a criminal past are perceived in terms of first impressions. More importantly, we were able to reveal the underlying structure of first impressions based on these most commonly used traits. There is an overall agreement within the first impressions literature that there are two or three fundamental dimensions of social evaluation – trustworthiness (valence or approachability), dominance and youthful-attractiveness, the first two of which capture our evaluation of someone's intentions to help us or harm us and their ability to carry out these intentions (Oosterhof & Todorov, 2008;

Sutherland et al., 2013). Analysing these data with a commonly used dimension-reduction technique (e.g., PCA or EFA) will therefore reveal any differences within the underlying first impressions structure brought about by prison sentence awareness, providing evidence for a more complex interplay between bottom-up visual and top-down semantic cues in social evaluation that follow the predictions of recent first impression theories (Freeman et al., 2020).

Method

Participants

A total of 226 participants (31 male, mean age = 22, age range = 18-64) were recruited through the University of X Psychology Participation Pool as well as through social media. Initially, data from 79 participants (14 male, mean age = 25.9, age range = 18-64) was used for Study 2. Each one of these participants provided ratings for 30 face images (15 male). These data were analysed with a PCA, however following a peer-review process, it became clear that 30 images might not be sufficient for a stable and reliable PCA. Therefore, we recruited an additional sample of participants who only provided ratings for a new set of 70 images, making our total number of images 100. Each participant was randomly assigned to either the prison sentence condition (total of 115 participants, 17 male) or to the control condition (total of 111 participants, 14 male). All participants were over the age of 18 and had normal or corrected-to-normal vision. Participants provided informed consent to procedures which were approved by the School of Psychology Ethics Committee at the University of X.

Materials & Procedure

A total of 100 face images (50 male) were used in Study 2. These included the same 30 images used in Study 1 and an additional set of 70 images obtained in the same way and following the same criteria. Participants were also asked to complete the same questionnaire as the one in Study 1. The experiment was created and hosted on Qualtrics (Provo, UT, USA). Similarly to Study 1, participants were randomly assigned to one of two conditions: the prison sentence awareness condition, whereby they were instructed that they would be presented with images of people currently serving time in prison or the control condition, whereby they were not provided with any additional semantic information about the identities depicted in the face images. Participants in both conditions were sequentially presented with a number of face images and asked to rate each face for the

12 traits identified in Study 1 (attractive, trustworthy, dominant, warm, intelligent, confident, friendly, kind, successful, shy, driven, and threatening) using a scale from 1 (not at all) to 5 (very). Each trait was rated in a separate block to minimise carryover effects (Rhodes, 2006). On average, each image was rated by 29 participants for each of the 12 traits. Trait block order and image presentation order within each block were randomised individually for each participant. The task was not timed but participants were encouraged to rely on their initial gut feeling. Following the rating task, all participants were invited to complete the same 11-item questionnaire administered in Study 1.

Results & Discussion

Rater Agreement

Inter-rater reliability was high in both conditions and for each first impression trait (all Cronbach's alphas > .793). In addition to Cronbach's alpha, we also calculated Intraclass Correlation Coefficients (ICCs) as a more appropriate measure of rater agreement (Cortina, 1993; Kramer et al., 2018). Since images were rated by varying numbers of participants for each of the 12 traits, it was not possible to calculate ICCs of the entire dataset. Therefore, Table 1 shows average ICCs for ratings of each trait, separately for participants assigned to the prison sentence and control conditions. Every ICC presented in the table is an average of 10 ICC analyses, which were performed with a random selection of participants to ensure that the same number of raters were included for every image. For example, in the control condition, between 25-39 participants rated images for kindness. We then sampled 10 random sets of 25 participants for each of the 100 images and calculated the ICCs for every iteration[3]. A One-Way Random model was used separately for each trait in each condition and we report the reliability of the average rating. These analyses showed significant rater agreement for all traits in the two conditions. Therefore, first impression ratings for each identity were averaged across all participants, separately for each condition (prison sentence and control).

[3] ICCs for all iterations are available at https://osf.io/6jrx4/

Table 1

Average ICCs for Each Trait in the Prison Sentence and the Control Conditions. All ps < .001.

Trait	Prison Sentence ICC [95% Confidence Intervals]	Control ICC [95% Confidence Intervals]
Attractiveness	.948 [.932, .962]	.945 [.928, .959]
Confidence	.836 [.786, .879]	.836 [.786, .879]
Dominance	.787 [.723, .843]	.798 [.736, .851]
Drive	.787 [.722, .843]	.831 [.779, .875]
Friendliness	.884 [.849, .915]	.859 [.816, .896]
Intelligence	.873 [.834, .906]	.827 [.774, .872]
Kindness	.859 [.816, .896]	.904 [.875, .929]
Shyness	.774 [.706, .834]	.765 [.693, .827]
Success	.876 [.838, .909]	.911 [.884, .934]
Threat	.817 [.761, .865]	.834 [.784, .878]
Trustworthiness	.860 [.817, .897]	.828 [.775, .873]
Warmth	.878 [.841, .910]	.884 [.848, .914]

