

Facial Composite Production

The Usability of the Reverse Correlation Image Classification Technique for Perpetrator
Recognition

Marlyn Bijvank

Faculty of Behavioral Sciences, University of Twente

Master thesis

1st supervisor: Dr. Elze G. Ufkes

2nd supervisor: Dr. Ir. Peter W. de Vries

Specialization: Psychology of Conflict, Risk & Safety

Twente University

Date: 27 June 2014

Abstract

Current identification techniques are based on the reproduction of memory in two ways; (1) By letting an eyewitness actively describe and select features (recall) to create a face composite or (2) by letting an eyewitness recognize someone in a lineup (recognition). As a result of holistic processing, recall of features is poor. This research tries to construct a composite by matching the preferred retrieval method to the holistic information. We test the usability of the reverse correlation image classification technique for composite purposes. In the first part of this study participants constructed face composites of either a Caucasian or a Moroccan target perpetrator. Resemblance judgments of independent participants showed that the constructed face composites resembled their targets well. Furthermore, in a line-up the majority of the participants selected the correct perpetrator. Recommendations for future research are suggested.

The Usability of the Reverse Correlation Image Classification Technique for Perpetrator Recognition

On April, 4, 2011, Derrick Williams was exonerated after serving nearly 17 years in prison for a crime he did not commit. He was convicted based on eyewitness identification, this is his story: On August, 6, 1992 a 25-year old Caucasian woman arrived at her home and saw an African American man standing on her front porch. When she tried to exit her car the man forcibly entered her car. He pushed her over to the passenger seat and began to drive. Once parked at an abandoned place, the man sexually assaulted her. After the attacker left the vehicle, the woman managed to crawl back to the driver's seat and drove away. In the following police investigation, the victim was exposed to a photo lineup that contained two pictures of Williams. After the lineup, the victim said that she was 80% certain that Williams attacked her. After a subsequent live lineup, the victim said she was 100% certain that Williams was her attacker. The victim's eyewitness testimony was key evidence at trial. Despite all efforts of family and friends providing an alibi, Williams was sentenced to life in prison. On April, 4, 2011 Williams was set free and officially exonerated because DNA evidence had excluded Williams as the perpetrator (Innocence Project, n.d.-a).

Unfortunately, the case described above is not exceptional. It is just one of the many cases described by the non-profit organization "The Innocence Project" that tries to exonerate wrongfully convicted individuals. According to their website (www.innocenceproject.org), eyewitness misidentification is the major cause of wrongful convictions. In almost 75 percent of the DNA exoneration cases handled by this organization, wrongful eyewitness identification played a role (Innocence Project, n.d.-b). An eyewitness misidentification can lead investigators away from the real perpetrator and toward an innocent person. Errors in recognizing persons are usually made during one of these moments; (1) perception, when the physical appearance is encoded. (2) The retention period, when the witness has to store the

physical appearance of the perpetrator. (3) Retrieval, when the witness has to reproduce information about the perpetrator (van Koppen & Wagenaar, 2010). The focus of this paper lies on the retrieval part of memory.

Solving crimes often involves eyewitnesses. Eyewitnesses can help investigators by making a face composite of the perpetrator. Current composite programs have their limits; the resemblance of composites is often low. This study aims to establish whether a new computer technique can be used to produce perpetrator composites. The technique is based on recognition and steps away from conscious recall processes. It creates stimuli by adding or subtracting noise to a base face. Participants decide to what extent the stimuli represent a target face. By aggregating the faces that are selected to be representative of the target face, a representation of the perpetrator can be created. This technique is called the reverse correlation image classification technique (Mangini & Biederman, 2004). Dotsch and Todorov (2012) used this approach to model social perception of faces. Participants successfully created images of intended personality traits (untrustworthy or trustworthy, dominant or submissive). Mangini and Biederman (2004) used this technique to create representations of famous individuals. In this study we try to construct a representation of an individual as well. Besides, as is the case in the first paragraph, there seems to exist a cross-race effect in eyewitness identification evidence. Eyewitnesses seem to perform worse in identifying perpetrators from another race (Chiroro & Valentine, 1995). This study carefully tries to establish whether perpetrator group membership influences the performance of participants on the reverse correlation image classification task. In addition, research has suggested that combining input from multiple witnesses holds better potential in identification cases than using the input from all witnesses separately (Wells & Hasel, 2007). Merging individual composites together should reinforce correct aspects and minimize incorrect aspects. In a study by Bruce, Ness, Hancock, Newman and Rarity (2002) a morph

of four face composites, was rated as a better likeness than average individual composites and as good as the best composite. We wonder whether this effect also applies to the reverse correlation image classification task. Therefore the research questions central for the current work are:

To what extent can produced classification images be used as a composite, indicated by resemblance ratings?

