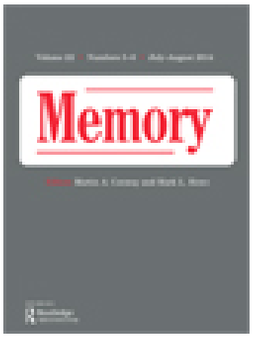

Memory

ISSN: 0965-8211 (Print) 1464-0686 (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

Valenced appearance-behavior cues affect the extent of impression memory

Brittany S. Cassidy

To cite this article: Brittany S. Cassidy (2020): Valenced appearance-behavior cues affect the extent of impression memory, Memory, DOI: 10.1080/09658211.2020.1756337

To link to this article: https://doi.org/10.1080/09658211.2020.1756337

Published online: 23 Apr 2020.

MEMORY
https://doi.org/10.1080/09658211.2020.1756337

Routledge
Taylor & Francis Group

Valenced appearance-behavior cues affect the extent of impression memory

Brittany S. Cassidy

Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC, USA

ABSTRACT
Prior work has often shown higher memory for impressions of valenced verbal cues when such valence is congruent with valence conveyed by actors' facial characteristics. The current work examined specific valence contributions to appearance-congruent memory advantages. Untrustworthy and trustworthy faces were paired with positive, negative, and neutral behaviours (Study 1) and traits (Study 2). Negative versus positive trust-related behaviours are more highly diagnostic and weigh more heavily into impressions, suggesting that impressions of negative behaviours should also be more memorable. Consistent with this possibility, an appearance-congruity advantage for memory of impressions formed from behaviours was larger for untrustworthy versus trustworthy faces after correcting for appearance-congruent response biases (Study 1). When forming impressions from traits, verbal cues less contextualised than behaviours, a larger appearance-congruity advantage in impression memory for untrustworthy versus trustworthy faces could be attributed to appearance-congruent responding (Study 2). Across studies, more extremely valenced impressions were better remembered than more neutral impressions regardless of facial trustworthiness. True appearance-congruity advantages in impression memory may thus be larger for untrustworthy faces when verbal cues can be more contextualised. Further, forming impressions of more extremely valenced verbal cues may enhance impression memory regardless of whether cues are incongruent with facial characteristics.

ARTICLE HISTORY
Received 3 June 2019
Accepted 7 April 2020

KEYWORDS
Trustworthiness; impression formation; valence; impression memory

People perceive countless new faces each day. Because people cannot exhaustively process every stimulus (Desimone & Duncan, 1995), not all of these faces will be remembered. Some faces, however, have intrinsic memorability (Bainbridge et al., 2013). Trustworthiness, a core dimension of face evaluation reflecting relative valence (Oosterhof & Todorov, 2008) and moral character (Krumhuber et al., 2007), contributes to this memorability. When incidentally encoded, faces appearing untrustworthy versus trustworthy are better recognised (Rule et al., 2012). People spontaneously evaluate facial trustworthiness (Todorov et al., 2015), showing consensus in impressions (Rule et al., 2013) and informing downstream behaviour (e.g., avoiding someone who looks untrustworthy; Todorov, 2008). Ecological theories of social perception (Zebrowitz & Collins, 1997) suggest that untrustworthy faces are salient because they signal threat (Buchner et al., 2009; but see Suzuki & Suga, 2010), potentially making them memorable. Indeed, people recognise more untrustworthy than trustworthy faces when trustworthiness is determined by norms and by their own ratings (Rule et al., 2012). Whether untrustworthiness confers advantages beyond recognition is unknown. Understanding this possibility is critical because remembering what a person *has* or *does* contextualises the trust cue signalled by the face and is important to make optimal decisions. The current work tested how facial trustworthiness affects memory for impressions, defined here as feelings that actors are positive or negative, of actors' behaviours and traits.

Although people spontaneously form impressions from facial trustworthiness (e.g., a person who looks untrustworthy has negative traits; Willis & Todorov, 2006), they also do so from verbally presented cues (e.g., trait adjectives or behaviours; Todorov & Uleman, 2002). Impressions from verbal cues are bound to actors' faces (Todorov & Uleman, 2003) and reference identities rather than index co-occurrence (Todorov & Uleman, 2004). Like appearance-based impressions, verbally-based impressions guide behaviours (e.g., avoiding someone who behaves in an untrustworthy way; McCarthy & Skowronski, 2011). Although impressions from appearance and verbal cues have largely been examined in isolation, their interplay affects impression formation. Appearance (e.g., looking untrustworthy), for example, affects the interpretation of ambiguous behaviours (Hassin & Trope, 2000). People also more likely agree with arguments from baby-faced versus mature-faced speakers, perhaps because child-like faces seem honest (Brownlow, 1992). That appearance affects verbal cue evaluation suggests that it could also

CONTACT Brittany S. Cassidy bscassid@uncg.edu Department of Psychology, University of North Carolina at Greensboro, 296 Eberhart, PO Box 26180, Greensboro, NC 27412, USA

affect memory for impressions of verbal cues. The present work examined how looking untrustworthy versus trustworthy affects memory for such impressions.

Social categories affect memory for associated information (attributing a statement matching a "Democrat schema" to a Democrat; Mather et al., 1999). Facial characteristics can function in the same manner. Specifically, people often remember appearance-congruent over appearance-incongruent information, including impressions of verbal cues, when the outcomes of behaviours do not immediately affect them (Cassidy & Gutchess, 2015; Cassidy et al., 2012; Kleider et al., 2012; Nash et al., 2010). This appearance-congruity advantage in memory emerges without mentioning facial trustworthiness (e.g., Nash et al., 2010), supporting its spontaneous evaluation and encoding (e.g., Rule et al., 2012). This memory advantage is also consistent with outcome dependency hypothesis (Erber & Fiske, 1984), which would posit that people remember appearance-congruent impressions in contexts when behavioural outcomes do not directly affect them because they are motivated to maintain schema-relevant expectations (e.g., an untrustworthy-looking person *is* untrustworthy).

Appearance-congruity advantages in impression memory may differ, however, based on relative facial trustworthiness. Untrustworthy faces are perceived as rare (e.g., Barclay, 2008), contributing to their high ecological value (e.g., Buchner et al., 2009; but see van't Wout & Sanfey, 2008) and their advantages in recognition memory (Rule et al., 2012). At the same time, negative versus positive trust-related behaviours are perceived as rare (Mende-Siedlecki et al., 2013), which contributes to their heavier weighting into impressions (e.g., Coovert & Reeder, 1990; Fiske, 1980; Skowronski & Carlston, 1987; Wojciszke et al., 1998, 1993) and their being better remembered (Skowronski & Carlston, 1987). Such heavier weighting reflects a broader literature showing that negative versus positive stimuli produce "more cognitive activity" (Taylor, 1991) that often elicits negative versus positive stimuli as being more memorable (e.g., Ohira et al., 1998). As such, people more strongly update initial impressions given incoming negative (versus positive) trust-related behaviours (Mende-Siedlecki et al., 2013).

Valence-asymmetric updating may extend to how verbal cues affect the updating of appearance-based impressions and thus how facial trustworthiness affects appearance-congruity advantages in impression memory. People have more extreme impressions, for example, given negative versus positive behaviours learned after perceiving untrustworthy faces relative to positive versus negative behaviours learned after perceiving trustworthy faces (Study 2; Cassidy & Gutchess, 2015). Stronger impressions of negative trust-related behaviours may then drive appearance-congruity advantages for untrustworthy faces and weaken them for trustworthy faces. Such a pattern would reflect work showing negative versus positive stimuli to be more memorable (e.g., Robinson-Riegler & Winton, 1996). Positive trust-related behaviours of untrustworthy actors could also be more contextualised (e.g., attributed to situations) to maintain schema-consistency (Gawronski et al., 2014), thus weakening positive impressions (Gawronski et al., 2010) and their memory. Indeed, stronger impressions and expectations enhance memory for congruent verbal cues (Stangor & McMillan, 1992; Stangor & Ruble, 1989). Appearance-congruity advantages in impression memory should emerge when faces are untrustworthy because both appearance and verbal cues are highly salient in this case.

Negative trust-related behaviours, however, should remain salient (e.g., Wojciszke et al., 1993) if associated with trustworthy faces. Given their higher salience and memorability, negative versus positive impressions should be better remembered regardless of facial trustworthiness. This possibility is consistent with a "cheater detection module", in which threatening information is better remembered (e.g., Bell & Buchner, 2012). Indeed, enhanced memory for trustworthy-looking actors who were uncooperative over a series of economic games often emerges (e.g., Suzuki & Suga, 2010), a finding inconsistent with an appearance-congruity advantage in impression memory. As such, an appearance-congruity advantage might not emerge for trustworthy faces. Rather, an attenuated or reversed advantage could emerge in for trustworthy faces. Study 1 thus examined the possibility that appearance-congruity advantages in impression memory are larger for untrustworthy than for trustworthy faces.