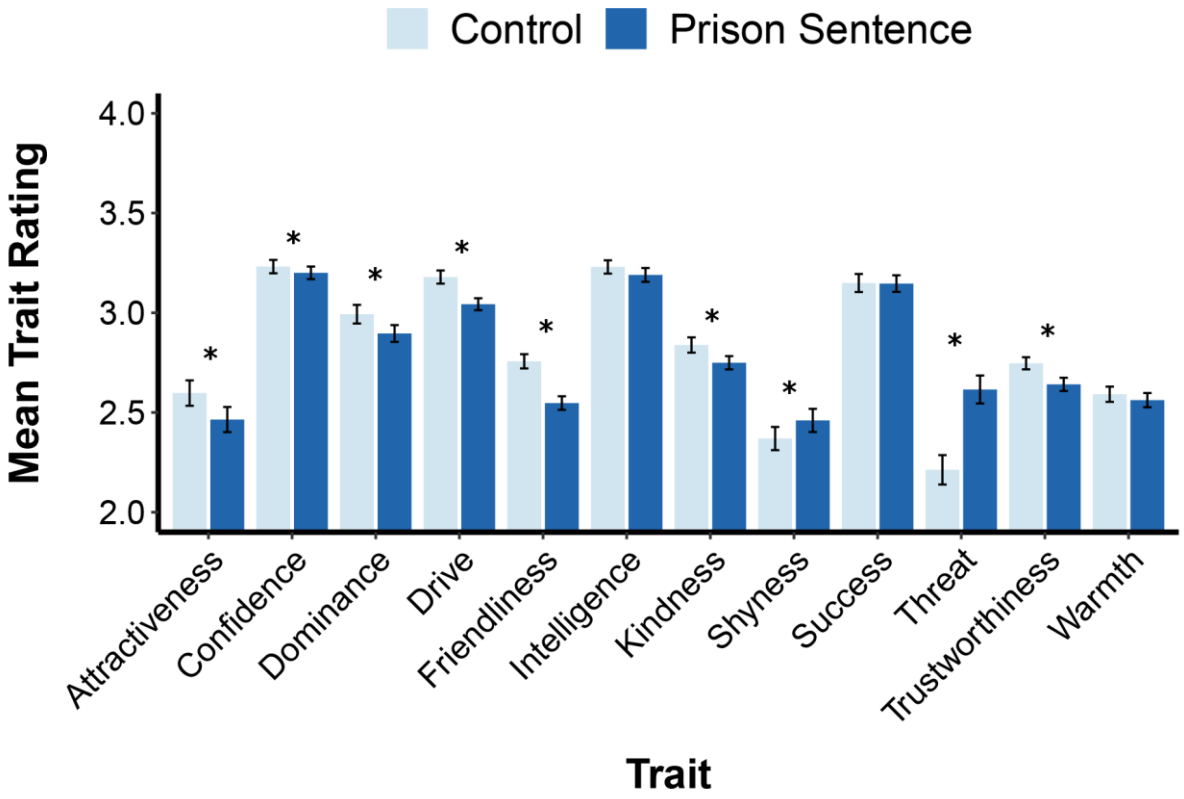
First Impression Differences

Mean trait ratings in the prison sentence and in the control condition, where participants did not receive any additional information about the presented identities, are shown in Figure 4. Data were analysed by-item with a 2 x 12 within subjects ANOVA (factors: condition – prison sentence vs control and trait – attractiveness, confidence, dominance, drive, friendliness, intelligence, kindness, shyness, success, threat, trustworthiness and warmth). Bayes factors in favour of the alternative hypothesis (BF_{10}) were calculated using the BayesFactor package in R (Morey et al., 2016) with a “default” prior, scale = 0.707. The Superpower package in R (Lakens & Caldwell, 2021) was used to provide an estimate of the minimum detectable effect size. More specifically, we used the plot_power function in order to explore whether effect sizes between 0-0.3 can be detected with the study

design (2x12 within-subjects), a sample size of 100 (which is the total number of images, since our analysis was a by-items one), an estimated standard deviation of 0.5 and an estimated correlation between first impression ratings of 0.8. The resulting plots are shown in Supplementary Figure 2. The analysis revealed that we could detect effect sizes of just over 0.3 for the main effect of condition and effect sizes in-between 0.125-0.15 for the main effect of trait with 80% power. Most importantly, we could detect effect sizes in-between 0.125-0.15 for the interaction between condition and trait, which was the main result of interest.

Figure 4

Mean Impression Ratings for Each Social Trait in the Prison Sentence and Control Conditions. Error Bars Show Within-Subjects Standard Error. * *p* < .05.



A 2 x 12 within-subjects ANOVA revealed significant main effects of both condition and trait, F (1, 99) = 12.69, p = .001, η_p^2 = .11, BF_{10} = 0.10 and F (11, 1089) = 44.39, p < .001, η_p^2 = .31, BF_{10} = 2.16 x 10^{156} respectively, as well as a significant interaction between them, F (11, 1089) = 41, p < .001, η_p^2 = .29, BF_{10} = 4.63 x 10^5. Simple main effects showed that when identities were presented together with the semantic information that

they have received a prison sentence, they were perceived as significantly more threatening (F (1, 1188) = 274.81, p < .001, η_p^2 = .19, BF_{10} = 3.33 x 10^{26}) and shy (F (1, 1188) = 11, p = .001, η_p^2 = .01, BF_{10} = 6.30). They were also perceived as significantly less friendly (F (1, 1188) = 74.64, p < .001, η_p^2 = .06, BF_{10} = 5.13 x 10^{11}), driven (F (1, 1188) = 31.11, p < .001, η_p^2 = .03, BF_{10} = 2.84 x 10^5), attractive (F (1, 1188) = 19.92, p < .001, η_p^2 = .02, BF_{10} = 1.55 x 10^4), dominant (F (1, 1188) = 18.25, p < .001, η_p^2 = .02, BF_{10} = 62.91), kind (F (1, 1188) = 8.38, p = .004, η_p^2 = .01, BF_{10} = 3), trustworthy (F (1, 1188) = 10.86, p = .001, η_p^2 = .01, BF_{10} = 7.33), and confident (F (1, 1188) = 4.77, p = .029, η_p^2 < .01, BF_{10} = 0.85). No significant differences were found for ratings of warmth (F (1, 1188) = 1.33, p = .248, η_p^2 < .01, BF_{10} = 0.29), intelligence (F (1, 1188) = 1.62, p = .203, η_p^2 < .01, BF_{10} = 0.25), and success (F (1, 1188) = 0.01, p = .933, η_p^2 < .01, BF_{10} = 0.11). Thus, our results provide evidence for substantial differences in social perception purely driven by the additional contextual knowledge of a prison sentence and not any physical information in the face.

