

# 12

## Who's That girl? Facial Appearance Based Inferences

**Ines Sučić, Anja Wertag**

*Institute of Social Sciences Ivo Pilar, Zagreb, Croatia*

**Tiffany Matej Hrkalović, Vanja Oparnica, Dora Trgovec, Marta Šojat,  
and Jelena Žipovski**

*University Department of Croatian Studies, University of Zagreb, Croatia*

### Abstract

*People automatically make inferences from other's appearance, and there is evidence that personality judgements have substantial accuracy for some traits, such as trustworthiness. However, interpersonal intuition is prone to error, so the validity of appearance-based inferences has been repeatedly questioned, and research has never conclusively established a relationship between facial appearance and participation in criminal activities. Therefore, the aims of this study were to examine: 1. participants' competence to differentiate between criminals and non-criminals based on the facial appearance, 2. gender differences in the ability to distinguish between criminals and non-criminals, and 3. associations between confidence in judging others and accuracy of differentiating between criminals and non-criminals. The study was conducted on-line. Convenience sample of 832 adult internet users took part in this study. Participants' average age was 24, they were mostly females (81%), with at least a high school diploma (55%). Participants had to estimate how much confidence they have in their own judgements of other people, and then if each of 20 presented faces looks like a criminal's. The average accuracy of (non)criminality ratings for all photographs was 57%. The accuracy of the estimated non-criminality was higher (70%) than the accuracy of the estimated criminality (45%). There were no significant gender differences in the accuracy of (non)criminality ratings. Finally, women were less confident than men in judging unfamiliar faces, and participants who were more confident in their own judgements had a higher accuracy in ratings of criminals and lower accuracy in ratings of non-criminals, but the effect sizes were small.*

*Keywords: facial appearance, stereotypes, gender differences, criminality*

## Introduction

Throughout history the link between physical appearance, especially facial characteristics, behavioural and psychological characteristics of people was of great interest to many researchers. One of these characteristics, that captured special attention of many researchers investigating facial appearance-based inferences for many years, was criminality. For example, in the 19th century Lombroso hypothesized that compared with „normal“ women, female criminals more often have “receding foreheads, over jutting brows, large lower jaws, and prominent cheekbone“ (Zebrowitz, 1997). Even today, researches argue that specific facial features may serve as predictors of propensity for prosocial and antisocial behaviour (e.g. Haselhuhn & Wong, 2012; Stirrat & Perrett, 2012; ten Brinke & Porter, 2012). For example, greater facial width-to-height ratio is hypothesised to add to the impression of person’s aggressiveness and dangerousness (Hehman, Leitner, Deegan, & Gaertner, 2013).

For creating impressions about others, people need very little cognitive effort, very little time (less than 100ms) and minimum information (e.g. a very brief interaction or even just exposure to a photo) (Klatt et al, 2016; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015; Todorov, Said, Engell, & Oosterhof, 2008; Valla, Ceci, & Williams, 2011; Willis & Todorov, 2006; Zebrowitz, Voinescu, & Collins, 1996). The usage of those heuristics helps us get around in the social world (Willis & Todorov, 2006), and it seems that some appearance-based trait judgments are quite accurate (e.g. Roney, Hanson, Durante, & Maestripieri, 2006; Shevlin, Walker, Davies, Banyard, & Lewis, 2003).

However, evidence that facial qualities are enough to enable accurate judgement is scarce, and there is even less evidence which appearance qualities are necessary for those judgements to be made (Zebrowitz, 1997). One of the main models that attempts to explain the relationship between individuals’ facial features and the perception of that individual by others is the two-dimensional model which proposes that people have a tendency to assess faces along two fundamental dimensions: trustworthiness and dominance (Todorov, Baron, & Oosterhof, 2008). Moreover, the perception of emotional expressions if influenced by the structural facial features (Marsh, Adams, & Kleck 2005; Neth & Martinez 2009; Oosterhof & Todorov 2008; Sacco & Hugenberg 2009), as well as resemblance to emotional expressions influence social attributions from faces (Montepare & Dobish, 2003; Oosterhof & Todorov 2008; Said, Sebe, & Todorov, 2009; Todorov & Duchaine 2008; Todorov et al., 2015; Zebrowitz, Kikuchi, & Fellous, 2010).

Generally, studies investigating personality judgements found substantial accuracy in assessing trustworthiness from photographs that were presented in a trial-by-trial manner (e.g. Todorov, Baron, & Oosterhof, 2008). Moreover, other studies suggest that face characteristics shown in photographs may also provide additional cues about behavioural tendencies of the target, such as cheating in a Prisoner’s Dilemma game (Verplaetse, Vanneste, & Braeckman, 2007).

People usually relate certain categories of people to stereotypic notions about how they are expected to look, and there are also present stereotypes about criminality, as well as types of faces prone to particular crimes (Bull, 1992; Bull & McAlpine, 1998; Dumas & Testé, 2006; MacLin & Herrera, 2006; Shoemaker, South, & Lowe, 1973). Raters are usually concordant in their estimates of the criminal appearance of faces (Flowe, 2012; Funk, Walker & Todorov, 2017; MacLin & Herrera, 2006), where dominance is referred to as the main factor contributing to the assessment of a person as someone who has committed a criminal offense (Todorov et al., 2008). Moreover, studies confirmed high correlations between criminal appearance, observed untrustworthiness and dominance, as well as greater masculinity of faces (Flowe, 2012; Funk et al., 2017; Porter, ten Brinke, Gustaw, 2010; Ward, Flowe, & Humphries, 2012). When it comes to emotional expression, angry faces are perceived as emotionally more dominant, and if the the eyes and the mouth are relatively close to each other even neutral faces are rated as having an angry appearance (Hess, Blairy, & Kleck, 2000; Knutson 1996; Montepare & Dobish 2003). In addition, unattractiveness, facial maturity and

asymmetry were frequently found to be positively associated with perceived criminality (Berry & Zebrowitz-McArthur, 1988; Bull, 2006; Dumas & Testé, 2006; Todorov, 2008; Zebrowitz et al., 1996; Zebrowitz & McDonald, 1991).