- a. Is this influenced by in-group- versus out-group membership of the target?*
- b. What is the effect of combined input from multiple witnesses on the resemblance ratings?*

Eyewitness memory

Judges and juries consider eyewitness testimony and eyewitness identifications as a strong piece of evidence (Brigham & Wolfskeil, 1983). Research has shown that both lay people and jurors do not take the circumstances and factors influencing eyewitness identification evidence into account. They consider only one variable, the witness confidence as a significant predictor of accuracy (Cutler, Penrod, & Dexter, 1990; Cutler, Penrod, & Stuve, 1988). As demonstrated in the opening case, witnesses' confidence is not necessarily a reliable factor. Research established there are problems with eyewitness identifications. Various factors have been demonstrated to affect memory accuracy (Clifford & Scott, 1978; Loftus, 1979; Sporer, 1992; Wright, Memon, Skagerberg, & Gabbert, 2009). It seems our memory is not always an exact representation of what happened (van Koppen & Wagenaar, 2010). When retrieving information, all available information is used to reconstruct an event. We tend to fill gaps in our memory with our own prior knowledge, expectations and assumptions of what seems logical to have happened (Valentine, 2012). Human memory is not like a tape recorder. We are not capable to record events exactly how they happen, nor are we capable of recalling them like a videotape that starts over. Memory is an active process

that is affected by the creativity of the person that is trying to remember an event. It is vulnerable to suggestion and bias (van Koppen & Wagenaar, 2010).

Memory in police practice. Current identification techniques are based on the reproduction of memory in two ways; (1) Letting an eyewitness actively describe and select features using a police sketch-artist or computer composite technology (recall) or (2) letting an eyewitness recognize someone in a lineup (recognition). When no suspect is apprehended at the scene of the crime, the police can rely on eyewitnesses to give a detailed description of the perpetrator. In this case the eyewitness is asked to recall the perpetrator. Police investigators have an eyewitness sit down with a sketch artist to create a feature-based composite of the perpetrator. Despite technological advances, research has shown that contemporary composite production systems produce poor likeness of target faces and constructing a composite may harm later recognition performance (Wells & Hasel, 2007). When an offense is committed in the presence of an eyewitness and the police have enough information to implicate a suspect in an offense, they can ask the eyewitness to point out the suspect in a photo lineup or a live lineup. In this procedure, the eyewitness is asked to recognize the perpetrator. Studies have shown recognition would lead to more reliable information, particularly when recollection of physical appearance of suspects is involved, (Valentine, 2012).

Problems with eyewitness evidence. Even though people assign a high level of accuracy and confidence to eyewitness identifications, some major implications must be addressed. First of all, studies have shown that giving a verbal description of a person can harm later recognition of that person. This is called the verbal-overshadowing effect (Schooler & Engstler-Schooler, 1990). As the eyewitness is trying to construct a verbal description of the perpetrator, his or her mental picture of the perpetrator will change. Second, when a composite is produced, the eyewitness is committed to the produced

composite and this can harm the subsequent identification in a lineup. Because of the witness' commitment to the composite, he or she may choose the person that most closely resembles that composite. This may increase the eyewitness' confidence but not necessarily his or her accuracy (Wells, Olson, & Charman, 2002). This effect is called the commitment effect and poses a serious problem in eyewitness identification cases (Dysart, Lindsay, Hammond, & Dupuis, 2001). Preferably, eyewitnesses are only exposed to direct recognition in a lineup, without interference of the production of composites. Still, a problem with lineup identification is that sometimes an eyewitness is too eager to identify a suspect or they are afraid the investigation will stop if they do not make an identification (Innocence Project, n.d.-c). A person could recognize an innocent person because they were present at the scene of the crime (or they seem familiar from another context), and therefore they have a memory of that person. This is called unconscious transference. (Egeth, 1993; Ross, Ceci, Dunning, & Toglia, 1994). The eyewitness transfers the identity of the innocent person to that of the perpetrator. In the case described in the first paragraph, the police admitted two pictures of the accused in a photo lineup. This was harmful because the eyewitness could have recognized the accused, from another context, such as the photo lineup.

Another problem introduced in the opening case is the problem of cross racial identification. Studies indicate people are better at recognizing faces from their own race. Identification performance is worse when identifying someone from another race. A study conducted by Chiroro and Valentine (1995) showed that Caucasian witnesses are better in recognizing Caucasian faces and African American witnesses are better in recognizing African American faces. This could have been the case in the opening case as well. The victim's identification accuracy could have been lower because she was Caucasian and her assailant was African American. A common explanation of this own-race bias is the assumption that we perceive other-race persons as having physiognomic homogeneity (out-

