The face of an angel: Effect of exposure to details of moral behavior on facial recognition memory

Alysha Baker a, Leanne ten Brinke b, Stephen Porter a,∗

a University of British Columbia – Okanagan, Centre for the Advancement of Psychological Science and Law (CAPSL), Canada
b UC Berkeley, Haas School of Business, United States

A R T I C L E  I N F O

Article history:
Received 17 January 2013
Received in revised form 21 March 2013
Accepted 24 March 2013
Available online 6 April 2013

Keywords:
Memory
Biased decision-making
Facial recognition

A B S T R A C T

Although eyewitness recollections are key evidence in the courtroom (e.g., Brewer & Wells, 2011; Wells, Memon, & Penrod, 2006), much research has highlighted their fragility (e.g., Pezdek, 2012; Wells & Olson, 2003). Of the 284 wrongful convictions overturned in part by efforts of The Innocence Project (2012) in the US, 75% of the flawed findings of guilt had relied on eyewitness testimony (subsequently refuted by DNA evidence). A focus of the current research was the potential influence of biasing “moral” information about a target on an observer’s memory for that target, and specifically, the target’s facial appearance.

1. First impressions

First impressions inform observers’ judgments about a target stranger’s character, personality characteristics, and intentions; these intuitive, split-second evaluations reliably are predicated on specific appearance-related features, including facial features, emotional expression, attractiveness, posture, and race (e.g., Andreoni & Petrie, 2008; Bull & Ramsey, 1988; Callan, Powell, & Ellard, 2007; Stewart et al., 2012). Although interpersonal intuition is prone to error, judgments informed by particular physical traits – facial width-to-height ratio, for example – may contain a “kernel of truth” in the prediction of that individual’s propensity for pro-social and anti-social behavior (e.g., Haselhuhn & Wong, 2011; Stirrat & Perrett, 2012; ten Brinke & Porter, 2012).

Research suggests that first impressions are pervasive and powerful in a variety of social situations, including hiring decisions, mate selection, and legal decision-making (e.g., Gilron & Gutchess, 2012; Korva, Porter, O’Connor, Shaw, & ten Brinke, 2012; Langlois et al., 2000; Porter, ten Brinke, & Gustaw, 2010b). Willis and Todorov (2006) demonstrated that trustworthiness judgments based on the face – made within about 100 ms – are enduring, and that confidence in these assessments increases over time (see also Bar, Neta, & Linz, 2006; Porter & ten Brinke, 2009). Such first impressions – particularly, those based on facial appearance – appear to be the result of an instantaneous and subconscious process drawing upon structural characteristics such as higher eyebrows, more pronounced cheekbones, shallow nose sillion, and wider chins as signals of trustworthiness (Bar et al., 2006; Todorov, 2008; Todorov, Baron, & Oosterhof, 2008; Vartanian et al., 2012; Willis & Todorov, 2006). Subjectively, impressions of trustworthiness are positively related to ratings of ‘babyfacedness’, symmetry, and attractiveness (e.g., Bull, 2006; Bull & Vine, 2003; Zebrowitz, Vojnovscu, & Collins, 1996).

2. Dangerous decisions

The Dangerous Decisions Theory (DDT; Porter & ten Brinke, 2009) posits that an interpersonal judgment of trustworthi

© 2013 Society for Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.
occurs instantaneously upon seeing a stranger’s face and that this initial evaluation will be enduring, subsequently influencing the manner in which new information concerning the target is assimilated and interpreted. Additional information about the target will be assimilated to fit the original (sometimes erroneous) initial inference, potentially resulting in inaccurate evaluations. In the first empirical test of this hypothesis, Porter et al. (2010b) found that mock jurors convicted untrustworthy-looking defendants of serious crimes based on less evidence, and with greater confidence, than comparatively trustworthy-looking defendants in the same crime scenario. Further, observers who believe that intuition is a reliable mechanism for guiding decisions are most affected by defendant appearance (Korva et al., 2012). Indeed, the impact of certain facial characteristics have major effects within the criminal justice system: baby-faced-looking individuals are attributed less blame for their actions and receive more lenient sentences than those with mature faces (Berry & Zebrowitz-McArthur, 1988). Further, certain (presumably untrustworthy) faces are perceived to be congruent with certain crimes, with some targets “looking like” a rapist or murderer (e.g., Bull & McAlpine, 1998; Dumas & Teste, 2006). Defendants with such faces, when charged with a congruent offense, are more likely to be convicted of that crime (e.g., Macrae & Shepherd, 1989; Shoemaker, South, & Lowe, 1973) regardless of the strength of the prosecution’s evidence (Dumas & Teste, 2006). Race and facial features also can influence legal decision-making in an alarming way; for example, Eberhardt, Davies, Purdie-Vaughns, and Johnson (2006) discovered that in criminal cases involving a Black defendant and White victim, Black defendants with “stereotypically” Black facial features were more likely to receive the death penalty in the United States than those with less stereotypical features.