First Impression Structure

In order to identify the underlying dimensions of face evaluation in the prison sentence and control conditions, mean trait ratings were submitted to an Exploratory Factor Analysis (EFA), separately for the two conditions. In addition to the high inter-rater reliability, Bartlett's test of sphericity indicated that the correlations between the different traits were large enough to make EFA an appropriate approach (χ^2 (66) = 1398.80, p < .001 for ratings in the prison sentence condition and χ^2 (66) = 1414.47, p < .001 for ratings in the control condition). The Kaiser-Meyer-Olkin test showed adequate levels of systematic variance – .86 for ratings in the prison sentence condition and .83 for ratings in the control condition.

Following the guidelines provided by Costello and Osborne (2005) as well as recent studies using EFA to determine the underlying structure of first impressions (e.g., Jones et al., 2021), an EFA (Maximum Likelihood approach) with no rotation was first used to indicate the overall number of dimensions in each condition. This number was determined based on four criteria in order to address some of the criticisms of the most commonly used practices (Fabrigar et al., 1999; O'Connor,
2000). These included Kaiser's criterion (eigenvalues larger than 1), the scree test (Fabrigar et al. 1999), parallel analysis (Horn, 1965), and the minimum average partial

analysis (MAP, Velicer, 2000). Parallel analysis and MAP were implemented in SPSS (see O'Connor, 2000 for further details), with 9000 random datasets generated for the parallel analysis and comparing the 95th percentile eigenvalues with the original data. For the prison sentence data, all four analyses indicated that 3 factors should be extracted. The analyses on the control data were somewhat more inconsistent – Kaiser's criterion indicated 2 factors (though the eigenvalue of the third factor was .964), parallel and the scree plot tests indicated 3 factors and MAP indicated 4 factors. Therefore, 3 factors were extracted from the control dataset too.

After the number of components was determined, data were analysed using the same EFA (Maximum Likelihood approach), but this time with a direct oblimin rotation to determine the component structure and loadings for each condition while allowing for components to remain oblique. This was considered an appropriate approach given the high correlations among first impression traits reported in the literature (Mileva et al., 2019). The resulting two pattern matrices (one for each condition) were then examined with loadings below .40 disregarded from the analyses. Tables 1 and 2 show these matrices for the control and the prison sentence impressions respectively. Since the proportion of variance explained by the rotated factors cannot be estimated following an oblique rotation, we only report the values from the first EFA, with no rotation.

In the control condition where no additional semantic information was available to perceivers, the first dimension explained 49.42% of the total variance, the second dimension explained 29.44% and the final third dimension explained 8.03% of the total variance. As expected, the first two dimensions appear to replicate Oosterhof and Todorov's (2008) valence and dominance dimensions, with the first dimension having high positive loadings from positive traits – kindness, warmth, friendliness and trustworthiness, and a high negative loading from threat. It also had high positive loadings from attractiveness. The second dimension captures perception of dominance and confidence. However, unlike Sutherland et al.'s third youthful-attractiveness dimension (2013), the final dimension seems to capture the evaluation of intelligence, ambition, and success, similar to the conscientiousness dimension in personality research. The first (approachability) and second (dominance) dimensions seem to be independent of one another (dimensions correlation = .005), consistent with Oosterhof and Todorov (2008). However, the third dimension (intelligence) is more closely related to the remaining two dimensions

(correlation between the intelligence and approachability dimensions = .293 and correlation between the intelligence and dominance dimensions = .400).

In the prison sentence condition, where perceivers were informed that all identities were serving a prison sentence, the first dimension explained 50.54% of the total variance, the second dimension explained 27.32% and the final third dimension explained 9.25% of the total variance. Overall, the three dimensions here follow closely the ones in the control condition. The first dimension reflects an approachability (or valence) evaluation, with high positive loadings from friendliness, warmth, kindness and trustworthiness and high negative loadings from threat. Similar to the control analysis, attractiveness is also a key trait for this first dimension. The second dimension captures evaluations of confidence and dominance, whereas the final third dimension reflects evaluations of intelligence and success and ambition. This analysis reveals the same pattern of dimension correlations with the first and second dimensions being independent from one another (dimension correlation = .045), whereas the third (intelligence) dimension seems to be more closely related to the remaining two (correlation between the intelligence and approachability dimensions = .308 and correlation between the intelligence and dominance dimensions = .333).

Table 2

Pattern Matrix from the Control Condition. Positive Loadings are Displayed in Red and Negative Loadings are Displayed in Blue. Darker Colours Correspond to Stronger Loadings.

Trait	Dimension 1 (Approachability)	Dimension 2 (Dominance)	Dimension 3 (Intelligence)
Warmth	.95	.00	.06
Kindness	.93	-.09	.12
Friendliness	.90	-.15	.10
Threat	-.83	.22	-.06
Trustworthiness	.79	-.09	.29
Attractiveness	.78	.29	-.22
Dominance	-.14	.82	.21
Shyness	.22	-.74	-.13

Confidence	.40	.72	.18
Intelligence	.04	-.02	.93
Drive	.10	.24	.76
Success	.07	.25	.75

Table 3

Pattern Matrix from the Prison Sentence Condition. Positive Loadings are Displayed in Red and Negative Loadings are Displayed in Blue. Darker Colours Correspond to Stronger Loadings