group homogeneity bias). We have difficulties discriminating between other-race persons as we do not see differences in their appearance; they look all the same to us. Whereas when we are discriminating between same-race persons we see variation in facial features. Ackerman et al. (2006) researched this homogeneity bias, they exposed participants to own-race and other-race faces with either a neutral or an angry facial expression, participants identified previously seen faces. Results showed recognition accuracy was worse for neutral other-race faces than for neutral same-race faces. However, recognition accuracy was higher for angry other-race faces than for angry same-race faces. This effect can perhaps be explained by evolutionary principles. As our cognitive resources are limited, we cannot attend to all available information in our environment. In social situations, we focus our attention more closely to circumstances whose physical appearance suggests better benefits for our own survival (e.g. focusing on a snake rather than focusing on a tree). Another popular explanation for this own-race bias is the contact hypothesis; the less contact we have with other-race persons, the less opportunities and experience we have to distinguish between faces of other races. People are more skilled in the recognition of faces they have more experience and contact with (Tanaka, Kiefer, & Bukach, 2004). Our perceptual expertise of in-group members is higher. Another take on this contact hypothesis is that since we have less contact with other-race persons, our attitudes towards them are more negative (Chiroro & Valentine, 1995). This in turn, means that we are less interested and motivated to recognize and identify people from other races. It seems group membership is an important variable in this research field, it affects motivation. A more recent study confirms this motivational approach by showing that mere categorizing participants to a certain in-group or out-group is sufficient to elicit the own-group recognition bias (Bernstein, Young, & Hugenberg, 2007). In their study, recognition performance was better for targets categorized as in-group than out-group, despite the fact that participants' perceptual expertise was equal for both groups.

We now covered the basic problems of eyewitness identifications, question remains why these problems occur. As we described above, retrieval of human memory is fallible. These principles apply to the retrieval of facial information as well. However, memory of faces is regarded as a “special” type of visual information, as faces are perceived differently than other stimuli (Farah, Wilson, Drain, & Tanaka, 1998).

Memory for faces. Faces are remembered and stored in memory as a complete picture; they are processed and stored holistically (Tanaka & Farah, 1993). The general conception of holistic processing is that faces are processed not as sets of different features but as a system that includes the properties between the features, such as distance, sizes and other types of information (Wells, 1993; Wells & Hryciw, 1984). So, specific details of features of a face are not essential for encoding. The memory of the face does contain information about the ears, eyes, mouth and the relationship between these features, but the overall impression is most important (Farah et al., 1998; Schmidt & Tredoux, 2006). Tanaka and Farah (1993) demonstrated this holistic processing of faces. They showed that recognition of individual facial features was worse when they were shown independently than when the features were shown as part of a whole face. As a result of holistic processing, recall of details is rather poor; yet current face composite production systems require the eyewitness to recall individual facial features (Wells & Hasel, 2007). This mismatch between holistic processing and detailed retrieval is the main problem in current police procedures and an important cause for mistakes in eyewitness identification cases. For composite production, it would be better for the witness to give a description of the person the same way he or she remembers it (holistically). However, we do not have the verbal capacity to describe persons as complete pictures.

Reverse correlation image classification technique

Recently, a new concept of face recall systems has emerged. This new technique uses the reverse correlation image classification technique as described by Mangini and Biederman (2004). This technique creates stimuli by adding or subtracting noise to a base face. It may offer potential for future composite production. Participants have to decide to what extent the stimuli represent the target face. By merging the faces that are selected to be most representative of the target face, a representation of the perpetrator can be created. Mangini and Biederman (2004) used this technique to produce two different faces out of one single base face. They projected noise over the base face, this produced stimuli that participants had to classify. Throughout all trials, the base face did not change. The only thing that changed was the noise that was projected over the base face. The researchers had participants classify high noise faces as female/male as happy/unhappy or as Tom Cruise/John Travolta. The reverse correlation task offers potential because of the holistic way of invoking recognition. A variation of the reverse correlation task is currently being used by Dotsch and Todorov. In their research, participants selected one out of two stimuli that looked the most trustworthy, untrustworthy, dominant or submissive. Classification images were calculated by averaging all selected images. Subsequent trait judgment of independent participants showed that the calculated classification images displayed the intended traits well. The authors suggest that reverse correlation techniques work well in extracting psychologically meaningful images (Dotsch & Todorov, 2012). This research tries to elaborate on whether the reverse correlation imaging classification technique can be used to create composites of individual in-group and out-group members. It moves away from the feature selection method and uses recognition of whole faces to create a likeness of a person.

Hypotheses

Extensive research has to be done in the field of eyewitness misidentifications. Many studies have suggested there exists a cross-race effect in eyewitness identification. This research aims to explore this effect further whereby the most standard eyewitness identification problems (verbal overshadowing, feature based composite production, commitment effects) are controlled. This experiment consists of two steps to examine whether the classification images created by the Reverse Correlation Image Classification Technique can be used as composites.

We expect that if the technique holds potential, Caucasian classification images (CI) will most resemble the Caucasian perpetrator and Moroccan classification images will most resemble the Moroccan perpetrator.

Hypothesis 1a: *The resemblance ratings of the CI of the target perpetrator are higher than the resemblance ratings of the CI of the nontarget perpetrator.*

By target we mean the intended perpetrator. So when the participant of the first part of the study constructed the Caucasian perpetrator, the Caucasian male is the target perpetrator and the Moroccan male is the nontarget perpetrator. When the participant of the first part of the study constructed the Moroccan perpetrator, it is vice versa.