### 3. Eyewitness memory and processing biases

Collectively, studies have highlighted the powerful influence of physical appearance on decisions in the legal system. However, little attention has been paid to the possibility of an inverse relationship of decisions on impressions; that is, it is possible that exposure to moral information (e.g., having allegedly committed a violent crime) about a target might influence memory for the target’s appearance. While DDT suggests that biased first impressions based on facial appearance influence subsequent decision-making, we hypothesized that post-assessment information would distort observers’ memories for the target’s face in line with the moral valence of the information. Specifically, it was predicted that information concerning one’s moral actions would influence observers’ recall for his/her facial trustworthiness such that a target who has engaged in an immoral/criminal act will be recalled as appearing less trustworthy than he/she actually appears. If validated, this pattern would suggest that the current DDT model should be expanded to include a bi-directional relationship between decision-making and impressions.

Generally, biased recall may be attributed to the tendency to remember information that is consistent with one’s expectations. Indeed, Macrae, Milne, and Bodenhausen (1994) found that observers who were presented with a target individual and a stereotype label, such as doctor, recalled twice as many stereotype consistent features (e.g., upstanding, responsible, caring) than other participants presented with a non-labeled target. Similarly, in a mock jury paradigm, negative stereotypes, such as ‘hard drug addict who had spent large amounts of time in prison’, elicit more guilty verdicts, harsher sentences, and heightened memory of incriminating evidence from participants than positive stereotypes—particularly when mock jurors were under high cognitive load (Van Knippenberg, Dijksterhuis, and Vermeulen, 1999). Further, this effect extends to the perception of facial features; Hilliar and Kemp (2008) found that ratings of Asian-Australian and European-Australian targets’ racial facial appearance were related to the name assigned to the target by the researchers (e.g., Ng Yat Ho vs. Andrew Brown).

Such effects may be exacerbated in the context of negative emotion. The Paradoxical Negative Emotion (PNE) hypothesis (Porter, Bellhouse, McDougall, ten Brinke, & Wilson, 2010) suggests that negatively valenced – relative to positively valenced – information and events are recalled with greater accuracy but also are more susceptible to misinformation. Memory for negative images is particularly vulnerable to distortion following exposure to misinformation (Brainerd, Stein, Silveira, Rohenkohl, & Reyna, 2008; Hess, Popham, Emery, & Elliott, 2012; Porter, Taylor, & ten Brinke, 2008). Similarly, false memories for negative public events are more likely to be generated than for similar positively valenced events (Porter et al., 2008). This pattern also could extend to memory for faces. That is, while threatening, negatively valenced faces may be recalled with greater accuracy in general (e.g., Cosmides, Tooby, Fiddick, & Bryant, 2005), memory for such faces/stimuli may be highly susceptible to misinformation. For example, Kinzler and Shutts (2008) found an enhanced facial recognition memory for particularly ‘mean-looking’, threatening faces versus ‘nice-looking’ faces. In the present study, we predicted that following the presentation of a face, an ‘untrustworthy’ description of his/her character would distort recalled facial characteristics more than subsequently presented ‘trustworthy’ details.