Trait	Dimension 1 (Approachability)	Dimension 2 (Dominance)	Dimension 3 (Intelligence)
Friendliness	.94	-.05	.05
Warmth	.92	-.00	.09
Kindness	.90	-.13	.10
Trustworthiness	.83	-.02	.16
Threat	-.81	.34	-.22
Attractiveness	.78	.30	-.25
Confidence	.41	.80	.12
Dominance	-.17	.77	.26
Shyness	.13	-.76	-.11
Intelligence	.12	-.01	.91
Success	.03	.20	.81
Drive	.11	.35	.71

Overall, Study 2 demonstrates a substantial negative shift in person evaluation following the awareness of an existing prison sentence, consistent with the free descriptor findings from Study 1. However, despite these differences, the additional contextual information did not have an impact on the key dimensions of social evaluation and their underlying structure. We find the same three key dimensions that capture judgements of approachability, dominance and intelligence in both conditions, which align well with previous first impression models. This suggests that while additional information can

definitely guide and bias our overall person impressions, it might not have the scope to alter the core processes and fundamental dimensionality of these impressions.

The analyses and main findings from the prisoner rehabilitation questionnaire can be found in the Supplementary Materials. Briefly, the thematic analysis applied to the open-ended questions revealed five main themes that captured issues surrounding the severity of the committed crime, trust, the offender's current behaviour, the perceiver's attitudes towards rehabilitation and their previous personal experiences. In order to allow for a more direct comparison between these facial impression models and the ones suggested by previous work (Oosterhof & Todorov, 2008), an additional analysis based on PCA with no rotation is also reported in the Supplementary Materials. The results of this analysis are also discussed below.

General Discussion

The present set of experiments sought to explore how the prior knowledge of a prison sentence can affect subsequent social evaluations and the underlying structure of first impressions, thus testing the idea of the interplay between bottom-up (visual) and top-down (contextual, semantic or affective) cues in the formation of first impression judgements. Our results show that the inferences brought about by criminal sentence awareness led to a negative shift in first impressions, both in terms of spontaneous free descriptors and specific trait ratings, independently from any physical facial differences. When target identities were thought to have received a prison sentence, they were perceived as significantly less trustworthy, attractive, friendly, kind, confident, dominant and driven as well as significantly more threatening and shy than when no additional information was available for the exact same set of face images. This demonstrates the pervasive biases associated with criminal sentencing, which could undoubtedly have important negative consequences on re-integration attempts. Some of the most affected traits, such as attractiveness and trustworthiness, have been repeatedly shown to guide a range of different behaviours and decisions such as the severity of a sentence if found guilty of a crime (Wilson & Rule, 2015), the likelihood of receiving a money loan (Duarte et al., 2012) and the likelihood of being hired as well as the salary one might receive (Fruhen et al., 2015).

The differences brought about by prison sentence awareness, however extensive, did not affect the underlying structure of first impressions, with judgements of approachability, dominance, and intelligence being the key evaluative dimensions both when additional semantic information was available and when it was not. This structure fits well with existing models of first impressions based on both unfamiliar and familiar identities (Oosterhof & Todorov, 2008; Rosenberg et al., 1968) as well as more general social perception models (Cuddy et al., 2008), all of which present evidence for two independent key dimensions that reflect our tendency to track others' intentions (e.g., to help or to cause harm) and ability (i.e., dominance or competence) to fulfil these intentions. While there is a one-to-one correspondence between these existing models and our data when it comes to the first dimension (here labelled approachability), the traits that make up our second and third dimensions have all been associated with the dominance (or competence) dimension from previous work. It seems like setting a specific context for first impression evaluations, such as that of criminal sentencing, might result in two separate judgements of someone's dominance (or confidence) and their competence. This is consistent with previous work that has shown changes to the two fundamental dimensions when priming a specific context (Wojciszke, 2005). Here, the added context could have introduced some further subtleties to the underlying structure of first impressions, where perceivers are first making a judgement about someone's intentions and whether to approach or avoid them, followed by a judgement of how likely they are to act or impose these intentions (i.e., a judgement of how confident or dominant someone might be) and finally, a judgements of how well they can accomplish these intentions (e.g., competence or intelligence). Therefore, judgements that have direct consequences for the perceiver are prioritised, whereas the final judgement of intelligence has more important consequences for the person being judged.

The differences between the underlying structure reported here and other two-dimensional models could also be due to differences in the dimension reduction approach used. While Oosterhof and Todorov (2008) use a PCA that forces orthogonal dimensions, we used an EFA with a rotation that allows for dimensions to remain oblique. In fact, when applying the same analysis to our data (see Supplementary Materials), we find two fundamental dimensions capturing evaluations of valence and dominance that replicate this previous work. Our third, intelligence, dimension was somewhat more highly correlated to the first

two dimensions which were more independent of one another. This might explain why this additional dimension was not extracted with an analysis that is asking for independent, orthogonal dimensions only.

Crucially, trait ratings of perceivers in both the prison sentence and the control conditions, followed the exact same underlying structure, showing no evidence that additional semantic information can change or re-prioritise the fundamental dimensions in any way. This is in contrast to previous reports of such a re-structuring with access to further demographic information. For example, it has been shown that when evaluating the faces of older adults, the dominance dimension adopts a somewhat different connotation, more in line with judgements of intelligence or wisdom rather than social and/or physical dominance and aggression which follows age stereotypes of older adults perceived as being more frail and less physically strong (Hehman et al., 2014). There is also evidence that female identities perceived as dominant are also perceived as less trustworthy, possibly driven by gender stereotype expectations of women to be more submissive (benevolent sexism, Glick & Fiske, 1996), while no such relationship exists for male identities (Mileva et al., 2019; Sutherland et al., 2015). This means that the two fundamental dimensions, trustworthiness and dominance, are no longer independent of one another when specifically considering female identities. None of these studies, however, use the same data-driven approach to explore any potential changes to the fundamental structure of first impressions as the one that was used to establish this structure in the first place. It is therefore, worth pointing out that by adopting the same data-driven sampling method here, we offer a stronger empirical test of the role of semantic or contextual information in social evaluation and the key evaluative dimensions, in particular. However, given the existing gender differences both when it comes to the people being judged and those forming first impressions (Mattarozzi et al., 2015; Mileva et al., 2019), we should also acknowledge the much higher proportion of female perceivers in our dataset which means that these data might not fully represent the impressions formed by male perceivers. Along similar lines, the faces we have used throughout these studies depicted White identities only so a different pattern might be observed with faces from different ethnical or racial backgrounds. While there is some evidence that impressions based on own- and other-race faces have a surprisingly similar underlying structure (Sutherland et al., 2018), a more forensic context might activate certain attitudes and

stereotypes against certain demographic groups that can then lead to a different pattern of results.