To further test the first hypothesis we want to know whether the differences in resemblance ratings are based on neutral answers. If the resemblance ratings are not based on neutral answers the ratings should significantly differ from 4, which is the scale midpoint.

Hypothesis 1b: *The mean resemblance ratings of the classification images are significantly different from 4, which is the scale midpoint.*

Research suggests that combining input of multiple independent witnesses might hold better potential for identification evidence. A morph will reduce incorrect details because those are uncorrelated between participants (Wells & Hasel, 2007). To test this, the

individually produced composites will be aggregated into one. We expect the mean resemblance ratings of this aggregated composite to be higher than the mean resemblance ratings of the individual composites.

Hypothesis 2: Mean resemblance ratings for the aggregated composites are higher than mean resemblance ratings for the individual composites.

To further explore the usability of the reverse correlation image classification task for composite production, we expect that the majority of participants will select the target perpetrator in a target present lineup above chance level.

Hypothesis 3: Based on the aggregated composite, the chance that the correct perpetrator is selected in a target present photo lineup lies above chance level.

Based on available research, we expect that people are better able to distinguish between in-group members. They are less likely to recognize the constructed Caucasian classification image as the perpetrator. Conversely they are less capable to see distinguishing features of out-group members. They are more likely to perceive out-group members as ‘all the same’. Therefore they are more likely to recognize the constructed Moroccan classification image as the perpetrator. We expect that mean resemblance ratings of out-group (Moroccan) target classification images are higher than mean resemblance ratings of in-group (Caucasian) target classification images.

Hypothesis 4: Mean resemblance ratings of Moroccan target classification images are higher than mean resemblance ratings of Caucasian target classification images.

Method

This study consisted of two steps. In the first part of the study, participants produced composites of a target perpetrator (either Caucasian or Moroccan) from memory. In the second part of this study, participants rated the produced composites for resemblance against

their targets and against their nontargets. Methods of each study part will be discussed separately.

1.1: Generating composites using reverse correlation

Participants and design

The design of this study is a 1 factor (group membership) between participants design. Participants were 21 students (13 women and 8 men) of the University of Twente. All participants received 1 course credit for their participation. Age varied from 18 to 25 ($M=20.62$, $SD=1.63$). 16 participants were psychology students, 4 participants studied communication sciences and one participant graduated recently. 11 participants were Dutch, 10 participants were German. We randomly assigned all participants to one of two conditions: an in-group target perpetrator condition (Caucasian) or an out-group target perpetrator (Moroccan) condition.

Procedure

After signing in, the researcher directed the participants to a small room with a computer/laptop. Before the participants could begin, the researcher asked for their informed consent. The researcher instructed the participants the task was challenging and they had to try to stay focused because results would be meaningful for practice. To motivate participants, instructions also said the participant had the opportunity to receive their own constructed result. The experiment itself provided further instructions. Participants read there had been a robbery in a local night store; they had the opportunity to take a good look at the perpetrator (initial exposure). Participants could scroll back and forth between different viewing points of the perpetrator. First, the participants did a practice block of 30 trials to get

acquainted with the task. Participants selected one of two faces that most closely resembled the perpetrator shown in the beginning of the experiment (initial exposure, Figure 1).

Completing the 800 trials took the participants approximately 55 minutes.

Materials

We launched a pilot study to select the stimuli. Thirty-six participants rated eight faces from the Radboud Face Database (Langner et al., 2010), four Caucasian faces and four Moroccan faces, all faces were male. As a result of the pilot study, we selected two male faces as the perpetrator, one Caucasian male and one Moroccan male. We selected the most ‘average’ faces as the perpetrator; the more ‘extreme’ faces were not suitable for selection and subsequent morphing. We used both a three-quarter view from both sides and a full-face view of the perpetrator as initial exposure to the participants (Figure 1). We used the morph of the Caucasian and the Moroccan perpetrator as the base face (Figure 2).