Both perceiver and target characteristics contribute to the accuracy of inference-based impressions (e.g. Ambady, Hallahan, & Rosenthal, 1995). Accuracy is moderated by our knowledge gained from prior experiences, our attitudes and experiences with the people who show particular face-trait correlation, the culture to which we belong (Todorov et al., 2015) and a disposition for particular face-trait correlations and overgeneralized response to facial features (Zebrowitz, 1997). The inner-group bias in recognizing faces was generally confirmed in both genders (e.g. Rehnman & Herlitz, 2006, 2007; Wright & Sladden, 2003). However, some authors found that females perform better of regardless of target's gender (e.g. McBain, Norton, & Chen, 2009; Rehnman & Herlitz, 2007; Shapiro & Penrod, 1986), while other found males are better in recognizing females faces (McKelvie, Standing, Jean, & Law, 1993; Rehnman & Herlitz, 2007). Nevertheless, females were also found to make more errors than males in the recognition task (Shapiro & Penrod, 1986) and making more false positive judgements in criminality ratings for all faces (Hirst et al., 2016). This could be attributed to the females' tendency to overemphasize their probability of being victimized and thus maintaining a high fear of crime (Miller, Rossi & Simpson, 1991) and a higher sensitivity in rating the degree of criminality from facial expressions. Moreover, it was found that females have more difficulty in criminality ratings for a specific type of criminal offenses (Valla et al., 2011) although people's expectations about a criminal perpetrator's physical appearance vary in relation to the type of criminal offense (Dumas & Testé, 2006; Macrae & Shepherd, 1989; Shoemaker et al., 1973; Skorinko & Spellman, 2013; Yarmey, 1993). In addition to the targets' features and participants' characteristics, appearance-based trait judgments also depend on the conditions of the research situation (e.g. the length of exposure, image quality) (Todorov & Porter, 2014; Willis & Todorov, 2006).

Since interpersonal intuition is prone to error, validity of appearance-based inferences has been repeatedly questioned, and research has never conclusively proved an association between facial appearance and participation in criminal activities (e.g. Olivola & Todorov 2010; Zebrowitz, 1997). Negative consequences of criminality-based inferences also contributed to the long-lasting stigma assigned to the field (Gould, 1981). These stereotypes may influence adjudications at all stages of the legal process (Bull & Rumsey, 1988; Shepherd, 1989). People more often remember criminal-looking faces (MacLin & MacLin, 2004) and chose them more often in police line-ups (Flowe & Humphries, 2011; Flowe, Klatt, & Colloff, 2014; McQuiston & Malpass, 2002). Behavioural confirmation model speaks in favour of self-fulfilling prophecy effect - people whose faces induce strong expectations about their dispositional characteristics may be treated as though they indeed possess them (e.g. Snyder, Tanke, & Berscheid, 1977). Thus, if a person charged with an offence has a face that is "representative" of that offence, he/she is more likely to be found guilty than a person who does not have such a face (Macrae & Shepherd, 1989; Shoemaker et al., 1973) regardless of the strength of the prosecution's evidence (Dumas & Testé, 2006). In addition, according to Dangerous Decisions Theory (DDT; Porter & ten Brinke, 2009), intuitive evaluations of trustworthiness are lasting and may strongly direct assimilation and interpretation of the new data and decision about a person. Assessment of credibility and of guilt or innocence could be biased to such a degree that the subsequent judgements are largely irrational (Dumas & Testé, 2006; Kahneman & Tversky, 1982).

Consequently, this study was prompted by the: (a) inconclusiveness of support in favour of the relationship between facial appearance and participation in criminal behaviours; (b) potential negative outcomes of trait inferences of perceived criminality and today's pressure for quick impersonal solutions in detecting potential threat to people's security (Wu & Zang, 2016); (c) inconsistent findings on significant gender differences in face perception, recognition and classification (Megreya, Bindemann, & Havard, 2011); (d) lack of research on association between criminal stereotypes and identification for female suspects and

assessments of a female suspect's guilt (Ward et al., 2012); (e) and even less studies on how well people can assess their own (in)ability to make valuable inferences from faces (Todorov et al., 2015), and how this prospective confidence relates to the accuracy of their judgements.

## **Study aims and hypothesis**

1. To examine participants' ability to differentiate between criminals and non-criminals. Built on previous research (e.g. Johnson, Anderson, Westra, & Suter, 2018; Porter et al., 2008; Valla et al., 2011), it was hypothesised that accuracy of detecting non-criminals will be higher than of detecting criminals, and that the accuracy of detecting both non-criminals and criminals will be greater than the chance.

2. To examine gender differences in competence to distinguish amongst criminals and non-criminals. Previous studies indicated that females were generally more accurate in processing face-specific information (e.g. Rehnman & Herlitz, 2006, 2007; McBain et al., 2009), so it was expected that females will be more accurate in their judgements than males. However, because in this study (non)criminality of only female faces was assessed, and women tend to judge faces of other women as slightly more trustworthy than male faces (Mattrozzi et al., 2015), it was hypothesised that gender differences will be small.

3. To examine the association between the confidence in judging others and accuracy of distinguishing between criminals and non-criminals. Based on the findings from eyewitness identification field it was hypothesised that the association between confidence in judging others and the accuracy of distinguishing between criminals and non-criminals will be small (e.g. Cutler & Penrod, 1989), and that respondents will overestimate their ability (e.g. Brewer & Wells, 2006).

## **Method**

### **Participants**

Data were collected as a part of a larger on-line research , on a convenient sample of 1312 adult internet users. Only data from the participants who fully completed the survey (N=832) were analysed. Participants were mainly young (Mage = 23.95; SDage = 6.76), females (N = 675), with at least a high school diploma (55%). From the analysis answers from the respondents who partially filled survey (N=458) and from those who responded using obvious patterns (e.g. claiming that every photo depicts a (non)criminal) (N = 22) were excluded (see Johnson et al., 2018).

### **Procedure**

#### *Photographs selection (pilot study 1 and 2)*

In order to choose the initial set of photographs, we used the following criteria: coloured, forward-facing photographs with adequate lighting of the face, standard ID photo (photos taken by others for ID documents, excluding police mugshots and selfies), Caucasian, female, without scars or other marks on the face (i.e. tattoos, noticeable make-up) (e.g. Funk & Todorov, 2013), diverse in terms of the type of crime they are suspected of (e.g. homicide, fraud, drug trafficking, terrorism, human trafficking), socio-economic status, age, occupation. Criminals' photographs (N = 59) were chosen from Interpol database of wanted criminal suspects (<https://www.interpol.int/notice/search/wanted>) which is openly accessible. The researchers' female friends and acquaintances' photographs were used as non-criminals' photographs (N =

40), after they gave their written consent for using their photographs in the study, and were acquainted with the purpose and aims of the study. To ensure the photograph uniformity, other body parts except face, neck and shoulders, most of the background were removed, and they were adjusted to the same size and scaling (resized to 360 pixels).

The aim of the first pilot study was to assess the quality (e.g. contrast, acceptable face' lightening) of every photography on a 3-point scale (good, average, bad). Photographs were presented online, one by one and sequentially, in randomized order. In each viewing session, the quality of 99 photographs were estimated by independent raters (between 12 and 21). After the first pilot study, only photographs with a good quality were chosen (N = 35) (Table 1a and Table 1b).