Our findings align well with previous research on more general spontaneous trait inferences as well as recent theories on facial impressions specifically, such as Dynamic Interactive Theory (DIT). We show evidence that social evaluation based on face images is a result of the interplay between visual cues and any pre-existing social-conceptual knowledge (e.g., stereotypes) that perceivers might be introducing to the perceptual process. Thus, it is possible that the additional information provided to participants in the prison sentence condition activated stereotypical semantic or affective associations which were able to substantially influence the overall person perception by setting certain implicit expectations or biases for the observers (Freeman & Ambady, 2011). Such results reaffirm the need to acknowledge that, in addition to visual cues, impressions could also be informed by direct observations of behaviour or by second hand evidence (e.g., receiving information from others) which is particularly relevant for the present studies (see also Wildman & Ramsey, 2021). We also demonstrate that there are certain limits to the effect of top-down contextual information on impression formation processes – while knowledge of an existing prison sentence led to a much more negative overall evaluation, encompassing many important social traits, these changes were not sufficient to affect the underlying structure of first impressions. This suggests that universal dimensions underpin person evaluation, regardless of whether they are based on purely visual information or on the dynamic interaction of visual and semantic cues. Altogether, this highlights the need to break away from the standard approach where facial first impressions are conceptualised as only being driven by facial features and structure and rather consider that they might also be shaped more dynamically by context or pre-existing stereotypes and attitudes. In doing so, we are likely to get a more accurate and generalisable representation of social evaluation processes and how we form facial first impressions of unfamiliar others in everyday life.

While it is clear that additional contextual or inferential information can produce substantial changes in social evaluation, we should acknowledge that this could be manifested via at least three different routes: 1) perceptual, where the information directly affects the perception of a face, 2) attentional, where it might guide attention to different parts (or features) of the face or attribute different weights to them, and 3) interpretational, where

contextual information might change the way certain traits are interpreted (e.g., prison sentence awareness shifting the interpretation of dominance from social dominance, relating to leadership to physical dominance, relating to aggressiveness). The first two possible mechanisms can be related to the “reading into faces” hypothesis suggested by Hassin and Trope (2000), according to which information about someone’s character can change the way their face is physically perceived and consequently evaluated. It is likely that multiple or all of these routes are operating together in order to guide perception and the resulting social judgements.

The substantial decrease in the overall ratings of trustworthiness following the awareness of a criminal conviction could also be related to Dangerous Decisions Theory (Porter & ten Brinke, 2009) where an initial negative evaluation (i.e., low trustworthiness), irrespective of facial features, could trigger a tunnel-vision approach to the ensuing decisions relating to other personality traits (Korva et al., 2012). Any subsequent judgements might, therefore, reflect a type of a confirmation bias guided by the need to justify the initial negative evaluation (Porter et al., 2010). This suggests a potential route for automatic evaluations (e.g., of trustworthiness) to influence any further, more explicit, reasoning about other traits by imposing stricter criteria (i.e., we might need more evidence to change our minds following an initial negative perception of someone being untrustworthy, Over & Cook, 2018). The fact that we find significantly lower ratings of positive traits such as kindness and intelligence and significantly higher ratings of negative traits such as threat following the knowledge of a criminal conviction might also suggest the presence of a negative halo effect (also referred to as the horns or devil effect, Forgas & Laham, 2017). Here, global evaluations of a person can reflect the negative impression of one single attribute (Guillory & Hancock, 2015), which in this case was likely the presence of a criminal conviction. Within this more forensic context, it is also worth acknowledging that while the more ambiguous manipulation used here was sufficient to produce a widespread bias in face impressions, information about the specific crime committed (which will likely be accessible in everyday life) will have the scope to provide a more clear-cut and detailed account of the effect of prison sentence awareness on facial first impressions. This is particularly likely, given research showing relationships between certain facial features and judgements of criminality (Funk et al., 2017) and even specific types of crimes (Avery et al., 2021). What is more, many of the participants who opted to complete an additional survey on their attitudes towards rehabilitation (see Supplementary Materials), mentioned

that the type of crime committed is a very important consideration. Therefore, exploring the effects of specific types of crimes (e.g., ones involving violence or not) as well as the effect of perceiver's attitudes towards rehabilitation would provide an interesting direction for future work.

In addition to the more affective influence, the bias produced by the criminal sentence awareness introduced here, could also be based on changes in semantics (by making criminality stereotypes more accessible). Previous work has provided evidence for both of these mechanisms (Asch, 1946; Fazio et al., 1986; Higgins et al., 1977) and there is even evidence that their influence can be systematically manipulated by priming the more cognitive or the more affective aspects of the information provided to perceivers, resulting in different attitudes being formed (van den Berg et al., 2006). Given that we found significant differences in traits that fall under the valence dimension (friendliness, kindness) and the more competence- or dominance-related dimensions (confidence, dominance, drive), following the additional top-down cues, it is likely that both mechanisms are at play in this context. However, there is evidence from emotion categorisation studies that valence- and stereotype-related processes can be dissociated. For example, Bijlstra et al. (2010) show that when discriminating between positive and negative emotions, perceivers are quicker to categorise negative emotions (sadness) in the more negatively evaluated male faces (i.e., they act on valence information), whereas when discriminating between two different negative emotions, perceivers' response times are guided by stereotype associations leading to faster categorisation of anger rather than sadness in male faces. Therefore, identifying the exact nature of this bias in facial first impressions and the relative importance of semantic and affective influences would be worth exploring in future work.