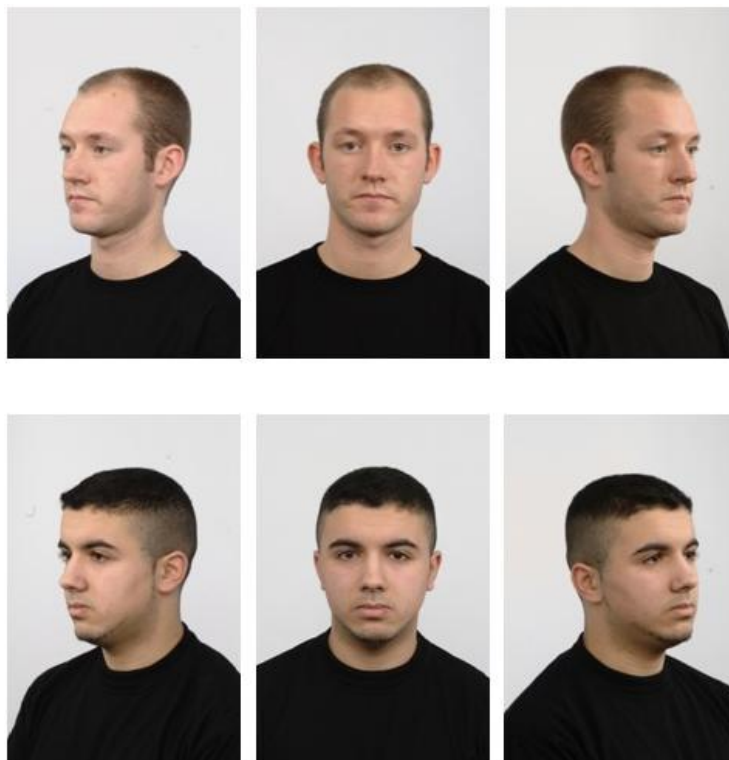


FIGURE 1: INITIAL EXPOSURE TO PARTICIPANT: CAUCASIAN (ABOVE) AND MOROCCAN (BELOW)



FIGURE 2: BASE FACE, MORPH OF BOTH PERPETRATORS

We programmed the reverse correlation task in Inquisit. Within each condition participants viewed a picture of the target perpetrator (either Caucasian or Moroccan). Each participant completed 800 trials, (30 practice trials followed by 770 trials). Each trial consisted of two stimuli (A and B) and each stimulus consisted of a base face with noise. We generated noise by randomly calculating one set of parameters for each stimulus. The resultant noise pattern was used as the first stimulus (A), while the inverted noise pattern was used as the second stimulus (B) (Figure 3). Participants decided to what extent the stimulus resembled the target perpetrator, shown in the beginning of the experiment. Participants had four options to choose from (1: Clearly A, 2: Probably A, 3: Probably B, 4: Clearly B). All stimuli were shown in grayscale.



FIGURE 3: EXAMPLE OF ONE TRIAL

Results

We analyzed the responses of the 21 participants with Python. We used the script from Dotsch and Todorov (2012) to calculate the constructed classification images for the second part of the study. The analysis yielded three types of classification images per respondent; (1) based on both response options (probably and clearly), (2) based on the clearly response option and (3) based on the probably response option (Figure 4). It must be mentioned that some respondents did not use all response options; therefore the amount of classification images per perpetrator differs. Aggregating these individual classification images yielded two multiple witness classification images based on both options (clearly and probably) (Figure 5).



FIGURE 4: THREE TYPES OF INDIVIDUAL CLASSIFICATION IMAGES; (LEFT) BASED ON BOTH, (CENTER) BASED ON CLEARLY, (RIGHT) BASED ON PROBABLY. ABOVE: CAUCASIAN, BELOW: MOROCCAN



FIGURE 5: AGGREGATED MULTIPLE WITNESS CLASSIFICATION IMAGES; (LEFT) FROM BOTH OPTIONS, (RIGHT) FROM THE CLEARLY OPTION. ABOVE: CAUCASIAN, BELOW: MOROCCAN

1.2: Rating the generated composites

Participants and design

The design of the study was a 2 factor within subject online study design (group membership x target/nontarget). A total of 104 Dutch and German men ($n=42$) and women ($n=62$) participated. 35 participants received 0.5 course credits for their participation. Age varied from 18 to 70 ($M= 32.39$, $SD= 15.18$). The experiment exposed each participant to both the out-group perpetrator as the in-group perpetrator in a randomized order. It took approximately 15 minutes to complete the survey.

In total, 50 persons were approached by email. Thirty-five persons signed up via sona-systems. Participants started 129 surveys and completed 104 surveys. We omitted the 25 uncompleted surveys from analysis.

Procedure

Before the participants could start the survey, they had to give their informed consent. After this, instructions told them there had been a robbery in the local night store. Eyewitnesses helped to make a sketch of one of the two perpetrators. Based on that sketch, participants made an identification by selecting one of six perpetrators in the lineup. Instructions told the participants that noise was added to the sketch to make it more difficult for them. After making an identification, participants had to take good look at the perpetrator. As in the first study, they had the opportunity to scroll back and forth between different viewing points of the perpetrator (Figure 1). Participants had 20 seconds before the 'next' button appeared. After this, participants judged resemblance between the perpetrator and classification images. The experiment was programmed in a way that all participants were

randomly assigned to judge 20 of 80 classification images per perpetrator. In addition, all participants judged the same four aggregated classification images (Figure 5).