In the legal system, these natural biases – influenced by characteristics of the crime at hand – may influence the content of an eyewitness’ subsequent recollection. Osbourne and Davies (2012) found that people hold stereotypes about a perpetrator’s appearance based on the particular crime of which they are accused and that these stereotypes contaminate subsequent attempts at eyewitness identification. Participants who witnessed an individual whom they were told was suspected of a “stereotypically” Black crime (e.g., drive by shooting) recalled him as having more stereotypically Black facial features (e.g., darker skin tone, larger nose, and lips), relative to participants who were told that the same man was suspected of a stereotypically White crime (e.g., Internet hacking).

In sum, it appears that perceptions of trustworthiness based on facial features may play an important role in legal decision-making and we hypothesized that the activation of a (un)trustworthy schema would influence eyewitness memory in a similar manner that has been observed with racial stereotypicity (Osbourne & Davies, 2012). However, research to date has not directly tested the influence of encountering biased information on subsequent memory for facial trustworthiness.

### 4. The present study

The current study investigated the influence of information about a target’s moral actions on observers’ memory for target facial features, using a between-subjects design. Face-vignette pairings were manipulated to determine the influence of immoral, morally neutral, or altruistic information on observers’ memory for a previously viewed target face. It was predicted that the nature of the vignette would influence observers’ memory for the target’s facial features of trustworthiness. Specifically, given the susceptibility of negative emotional memories to distortion, this effect was expected to be most evident after exposure to an immoral vignette such that a less trustworthy-looking face (relative to the actual face) would be recognized as the target. These findings would provide bi-directional support for the DDT model. In other words, not only would observers’ initial impression of trustworthiness influence the interpretation of evidence about a target (Porter, Bellhouse,
et al., 2010; Porter, ten Brinke, et al., 2010), but the nature of the evidence also would color memory for the trustworthiness of the target’s face.

5. Methods

5.1. Participants

Undergraduate participants (N = 141) attending a Canadian university were recruited through an online research participant pool and received course credit for completing the study. The sample consisted of 102 women and 38 men (one participant chose not to report his/her gender) with a mean age of 19.45 years (SD = 2.16 years). The immoral, neutral, and altruistic conditions included 37, 34, and 31 female participants, respectively. Fifteen, 16, and seven males completed the immoral, neutral, and altruistic conditions, respectively. A further 22 pilot study participants were recruited by convenience sampling to rate faces and vignettes for main study inclusion.

5.2. Materials

Faces. Images of four Caucasian, male faces were chosen from the Karolinska Set of Images (Lundqvist, Flykt, & Ohman, 1998) to create facial stimuli for the current study. This image set includes 20- to 30-year-old actors exhibiting neutral expressions. Faces were selected for the current study based on their previously rated level of perceived trustworthiness (see Porter, Bellhouse, et al., 2010; Porter, ten Brinke, et al., 2010). Specifically, the selected faces were those that were rated as falling near the middle of the trustworthiness spectrum [i.e., scoring approximately 4 on a scale ranging from 1 (not at all) to 7 (extremely)]. These faces were chosen so that they could easily be manipulated to increase, and decrease, trustworthiness (as opposed to faces that already were extreme in perceived trustworthiness) via manipulation of known facial structural features (e.g., Stewart, et al., 2012).

We planned to select two male faces from a total of four to manipulate in the main study. The four potential faces were uploaded into FaceGen Modeller (Version 3.2; Singular Inversions, 2009), a computer program allowing manual manipulation of features, including emotional expression and facial shape/structure. Once imported to FaceGen Modeller, each face was altered to create a “trustworthy” and “untrustworthy” version (e.g., Porter, ten Brinke, & Mathiessiu, in preparation; Stewart, et al., 2012; Todorov et al., 2008). Based upon previously identified facial characteristics of trustworthiness, the following features were altered to produce the desired effect with the latter representative of high trustworthiness: brow ridge inner (down/up), cheekbones (shallow/low/pronounced), cheeks (gaunt/round), chin (thin/wide), face (gaunt/round), mouth (lips thin/thick and lips small/large), nose (nostrils wide/thin), and nose-sellion (deep/shallow). Progressively trustworthy and untrustworthy versions were created by manipulating particular facial characteristics (e.g., inner brow ridge) in the appropriate direction by two points of measurement on the FaceGen scales, which correspond to standard deviations of the trait within the population. In this way, each original face was changed to appear more and less trustworthy, but otherwise retained their identifying facial characteristics (see Appendix A for an example of an original, high, and low trustworthiness face).