Regardless of the exact mechanism, however, these substantial differences in the positivity of first impressions imply that they are not only important for decisions within the criminal justice system (Wilson & Rule, 2015), but could also play an important role for a successful rehabilitation following a criminal conviction. This negative evaluation could be severely detrimental to the lives of those trying to reintegrate back into society, especially considering the pervasive nature of first impression consequences. Thus, facial first impressions might reflect a critical aspect of recidivism that warrants further investigation.

Open practices

All materials and data are available at https://osf.io/6jrx4/?view_only=c7c6698720164b729507a4233e3ba097.

References

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology, 41*(3), 258–290. https://doi.org/10.1037/h0055756

Atkin, C. A., & Cramer, R. J. (2012). Ex-offenders on the stand: Steps toward eliminating jury bias. *Journal of Forensic Psychology Practice, 12*(3), 211–226. https://doi.org/10.1080/15228932.2012.674468

Austin, J., & Hardyman, P. L. (2004). The risks and needs of the returning prisoner population. *Review of Policy Research*, 21(1), 13–29. https://doi.org/10.1111/j.1541-1338.2004.00055.x

Avery, J. J., Oh, D., & Cooper, J. (2021). Race and perceived immorality in stereotypes of criminal subtypes. *Basic and Applied Social Psychology, 43*(5), 307-318. https://doi.org/10.1080/01973533.2021.1931220

Bijlstra, G., Holland, R. W., & Wigboldus, D. H. (2010). The social face of emotion recognition: Evaluations versus stereotypes. *Journal of Experimental Social Psychology, 46*(4), 657-663. https://doi.org/10.1016/j.jesp.2010.03.006

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10*(1), 7. https://doi.org/10.7275/jyj1-4868

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology,* 40, 61–149. https://doi.org/10.1016/S0065-2601(07)00002-0

Dotsch, R., Wigboldus, D. H., 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

Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies, 25*(8), 2455–2484. https://doi.org/10.1093/rfs/hhs071

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology, 87*(6), 876–893. https://doi.org/10.1037/0022-3514.87.6.876

Edens, J. F. (2009). Interpersonal characteristics of male prisoners: Personality, psychopathological, and behavioural correlates. *Psychological Assessment, 21*(1), 89–98. https://doi.org/10.1037/a0014856

Efran. M. G. (1974). The effect of physical appearance on the judgment of guilt, interpersonal attraction, and severity of recommended punishment in a simulated jury task. *Journal of Research in Personality, 8*(1), 45–54. https://doi.org/10.1016/0092-6566(74)90044-0

Eisenberg, T., & Hans, V. P. (2009). Taking a stand on the stand: The effect of a prior criminal record on the decision to testify and on trial outcomes. *Cornell Law Review, 94*, 90.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. Behavior Research Methods, Instruments, & Computers, 28(1), 1-11. https://doi.org/10.3758/BF03203630

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*(2), 229–238. https://doi.org/10.1037/0022-3514.50.2.229

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77-83. https://doi.org/10.1016/j.tics.2006.11.005

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. https://doi.org/10.1002/acp.1673

Forgas, J. P., & Laham, S. M. (2017). Halo effects. In R. F. Pohl (ed.) *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory,* (pp. 276–290). Routledge.

Funk, F., Walker, M., & Todorov, A. (2017). Modelling perceptions of criminality and remorse from faces using a data-driven computational approach. *Cognition and Emotion, 31*(7), 1431-1443. https://doi.org/10.1080/02699931.2016.1227305

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118*(2), 247–279. https://doi.org/10.1037/a0022327

Freeman, J. B., Stolier, R. M., & Brooks, J. A. (2020). Dynamic interactive theory as a domain-general account of social perception. *Advances in Experimental Social Psychology, 61*, 237–287. https://doi.org/10.1016/bs.aesp.2019.09.005

Fruhen, L. S., Watkins, C. D., & Jones, B. C. (2015). Perceptions of facial dominance, trustworthiness and attractiveness predict managerial pay awards in experimental tasks. *The Leadership Quarterly, 26*(6), 1005–1016. https://doi.org/10.1016/j.leaqua.2015.07.001

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

Funk, F., Walker, M., & Todorov, A. (2017). Modelling perceptions of criminality and remorse from faces using a data-driven computational approach. *Cognition and Emotion, 31*(7), 1431–1443. https://doi.org/10.1080/02699931.2016.1227305

Gart, J. J., & Zweifel, J. R. (1967). On the bias of various estimators of the logit and its variance with application to quantal bioassay. *Biometrika, 54*(1), 181–187. https://doi.org/10.1093/biomet/54.1-2.181

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*(3), 491–512. https://doi.org/10.1037/0022-3514.70.3.491

GOV.UK. (2022). *Telling an employer, university or college about your criminal conviction*. GOV.UK. https://www.gov.uk/tell-employer-or-college-about-criminal-record/print

Graham, L., Fischbacher, C. M., Stockton, D., Fraser, A., Fleming, M., & Greig, K. (2015). Understanding extreme mortality among prisoners: A national cohort study in scotland using data linkage. *European Journal of Public Health*, *25*(5), 879–885. https://doi.org/10.1093/eurpub/cku252

Greene, E., & Dodge, M. (1995). The influence of prior record evidence on juror decision making. *Law and Human Behavior, 19*(1), 67–78. https://doi.org/10.1007/BF01499073

Guillory, J. E., & Hancock, J.T. (2015). Effects of network connections on deception and halo effects in Linkedin. In G. Riva, B. K. Wiederhold, & P. Cipresso (Eds.), *The psychology of social networking: Personal experience in online communities* (pp. 67–77). Berlin: De Gruyter Open Ltd.