Materials

Twenty-one participants of study 1 provided 38 classification images. Additionally, we made two aggregated classification images from each perpetrator, derived from the responses ‘clearly’ and the combined responses ‘clearly and probably’. So, per perpetrator, we programmed 42 produced classification images as trial stimuli (84 in total, 2 conditions). For programming the survey, we used the online research tool Qualtrics. Each participant rated 48 classification images (24 per condition). With each classification image, participants answered the same question “to what extent does this picture resemble the perpetrator? Participants could respond on a scale from 1 (not at all) to 7 (entirely). At the beginning of each condition (perpetrator) participants viewed the aggregated composite of both response options (‘clearly’ and ‘probably’). Based on this composite, they chose one of six perpetrators from a lineup (Figure 6 and 7).



FIGURE 6: CAUCASIAN LINEUP



FIGURE 7: MOROCCAN LINEUP

Results

Group membership

To establish whether all participants were part of the intended in-group, participants declared their nationality. All participants indicated they had either Dutch nationality ($n=87$) or German nationality ($n=17$).

Hypothesis 1a: The resemblance ratings of the CI of the target perpetrator are higher than the resemblance ratings of the CI of the nontarget perpetrator.

To compare the resemblance ratings of the individual Caucasian and Moroccan CI's, we conducted a repeated measures ANOVA with mean Caucasian resemblance scores and mean Moroccan resemblance scores as within-subject factors and condition as a between subjects factor. There was a significant interaction effect of condition on target perpetrator, Wilks' Lambda = .42, $F(1, 36) = 48.95$, $p < .001$. A simple main effect analysis revealed that, in the Caucasian target condition resemblance ratings for Caucasian classification images were higher ($M = 3.79$, $SD = .64$), than resemblance ratings for Moroccan classification images ($M = 2.44$, $SD = .49$). In the Moroccan target condition resemblance ratings for Moroccan classification images were higher ($M = 3.76$, $SD = .71$), than resemblance ratings for Caucasian classification images ($M = 2.59$, $SD = .56$). Figure 8 shows the distribution of the resemblance scores of the individual classification images. The results support the hypothesis that resemblance ratings of the classification image of the target perpetrator are higher than the resemblance ratings of the classification images of the nontarget perpetrator.

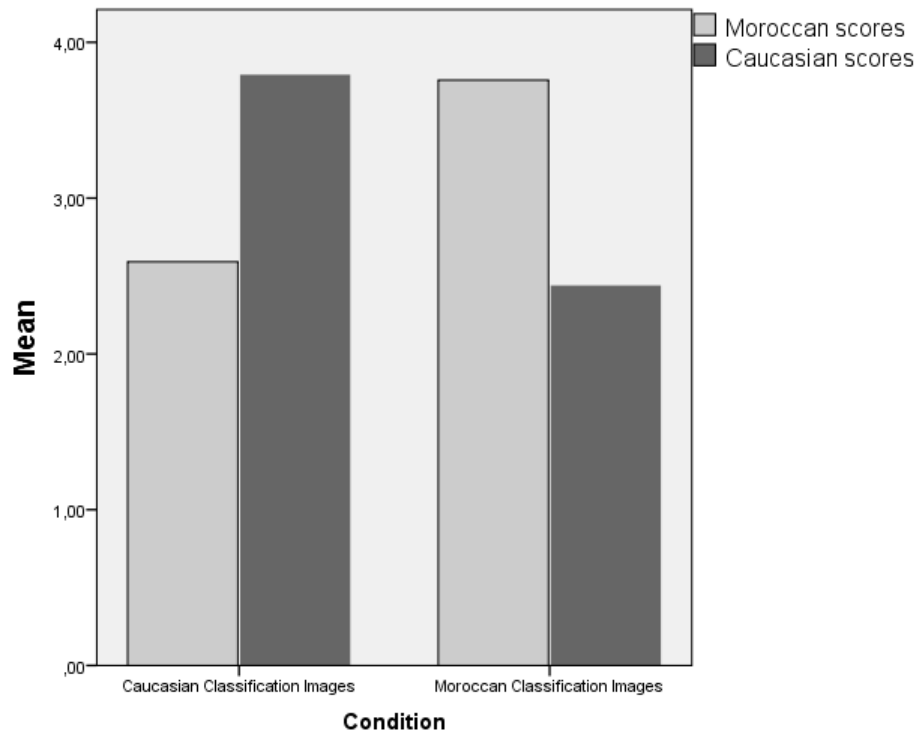


FIGURE 8: DISTRIBUTION OF THE RESEMBLANCE SCORES OF INDIVIDUAL CLASSIFICATION IMAGES

Hypothesis 1b: The mean resemblance ratings of the individual CI of the target perpetrator is significantly higher than 4, which is the scale midpoint.