Study 1 expanded work on overall appearance-congruity advantages in memory (e.g., Cassidy & Gutchess, 2015) by testing how facial trustworthiness affects these advantages. Study 1 also examined memory for neutral (with regard to trustworthiness) versus positive and negative impressions. It is unclear if receiving any diagnostic (i.e., more extremely valenced) behavioural cues, even if they are appearance-incongruent, enhances impression memory beyond when only one more diagnostic cue is received (i.e., the behavioural cue is relatively neutral). Including neutral impressions is important because it suggests boundaries to the extent to which facial untrustworthiness might afford an appearance-congruity advantage in impression memory. Moreover, although behavioural impressions are bound to actors (Todorov & Uleman, 2002), other verbal cues like trait adjectives also elicit impressions (Wojciszke et al., 1998). Traits and behaviours, however, are differentially processed (Stangor & McMillan, 1992). To further examine the boundaries of an expected larger appearance-congruity advantage in impression memory given untrustworthy faces, Study 2 tested how facial trustworthiness affects impression memory for traits.

Study 1

Untrustworthy faces and negative trust-related behaviours are highly salient (Wojciszke et al., 1998; Zebrowitz &

Collins, 1997) and perceived as rare (Barclay, 2008; Mende-Siedlecki et al., 2013), eliciting their heavier weighting in impressions relative to more positive trust-relevant cues. Here, larger appearance-congruity advantages in impression memory were expected for untrustworthy versus trustworthy faces because having stronger impressions enhances impression memory (Stangor & Ruble, 1989).

Beyond memory for positive and negative impressions, impression memory for more neutral behaviours (with regard to trustworthiness) was also examined. Valenced versus neutral emotional information is better remembered (e.g., Kensinger & Corkin, 2003). Two possibilities emerged for how memory for neutral impressions would compare to negative and positive impression memory of untrustworthy faces.

First, whereas negative impressions associated with untrustworthy faces should be remembered more than positive and neutral impressions, memory for positive and neutral impressions could be similar. Positive versus negative trust-related behaviours are less diagnostic (Wojciszke et al., 1993, 1998). To remain schema-consistent, people could make contextualised attributions for positive behaviours of untrustworthy-looking actors to the extent that impression memory is weakened to the level of more neutral impressions. Engaging emotional processing, however, increases memory for contextual information associated with faces (Mattarozzi et al., 2015). A second possibility was thus that negative and positive impressions associated with untrustworthy faces could be remembered more than neutral impressions because more emotional processing should engage for more extremely valenced verbal cues. Lower memory for neutral impressions associated with trustworthy faces were expected relative to memory for positive and negative associated impressions. Here, negative impressions should be remembered because negative trust-related behaviours are highly diagnostic (Mende-Siedlecki et al., 2013). At the same time, positive impressions should be remembered because they are consistent and because receiving two streams versus one stream of valenced information enhances impression memory (Cassidy & Gutchess, 2015).

A more exploratory goal was to examine boundaries of impression memory advantages by varying the number of behaviours associated with faces. Some work suggests that more same-valenced verbal cues elicit stronger impressions (i.e., an additive model; Triandis & Fishbein, 1963). Because stronger impressions increase memorability, receiving more (versus fewer) negative trust-related behaviours could potentially increase appearance-congruity advantages for untrustworthy versus trustworthy faces. Other work, however, suggests that people average cue valence when forming impressions (Anderson, 1965). If people average behavioural valence, a larger appearance-congruity advantage for untrustworthy versus trustworthy faces should be robust to the number of behaviours associated with faces. Albeit exploratory, testing for these possibilities suggests a boundary to expected patterns in memory.

Method

Participants

Power analyses (PANGEA; for details see www.jakewestfall.org/pangea/) using $d = .33$ ($r = .16$) and alpha = .05 targeted 120 participants for 80% power to detect a Behavioural Valence × Facial Trustworthiness interaction. One-hundred twenty-six undergraduates participated. Six participants with below chance hit rates (.33) were excluded. The final sample comprised 120 participants ($M_{age} = 18.92$ years, $SD = .84$, 91 female). All studies were approved by the Indiana University and University of North Carolina at Greensboro Institutional Review Boards.

Stimuli

Faces

Sixty neutrally expressive Caucasian male faces (30 trustworthy and 30 untrustworthy) were selected from the PAL database (Minear & Park, 2004) on the basis of norms (for details on norms, see Cassidy & Gutchess, 2015). Trustworthy ($M = 4.36$, $SD = .25$) and untrustworthy ($M = 3.59$, $SD = .27$) faces differed with respect to the perceived trustworthiness of their facial characteristics, $t(58) = 11.52$, $p < .001$, $d = 2.96$. Trustworthy and untrustworthy faces were similar in their distinctiveness ($M_{trustworthy} = 3.75$, $SD = .44$, $M_{untrustworthy} = 3.75$, $SD = .39$, $t(58) = .04$, $p = .97$, $d < .01$) and attractiveness ($M_{trustworthy} = 3.18$, $SD = .56$, $M_{untrustworthy} = 3.03$, $SD = .29$, $t(58) = 1.23$, $p = .22$, $d = .34$.).

Behaviours

One-hundred eighty behaviours (60 positive [e.g., "Told the owner of a store that she gave him too much change"], 60 neutral [e.g., "Mailed a letter at the post office"], and 60 negative [e.g., "Turned in someone else's project under his own name"]) were selected from an database of 400 behaviours (Fuhrman et al., 1989) used in impression formation work (e.g., Mende-Siedlecki et al., 2013). These behaviours were pre-normed for goodness (i.e., valence) and kindness (i.e., morality). Sixteen undergraduates rated these behaviours on arousal (1 = not at all arousing to 7 = extremely arousing).

Task versions varied by whether one, two, or three same-valence behaviours were presented at the same time in each trial. One version comprised 60 behaviours (20 positive, 20 neutral, and 20 negative). A second version comprised 120 behaviours (the 60 behaviours used in the first version plus an additional 60 behaviours, totalling 40 positive, 40 neutral, and 40 negative behaviours). A third version comprised 180 behaviours (the 120 used in the second version plus an additional 60 behaviours, totalling 60 positive, 60 neutral, and 60 negative

behaviours). Initial behaviour selection to each version was random. The same combinations of one, two, or three behaviours paired with faces were used across participants, however.

Valence, Morality, and Arousal ratings were entered into two 3 (Behavioural Valence: positive, neutral, negative) × 3 (Behaviour Number: one, two, three) ANOVAs. For Valence, a main effect of Trait Valence emerged, $F(2, 171) = 4031.58$, $p < .001$, $\eta_p^2 = .98$. Positive behaviours ($M = 7.80$, $SD = .29$) were more positive than negative ($M = 1.82$, $SD = .52$), $t(118) = 77.54$, $p < .001$, $d = 14.20$, and neutral ($M = 5.71$, $SD = .20$), $t(118) = 45.24$, $p < .001$, $d = 8.39$, behaviours. Neutral behaviours were also more positive than negative behaviours, $t(118) = 53.94$, $p < .001$, $d = 9.87$. There was no effect of Behaviour Number and no interaction, $ps > .81$. For Morality, a main effect of Trait Valence emerged, $F(2, 171) = 857.97$, $p < .001$, $\eta_p^2 = .91$. Positive behaviours ($M = 7.43$, $SD = 1.11$) were more moral than negative ($M = 1.76$, $SD = .61$), $t(118) = 34.59$, $p < .001$, $d = 6.33$ and neutral ($M = 5.53$, $SD = .36$), $t(118) = 12.56$, $p < .001$, $d = 2.30$, behaviours. Neutral behaviours were also more moral than negative behaviours, $t(118) = 41.12$, $p < .001$, $d = 7.53$. There was no effect of Behaviour Number and no interaction, $ps > .41$. For Arousal, a main effect of Behaviour Valence emerged, $F(2, 171) = 94.01$, $p < .001$, $\eta_p^2 = .52$. Neutral behaviours ($M = 3.38$, $SD = .84$) were less arousing than negative ($M = 4.93$, $SD = .60$), $t(118) = 11.64$, $p < .001$, $d = 2.12$, and positive ($M = 5.08$, $SD = .78$), $t(118) = 11.51$, $p < .001$, $d = 2.10$, behaviours. Negative and positive behaviours did not differ in their arousal, $t(118) = 1.18$, $p = .24$, $d = .21$. There was no effect of Behaviour Number and no interaction on arousal ratings, $ps > .55$.

Face-behaviour pairs

Trustworthy and untrustworthy faces were paired evenly across positive, neutral, and negative behaviours (Figure 1(A)). There were nine task versions. For each Behaviour Number version (one, two, or three same-valence behaviours paired with faces), the positive, neutral, and negative behaviours were counterbalanced across the trustworthy and trustworthy faces. This organisation means that each face was paired with one, two, or three positive, negative, or neutral behaviours in each of nine versions. Overall, there were 60 face-behaviour(s) pairings (10 each of trustworthy/positive, trustworthy/neutral, trustworthy/negative, untrustworthy positive, untrustworthy/neutral, and untrustworthy/negative). Behaviour Number was manipulated between-subjects and randomly assigned.