These four sets of male images (the original faces, the manipulated to be high trustworthiness faces, and the manipulated to be low trustworthiness faces) were viewed by a pilot sample (N = 22) to examine whether the manipulated faces achieved the desired effect (e.g., appropriately rated as untrustworthy and trustworthy). Participants rated each face on trustworthiness on a seven-point scale, ranging from 1 (not at all) to 7 (very). Of the four sets, two matched our criteria for inclusion (i.e., all three faces in each set were significantly different from each other in the predicted direction, ps < .05) and were used in the main study.

Facial morphs. FantaMorph 4.0 (Abrosoft, 2007) was used to develop video (i.e., moving) morphs of those two sets of faces selected for inclusion in the main study (see Osbourne & Davies, 2012 for similar morphing). FantaMorph is a software program that seamlessly transforms (i.e., morphs) multiple photographs into a series of naturally changing intermediate exemplars. The facial morphs were created from the three versions (original, high, and low trustworthiness) of each target’s face. Each facial morph transitioned at a speed of 100 frames in 10 s (ranging from low, to original, to high in trustworthiness facial characteristics, or changing in the opposing direction; randomly assigned to control for potential order effects), and then repeated itself infinitely. This speed was chosen to ensure that the morphs transitioned seamlessly but also allowed sufficient time for the participant to accurately choose a face.

Vignettes. Variations of four vignettes were created for this study. Specifically, the variations included similar details but either described an immoral (criminal), morally neutral, or altruistic act. For example, in one series of vignettes, the target either murdered a baby, encountered a baby and walked past without incident, or saved a baby from a harm (see Appendix B for the full example). These twelve vignettes (four immoral, four morally neutral, and four altruistic) were reviewed by a pilot sample to examine whether the vignettes were perceived to be maximally immoral, neutral, or altruistic as intended. Of the various sets of three vignettes, the two sets that most closely fit our desired pattern (maximally immoral, neutral, and maximally altruistic) were chosen for inclusion in the main study. Pilot study participants rated the morality of the target’s behavior for each vignette on a seven-point scale, ranging from 1 (very anti-social) to 7 (very pro-social). For each vignette set, the mean difference ratings between the altruistic and immoral conditions were calculated to determine the set that best fit the intended pattern. The two vignette sets describing an immoral, neutral, or altruistic action (see Appendix B for example) having the largest difference in mean ratings of morality across the altruistic and immoral vignettes were chosen for inclusion in the study. The child vignette described above was chosen and the second vignette chosen for inclusion was similar in structure but specifically described the target either donating to or stealing from a charity. For the vignette involving a child, the means were 6.78 (SD = .42) and 1.57 (SD = 1.27) for the altruistic and immoral condition, respectively, resulting in a mean difference of 5.21. For the vignette involving a charity, the means were 6.57 (SD = .66) and 1.48 (SD = .73) for the altruistic and immoral condition, respectively, resulting in a mean difference of 5.09.1 Mean ratings of neutral vignette versions involving the child and charity were 3.96 (SD = .77) and 4.13 (SD = 1.1), respectively.

5.3. Procedure

Each participant was provided a website address that he/she could follow to complete the study online. Following the completion of a brief demographic survey, participants viewed one target face (the original, unaltered, emotionally neutral FaceGen face) for...
15 s and then read a vignette (on a following page) about the target’s actions (randomly assigned to be either immoral, morally neutral, or altruistic). Next, they were asked to decide whether the target should be punished, rewarded, or experience no repercussions for his actions. This consequence decision served as an attention and comprehension check. Following a half hour delay, during which time individual difference questionnaires were completed, participants were given a memory task. During this phase of the study, they were presented with a constantly changing, and repeating, 10 s morph of the target and were asked to indicate the exact point at which they recognized the face they had viewed previously (i.e., the target face) by pressing the ‘next’ button.