To further test hypothesis 1, we conducted a one sample t-test to establish whether the mean resemblance ratings of the individual Classification Images are significantly different from 4, the scale midpoint. In the Caucasian condition, the sample mean of the Caucasian resemblance scores ($M = 3.79$, $SD = .64$) was not significantly higher than 4, $t(21) = -1.515$, $p = .145$. The sample mean of the Moroccan resemblance scores ($M = 2.59$, $SD = .56$) was significantly lower than 4, $t(21) = -11.785$, $p < .001$. In the Moroccan condition, the sample mean of the Moroccan resemblance scores ($M = 3.76$, $SD = .71$) was not significantly higher than 4, $t(15) = -1.37$, $p = .19$. The sample mean of the Caucasian resemblance scores ($M = 2.44$, $SD = .49$) was significantly lower than 4, $t(15) = -12.819$, $p < .001$. Based on these analyses we found that in the nontarget condition, answers are not based on neutral answers,

whereas for the target condition we could not reject the assumption that scores are based on neutral answers.

Hypothesis 2: Resemblance ratings for the aggregated composites will be higher than the mean resemblance ratings for the individual composites.

To test hypothesis 2, we conducted a one sample t-test to test whether the aggregated classification images were rated higher on resemblance compared to the mean individual classification images (Caucasian, $M = 3.79$; Moroccan $M = 3.76$). The sample mean of the Caucasian aggregated classification image derived from both response options ('clearly' and 'probably') ($M = 5.02$, $SD = 1.50$), was significantly higher than 3.79, $t(103) = 8.35$, $p < .001$. The sample mean of the Caucasian aggregated classification image derived from only one response option ('clearly') ($M = 3.31$, $SD = 1.56$), was significantly lower than 3.79, $t(103) = -3.156$, $p = .002$. The sample mean of the Moroccan aggregated classification image derived from both response options ('clearly' and 'probably') ($M = 4.92$, $SD = 1.56$), was significantly higher than 3.76, $t(103) = 7.59$, $p < .001$. The sample mean of the aggregated Moroccan classification image derived from one response option ('clearly') ($M = 4.74$, $SD = 1.49$), was significantly higher than 3.76 as well, $t(103) = 6.69$, $p < .001$. Based on this analyses we found support for the second hypothesis in the Moroccan condition, the aggregated composites were rated as a better resemblance than the best individual classification image ($M = 4.70$, $SD = 1.68$). However one type of aggregated classification image (based on 'clearly') in the Caucasian condition seems to score lower on resemblance than mean individual Caucasian classification images.

Hypothesis 3: Based on the aggregated composite, the chance that the correct perpetrator is selected in a target present photo lineup lies above chance level.

To test hypothesis 3, we performed a chi-square test of goodness-of-fit to determine whether the six persons in the photo-lineup were equally chosen by the participants. With a sample size of 104 participants and 6 persons to choose from in the lineup, we would expect that, on chance level, 17.3 participants selected the correct perpetrator. Analysis shows that selection of the six persons in the Caucasian lineup were not equally distributed, $\chi^2(5, n= 104) = 207.598, p < .001$. 68.3 % of 104 participants selected the correct perpetrator. Frequencies per photo option are displayed in Figure 9. Analysis of the Moroccan lineup shows that selection of the six persons in the photo-lineup were not equally distributed, $\chi^2(5, n= 104) = 369.44, p < .001$. 86.5 % of 104 participants selected the correct perpetrator. Frequencies per photo option are displayed in Figure 10.