Procedure

Impression formation task

Participants were told they were participating in a task assessing reaction times and social cognition. Participants were unaware of an upcoming memory test or that the faces systematically varied in their trustworthiness. Stimuli were presented via E-Prime (Psychology Software Tools, Pittsburgh, PA). Participants viewed instructions that read,

> Welcome to the impression formation task. In this task, you will be looking at pictures of people and learning about their personalities. We are assessing how people quickly react to getting to know others. First, you will see a person alone on the screen for two seconds. Press 1 when you have seen the face. Then, you will see the same face paired with one or more behaviors describing that person. Press 2 when you have formed a behavior-based impression.

Like prior work (Cassidy & Gutchess, 2015; Cassidy et al., 2012), each trial began with participants seeing a face alone for two seconds. Each trial continued with participants seeing the same face paired with one, two, or three behaviour(s) for five seconds (Figure 1(A)).

This design affords ecological validity in that it reflects people meeting others and spontaneously evaluating their faces (Todorov, 2008; Zebrowitz & Montepare, 2008) before learning individuating behaviours. Participants were told their responses would not advance stimuli, but that they should respond as soon as they formed an impression. To increase the likelihood of above chance memory, each face-behaviour(s) pair appeared twice, once per run over two runs. Trials within runs were presented at random. To reduce the possibility of recency effects, participants named a number and counted backwards from that number by 7s for 30 s after the impression formation task.

Impression memory task

Participants completed a self-paced impression memory task. Participants viewed instructions that read,

> Welcome to the memory task. In this task, you will be viewing all 60 people you saw in the previous task. There will be three choices listed below each face. These choices will be 'positive,' 'neutral,' or 'negative.' Select 'positive,' 'neutral,' or 'negative' based on what you remember about that person's behavior.

Participants viewed all faces (with no new faces) from the impression formation task presented one at a time in a random order (Figure 1(B)). Participants indicated their behaviour-based impression associated with each face as positive, neutral, or negative via button press. Like past work (e.g., Nash et al., 2010), memory accuracy was determined based on norms (described above) indicating the behaviours as positive, neutral, or negative.

Measuring impression memory

Hit rate

Some past work (Cassidy et al., 2012; Nash et al., 2010) has examined memory accuracy to illustrate appearance-congruity advantages in impression memory. To this end, we examined hit rates in impression memory as a function of facial trustworthiness. Hit rates were defined as being the frequency of accurate responses divided by the frequency of stimuli of that target type. A hit rate for

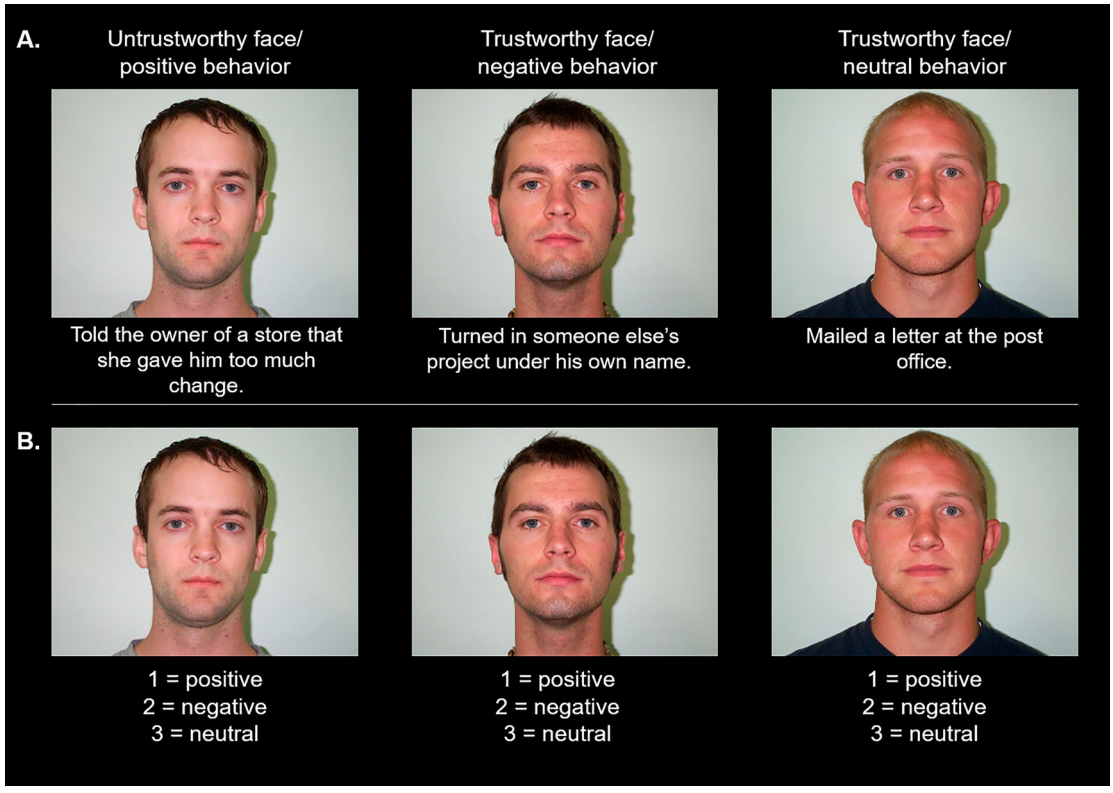

Figure 1. Example face-behaviour pairs in Study 1 (A). No words were listed above the face in actual trials. Example impression memory trials (B).

trustworthy/positive face-behaviour pairs, for example, would be calculated as the number of times "positive" was accurately endorsed as an impression of trustworthy faces that had been paired with positive behaviours divided by 10 (the total number of trustworthy/positive face-behaviour pairs).

Unbiased hit rate

The nature of the impression memory task raised the possibility that responding in an appearance-congruent way (e.g., that trustworthy-looking actors always performed a positive behaviours) might inflate appearance-congruity advantages in memory defined solely by hit rates. To more precisely gauge appearance-congruity advantages as a function of facial trustworthiness, we also examined unbiased hit rates proposed by Wagner (1993) and used in related work (e.g., Cassidy & Gutchess, 2015; Suzuki & Suga, 2010) to control for tendencies to respond in appearance-congruent ways. The unbiased hit rate was calculated by multiplying hit rates by response bias terms. Response biases were defined as the frequency of accurate responses divided by the frequency that a participant endorsed an appearance-congruent impression valence (i.e., positive impressions of trustworthy faces). The response bias term for trustworthy/positive face-trait pairs, for example, would be calculated as the number of times positive was accurately endorsed as an impression of behaviour associated with a trustworthy face divided by the number of times "positive" was endorsed for trustworthy faces overall. We calculated unbiased hit rates for trustworthy/positive, trustworthy/neutral, trustworthy/negative, untrustworthy positive, untrustworthy/neutral, and untrustworthy/negative face-behaviour pairs.

Results

Hit rate

Participants' hit rates were above chance ($M = .49$, $SD = .09$). Hit rates were entered into a 3 (Behaviour Number: one, two, three) × 3 (Behavioural Valence: positive, neutral, negative) × 2 (Facial Trustworthiness: trustworthy, untrustworthy) mixed ANOVA. See Table 1(A). A Behavioural Valence × Facial Trustworthiness interaction emerged (Figure 2(A)), $F(2, 234) = 14.83$, $p < .001$, $\eta_p^2 = .11$. Appearance-congruent memory advantages are suggested if hit rates are higher when relative facial trustworthiness is congruent versus incongruent with the valence of the original behaviour(s). Higher hit rates emerged for positive versus negative, $F(1, 117) = 6.90$, $p = .01$, $\eta_p^2 = .06$, and positive versus neutral, $F(1, 117) = 13.47$, $p < .001$, $\eta_p^2 = .10$, impressions of trustworthy faces. Similar hit rates emerged for negative and neutral impressions of trustworthy faces, $F(1, 117) = 1.75$, $p = .19$, $\eta_p^2 = .02$. Higher hit rates emerged for negative versus positive, $F(1, 117) = 14.39$, $p < .001$, $\eta_p^2 = .11$, and negative versus neutral, $F(1, 117) = 13.97$, $p < .001$, $\eta_p^2 = .11$, impressions of untrustworthy faces. Similar hit rates emerged for positive and

Table 1. Means (standard deviations) for hit rates (A) and unbiased hit rates (B) reflecting impression memory of behaviours associated with trustworthy and untrustworthy faces in Study 1.