Table 1a *Estimated quality of photographs and age of the photographed persons*

	Whole set	Excluded photos	First selection	Final selection
N <sub>photos</sub>	99	64	35	20
Average age range	18.08 – 67.92	18.08 – 67.92	19.80 – 62.67	21.10 – 62.67
M <sub>age</sub> (SD <sub>age</sub> )	36.38 (11.58)	37.47 (11.53)	34.40 (11.56)	37.25 (12.32)
% good quality photos	0.26	0.23	0.33	0.31
% average quality photos	0.49	0.47	0.52	0.57
% bad quality photos	0.25	0.30	0.15	0.12

Table 1b *Estimated quality of photographs and age of the photographed persons in the final set*

Category	Characteristics	Criminals		Non-criminals		t(18)	p
		M	SD	M	SD		
Photo quality	good	0.25	0.15	0.36	0.23	1.63	.12
	average	0.61	0.10	0.47	0.22	1.83	.08
	poor	0.17	0.13	0.17	0.25	0.03	.99
	Age	37.35	11.38	37.15	13.82	0.19	.09

The aim of the second pilot study was to reduce the number of stimuli (from N=35 to N=20 photographs with satisfying quality) and to collect additional information about photograph properties. There is always a possibility that accuracy of judgements for the set of photographs (criminal/non-criminal) is a product of specific characteristic(s) of a certain photograph within the set (e.g. Todorov & Porter, 2014). Thus, it was examined whether criminal and non-criminal photographs have pronounced characteristics related to social attributions (e.g. Flowe & Humphries, 2011; Klatt et al., 2016; Langlois et al., 2000; Rule & Ambady, 2008; Zebrowitz & Montepare 1992). Using protocols in paper-pencil format, a small group of university students (between 24 and 39) provided their ratings of sequentially presented photographs on a PowerPoint presentation. They rated every face on several dimensions. They were instructed to make their estimates fast, and to rely on their first impressions. Before this task they saw a sample photo and every dimension was briefly orally described to them to reduce interpretative subjectivity. After estimating persons' age, participants rated persons' facial characteristics (e.g. symmetry, femininity, attractiveness, averageness, baby faceness, emotionality) on a 7-point semantic differential type of scale. After that, participants estimated to what degree a person's face expresses specific emotions (e.g. remorse, sadness, anger, happiness,

guilt), and personality traits and competencies (e.g. neurotic, responsible, evil, competent) on a 7-point scale (ranging from 1 - not at all to 7 – completely). The estimates of characteristics for the initial set of 35 photographs (those with satisfying photo quality) and the characteristics for the final set of 20 photographs, used in the main study by the independent raters are shown in Table 2. The direction and significance of the differences in estimates for criminals and non-criminals were consistent in the initial as well as final set of photographs.

Table 2 Initial comparison of a set of criminal and non-criminal photographs after the first and final selection

Category	Characteristics	First selection (19 criminals – 16 non-criminals)		Final selection (10 criminals – 10 non-criminals)	
		<i>t-test</i>	<i>p</i>	<i>t-test</i>	<i>p</i>
Facial characteristics	symmetric - asymmetric	.83	.41	2.08	.05
	feminine - masculine	1.98	.06	.27	.79
	non-attractive- attractive	1.64	.11	1.19	.25
	average - unusual	.65	.52	.08	.94
	baby-face - mature face	.16	.87	.33	.74
	neutral - emotional	1.67	.10	.38	.71
Emotional expressions	anger	.08	.94	.17	.87
	happiness	.28	.78	.07	.95
	remorse	.05	.96	.59	.56
	sadness	.08	.94	.64	.53
	guilt	.59	.56	.48	.64
Personality traits and competencies	criminal	1.32	.20	2.06	.05
	intelligent	.84	.41	1.12	.28
	trustworthy	.79	.43	1.47	.16
	threatening	.02	.99	.87	.40
	social	.83	.41	1.26	.22
	caring	.97	.34	1.72	.10
	evil	.19	.85	1.19	.25
	responsible	1.16	.26	1.55	.14
	nice	.43	.67	1.32	.21
	aggressive	.78	.44	1.63	.12
	capable	1.14	.26	1.27	.22
	pleasant	.50	.62	1.23	.24
neurotic	1.45	.16	1.60	.13	

In order to further examine whether, and to what extent, the photographs of criminal and non-criminal females provoke different impressions in respondents (see Porter, England, Juodis, ten Brinke, & Wilson, 2008; Valla et al., 2011; Todorov & Porter, 2014) a larger sample of participants - a convenient sample of N=252 students at University of Zagreb (60.7% female; Mage=21.27; SDage=2.152) provided assessment of those photographs in the additional pilot study. The same procedure was used as in the previous pilot study but in order to avoid fatigue, each of the students estimated characteristics for only 10 female photographs (5 depicting a criminal and 5 non-criminal) on paper-pencil format protocols. With this additional analysis it was shown that females who were involved in criminal activity were rated as appearing more criminal, threatening, aggressive, more dominant and less trustworthy than females who did not commit any crime (see Table 3).

Table 3 *Additional comparison of a set of criminal and non-criminal photographs in the final set*

Characteristics	Criminals		Non-criminals		<i>t-test</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
criminality	2.98	1.05	2.61	.99	4.96	.00
trustworthiness	3.09	.82	3.20	.83	2.04	.04
threatening	2.79	.94	2.55	.87	3.45	.00
evil	2.58	.92	2.48	.91	1.73	.08
aggressiveness	2.82	1.00	2.56	.95	3.43	.00
dominance	3.33	1.06	3.10	.99	3.65	.00

## Main study

The main study was conducted on-line. In order to mask the motivation behind the study, participants were told they will participate in a study which aims to assess how judgments about ourselves and others are formatted. Every participant was invited either via the mailing list or social network. The main task was to estimate first how much confidence they have in their own judgements' of other people, on a 5-point Likert scale (ranging from 1 - not at all to 5 - completely) ("When I see a person I know with whom I am dealing with"), and then to evaluate females' criminality based on their physical appearance. They were instructed to make their judgments spontaneously and as fast as possible. Before being presented with 20 photographs in a randomized order participants saw the example of viewing and judging task. Each photograph was presented at the screen for 3000 ms. Participants were asked to rate on a 4-point scale (ranging from 1 - not at all to 4 - completely) the degree to which every presented female photograph looks like a criminal. For those photographs which were judged as depicting a criminal, participant had to link the female face with a type of criminal offence chosen from a list (terrorism, fraud, drug trafficking, homicide) (see Valla et al., 2011; Yarmey, 1993). Participants were not familiarized with the number of photographs they would be presented with, nor with the ratio of criminal and non-criminal photographs. After judgments, participants indicated their age and level of education. They had the opportunity for receive feedback on the accuracy of their judgements via e-mail. Ethical Board of the Institute of Social Sciences Ivo Pilar permitted all the aspects of the study.

## Statistical analyses

T-tests were used to assess gender differences in the accuracy and confidence of the appearance-based criminality judgements, and zero-order correlations to assess the link between the appearance-based criminality judgements and the confidence of the judgements.