Hassin, R., & Trope, Y. (2000). Facing faces: studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology, 78*(5), 837–852. https://doi.org/10.1037/0022-3514.78.5.837

Hehman, E., Leitner, J. B., & Freeman, J. B. (2014). The face–time continuum: Lifespan changes in facial width-to-height ratio impact aging-associated perceptions. *Personality and Social Psychology Bulletin, 40*(12), 1624–1636. https://doi.org/10.1177/0146167214552791

Hehman, E., Sutherland, C. A., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology, 113*(4), 513–529. https://doi.org/10.1037/pspa0000090

Henderson, L., Gilbert, P., & Zimbardo, P. G. (2014). Shyness, social anxiety and social phobia. In S. G. Hofmann, & P. M. DiBartolo (Eds.), *Social anxiety* (pp. 95–115). Academic Press.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13*(2), 141-154. https://doi.org/10.1016/S0022-1031(77)80007-3

Hirschfield, P. J., & Piquero, A. R. (2010). Normalization and legitimation: Modelling stigmatizing attitudes toward ex-offenders. *Criminology: An Interdisciplinary Journal, 48*(1), 27–55. https://doi.org/10.1111/j.1745-9125.2010.00179.x

Home Office. (2013). *Guidance: Find Out If a Person Has a Record for Child Sexual Offences.* Retrieved at: https://www.gov.uk/guidance/find-out-if-a- person-has-a-record-for-child-sexual-offences

Jaeger, B., Todorov, A. T., Evans, A. M., & van Beest, I. (2020). Can we reduce facial biases? Persistent effects of facial trustworthiness on sentencing decisions. *Journal of Experimental Social Psychology, 90*, 104004. https://doi.org/10.1016/j.jesp.2020.104004

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). *Analyzing and improving the image quality of styleGAN*. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 8110–8119.

Korva, N., Porter, S., O'Connor, B. P., Shaw, J., & ten Brinke, L. (2012). Dangerous decisions: Influence of juror attitudes and defendant appearance on legal decision-making. *Psychiatry, Psychology & Law*, 20(3), 384–398. https://doi.org/10.1080/13218719.2012.692931

Kramer, R. S. S., Mileva, M., & Ritchie, K. L. (2018). Inter-rater agreement in trait judgements from faces. *PloS One, 13*(8), e0202655. https://doi.org/10.1371/journal.pone.0202655

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. Advances in Methods and Practices in Psychological Science, 4(1). https://doi.org/10.1177/2515245920951503

Lavan, N., Mileva, M., Burton, A. M., Young, A. W., & McGettigan, C. (2021). Trait evaluations of faces and voices: Comparing within- and between-person variability. *Journal of Experimental Psychology: General, 150*(9), 1854–1869. https://doi.org/10.1037/xge0001019

Levenson, J. S., & Cotter, L. P. (2005). The effect of Megan's law on sex offender reintegration. *Journal of Contemporary Criminal Justice*, 21(1), 49–66. https://doi.org/10.1177/1043986204271676

Linke, L., Saribay, S. A., & Kleisner, K. (2016). Perceived trustworthiness is associated with position in a corporate hierarchy. *Personality and Individual Differences, 99*, 22–27. https://doi.org/10.1016/j.paid.2016.04.076

Martin, L. L., Seta, J. J., & Crelia, R. (1990). Assimilation and contrast as a function of people's willingness and ability to expend effort in forming an impression. *Journal of Personality and Social Psychology, 59*, 27–37. https://doi.org/10.1037/0022-3514.59.1.27

Milazzo, C., & Mattes, K. (2016). Looking good for election day: Does attractiveness predict electoral success in Britain? *The British Journal of Politics and International Relations, 18*(1), 161–178. https://doi.org/10.1111/1467-856X.12074

Mileva, M., Kramer, R. S. S., & Burton, A. M. (2019). Social evaluation of faces across gender and familiarity. *Perception, 48*(6), 471–486. https://doi.org/10.1177/0301006619848996

Morey, M., & Crewe, B. (2018). Work, intimacy and prisoner masculinities. In M. Maycock & K. Hunt (Eds.), *New perspectives of prison masculinities.* (pp. 17–41). Springer International Publishing.

Nacro. (2022). *Advice for people convicted for sex offences.* Retrieved at: https://www.nacro.org.uk/criminal-record-support-service/support-for-individuals/advice-prisoners-people-licence-sex-offenders-mappa/advice- people-convicted-sex-offences/

Newman, L. S., Duff, K. J., Hedberg, D. A., & Blitstein, J. (1996). Rebound effects in impression formation: Assimilation and contrast effects following thought suppression. *Journal of Experimental Social Psychology, 32*(5), 460-483. https://doi.org/10.1006/jesp.1996.0021

Newman, L. S., & Uleman, J. S. (1990). Assimilation and contrast effects in spontaneous trait inference. *Personality and Social Psychology Bulletin, 16*(2), 224-240. https://doi.org/10.1177/0146167290162004

Oh, D., Walker, M., & Freeman, J. B. (2021). Person knowledge shapes face identity perception. *Cognition, 217*, 104889. https://doi.org/10.1016/j.cognition.2021.104889

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior, 34*, 83–110. https://doi.org/10.1007/s10919-009-0082-1