A. Hit rate		*Behavioural Valence*		
Facial Trustworthiness	*Behaviour Number*	Positive	Negative	Neutral
Trustworthy	1	.58 (.17)	.50 (.19)	.51 (.23)
	2	.54 (.20)	.51 (.22)	.46 (.19)
	3	.54 (.15)	.46 (.20)	.39 (.17)
	Overall	.55 (.17)	.49 (.20)	.45 (.21)
Untrustworthy	1	.47 (.20)	.52 (.18)	.42 (.18)
	2	.44 (.17)	.56 (.16)	.52 (.22)
	3	.46 (.19)	.53 (.20)	.44 (.17)
	Overall	.46 (.19)	.54 (.19)	.39 (.25)
Overall valence		.50 (.15)	.52 (.16)	.45 (.19)
B. Unbiased hit rate		*behavioural Valence*		
Facial Trustworthiness	*Behaviour Number*	Positive	Negative	Neutral
Trustworthy	1	.33 (.15)	.35 (.17)	.25 (.16)
	2	.27 (.14)	.34 (.20)	.24 (.16)
	3	.25 (.11)	.28 (.15)	.18 (.12)
	Overall	.28 (.14)	.32 (.17)	.22 (.15)
Untrustworthy	1	.27 (.19)	.33 (.18)	.25 (.15)
	2	.23 (.14)	.33 (.15)	.20 (.11)
	3	.22 (.13)	.30 (.16)	.18 (.15)
	Overall	.24 (.15)	.32 (.16)	.21 (.14)
Overall valence		.26 (.12)	.32 (.14)	.22 (.12)

neutral impressions of trustworthy faces, $F(1, 117) = .08$, $p = .78$, $\eta_p^2 < .01$.

Suggesting the above findings were robust to the number of behaviours paired with faces, there was no interaction between Behaviour Number, Behavioural Valence, and Facial Trustworthiness, $F(4, 234) = .30$, $p = .88$, $\eta_p^2 < .01$. Suggesting no overall advantage for untrustworthy faces, no Facial Trustworthiness effect emerged, $F(1, 117) = 2.33$, $p = .13$, $\eta_p^2 = .02$.

A Behavioural Valence effect emerged, $F(2, 234) = 5.07$, $p = .007$, $\eta_p^2 = .04$. Hit rates were higher given positive versus neutral, $F(1, 117) = 4.99$, $p = .03$, $\eta_p^2 = .04$, and negative versus neutral, $F(1, 117) = 8.29$, $p = .005$, $\eta_p^2 = .07$, behaviours. Hit rates were similar given positive versus negative behaviours, $F(1, 117) = .41$, $p = .52$, $\eta_p^2 < .01$.

There was an unexpected Behaviour Number effect, $F(2, 117) = 3.22$, $p = .04$, $\eta_p^2 = .05$. Hit rates were higher given one ($M = .52$, $SD = .11$) versus three ($M = .46$, $SD = .07$) behaviours, $t(78) = 2.54$, $p = .01$, $d = .65$. No differences emerged given one and two ($M = .49$, $SD = .09$), $t(78) = 1.04$, $p = .30$, $d = .30$, or two and three, $t(78) = 1.55$, $p = .13$, $d = .37$, behaviours. No other effects were significant, $ps > .17$.

Unbiased hit rate

To disentangle if the above effects could be attributed to response biases, unbiased hit rates were entered into a 3 (Behaviour Number: one, two, three) × 3 (Behavioural Valence: positive, neutral, negative) × 2 (Facial Trustworthiness: trustworthy, untrustworthy) mixed ANOVA. See Table 1(B). True appearance-congruent memory advantages are suggested if unbiased hit rates are higher when relative facial trustworthiness is congruent versus incongruent with the valence of the behaviour(s).

A Behavioural Valence × Facial Trustworthiness interaction emerged (Figure 2(B)), $F(2, 234) = 3.16$, $p = .04$, $\eta_p^2 = .03$. A pattern for trustworthy faces that contrasted hit rate analyses emerged when correcting for appearance-congruent response biases. Higher unbiased hit rates emerged for negative versus positive impressions of

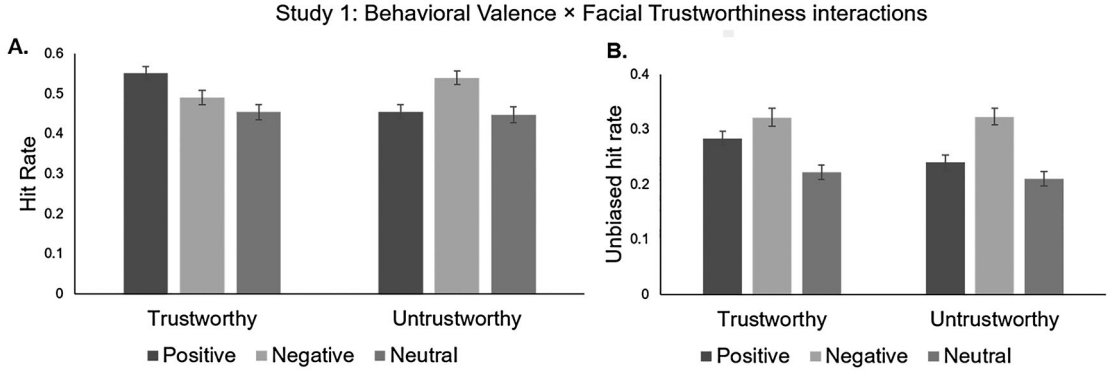

Figure 2. Study 1: When forming impressions of behaviours, appearance-congruity advantages in impression memory emerged for hit rates of trustworthy and untrustworthy faces (A). An appearance-congruity advantage in impression memory only emerged for untrustworthy faces after correcting for appearance-congruent responding (B).

trustworthy faces, $F(1, 117) = 5.43$, $p = .02$, $\eta_p^2 = .04$. Higher unbiased hit rates also emerged for positive, $F(1, 117) = 20.00$, $p < .001$, $\eta_p^2 = .15$, and negative, $F(1, 117) = 28.39$, $p < .001$, $\eta_p^2 = .20$, versus neutral impressions of trustworthy faces. A pattern consistent with hit rate analyses emerged for untrustworthy faces. Higher unbiased hit rates emerged for negative than positive impressions of untrustworthy faces, $F(1, 117) = 36.80$, $p < .001$, $\eta_p^2 = .24$. Higher unbiased hit rates also emerged for positive, $F(1, 117) = 4.61$, $p = .03$, $\eta_p^2 = .04$, and negative, $F(1, 117) = 65.50$, $p < .001$, $\eta_p^2 = .36$, versus neutral impressions of untrustworthy faces. Breaking down the Behavioural Valence × Facial Trustworthiness interaction another way, unbiased hit rates for neutral, $F(1, 117) = .69$, $p = .41$, $\eta_p^2 < .01$, and negative, $F(1, 117) = .01$, $p = .93$, $\eta_p^2 < .01$, impressions did not differ by facial trustworthiness. Unbiased hit rates were higher for positive impressions of trustworthy versus untrustworthy faces, $F(1, 117) = 8.87$, $p = .004$, $\eta_p^2 = .07$.

Consistent with hit rate analyses, there was no interaction between Behaviour Number, Behavioural Valence, and Facial Trustworthiness, $F(4, 234) = .61$, $p = .66$, $\eta_p^2 = .01$, and Facial Trustworthiness effect, $F(1, 117) = 2.72$, $p = .10$, $\eta_p^2 = .02$.

A Behavioural Valence effect emerged that contrasted hit rate analyses, $F(2, 234) = 40.22$, $p < .001$, $\eta_p^2 = .26$. Unbiased hit rates were higher for negative versus positive impressions, $F(1, 117) = 27.68$, $p < .001$, $\eta_p^2 = .19$. Unbiased hit rates were again higher for positive versus neutral, $F(1, 117) = 16.61$, $p < .001$, $\eta_p^2 = .12$, and negative versus neutral, $F(1, 117) = 68.14$, $p < .001$, $\eta_p^2 = .37$, impressions.

The unexpected Behaviour Number effect emerged again, $F(2, 117) = 3.50$, $p = .03$, $\eta_p^2 = .06$. Unbiased hit rates were higher given one ($M = .30$, $SD = .12$) versus three ($M = .24$, $SD = .07$) behaviours, $t(78) = 2.65$, $p = .01$, $d = .61$. No differences emerged given one and two ($M = .27$, $SD = .10$), $t(78) = 1.14$, $p = .26$, $d = .27$, or two and three, $t(78) = 1.57$, $p = .12$, $d = .35$, behaviours. No other effects were significant, $ps > .62$.

Discussion

A larger appearance-congruity advantage in impression memory emerged for untrustworthy versus trustworthy faces when correcting for appearance-congruent response biases. Untrustworthy faces evoke threat (Buchner et al., 2009; Todorov, 2008) and are perceived as rare (Barclay, 2008). Likewise, negative trust-related behaviours are highly diagnostic (Wojciszke et al., 1993, 1998) and perceived as rare (Mende-Siedlecki et al., 2013). As such, people more strongly evaluate negative versus positive behaviours of untrustworthy faces than positive versus negative behaviours of trustworthy faces (Cassidy & Gutchess, 2015). Because stronger impressions enhance memory (Stangor & Ruble, 1989), a larger appearance-congruity advantage for untrustworthy faces naturally expands this work by suggesting that these stronger impressions are more memorable.