To examine the accuracy of observers’ memories for the target’s face, the point at which they reported identifying the target on the video morph of trustworthiness characteristics was examined. When participants identified the face, a time stamp was associated with this response. This time was adjusted for viewing repetitions and order (response times for participants assigned to view morphs beginning with trustworthy and ending with untrustworthy face versions were reverse scored). Target face recognition scores ranged from 1 (low trustworthiness version) to 5 (actual version; accurate identification) to 10 (high trustworthiness version) seconds.

6. Results

To examine the effect of moral behavioral information on later recall for a target’s facial characteristics, a one-way ANOVA was conducted comparing facial recognition scores across each vignette type (immoral, neutral, altruistic). A significant main effect was found, F(2, 138) = 3.15, p = .046, η² = .04. Mean face identification following presentation of immoral, neutral, and altruistic vignettes occurred at 3.90 (SD = 2.88), 5.25 (SD = 2.98), and 5.03 (SD = 2.70) seconds, respectively. Follow-up analyses revealed a significant difference between the neutral and immoral condition, t(101) = 2.33, p = .02. Facial recognition scores following altruistic information were not different from recognition scores following immoral or morally-neutral vignettes, ps > .05. The relationship between the altruistic and immoral condition was approaching significance, p = .065.

A series of one-sample t-tests was conducted to examine whether facial recognition scores for each condition differed from the correct identification value of 5. Analyses revealed that individuals in the immoral vignette condition (M = 3.90, SD = 2.88) provided ratings significantly different from 5, t(49) = -2.69, p = .01, such that exposure to the immoral vignette led to identifications of faces that were significantly less trustworthy than the actual target face. Mean face identifications in the altruistic vignette and neutral vignette condition were not significantly different from 5 (ps > .05).

7. Discussion

Akin to the well-studied misinformation effect (e.g., Frenda, Nichols, & Loftus, 2011), research suggests that an observer’s memory for a target face may be influenced by ensuing information about that target, distorting subsequent identification. Here, specifically, it was hypothesized that memory for the target’s facial features may be distorted in a predictable manner following exposure to information about the target’s morality/criminality. While this trend has been investigated in relation to race (Osbourne & Davies, 2012), the influence of potentially biasing information encountered post-facial encoding on memory for facial trustworthiness had yet to be examined. Here, we sought to address this gap in the literature by examining the influence of subsequent information about a target’s actions on memory for his/her facial trustworthiness. That is, does hearing about a target’s morality after an initial impression influence the way his/her facial appearance is remembered by an observer? It was predicted that such a result would validate a bi-directional DDT model (Porter & ten Brinke, 2009; Porter, Bellhouse, et al., 2010; Porter, ten Brinke, et al., 2010) with facial trustworthiness impressions influencing subsequent interpretations of evidence about that target, and morality information distorting recall for his/her facial features. Further, this finding would be consistent with cognitive consistency theories that suggest that evaluations are guided by a tendency toward coherence/congruency in a bi-directional manner, such that encountering subsequent information results in a revision of initial impressions to generate coherence (Holyoak & Simon, 1999; Simon, Pham, Le, & Holyoak, 2001).

Our results supported the predicted pattern: encountering subsequent information about a target’s moral actions resulted in distorted memory for his/her facial features (i.e., relating to facial trustworthiness) in line with the morality valence. In particular, observers presented with the immoral vignette mistakenly recognized a less trustworthy version of the target than the actual face they had viewed. The opposite effect, however, was not evident after reading an altruistic vignette, suggesting that memory distortion in this context may be uniquely related to negative information about a target, consistent with the PNE hypothesis (Porter, Bellhouse, et al., 2010; Porter, ten Brinke, et al., 2010). Recognition scores associated with the altruistic vignette actually were the most accurate of the three conditions. Although some past studies have identified that cheaters may be remembered with greater accuracy (e.g., Mealy, Daod, & Krage, 1996), other research suggests that altruists may be better recalled than cheaters (e.g., Barclay & Lalumiere, 2006). While future research is required to further elucidate this relationship, the current study provided support for the latter argument with targets associated with altruistic actions being recalled with near perfect accuracy even in the presence of potentially distorting information. This study is the first to show that post-event information about an individual’s moral activity can distort recognition in predictable ways, extending the effect reported by Osborne and Davies (2012).