## Results

The criminality ratings of photographs, made on a 4-point Likert scale, were recoded into binary criminality ratings: responses 1 and 2 were recorded in non-criminality ratings (0), and responses 3 and 4 into criminality ratings (1). After that, these responses were checked against factual criminality of the person at the photograph and recoded into binary accuracy ratings accordingly (accurate – inaccurate). We opted for this approach for two reasons: 1) we were interested in calculating rates of accurately recognized (non)criminals, since studies based on recognition tasks traditionally use binary scales (i.e. “yes” or “no”) in order to assess participants’ recognition abilities, and 2) we wanted to directly compare the rates of successful (non)criminal recognition to those reported in other studies (i.e. Valla et al., 2011). Accuracy ratings were calculated for the whole sample, and separately for males and females. Moreover, accuracy ratings were calculated for the whole set of photographs, and separately for criminals’ and non-criminals’ photos (Table 4).

Table 4 Accuracy of (non)criminality judgements – average number of correctly classified photographs

Judgements’ accuracy	$N_{photos}$	$M$	$SD$	$M_f$	$SD_f$	$M_m$	$SD_m$
Non-Criminals	10	6.99	1.72	7.02	1.70	6.78	1.82
Criminals	10	4.47	1.90	4.43	1.87	4.66	1.99
Total	20	11.45	1.97	11.45	1.96	11.45	2.05

Note.  $f$  – female (N = 675);  $m$  - male (N = 157)

The accuracy of (non)criminality ratings for all photographs was a little above the chance level, since 57% of all photos was correctly classified (Table 4). The classification of the non-criminals was more accurate (70%) than classification of the criminals (45%) indicating innocence bias among the respondents ( $r(832) = .41$ ;  $p < .01$ ;  $t(831) = 23.79$ ,  $p = .000$ ).

Both female and male participants classified non-criminals more accurately than criminals (Table 4) and there were no significant gender differences in the accuracy of criminality ratings for the whole set of photographs ( $t(830) = 2.51$ ;  $p = .973$ ), nor for criminals’ photographs ( $t(830) = 1.39$ ;  $p = .164$ ) and non-criminals’ photographs ( $t(830) = 1.58$ ;  $p = .115$ ) separately.

Table 5 Hit, false alarm, miss and correct rejection rates for criminal recognition accuracy

	Hit	Miss	False Alarm	Correct Rejection	$d'$
Non-Criminal	N/A	N/A	.30	.70	0.36
Criminal	.45	.55	N/A	N/A	

Note.  $d'$  - sensitivity Index

Judgement accuracy, in terms of discriminative sensitivity ( $d'$ ) which controls for response biases, showed that our participants were not very sensitive ( $d' = 0.36$ ) to detecting criminals (Table 5).

Table 6 *Confidence ratings*

Judgements' confidence	<i>M</i>	<i>SD</i>
Total sample	3.06	0.95
Females	3.02	0.93
Males	3.24	1.05

*Note.* Nfemale = 675; Nmale = 157

In general, participants were averagely confident in their capability to judge the person based on the first impression (Table 6). However, males were statistically more confident than females in the accuracy of their judgements ( $t(830) = 2.59$ ;  $p < .01$ ).

Table 7 *Correlations between the confidence and the accuracy of (non)criminality judgements*

	Judgements' accuracy		
	Total	Non-Criminals	Criminals
Confidence	-.02	-.11**	.08*

*Note.* \*  $p < .05$ ; \*\*  $p < .01$

Zero-order correlations indicated that participants who were more confident in their own judgements had higher accuracy in ratings of criminals and lower accuracy in ratings of non-criminals, but the effect size of these associations was very small (Table 7).

Table 8 *Accuracy of (non)criminality judgements with regard to extreme scores on confidence ratings (25% of scale results)*

	Judgment accuracy					
	Total		Criminals		Non-criminals	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Lowest 25%	11.31	1.99	4.28	2.14	7.04	1.76
Highest 25%	11.24	1.91	4.57	1.72	6.67	1.68
<i>t</i> -test	.40		1.64		2.28*	

*Note.* Nlowest 25% = 207; Nhighest 25% = 265;  $df = 470$ ; \*  $p < .05$ ; \*\*  $p < .01$

Those respondents who were the least confident in their accuracy of judging other people were more accurate in ratings of non-criminals than those who were most confident in their accuracy of judging other people (Table 8).

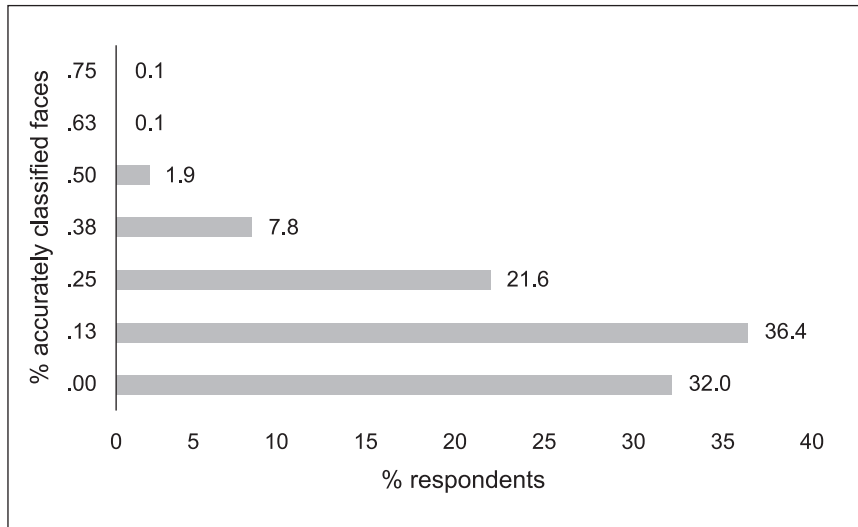


Figure 1 Distribution of participants with regard to the percentage of accurately linked photos to specific crimes

The accuracy of criminals' classification regarding specific crimes was very low (Figure 1). Specifically, 68.4% of participants classified female faces regarding specific crimes below the level of guessing (below 25%).

## Discussion

Although there is a prevailing opinion that appearance-based inferences on criminality should be rather easy, and that more and less trustworthy members of society could be easily discriminated by looking at their faces (e.g. Porter et al., 2008), according to most studies, judgement accuracy levels rarely exceeds chance level, especially when it is controlled for other features related to criminality judgements and/or for response bias (Todorov & Porter, 2014). Todorov and Porter (2014) also argued that even above-chance accuracy is "a rather feeble benchmark" and does not provide a lot of information about the validity of face-based social attributions.

Accuracy of inferences about criminality for the whole set of photographs in our study was 57%, which is concurrent with most studies showing that judgement accuracy is either around 50%, or slightly above (Olivola, Sussman, Tsetsos, Kang, & Todorov, 2012; Todorov et al., 2015). However, there are some findings that the accuracy of inferences on a sample of photographs that had been controlled for racial features and attractiveness is lower than 50% (Valla et al., 2011), indicating that the capability to distinguish some characteristics of people on photographs better than chance may have to do more with the photographs' selection rather than with authentic cues of criminality (Todorov et al., 2015).