Although hit rates suggested appearance-congruity advantages for untrustworthy and trustworthy faces, only the advantage for untrustworthy faces remained after correcting for response biases. Not all apparent appearance-congruity advantages may reflect true memory advantages. In fact, an appearance-incongruity advantage for trustworthy faces emerged after bias correction. This advantage is consistent with work showing higher memory for trustworthy-looking cheaters (e.g., Suzuki & Suga, 2010). Contrasting Study 1, some work has found overall appearance-incongruity advantages (e.g., Bell et al., 2015). Interestingly, explicit evaluation positively increases memory for incongruent person information (Stangor & McMillan, 1992). In line with this work, explicitly evaluating actors by rating likability at encoding elicits appearance-incongruity advantages when remembering their behaviours (Bell et al., 2015). This pattern could, speculatively, weaken in some conditions without explicit evaluations. An appearance-incongruity advantage for trustworthy faces and an appearance-congruity advantage for trustworthy faces may rather reflect a broader "cheater detection module" (e.g., Mealey et al., 1996), such that negative trust-related behaviours are prioritised in impressions (Rozin & Royzman, 2001). Consistent with this idea, negative impressions were best remembered after bias correction.

Impressions of untrustworthy versus trustworthy faces were not better remembered overall. A recognition advantage for untrustworthy faces (Rule et al., 2012) may not extend to impression memory. Negative impressions of untrustworthy versus trustworthy faces were similarly remembered, however, suggesting that negative trust-related behaviours enhance impression memory more than untrustworthy faces do alone. Advantages for untrustworthy faces may emerge in nuanced ways. Although negative behaviours are salient (Rozin & Royzman, 2001), one possibility is that untrustworthy faces could reinforce negative versus positive behavioural impressions to deepen encoding and enhance memory. Another possibility is that facial characteristics could sometimes provide schema-consistent contexts to encode and remember impressions. That is, it could sometimes be simply easier to encode and remember appearance-congruent impressions. Consistent with this possibility, positive impressions of trustworthy versus untrustworthy faces were better remembered. Here, trustworthy faces might provide a schema-consistent context to encode and remember positive impressions.

Although more negative versus positive impressions of untrustworthy faces were remembered, higher memory for positive and negative relative to neutral impressions of untrustworthy faces emerged after response bias correction. This pattern also emerged for trustworthy faces. Attributions of untrustworthy-looking actors' positive behaviours were thus not contextualised or situational to the extent that positive valence conferred no memory advantages. This finding is consistent with work showing higher associative memory for valenced versus neutral

information (e.g., Kensinger & Corkin, 2003) and that emotional context makes faces more memorable (Cassidy & Gutchess, 2015; Mattarozzi et al., 2015). One possibility, potentially reflecting a dilution effect (Nisbett et al., 1981), is that less trust-diagnostic information from verbal cues weakens their value, reducing impression memory relative to when behaviours are more diagnostic. Alternatively, receiving two more extremely valenced cues could enhance impression memory relative to when only one cue conveys more extreme valence (a face and a neutral behaviour). Indeed, faces are better remembered with increasing task emotionality (Mueller et al., 1983).

The larger appearance-congruity advantage for untrustworthy faces was robust to the number of behaviours paired with faces. This finding is consistent with an averaging impression formation model (Anderson, 1965). Increasing the ratio of incongruent to congruent behaviours, however, affects the extent to which congruent versus incongruent person information is advantaged in memory (Srull & Wyer, 1989; Stangor & McMillan, 1992). Appearance-congruity advantages in impression memory for untrustworthy faces could thus shift if behaviours varying in valence are associated with faces. Future work should include combinations of valenced behaviours to further characterise appearance-congruity advantages. Impression memory was also higher given one versus three behaviours. Although same-valence, behavioural content was unrelated. Speculatively, people might spend more time contextualising multiple unrelated behaviours than linking them to actors, potentially weakening actor-impression associations. People also might not have had enough time to read and integrate information from three behaviours versus one behaviour, which could also weaken impression memory.

Although much work has characterised impression memory using behaviours, other verbal cues elicit impressions. One cue, trait adjectives, creates congruity with facial trustworthiness because traits convey moral character (Fiske et al., 2007; Wojciszke et al., 1998). Behaviours require interpretation for impressions to be abstracted from them (Stangor & McMillan, 1992). Because traits do not require additional interpretation, they could be less rich than behaviours, thereby potentially affecting memory. It remains unclear if larger appearance-congruity advantages in impression memory for untrustworthy versus trustworthy faces generalise across verbal cues or are constrained to cues more likely to be interpreted. Study 2 tested how facial trustworthiness affected impression memory when verbal cues were traits.

Study 2

Larger appearance-congruity advantages in impression memory for untrustworthy versus trustworthy faces after correcting for appearance-congruent response biases could generalise across different verbal cues. Yet, people abstract traits from behaviours during impression formation, using traits as an organising principle for person memory (Hastie & Kumar, 1979). Trait abstraction from behaviours means that behaviours are interpreted more than traits during impression formation (for a meta-analysis, see Stangor & McMillan, 1992). Because of this differential processing, appearance-congruity advantages in impression memory as a function of facial trustworthiness could differ from Study 1 if using traits as verbal cues.

Because they elicit more contextualisation than traits, negative behaviours versus traits could be richer diagnostic cues traits (e.g., forming an impression after witnessing someone behave immorally versus hearing secondhand about someone's personality). Supporting this possibility, whereas Study 1 showed that negative behaviours elicited appearance-congruity advantages in impression memory for untrustworthy faces, trait adjectives are discounted (i.e., hold less weight) in impressions if an actor has undesirable facial characteristics (Lampel & Anderson, 1968). Moreover, simple emotion-evoking verbal cues (Somerville et al., 2006) reduce appearance-congruity advantages in impression memory for untrustworthy versus trustworthy faces (Study 1; Cassidy & Gutchess, 2015), contrasting the larger advantages when verbal cues are behaviours (Study 1). Because simpler verbal cues are more likely discounted, they should elicit smaller appearance-congruity advantages in impression memory for untrustworthy faces. If true, one possibility is that larger appearance-congruity advantages for untrustworthy versus trustworthy faces could reflect appearance-congruent response biases rather than a true advantage afforded by untrustworthy faces. Study 2 examined this possibility to further test the boundaries of an appearance-congruity advantage in impression memory for untrustworthy faces.

Method

Participants

132 undergraduates participated. Eleven participants with below chance hit rates (.33) were excluded. The final sample comprised 121 participants ($M_{age} = 18.92$ years, $SD = 1.09$, 73 female).

Stimuli and Procedure

Study 2 replicated Study 1 with one difference: In the five seconds of each trial during which faces were paired with verbal cues, traits were paired with faces. To select traits, fifteen undergraduates rated 555 words in the Anderson personality-trait word database (Anderson, 1968) for valence (1 = extremely negative to 7 = extremely positive). Thirteen undergraduates rated the 555 words for morality (1 = extremely immoral to 7 = extremely moral). Based on these ratings, 180 words were selected (60 positive, 60 neutral, and 60 negative). Fifteen undergraduates rated these 180 words for arousal (1 = not at all emotionally intense to 7 = extremely emotionally intense).

Table 2. Means (standard deviations) for hit rates (A) and unbiased hit rates (B) reflecting impression memory for traits associated with trustworthy and untrustworthy faces in Study 2.

A. Hit rate		*Trait Valence*		
Facial Trustworthiness	*Trait Number*	Positive	Negative	Neutral
Trustworthy	1	.58 (.23)	.58 (.22)	.37 (.18)
	2	.61 (.21)	.53 (.20)	.37 (.15)
	3	.55 (.22)	.54 (.21)	.36 (.16)
	Across traits	.58 (.22)	.55 (.21)	.37 (.16)
Untrustworthy	1	.54 (.25)	.56 (.21)	.42 (.18)
	2	.56 (.20)	.60 (.19)	.41 (.17)
	3	.49 (.24)	.60 (.19)	.43 (.19)
	Across traits	.53 (.23)	.59 (.20)	.42 (.18)
Overall valence		.56 (.19)	.57 (.17)	.39 (.13)
B. Unbiased hit rate		*Trait Valence*		
Facial Trustworthiness	*Trait Number*	Positive	Negative	Neutral
Trustworthy	1	.36 (.23)	.32 (.18)	.18 (.13)
	2	.35 (.19)	.30 (.15)	.18 (.12)
	3	.30 (.17)	.31 (.18)	.16 (.09)
	Overall	.34 (.17)	.31 (.18)	.18 (.12)
Untrustworthy	1	.33 (.20)	.31 (.17)	.22 (.14)
	2	.35 (.20)	.32 (.15)	.21 (.13)
	3	.31 (.21)	.32 (.15)	.22 (.14)
	Overall	.33 (.20)	.32 (.16)	.22 (.14)
Overall valence		.33 (.17)	.31 (.13)	.20 (.10)