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

Otten, S., & Moskowitz, G. B. (2000). Evidence for implicit evaluative in-group bias: Affect-biased spontaneous trait inference in a minimal group paradigm. *Journal of*

Experimental Social Psychology, 36(1), 77-89. https://doi.org/10.1006/jesp.1999.1399

Otten, S., & Stapel, D. A. (2007). Who is this Donald? How social categorization affects aggression-priming effects. *European Journal of Social Psychology, 37*(5), 1000-1015. https://doi.org/10.1002/ejsp.413

Over, H., & Cook, R. (2018). Where do spontaneous first impressions of faces come from? *Cognition, 170*, 190–200. https://doi.org/10.1016/j.cognition.2017.10.002

Porter, S., & ten Brinke, L. (2009). Dangerous decisions: A theoretical framework for understanding how judges assess credibility in the courtroom. *Legal and Criminological Psychology, 14*(1), 119–134. https://doi.org/10.1348/135532508X281520

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law, 16*(6), 477–491. https://doi.org/10.1080/10683160902926141

Pritikin, M. H. (2008). Is prison increasing crime? *Wisconsin Law Review, 6,* 1049.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57*(1), 199–226. doi: 10.1146/annurev.psych.57.102904.190208

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9*(4), 283–294. https://doi.org/10.1037/h0026086

Rule, N. O., & Ambady, N. (2010). First impressions of the face: Predicting success. *Social and Personality Psychology Compass, 4*(8), 506–516. https://doi.org/10.1111/j.1751-9004.2010.00282.x

Rule, N. O., & Ambady, N. (2011). Face and fortune: Inferences of personality from managing partners' faces predict their law firms' financial success. *Leadership Quarterly, 22*(4), 690–696. https://doi.org/10.1016/j.leaqua.2011.05.009

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

Secord, P. F. (1958). Facial features and inference processes in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and iterpersonal behavior* (pp. 300–315). Stanford, CA: Stanford University Press.

Sigall, H., & Ostrove, N. (1975). Beautiful but dangerous: Effects of offender attractiveness and nature of the crime on juridic judgment. *Journal of Personality and Social Psychology, 31*(3), 410–414. https://doi.org/10.1037/h0076472

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology, 37*(10), 1660–1672. https://doi.org/10.1037/0022-3514.37.10.1660

Sussman, A. B., Petkova, K., & Todorov, A. (2013). Competence ratings in US predict presidential election outcomes in Bulgaria. *Journal of Experimental Social Psychology, 49*(4), 771–775. https://doi.org/10.1016/j.jesp.2013.02.003

Sutherland, C. A., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Personality and Social Psychology Bulletin, 44*(4), 521-537. https://doi.org/10.1177/0146167217744194

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

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

Sutherland, C. A. M., & Young, A. W. (2022). Understanding trait impressions from faces. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12583

Tan, X. X., Chu, C, M., & Tan, G. (2016). Factors contributing towards stigmatisation of offenders in Singapore. *Psychiatry, Psychology & Law, 23*(6), 956–969. https://doi.org/10.1080/13218719.2016.1195329

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623–1626. doi: 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*, 519–545. doi: 10.1146/annurev-psych-113011-143831

Tuk, M. A., Verlegh, P. W., Smidts, A., & Wigboldus, D. H. (2009). Interpersonal relationships moderate the effect of faces on person judgments. *European Journal of Social Psychology, 39*(5), 757-767. https://doi.org/10.1002/ejsp.576

Van den Berg, H., Manstead, A. S., van der Pligt, J., & Wigboldus, D. H. (2006). The impact of affective and cognitive focus on attitude formation. *Journal of Experimental Social Psychology, 42*(3), 373-379. https://doi.org/10.1016/j.jesp.2005.04.009

Wigboldus, D. H., Dijksterhuis, A., & van Knippenberg, A. (2003). When stereotypes get in the way: stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology, 84*(3), 470–484. https://doi.org/10.1037/0022-3514.84.3.470

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37*(3), 395 – 412. https://doi.org/10.1037/0022-3514.37.3.395

Wildman, A., & Ramsey, R. (2022). Estimating the effects of trait knowledge on social perception. *Quarterly Journal of Experimental Psychology, 75*(5), 969–987. https://doi.org/10.1177/17470218211047447

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

Wilson, J. P., & Rule, N. O. (2016). Hypothetical sentencing decisions are associated with actual capital punishment outcomes: The role of facial trustworthiness. *Social Psychological and Personality Science, 7*(4), 331–338. https://doi.org/10.1177/1948550615624142

Wissler, R. L., & Saks, M. J. (1985). On the inefficacy of limiting instructions. *Law and Human Behaviour, 9*, 37–48. https://doi.org/10.1007/BF01044288

Wojciszke, B. (1994). Multiple meanings of behavior: Construing actions in terms of competence or morality. *Journal of Personality and Social Psychology, 67*(2), 222 – 232. https://doi.org/10.1037/0022-3514.67.2.222

Wormith, J. S., Althouse, R., Simpson, M., Reitzel, L. R., Fagan, T. J., & Morgan, R. D. (2007). The rehabilitation and reintegration of offenders: The current landscape and some future directions for correctional psychology. *Criminal Justice and Behavior, 34*(7), 879–892. https://doi.org/10.1177/0093854807301552

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

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*(3), 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109.x

Zebrowitz, L. A., & Montepare, J. M. (1992). Impressions of babyfaced individuals across the life span. *Developmental Psychology, 28*(6), 1143–1152. https://doi.org/10.1037/0012-1649.28.6.1143

Zebrowitz, L. A., Montepare, J. M., & Lee, H. K. (1993). They don't all look alike: Individual impressions of other racial groups. *Journal of Personality and Social Psychology, 65*(1), 85–101. https://doi.org/10.1037/0022-3514.65.1.85