When it comes to the classification of criminals vs. classification of non-criminals, the results showed that participants were more prone to label the target as a non-criminal since the classification of the criminals was significantly less accurate (45%) than the classification of the non-criminals (70%). This is in concordance with findings obtained by other studies showing that inferences of non-criminals are statistically more accurate than those of criminals (e.g. Johnson et al., 2018; Porter et al., 2008). This could be at least partly attributed to respondents considering the base criminality population rate. Namely, there is approximately 1% to 5% criminals in the general population, and assuming someone is not a criminal would seem to be a more probable outcome. Since the set of stimuli, presented in research such as this (e.g. Johnson et al., 2018; Porter et al., 2008; Valla et al., 2011), has a significantly higher proportion of criminals

(e.g. 50%) than expected in general population, more accurate identification of non-criminals than criminals could be expected. Also, equalizing groups of people by the majority of clues people use to discriminate between those groups, and exposing them to very specific examples of people from the groups does not allow us to conclude that people are unable to discriminate between them. Moreover, these results could be explained by the activation of the “presumed innocence” heuristics, which became activated because people in general believe that other people are “good”, especially when judgments are made in ambiguous circumstances (Johnson et al., 2018; Tamborini, Huang, Mastro, & Nabashi-Nakahara, 2007). However, study sensitivity index for detecting criminals was higher in some previous studies (i.e. Valla et al., 2011,  $d' = 0.5$ ) than in this study ( $d' = 0.36$ ). It is possible that the participants of the former study were directed toward making more criminality judgements with instruction on how many photographs they will see and with the instruction “that some men were convicted of crimes”.

Literature goes in favour of a relatively high degree of agreement in the connection of specific faces with particular offences (Dumas & Testé, 2006; Macrae and Shepherd, 1989; Shoemaker et al., 1973), but studies investigating the accuracy of linking a specific person to a specific crime consistently showed that accuracy is, also, at best on the chance level. For example, some studies indicated that the main effect of the type of crime on inference accuracy is non-significant (e.g. Funk & Todorov, 2013), while in other studies the accuracy of linking face to specific crimes is between 10% and 20% (Valla et al., 2011), similar to the results in this study. Specifically, 68.4% of participants classified faces regarding specific crimes below the chance level, and only 0.1% of respondents correctly matched 75% of persons to crimes they committed. However, when interpreting these results, several things should be acknowledged. Firstly, the accuracy of associating certain faces with specific offences is directly influenced by the level of congruency between target’s face and the specific offence. When asked to describe what they think a criminal looks like, people significantly more frequently report that the typical criminal perpetrator is “male” (Madriz, 1997; O’Connor, 1984; Reed & Reed, 1973), and distinguish between males’ crimes (e.g. violent crimes and drug offenses) and females crimes (e.g. prostitution and shoplifting) (Skorinko & Spellman, 2013). Since female photographs were used in this study, and the crimes they committed were not typically feminine (e.g. murder, drug dealing, terrorism), the respondents’ low accuracy was maybe caused by induced mismatch between the used stimulus and respondents’ criminality stereotypes. However, previous studies also showed that people couldn’t distinguish between types of criminals but are capable of accurate general inferences about criminality (Valla et al., 2011). These findings are in line with the notion that propensity for criminal behaviour is generalizable for most of the crimes and that the person who committed one type of a crime is likely to commit other types of crime as well (Gottfredson & Hirschi, 1994).

Regarding gender differences in ability to distinguish between criminals and non-criminals, our results indicated that female and male participants had a similar tendency of being more accurate in non-criminal classification than in criminal classification. Although some authors found general advantage for female observers in judging faces (e.g. McBain et al., 2009; Rehnman & Herlitz, 2007; Shapiro & Penrod, 1986; Mattarozzi, Todorov, Marzocchi, Vicari, & Russo, 2015), in our study there were no significant gender differences in the accuracy of general criminality ratings for neither criminal photographs, nor for non-criminal photographs. The obtained results could partly be attributed to general expectation of respondents that females less often commit crimes than males. However, neither some previous studies, in which only male photographs were used, confirmed gender differences in the ability to detect criminality (e.g. Johnson et al., 2018; Valla et al., 2011). Mattarozzi et al. (2015, p. 9) suggested that “gender is a factor that should be considered in studies on first impressions” although the obtained effects are rather small, and thus caution is required in interpreting gender effects.

Finally, similarly to previous studies in the field of recognition accuracy in forensic settings (e.g. Cutler & Penrod, 1989; Perfect, Watson, & Wagstaff, 1993), our results indicated that prospective confidence

is not a (strong) predictor of the accuracy of the appearance-based inferences. Although participants who were more confident in their own judgements had higher accuracy in the ratings of criminals, and lower accuracy in the ratings of non-criminals, the effect size of these associations was very small. This finding could be more related to the content of the question asked to estimate the confidence (“When I see a person I know who I am dealing with”) than to the diagnostic value of the prospective confidence for accuracy of appearance-based inferences. In addition, it could be assumed that prospective confidence in an (un)successful task completion is based on the anticipated type of the task and anticipated difficulty of the task. If there is a mismatch between expectation and the actual properties of the task, that could not be judged before its completion, it is not surprising that the prospective confidence is unrealistic and marginally related to the accuracy of the appearance-based inferences (Sporer, 1993). Moreover, as in previous studies, gender affected the confidence judgment, and it was confirmed that women were less confident than men in judging unfamiliar faces (Mattarozzi, et al., 2015).