Trait Valence, Morality, and Arousal ratings were entered into three 3 (Trait Valence: positive, neutral, negative) × 3 (Trait Number: one, two, three) ANOVAs. For Trait Valence, a main effect of Trait Valence emerged, $F(2, 171) = 5414.17$, $p < .001$, $\eta_p^2 = .98$. Positive traits ($M = 5.95$, $SD = .22$) were more positive than negative ($M = 1.98$, $SD = .17$), $t(118) = 111.94$, $p < .001$, $d = 20.19$, and neutral ($M = 3.56$, $SD = .23$), $t(118) = 58.27$, $p < .001$, $d = 10.62$, traits. Neutral traits were also more positive than negative traits, $t(118) = 42.97$, $p < .001$, $d = 7.81$. There was no effect of Trait Number and no interaction, $ps > .80$. For Morality, a main effect of Trait Valence emerged, $F(2, 171) = 313.71$, $p < .001$, $\eta_p^2 = .79$. Positive traits ($M = 5.15$, $SD = .55$) were more moral than negative ($M = 3.06$, $SD = .45$), $t(118) = 22.86$, $p < .001$, $d = 4.16$, and neutral ($M = 3.81$, $SD = .37$), $t(118) = 15.77$, $p < .001$, $d = 2.86$, traits. Neutral traits were also more moral than negative traits, $t(118) = 9.90$, $p < .001$, $d = 1.82$ There was no effect of Trait Number and no interaction, $ps > .19$. For Arousal, a main effect of Trait Valence emerged, $F(2, 171) = 17.35$, $p < .001$, $\eta_p^2 = .17$. Neutral traits ($M = 3.94$, $SD = .62$) were less arousing than negative ($M = 4.52$, $SD = .59$), $t(118) = 5.27$, $p < .001$, $d = .96$, and positive ($M = 4.44$, $SD = .53$), $t(118) = 4.77$, $p < .001$, $d = .87$, traits. Negative and positive traits did not differ in their arousal, $t(118) = .75$, $p = .46$, $d = .14$. There was no effect of Trait Number and no interaction, $ps > .55$. Participants formed trait-based impressions in the impression formation task.

Results

Hit rate

Overall hit rates were above chance ($M = .50$ $SD = .10$) and similar to Study 1, suggesting that using traits as verbal cues did not affect task difficulty. Hit rates were entered into a 3 (Trait Number: one, two, three) × 3 (Trait Valence: positive, neutral, negative) × 2 (Facial Trustworthiness: trustworthy, untrustworthy) mixed ANOVA. See Table 2 (A). A Trait Valence × Facial Trustworthiness interaction emerged (Figure 3(A)), $F(2, 236) = 6.12$, $p = .003$, $\eta_p^2 = .05$. Hit rates were similar for positive and negative impressions associated with trustworthy faces, $F(1, 118) = 1.73$, $p = .19$, $\eta_p^2 = .01$. Hit rates were higher for positive, $F(1, 118) = 79.96$, $p < .001$, $\eta_p^2 = .40$, and negative, $F(1, 118) = 53.16$, $p < .001$, $\eta_p^2 = .31$, versus neutral impressions associated with trustworthy faces. Like Study 1, hit rates were higher for negative versus positive impressions associated with untrustworthy faces, $F(1, 118) = 4.73$, $p = .03$, $\eta_p^2 = .04$. Hit rates were higher for positive, $F(1, 118) = 19.16$, $p < .001$, $\eta_p^2 = .14$, and negative, $F(1, 118) = 43.27$, $p < .001$, $\eta_p^2 = .27$, versus neutral impressions associated with trustworthy faces.

There was no interaction between Trait Number, Trait Valence, and Facial Trustworthiness, $F(4, 236) = .74$, $p = .56$, $\eta_p^2 = .01$, and no Facial Trustworthiness effect, $F(1, 118) = 1.22$, $p = .27$, $\eta_p^2 = .01$.

A Trait Valence effect emerged, $F(2, 236) = 49.56$, $p < .001$, $\eta_p^2 = .30$. Hit rates were higher for positive versus neutral, $F(1, 118) = 67.07$, $p < .001$, $\eta_p^2 = .36$, and negative versus neutral, $F(1, 118) = 70.06$, $p < .001$, $\eta_p^2 = .37$, impressions. Hit rates were similar for positive versus negative impressions, $F(1, 118) = .40$, $p = .53$, $\eta_p^2 < .01$. No other effects were significant, $ps > .56$.

Unbiased hit rate

To disentangle if the above effects could be attributed to appearance-congruent responding, unbiased hit rates were entered into a 3 (Trait Number: one, two, three) × 3

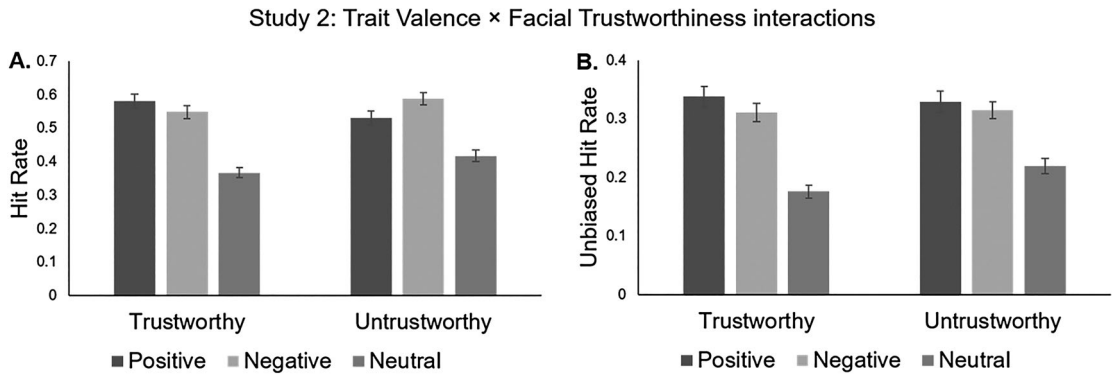

Figure 3. Study 2: When forming impressions of traits, an appearance-congruity advantage in impression memory emerged for hit rates of untrustworthy faces (A). No appearance-congruity advantages emerged after correcting for appearance-congruent responding (B).

(Trait Valence: positive, neutral, negative) × 2 (Facial Trustworthiness: trustworthy, untrustworthy) mixed ANOVA. See Table 2(B). A Trait Valence × Facial Trustworthiness interaction emerged (Figure 3(B)) that was inconsistent with hit rate analyses, $F(2, 236) = 3.12$, $p = .046$, $\eta_p^2 = .03$. Unlike Study 1, unbiased hit rates were similar for negative and positive impressions associated with untrustworthy faces, $F(1, 118) = .60$, $p = .44$, $\eta_p^2 < .01$. Like the hit rate analyses and Study 1, unbiased hit rates were higher for positive, $F(1, 118) = 40.13$, $p < .001$, $\eta_p^2 = .25$, and negative, $F(1, 118) = 42.42$, $p < .001$, $\eta_p^2 = .26$, versus neutral impressions associated with untrustworthy faces. Unlike Study 1, unbiased hit rates for negative and positive impressions were similar for impressions associated with trustworthy faces, $F(1, 118) = 2.62$, $p = .11$, $\eta_p^2 = .02$. Like the hit rate analyses and Study 1, unbiased hit rates were higher for positive, $F(1, 118) = 110.44$, $p < .001$, $\eta_p^2 = .48$, and negative, $F(1, 118) = 92.12$, $p < .001$, $\eta_p^2 = .44$, versus neutral impressions associated with trustworthy faces.

There was no interaction between Trait Number, Trait Valence, and Facial Trustworthiness, $F(4, 236) = .38$, $p = .83$, $\eta_p^2 < .01$. No Facial Trustworthiness effect emerged, $F(1, 118) = 1.09$, $p = .30$, $\eta_p^2 = .01$.

A Trait Valence effect emerged, $F(2, 236) = 73.51$, $p < .001$, $\eta_p^2 = .38$. Unbiased hit rates were higher given positive versus neutral, $F(1, 118) = 110.87$, $p < .001$, $\eta_p^2 = .48$, and negative versus neutral, $F(1, 118) = 104.19$, $p < .001$, $\eta_p^2 = .47$, impressions. Unbiased hit rates were similar given positive versus negative impressions, $F(1, 118) = 2.94$ $p = .09$, $\eta_p^2 = .02$. No other effects were significant, $ps > .54$.