While this study elucidates the role of subsequent, and potentially biasing, information on observers’ memory for facial trustworthiness, some consideration should be given to its limitations. First, undergraduate students were used as participants and potential features of real-life crimes were not present. For example, an eyewitness who encounters a robbery suspect as he runs to his vehicle could experience poor encoding of the suspect’s face because of limited exposure, poor lighting, etc. It is likely that by introducing these factors and impairing the initial impression of the face, the effects reported here would be exacerbated, such that subsequently presented information would be used to “fill in the gaps” created by poor encoding conditions. Future research should explore this possibility using a more naturalistic method to further examine the influence of subsequent information on memory for a suspect’s facial features. Further, future research should examine the cognitive processes occurring during the morph sequence to elucidate the underlying mechanisms leading to the pattern exhibited. For example, do observers stop the morph once they recognize the specific version of the target that was originally seen or do observers stop the morph as they become aware that the morph is progressively moving away from the version of the target they

---

2 Participants who provided the inappropriate response (e.g., indicated that the child murderer should be rewarded) were excluded from analysis.

3 A series of one-way ANOVAs was conducted to examine the influence of control variables (vignette, face, and morph order) on face recognition scores. The effects were not significant, ps > .05, suggesting that counterbalancing on these parameters was successful. Further, there was no evident effect of participant gender, p > .05.
remember? In sum, this study used a novel paradigm to highlight the influence of information about a target's character might have on actual memory for his/her facial trustworthiness and further support the fallibility of human memory.

8. Practical application

The overall support for a bi-directional DDT model highlights the negative impact of encountering crime relevant information on memory for previously viewed faces. Collectively, with the large body of research highlighting the fragility of memory, our results show how memory can be colored by emotionally charged, and potentially biasing, information encountered following first impression formation. Following further research, it may be important to educate those encountering eyewitnesses in the legal system, including jurors and judges, on the potentially biasing effect of the defendant's criminal history or offense details, for example, on subsequent eyewitness identifications.

Acknowledgements

This project was funded by grants to the second and third author from the Social Sciences and Humanities Research Council of Canada (SSHRC) and the Natural Sciences and Engineering Research Council (NSERC).

Appendix A. Example of low, original, and high trustworthiness face versions

![Original](image1.png)

![Low Trustworthiness](image2.png)

![High Trustworthiness](image3.png)

Appendix B. Example of Immoral, morally neutral, and altruistic vignettes

**Immoral vignette.** Brian was walking downtown on a Sunday afternoon and stopped at a newsstand to buy a coffee and a paper. As he walked up to the newsstand, he noticed a mother pushing a stroller stop at the stand. Brian walked up to the stand and gave a small candy to the child that he had in his bag. As the mother was ordering, he quickly grabbed the blonde haired girl and continued walking down the street. The moment he heard the mother's scream he ducked into the back alley of his apartment complex and entered through the back entrance. Two weeks later, the child's body was found in a nearby park and the autopsy determined she had died of suffocation.

**Morally neutral vignette.** Brian was walking downtown on a busy Sunday afternoon and stopped at a newsstand to buy a coffee and a paper. As he walked up to the newsstand, he noticed a mother pushing a stroller stop at the stand. Brian walked up beside the two, ordered a coffee and picked up a paper. After he paid, he continued on his walk around the block and into his apartment.

**Altruistic vignette.** Brian was walking downtown on a Sunday afternoon and stopped at a newsstand to buy a coffee and a paper. As he walked up to the newsstand, he noticed a mother pushing a stroller stop at the stand. Brian walked up to the stand and the child looked curiously at him as he started to find a newspaper. As he went to pay, he noticed the child wasn't making any noise, quickly realized the child was missing and saw a man hurriedly walking away with the blonde haired girl in his arms. Realizing what had happened, Brian dropped his paper and ran after the man who took the child. The man tried to turn into a dark alley to avoid detection but Brian caught up with the two and grabbed the girl out of the stranger's hands. Others who also realized what had happened restrained the stranger until the police arrived.

References