The present findings should be interpreted in the light of certain study limitations. Firstly, the conclusion that a person’s facial appearance is, or is not, a valid indicator of underlying characteristics, is directly influenced by the photography selection and quality (e.g. Jenkins, White, Van Montfort, & Burton, 2011; Todorov & Porter, 2014). The praxis regarding photograph selection in the previous research on criminality judgements was very different. For example, Valla et al. (2011) based on their photo selection criteria in the first study concluded that results “could not be attributed to differences in attractiveness, race, gender, age, facial hair, hairstyle, or photo quality between these groups.” In the second study, Valla et al. (2011) additionally controlled for happy, sad, angry, surprised, pleasant, and aroused emotional expression of the faces, while Johnson et al. (2018) while studying appearance-based criminality judgements did not control for any features related to trustworthiness. Quite differently, in the Porter’s et al. (2008), participants in the pilot study rated target faces on the following characteristics: attractiveness, baby-facedness, symmetry, kindness, aggression, age, ethnicity, and familiarity. They concluded that two target groups (criminals and non-criminals) did not differ on these pilot ratings, and that criminals were as likely to be judged as trustworthy as they were to be judged untrustworthy. Todorov and Porter (2014) argued that the ability to discriminate mug shots of arrested people from university students’ photographs (Valla et al., 2011), or America’s Most Wanted persons photographs from the Nobel Peace Prize winners’ photographs (Porter et al., 2008), on a better than chance level, may have to do more with the photographs’ selection rather than with the authentic signals of criminality since none of the control photographs in these studies were taken in the context of the police arrest. Regardless of the efforts made in controlling certain target characteristics in experiments, Adams et al. (2012) believe that eventually no one sees a face as neutral, but a face takes on an emotional tone due to particular properties such as race, age, and sex, and Todorov and Porter (2014) showed that even different images of the same individual can result in different social attributions. Moreover, since people seem to have pre-determined stereotypes about what a criminal looks like (e.g. have unkempt or long hair, scars, pockmarks, facial hair, and sharp eyes - MacLin & Herrera, 2006), activation of criminality stereotype also depends on the type of crime each respondent had in mind while making judgements and the gender of the criminal (Flowe, 2012). Furthermore, social attributions from faces are not just a product of facial physical structure, but also (a)typicality of the face, our implicit association of observed face to specific social category, our knowledge, resemblance of observed face to the faces we (dis)like, or to ourself (Todorov & Porter, 2014), so it would be good to include those variables as well in the future research. The second limitation of this study is related to a relatively small set of photographs ( $N = 20$ ) used. Although this attenuated participants’ fatigue while making repetitive judgements, upcoming studies should examine what is an optimal number of stimulus to produce most reliable judgements. In the future studies, retrospective confidence should be measured and related to the accuracy of the appearance-based inferences. It will be interesting to compare retrospective confidence in judgement accuracy separately for

criminal and non-criminal photographs. In addition, prospective studies should try to imitate as much as possible direct personal interaction between the eyewitness and the target to improve results' external validity. Moreover, as the study was conducted online and the process of assessing the photographs was not controlled, there is a possibility that some distractions which influenced the accuracy of the inferences were present. Finally, the participants in our study were mostly female. Hence, replication of the results in more controlled experimental conditions and with more gender-balanced samples would contribute to the findings validity.

## Conclusion

Females' facial appearance was not proven to be a valid indicator of their criminal behaviour in this study. Namely, the participants' ability to distinguish between criminals and non-criminals did not exceed significantly the chance level. Moreover, the capability of the majority of participants in relating specific faces to specific crimes was not better than chance. Male and female respondents showed the same level of competence in distinguishing between female criminals and non-criminals, and the confidence in judging others did not exert a significant impact on the appearance-based inferences' accuracy. However, limitations of this study prevent us from reaching a definitive conclusion about people's ability to discriminate between criminals and non-criminals based on their appearance, and this research could be considered only as a starting point in further investigation of this topic.

## References

- Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology*, *69*(3), 518-529. <http://dx.doi.org/10.1037/0022-3514.69.3.518>
- Berry, D. S., & Zebrowitz-McArthur, L. A. (1988). What's in a face? Facial maturity and the attribution of legal responsibility. *Personality and Social Psychology Bulletin*, *14*, 23-33. <http://dx.doi.org/10.1177/0146167288141003>
- Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*, 11-30. <http://dx.doi.org/10.1037/1076-898X.12.1.11>
- Bull, R. (1992). Physical appearance and criminality. *Current Psychological Reviews*, *2*, 269-281. <http://dx.doi.org/10.1007/BF02684461>
- Bull, R. (2006). Detecting lies and deceit: The psychology of lying and the implications for professional practice. *Journal of Community & Applied Social Psychology*, *16*, 166-167. <https://doi.org/10.1002/casp.828>
- Bull, R., & McAlpine, S. (1998). Facial appearance and criminality. In A. Memon, A. Vrij, & R. Bull (Eds.), *Psychology and law: Truthfulness, accuracy and credibility* (pp. 59-76). London: McGraw-Hill.
- Bull, R. & Rumsey, N. (1988). *The Social Psychology of Facial Appearance*. New York: Springer. <http://dx.doi.org/10.1007/978-1-4612-3782-2>
- Croatian Bureau of Statistics (2016). *Croatia in Figures, 2016*. Retrieved from [https://www.dzs.hr/Hrv\\_Eng/CroInFig/croinfig\\_2016.pdf](https://www.dzs.hr/Hrv_Eng/CroInFig/croinfig_2016.pdf)
- Croatian Bureau of Statistics (2018, December). *Statistical Yearbook of the Republic of Croatia, 2018*. Re-

- trieved from [https://www.dzs.hr/Hrv\\_Eng/ljetopis/2018/sljh2018.pdf](https://www.dzs.hr/Hrv_Eng/ljetopis/2018/sljh2018.pdf)
- Cutler, B. L., & Penrod, S. D. (1989). Forensically relevant moderators of the relation between eyewitness identification accuracy and confidence. *Journal of Applied Psychology, 74*(4), 650. <http://dx.doi.org/10.1037/0021-9010.74.4.650>
- Dumas R., & Testé, B. (2006). The influence of criminal facial stereotypes on juridic judgments. *Swiss Journal of Psychology, 65*, 237-244. <http://dx.doi.org/10.1024/1421-0185.65.4.237>
- Flowe, H. D. (2012). Do characteristics of faces that convey trustworthiness and dominance underlie perceptions of criminality? *PLoS ONE, 7*(6): e37253. <http://dx.doi.org/10.1371/journal.pone.0037253>
- Flowe, H. D. & Humphries, J. E. (2011). An examination of criminal face bias in a random sample of police lineups. *Applied Cognitive Psychology, 25*, 265-273. <http://dx.doi.org/10.1002/acp.1673>
- Flowe, H. D., Klatt, T., & Colloff, M. F. (2014). Selecting fillers on emotional appearance improves lineup identification accuracy. *Law and Human Behavior, 38*(6), 509-519. <http://dx.doi.org/10.1037/lhb0000101>
- Funk, F., & Todorov, A. (2013). Criminal stereotypes in the courtroom: Facial tattoos affect guilt and punishment differently. *Psychology, Public Policy, and Law, 466-478*. <http://dx.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*, 1431-1443. <http://dx.doi.org/10.1080/02699931.2016.1227305>
- Gould, S. J. (1981). *The Mismeasure of Man*. NY: Norton.
- Gottfredson, M. R., & Hirschi, T. (1994). A general theory of adolescent problem behavior: Problems and prospects. In R. D. Ketterlinus & M. E. Lamb (Eds.), *Adolescent problem behaviors: Issues and research* (pp. 41-56). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Haselhuhn, M. P., & Wong, E. M. (2012). Bad to the bone: facial structure predicts unethical behaviour. *Proceedings of the Royal Society of London B: Biological Sciences, 279*(1728), 571-576. <http://dx.doi.org/10.1098/rspb.2011.1193>
- Helman, E., Leitner, J. B., Deegan, M. P., & Gaertner, S. L. (2013). Facial structure is indicative of explicit support for prejudicial beliefs. *Psychological Science, 24*(3), 289-296. <http://dx.doi.org/10.1177/0956797612451467>
- Hess, U., Blairy, S., & Kleck, R. E. (2000). The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation. *Journal of Nonverbal Behavior, 24*(4), 265-283. <https://doi.org/10.1023/A:1006623213355>
- Hirst, R. B., Teague, A. M., Sodos, L. M., Wickham, R. E., Whittington, L. T., Mills, B. M., & Mitch, E. (2016). Determining Cannabis Use Status From a Photograph: An Assessment of the “Jay-dar” in Neuropsychologists. *Substance Use & Misuse, 401-410*. <http://dx.doi.org/10.1080/10826084.2016.1233564>
- Interpol. (2018). Wanted persons. <https://www.interpol.int/notice/search/wanted.01.11.17>.
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121*(3), 313-323. <http://dx.doi.org/10.1016/j.cognition.2011.08.001>
- Johnson, H., Anderson, M., Westra, H. R. & Suter, H. (2018). Inferences on criminality based on appearance. *Butler Journal of Undergraduate Research, 4*, 86-99.
- Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American, 246*(1), 160-173. <http://dx.doi.org/10.1038/scientificamerican0182-160>
- Klatt, T., Maltby, J., Humphries, J. E., Smailes, H. L., Ryder, H., Phelps, M. & Flowe, H. D. (2016). Looking bad: inferring criminality after 100 milliseconds. *Applied Psychology in Criminal Justice, 12*, 114-125.
- Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior, 20*(3), 165-182. <http://dx.doi.org/10.1007/BF02281954>
- Langlois, J. H, Kalakanis L., Rubenstein A. J., Larson A., Hallam M., & Smoot M. (2000). Maxims or myths of

- beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, 126, 390–423. <http://dx.doi.org/10.1037/0033-2909.126.3.390>
- MacLin, M. K., & Herrera, V. (2006). The criminal stereotype. *North American Journal of Psychology*, 8, 197–207.
- MacLin, O. H., & MacLin, M. K. (2004). The effect of criminality on face attractiveness, typicality, memorability and recognition. *North American Journal of Psychology*, 6(1), 145.
- Macrae, C. N., & Shepherd, J. W. (1989). Do criminal stereotypes mediate juridic judgements? *British Journal of Social Psychology*, 28, 189–191. <http://dx.doi.org/10.1111/j.2044-8309.1989.tb00860.x>
- Madriz, E. I. (1997). Images of criminals and victims: A study on women's fear and social control. *Gender & Society*, 11(3), 342–356. <http://dx.doi.org/10.1177/089124397011003005>
- Marsh, A. A., Adams Jr, R. B., & Kleck, R. E. (2005). Why do fear and anger look the way they do? Form and social function in facial expressions. *Personality and Social Psychology Bulletin*, 31(1), 73–86. <http://dx.doi.org/10.1177/0146167204271306>
- Mattarozzi, K., Todorov, A., Marzocchi, M., Vicari, A., & Russo, P. M. (2015). Effects of Gender and Personality on First Impression. *PLoS ONE*, 10(9): e0135529. <http://dx.doi.org/10.1371/journal.pone.0135529>
- McBain, R., Norton, D., & Chen, Y. (2009). Females excel at basic face perception. *Acta Psychologica*, 130(2), 168–173. <http://dx.doi.org/10.1016/j.actpsy.2008.12.005>
- McKelvie, S. J., Standing, L., Jean, D. S., & Law, J. (1993). Gender differences in recognition memory for faces and cars: Evidence for the interest hypothesis. *Bulletin of the Psychonomic Society*, 31(5), 447–448. <http://dx.doi.org/10.3758/BF03334958>
- McQuiston, D. E., & Malpass, R. S. (2002). Validity of the mockwitness paradigm: Testing the assumptions. *Law and Human Behavior*, 26(4), 439–453. <http://dx.doi.org/10.1023/A:1016383305868>
- Megreya, A. M., Bindemann, M., & Havard, C. (2011). Sex differences in unfamiliar face identification: Evidence from matching tasks. *Acta Psychologica*, 137(1), 83–89. <http://dx.doi.org/10.1016/j.actpsy.2011.03.003>
- Miller, J. L., Rossi, P. H., & Simpson, J. E. (1991). Felony punishments: A factorial survey of perceived justice in criminal sentencing. *The Journal of Criminal Law and Criminology*, 82(2), 396–422. <http://dx.doi.org/10.2307/1143812>
- Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior*, 27(4), 237–254. <http://dx.doi.org/10.1023/A:1027332800296>
- Neth, D., & Martinez, A. M. (2009). Emotion perception in emotionless face images suggests a norm-based representation. *Journal of Vision*, 9, 1–11. <http://dx.doi.org/10.1167/9.1.5>
- O'Connor, M. E. (1984). The Perception of Crime and Criminality: The Violent Criminal and Swindler as Social Types. *Deviant Behavior*, 5, 255–274. <http://dx.doi.org/10.1080/01639625.1984.9967644>
- Olivola C. Y., & Todorov A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, 46, 315–324. <http://dx.doi.org/10.1016/j.jesp.2009.12.002>
- Olivola, C. Y., Sussman, A. B., Tsetsos, K., Kang, O. E., & Todorov, A. (2012). Republicans prefer Republican-looking leaders: Political facial stereotypes predict candidate electoral success among right-leaning voters. *Social Psychological and Personality Science*, 3(5), 605–613. <http://dx.doi.org/10.1177/1948550611432770>
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087–11092. <http://dx.doi.org/10.1073/pnas.0805664105>
- Perfect, T. J., Watson, E. L., & Wagstaff, G. F. (1993). Accuracy of confidence ratings associated with general knowledge and eyewitness memory. *Journal of Applied Psychology*, 78(1), 144–147. <http://dx.doi.org/10.1037/0021-9010.78.1.144>