Discussion

Contrasting Study 1, a larger appearance-congruity advantage in impression memory for untrustworthy versus trustworthy faces did not emerge after correcting for appearance-congruent response biases. This finding is consistent with the possibility that verbal cues like negative traits do not weigh as heavily into impressions as negative trust-related behaviours (Cassidy & Gutchess, 2015; Lampel & Anderson, 1968; Stangor & McMillan, 1992). If they did, untrustworthy faces should enhance appearance-congruity advantages because both cues would be highly salient. Instead, memory advantages for impressions associated with untrustworthy faces could be attributed to appearance-congruent responding versus a true advantage. Larger appearance-congruity advantages in impression memory associated with untrustworthy versus trustworthy faces after bias correction thus did not generalise across verbal cues.

If negative versus positive trust-related traits were more heavily weighed into impressions, an overall memory advantage for negative impressions should emerge like in Study 1. This was not the case, as positive and negative impressions were similarly remembered in Study 2. Positive and negative impressions were still remembered over neutral impressions regardless of facial trustworthiness. More extremely versus less valenced cues thus continued to confer impression memory advantages as in Study 1 and past work (e.g., Mattarozzi et al., 2015). Like Study 1, impression memory was robust to the number of traits associated with faces. This finding again supports an averaging model of impression formation (e.g., Anderson, 1965). Unlike Study 1, memory was not higher when one versus three traits was associated with faces. Speculatively, less complexity of trait information (e.g., single adjectives versus multi-word behaviours) could make multiple traits easier to associate with actors than multiple behaviours.

General discussion

Two studies extended work on appearance-congruity advantages in memory (e.g., Cassidy & Gutchess, 2015; Cassidy et al., 2012; Kleider et al., 2012; Nash et al., 2010; Rule et al., 2012) by revealing nuanced effects of facial trustworthiness on how people remember impressions of verbal cues. After forming impressions of behaviours and correcting for appearance-congruent response biases, a congruity advantage in impression memory emerged for untrustworthy faces in Study 1. Study 1 also revealed an appearance-incongruity advantage for trustworthy faces after correcting for appearance-congruent response

biases, linking to work showing appearance-incongruity advantages in memory (Bell et al., 2015; Suzuki & Suga, 2010). Reasoning that these advantages could be due to the relatively rich diagnostic information gleaned from negative trust-related behaviours (Wojciszke et al., 1993, 1998), Study 2 tested if these effects generalised to impression memory for traits, which are less context-rich verbal cues. Contrasting Study 1, advantages could be attributed to appearance-congruent response biases in Study 2. Appearance-congruity advantages in impression memory for untrustworthy faces may more likely reflect true advantages given rich diagnostic verbal cues from which to form impressions.

Rather than overall recognition advantages for untrustworthy faces (Rule et al., 2012) extending to overall advantages in memory for information associated with them, a nuanced memory advantage for information associated with untrustworthy faces emerged in Study 1. Untrustworthy faces could provide a richer context in which to remember impressions of negative versus positive trust-related behaviours. People might contextualise or make more situational attributions for positive than negative trust-related behaviours because positive behaviours are less diagnostic of impressions (Wojciszke et al., 1993). Alternatively, because positive versus negative trust-related behaviours are perceived as more common (Mende-Siedlecki et al., 2013), it might be harder for people to associate actors with specific positive behaviours and associated impressions. These findings also suggest that valence asymmetries in impression updating affect impression memory. Indeed, appearance-incongruity advantages for trustworthy faces in Study 1 suggest that initial positive impressions of trustworthy faces were updated by impressions of negative behaviours to the extent that negative impressions were more likely to be remembered. This pattern did not emerge for untrustworthy faces, suggesting that impressions of positive behaviours did not similarly update negative appearance-based impressions.

Appearance-incongruity advantages for trustworthy faces and appearance-congruity advantages for untrustworthy faces emerged in Study 1, but not Study 2. An overall advantage for negative impressions in Study 1, but not Study 2, suggests that negative behaviours elicit more elaboration than traits because valence must be extracted from behaviours by integrating information comprised in behavioural descriptions. More elaboration of negative behaviours versus traits may elicit deeper processing that enhances memory similar to a generation effect (Bertsch et al., 2007). Alternatively, learning about negative behaviours (e.g., "Turned in someone else's project under his own name") could be more diagnostic for impressions than receiving secondhand trait information (e.g., "liar"). Not all negative behaviours share such high diagnosticity, however. In fact, positive versus negative competence-related behaviours are considered more diagnostic (Wojciszke et al., 1993) and more rare (Mende-Siedlecki et al., 2013). To further examine how facial trustworthiness affects impression memory for behaviours, future work should consider unique roles of valence and diagnosticity by using valenced behaviours normed whether they best apply to a morality versus a competence domain. Such work would elucidate if appearance-congruity advantages in impression memory for untrustworthy faces are larger if negative behaviours are more diagnostic for impressions.

Regardless of whether verbal cues were behaviours or traits, having more extremely valenced impressions afforded memory advantages. This finding conceptually replicates work (e.g., Mattarozzi et al., 2015) by showing that more streams of valenced information offers memory advantages, even if they are incongruent. Prior work showing this advantage used positive and negative verbal cues, but faces that were trustworthy, untrustworthy, and comparatively average with regard to trustworthiness (Cassidy & Gutchess, 2015). The current work suggests impression memory advantages with more extremely valenced cues emerge regardless of the modality of the neutral cue. Future work might directly compare these advantages to examine if the modality of the more neutral cue affects the extent of advantages.

These studies have wide-ranging practical implications in particular for the legal system. Having untrustworthy facial characteristics is associated with more extreme sentencing outcomes (Wilson & Rule, 2015), including greater likelihood of capital sentencing (Wilson & Rule, 2016). Appearance-congruity advantages in impression memory for untrustworthy faces (Study 1) suggest difficulty in remembering positive impressions of these actors' behaviours. As a consequence, people might misattribute negative behaviours to untrustworthy-looking actors. Indeed, appearance-congruent response biases for untrustworthy faces emerged across studies. If it is harder to remember what a person did, it could be easier to default to facial characteristics as a cue to guide decisions. These biases could undermine legal proceedings with effects on both eyewitness testimony and juror behaviour by eliciting unfair outcomes for defendants.

A limitation of the current studies is their focus on the association between appearance and verbal cues. Because memory for actors and verbal cues was not clearly separated, the impression memory data likely reflect combinations of face memory and behaviour/trait memory. Future work should use paradigms designed to replicate the present findings while also testing face memory and behaviour/trait memory. Such studies can highlight how valenced cues more specifically interact to affect impression memory. Because impression valence was determined based on norms and not participants' stated impressions, another limitation is that it is unclear if participants' initial impressions were always accurately identified. Although people show consensus in impressions formed from faces (e.g., Rule et al., 2013) and verbal cues (e.g., Fuhrman et al., 1989), mismatches in participants' actual impressions and norm-based impressions

would add noise to the data. To reduce noise in future work, researchers can use participant-generated impressions to more precisely characterise valence contributions to impression memory.

Together, these studies contribute to a growing understanding of appearance-congruity (and incongruity) advantages in impression memory. By showing conditions by which negative verbal cues elicit advantages in impression memory, the present work links appearance-behaviour impression research to a broader literature showing negative versus positive stimuli to evoke stronger responses (Taylor, 1991) and enhanced memory (e.g., Ohira et al., 1998; Robinson-Riegler & Winton, 1996). The current studies suggest that even subtle differences in facial trustworthiness affect how people remember impressions of verbal cues. Appearance-congruity affects behaviour and memory emerge across contexts from the legal system (Zebrowitz & McDonald, 1991) to parental discipline (Zebrowitz et al., 1991). These broad effects make relative appearance-congruity advantages in memory an important avenue of research to more comprehensively understand social cognition.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Brittany S. Cassidy http://orcid.org/0000-0002-1449-0886

References

Anderson, N. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology*, *70*(4), 394–400. https://doi.org/10.1037/h0022280

Anderson, N. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, *9*(3), 272–279. https://doi.org/10.1037/h0025907

Bainbridge, W., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General*, *142*(4), 1323–1334. https://doi.org/10.1037/a0033872

Barclay, P. (2008). Enhanced recognition of defectors depends on their rarity. *Cognition*, *107*(3), 817–828. https://doi.org/10.1016/j.cognition.2007.11.013

Bell, R., & Buchner, A. (2012). How adaptive is memory for cheaters? *Current Directions in Psychological Science*, *21*(6), 403–408. https://doi.org/10.1177/0963721412458525

Bell, R., Mieth, L., & Buchner, A. (2015). Appearance-based first impressions and person memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *41*(2), 456–472. https://doi.org/10.1037/xlm0000034

Bertsch, S., Pesta, B., Wiscott, R., & McDaniel, M. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210. https://doi.org/10.3758/BF03193441

Brownlow, S. (1992). Seeing is believing: Facial appearance, credibility, and attitude-change. *Journal of Nonverbal Behavior*, *16*(2), 101–115. https://doi.org/10.1007/BF00990325

Buchner, A., Bell, R., Mehl, B., & Musch, J. (2009). No enhanced recognition memory, but better source memory for faces of cheaters. *Evolution and Human Behavior*, *30*(3), 212–224. https://doi.org/10.1016/j.evolhumbehav.2009.01.004

Cassidy, B., & Gutchess, A. (2015). Influences of appearance-behaviour congruity on memory and social judgments. *Memory (Hove, England)*, *23*(7), 1039–1055. https://doi.org/10.1080/09658211.2014.951364

Cassidy, B., Zebrowitz, L., & Gutchess, A. (2012). Appearance-based inferences bias source memory. *Memory & Cognition*, *40*(8), 1214–1224. https://doi.org/10.3758/s13421-012-0233-1

Coovert, M., & Reeder, G. (1990). Negativity effects in impression formation: The role of unit formation and schematic expectations. *Journal of Experimental Social Psychology*, *26*(1), 49–62. https://doi.org/10.1016/0022-1031(90)90061-P

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Erber, R., & Fiske, S. (1984). Outcome dependency and attention to inconsistent information. *Journal of Personality and Social Psychology*, *47*(4), 709–726. https://doi.org/10.1037/0022-3514.47.4.709

Fiske, S. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Experimental Research in Personality*, *22*, 889–906. https://doi.org/10.1037/0022-3514.38.6.889

Fiske, S., Cuddy, A., & 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

Fuhrman, R., Bodenhausen, G., & Lichtenstein, M. (1989). On the trait implications of social behaviors: Kindness, intelligence, goodness, and normality ratings for 400 behavior statements. *Behavior Research Methods, Instruments & Computers*, *21*(6), 587–597. https://doi.org/10.3758/BF03210581

Gawronski, B., Rydell, R., Vervliet, B., & De Houwer, J. (2010). Generalization versus contextualization in automatic evaluation. *Journal of Experimental Psychology: General*, *139*(4), 683–701. https://doi.org/10.1037/a0020315

Gawronski, B., Ye, Y., Rydell, R., & De Houwer, J. (2014). Formation, representation, and activation of contextualized attitudes. *Journal of Experimental Social Psychology*, *54*, 188–203. https://doi.org/10.1016/j.jesp.2014.05.010

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

Hastie, R., & Kumar, P. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology*, *37*(1), 25–38. https://doi.org/10.1037/0022-3514.37.1.25

Kensinger, E., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition*, *31*(8), 1169–1180. https://doi.org/10.3758/BF03195800

Kleider, H., Cavrak, S., & Knuycky, L. (2012). Looking like a criminal: Stereotypical black facial features promote face source memory error. *Memory & Cognition*, *40*(8), 1200–1213. https://doi.org/10.3758/s13421-012-0229-x

Krumhuber, E., Manstead, A., Cosker, D., Marshall, D., Rosin, P., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, *7*(4), 730–735. https://doi.org/10.1037/1528-3542.7.4.730

Lampel, A., & Anderson, N. (1968). Combining visual and verbal information in an impression formation ask. *Journal of Personality and Social Psychology*, *1968*(9), 1–6. https://doi.org/10.1037/h0025716

Mather, M., Johnson, M., & Leonardis, D. (1999). Stereotype reliance in source monitoring: Age differences and neuropsychological test correlates. *Cognitive Neuropsychology*, *16*(3-5), 437–458. https://doi.org/10.1080/026432999380870

Mattarozzi, K., Todorov, A., & Codispoti, M. (2015). Memory for faces: The effect of facial appearance and the context in which the face is encountered. *Psychological Research*, *79*(2), 308–317. https://doi.org/10.1007/s00426-014-0554-8

McCarthy, R., & Skowronski, J. (2011). What will Phil do next? Spontaneousl inferred traits influence predictions of behavior. *Journal of Experimental Social Psychology*, *47*(2), 321–332. https://doi.org/10.1016/j.jesp.2010.10.015

Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology*, *17*(2), 119–128. https://doi.org/10.1016/0162-3095(95)00131-X

Mende-Siedlecki, P., Baron, S., & Todorov, A. (2013). Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. *The Journal of Neuroscience*, *33*(50), 19406–19415. https://doi.org/10.1523/JNEUROSCI.2334-13.2013

Minear, M., & Park, D. (2004). A lifespan database of adult facial stimuli. *Brain Research Methods, Instruments, & Computers*, *36*(4), 630–633. https://doi.org/10.3758/BF03206543

Mueller, J., Heesacker, M., Ross, M., & Nicodemus, D. (1983). Emotionality of encoding activity in face memory. *Journal of Research in Personality*, *17*(2), 198–217. https://doi.org/10.1016/0092-6566(83)90031-4

Nash, R., Bryer, O., & Schlaghecken, F. (2010). Look who's talking! Facial appearance can bias source monitoring. *Memory (Hove, England)*, *18*(4), 451–457. https://doi.org/10.1080/09658211003742706

Nisbett, R., Zukier, H., & Lemley, R. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, *13*, 248–277. https://doi.org/10.1016/0010-0285

Ohira, H., Winton, W., & Oyama, M. (1998). Effects of stimulus valence on recognition memory and endogenous eyeblinks: Further evidence for positive-negative asymmetry. *Personality and Social Psychology Bulletin*, *24*(9), 986–993. https://doi.org/10.1177/0146167298249006

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

Robinson-Riegler, G., & Winton, W. (1996). The role of conscious recollection in recognition of affective material: Evidence for positive-negative asymmetry. *The Journal of General Psychology*, *123*(2), 93–104. https://doi.org/10.1080/00221309.1996.9921263

Rozin, P., & Royzman, E. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Rule, N., Krendl, A., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology*, *104*(3), 409–426. https://doi.org/10.1037/a0031050

Rule, N., Slepian, M., & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition*, *124*(2), 207–218. https://doi.org/10.1016/j.cognition.2012.06.017

Skowronski, J., & Carlston, D. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, *52*(4), 689–699. https://doi.org/10.1037/0022-3514.52.4.689

Somerville, L., Wig, G., Whalen, P., & Kelley, W. (2006). Dissociable medial temporal lobe contributions to social memory. *Journal of Cognitive Neuroscience*, *18*(8), 1253–1265. https://doi.org/10.1162/jocn.2006.18.8.1253

Srull, T., & Wyer, R. (1989). Person memory and judgment. *Psychological Review*, *96*(1), 58–83. https://doi.org/10.1037/0033-295X.96.1.58

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, *111*(1), 42–61. https://doi.org/10.1037/0033-2909.111.1.42

Stangor, C., & Ruble, D. (1989). Strength of expectancies and memory for social information: What we remember depends on how much we know. *Journal of Experimental Social Psychology*, *25*(1), 18–35. https://doi.org/10.1016/0022-1031(89)90037-1

Suzuki, A., & Suga, S. (2010). Enhanced memory for the wolf in sheep's clothing: Facial trustworthiness modulates face-trait associative memory. *Cognition*, *117*, 224–229. https://doi.org/10.1016/j.cognition.2010.08.004

Taylor, S. (1991). Asymmetrical effects of positive and negative events: The mobilzation-minimilization hypothesis. *Psychological Bulletin*, *110*(1), 67–85. https://doi.org/10.1037/0033-2909.110.1.67

Todorov, A. (2008). Evaluating faces on trustworthiness: An extention of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, *1124*(1), 208–224. https://doi.org/10.1196/annals.1440.012

Todorov, A., Olivola, C., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*(1), 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Todorov, A., & Uleman, J. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*(5), 1051–1065. https://doi.org/10.1037/0022-3514.83.5.1051

Todorov, A., & Uleman, J. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, *39*(6), 549–562. https://doi.org/10.1016/S0022-1031(03)00059-3

Todorov, A., & Uleman, J. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*(4), 482–493. https://doi.org/10.1037/0022-3514.87.4.482

Triandis, H., & Fishbein, M. (1963). Cognitive interaction in person perception. *Journal of Abnormal and Social Psychology*, *67*(5), 446–453. https://doi.org/10.1037/h0038494

van't Wout, M., & Sanfey, A. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, *108*(3), 796–803. https://doi.org/10.1016/j.cognition.2008.07.002

Wagner, H. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*(1), 3–28. https://doi.org/10.1007/BF00987006

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wilson, J., & Rule, N. (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. https://doi.org/10.1177.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*(12), 1251–1263. https://doi.org/10.1177/01461672982412001

Wojciszke, B., Brycz, H., & Borkenau, P. (1993). Effects of information content and evaluative extremity on positivity and negativity biases. *Journal of Personality and Social Psychology*, *64*(3), 327–335. https://doi.org/10.1037/0022-3514.64.3.327

Zebrowitz, L., & Collins, M. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review*, *1*(3), 204–223. https://doi.org/10.1207/s15327957pspr0103_2

Zebrowitz, L., Kendall-Tackett, K., & Fafel, J. (1991). The influence of children's facial maturity on parental expectations and punishments. *Journal of Experimental Child Psychology*, *52*(2), 221–238. https://doi.org/10.1016/0022-0965(91)90060-6

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

Zebrowitz, L., & Montepare, J. (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