- org/10.1037/0021-9010.78.1.144
- 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*, 119-134. <http://dx.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*, 477-491. <http://dx.doi.org/10.1080/10683160902926141>
- Porter, S., England, L., Juodis, M., ten Brinke, L., & Wilson, K. (2008). Is the face a window to the soul? Investigation of the accuracy of intuitive judgments of the trustworthiness of human faces. *Canadian Journal of Behavioural Science, 40*(3), 171-177. <http://dx.doi.org/10.1037/0008-400X.40.3.171>
- Reed, J. P., & Reed, R. S. (1973). Status, images, and consequence: Once a criminal always a criminal. *Sociology and Social Research, 57*(4), 460-472.
- Rehman, J., & Herlitz, A. (2007). Women remember more faces than men do. *Acta psychologica, 124*(3), 344-355. <http://dx.doi.org/10.1007/s10610-008-9088-1>
- Rehman, J., & Herlitz, A. (2006). Higher face recognition ability in girls: Magnified by own-sex and own-ethnicity bias. *Memory, 14*(3), 289-296. <http://dx.doi.org/10.1080/09658210500233581>
- Roney, J. R., Hanson, K. N., Durante, K. M., & Maestripieri, D. (2006). Reading men's faces: women's mate attractiveness judgments track men's testosterone and interest in infants. *Proceedings of the Royal Society of London B: Biological Sciences, 273*(1598), 2169-2175. <http://dx.doi.org/10.1098/rspb.2006.3569>
- Rule, N. O., & Ambady, N. (2008). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science, 19*(2), 109-111. <http://dx.doi.org/10.1111/j.1467-9280.2008.02054.x>
- Sacco, D. F., & Hugenberg, K. (2009). The look of fear and anger: facial maturity modulates recognition of fearful and angry expressions. *Emotion, 9*, 39-49. <http://dx.doi.org/10.1037/a0014081>
- Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion, 9*, 260-264. <http://dx.doi.org/10.1037/a0014681>
- Schonlau, M., & Toepoel, V. (2015). Straightlining in Web survey panels over time. *Survey Research Methods, 9*(2), 125-137. <http://dx.doi.org/10.18148/srm/2015.v9i2.6128>
- Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*(2), 139. <http://dx.doi.org/10.1037/0033-2909.100.2.139>
- Shepherd, J. W. (1989). The face and social attribution. In A. A. Young & H. D. Ellis (Eds), *Handbook of Research in Face Processing*. Amsterdam: North Holland. <http://dx.doi.org/10.1016/B978-0-444-87143-5.50024-X>
- Shevlin, M., Walker, S., Davies, M. N., Banyard, P., & Lewis, C. A. (2003). Can you judge a book by its cover? Evidence of self-stranger agreement on personality at zero acquaintance. *Personality and Individual Differences, 35*(6), 1373-1383. [http://dx.doi.org/10.1016/S0191-8869\(02\)00356-2](http://dx.doi.org/10.1016/S0191-8869(02)00356-2)
- Shoemaker, D. J., South, D. R., & Lowe, J. (1973). Facial stereotypes of deviants and judgments of guilt or innocence. *Social Forces, 51*, 427-433. <https://doi.org/10.1093/sf/51.4.427>
- Skorinko, J. L., & Spellman, B. A. (2013). Stereotypic crimes: How group-crime associations affect memory and (sometimes) verdicts and sentencing. *Victims & Offenders, 8*(3), 278-307. <http://dx.doi.org/10.1080/15564886.2012.755140>
- Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35*(9), 656-666. <http://dx.doi.org/10.1037/0022-3514.35.9.656>
- Sporer, S.L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous

- and sequential lineups. *Journal of Applied Psychology*, 78, 22-23. <http://dx.doi.org/10.1037/0021-9010.78.1.22>
- Sučić, I., Wertag, A., & Juničić, N. (2019). Better the Devil you know than the Devil you don't: Dark traits and appearance-based inferences. *Psihološka istraživanja*, 22(1), 89-114. <https://doi.org/10.5937/psistra22-19029>
- Stirrat, M., & Perrett, D. I. (2012). Face structure predicts cooperation: Men with wider faces are more generous to their in-group when out-group competition is salient. *Psychological Science*, 23(7), 718-722. <http://dx.doi.org/10.1177/0956797611435133>
- Tamborini, R., Huang, R-H., Mastro, D., & Nabashi-Nakahara, R. (2007). The influence of race, heuristics, and information load on judgments of guilt and innocence. *Communication Studies*, 58(4), 341-358. <http://dx.doi.org/10.1080/10510970701648566>
- ten Brinke, L., & Porter, S. (2012). Cry me a river: Identifying the behavioral consequences of extremely high-stakes interpersonal deception. *Law and Human Behavior*, 36(6), 469-477. <http://dx.doi.org/10.1037/h0093929>
- The Sentencing Project. (2017, August). Retrieved from <https://www.sentencingproject.org/wp-content/uploads/2016/02/Facts-About-Prisons.pdf>
- Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. In A. Kingstone & M. Miller (Eds.), *The Year in Cognitive Neuroscience 2008*, *Annals of the New York Academy of Sciences*, 1124, 208-224. <http://dx.doi.org/10.1196/annals.1440.012>
- Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: a model based approach. *Social Cognitive and Affective Neuroscience*, 3(2), 119-127. <http://dx.doi.org/10.1093/scan/nsn009>
- Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology*, 25(3), 395-410. <http://dx.doi.org/10.1080/02643290802044996>
- 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, 19-45. <http://dx.doi.org/10.1146/annurev-psych-113011-143831>
- Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different facial images of the same person. *Psychological Science*, 25(7), 1404-1417. <http://dx.doi.org/10.1177/0956797614532474>
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, 12, 455-460. <http://dx.doi.org/10.1016/j.tics.2008.10.001>
- Valla, J. M., Ceci, S. J., & Williams, W. M. (2011). The accuracy of inferences about criminality based on facial appearance. *Journal of Social, Evolutionary, and Cultural Psychology*, 5, 66-91. <http://dx.doi.org/10.1037/h0099274>
- Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: the sequel: A kernel of truth in predictive cheating detection. *Evolution and Human Behavior*, 28(4), 260-271. <http://dx.doi.org/10.1016/j.evolhumbehav.2007.04.006>
- Ward, C., Flowe, H., & Humphries, J. (2012). The effects of masculinity and suspect gender on perceptions of guilt. *Applied Cognitive Psychology*, 26(3), 482-488. <https://doi.org/10.1002/acp.2823>
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17, 592-598. <http://dx.doi.org/10.1111/j.1467-9280.2006.01750.x>
- Wu, X., & Zhang, X. (2016). Automated inference on criminality using face images. *arXiv:1611.04135*.
- Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica*, 114(1), 101-114. [http://dx.doi.org/10.1016/S0001-6918\(03\)00052-0](http://dx.doi.org/10.1016/S0001-6918(03)00052-0)
- Yarmey, D. A. (1993). Stereotypes and recognition memory for faces and voices of good guys and bad guys.

- Applied Cognitive Psychology*, 7, 419-431. <https://doi.org/10.1002/acp.2350070505>
- Zebrowitz, L.A. (1997). *Reading Faces*. New York: Routledge. <http://dx.doi.org/10.4324/9780429493188>
- 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, 603–623. <http://dx.doi.org/10.1007/BF01065855>
- Zebrowitz, L. A., & Montepare J. M. (1992). Impressions of babyfaced individuals across the life span. *Developmental Psychology*, 28, 1143–1152. <http://dx.doi.org/10.1037/0012-1649.28.6.1143>
- Zebrowitz, L. A., Kikuchi, M., & Fellous, J. M. (2010). Facial resemblance to emotions: group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology*, 98(2), 175. <http://dx.doi.org/10.1037/a0017990>
- Zebrowitz, L. A., Voinescu, L., & Collins, M. A. (1996). “Wide-eyed” and “crooked-faced”: Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin*, 22, 1258–1269. <https://doi.org/10.1177/01461672962212006>