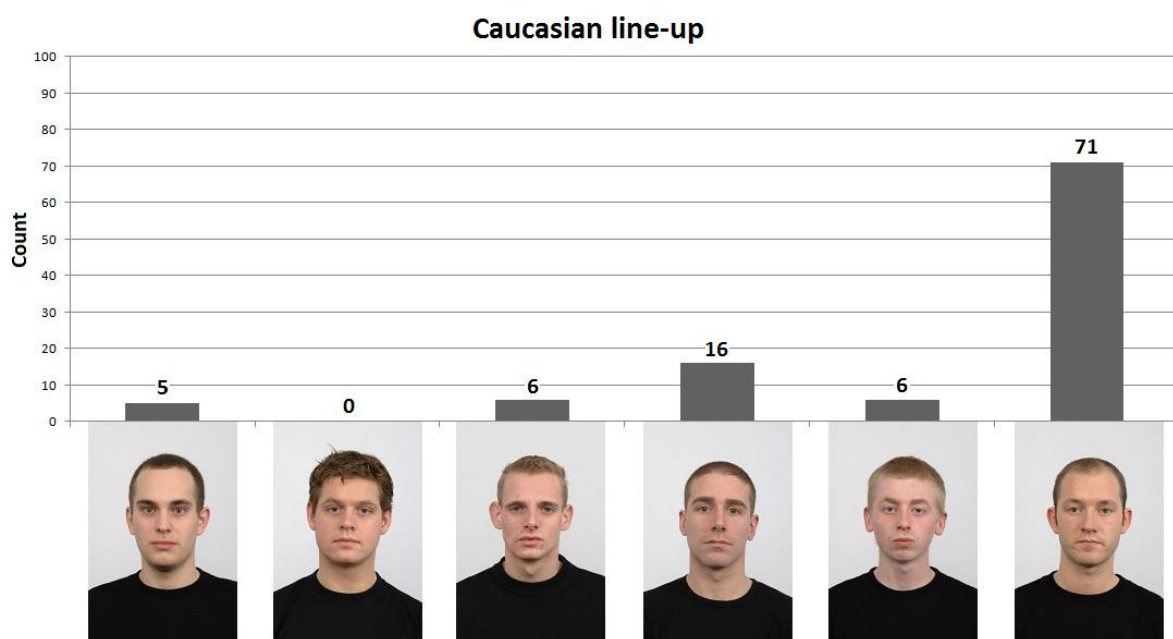


FIGURE 9: RESULTS CAUCASIAN LINEUP

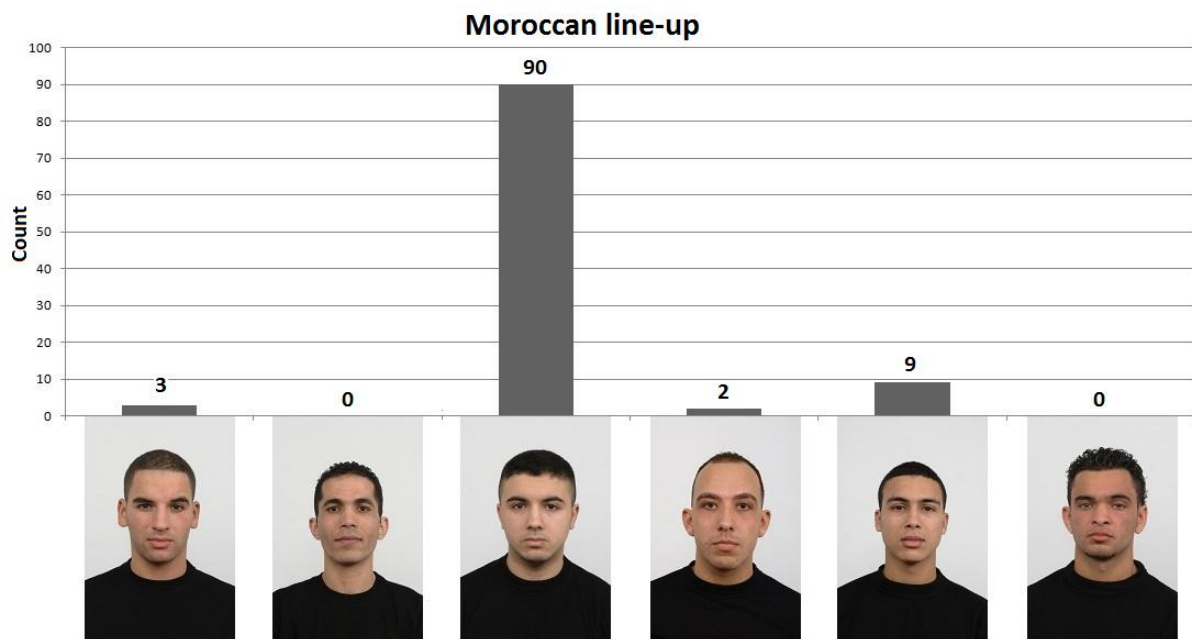


FIGURE 10: RESULTS MOROCCAN LINEUP

H4: *Mean resemblance ratings of Moroccan target classification images are higher than mean resemblance ratings of Caucasian target classification images.*

To test hypothesis 4, we conducted a one sample t-test to test whether the mean resemblance ratings of the Moroccan target condition were higher than the mean resemblance ratings of the Caucasian target condition ($M = 3.79$, $SD = .64$). The sample mean of the resemblance ratings of the Moroccan target condition ($M = 3.76$, $SD = .71$) was not significantly higher than mean resemblance ratings of the Caucasian target condition, $t(15) = -.184$, $p = .857$.

Discussion

Eyewitness identifications play an important role in many police investigations and courtroom decisions. Innocent project research showed eyewitness misidentification is the single greatest cause of wrongful convictions. In almost 75% of the DNA exoneration cases, wrongful eyewitness identification played a role (Innocence Project, n.d.-b). We conducted current research to gain insight in the usability of the reverse correlation image classification

technique for perpetrator composite purposes. In general, results suggest that the reverse correlation image classification technique can be used to create meaningful composites.

Analysis of the resemblance scores and the subsequent lineup were in line with our expectations. The participants of the first part of the study successfully used this technique to make a distinction between the target and the nontarget perpetrator.

Target versus nontarget

The first hypothesis described the expectation that the created composites looked more like their target than their nontarget. The results support this expectation, mean resemblance ratings of the target classification images are higher than mean resemblance ratings of the nontarget classification images. However, based on further analyses of these hypotheses we failed to completely reject the possibility that target resemblance scores are based on neutral answers. Participants seem to be more certain when the classification image does not represent the perpetrator (nontarget) and less certain when the classification image represents the perpetrator (target). This is in line with previous research, in a study by Hosch, Leippe, Marchioni, and Cooper (1984) nonchoosers were more certain of their decision than positive identifiers. The reasons for this are not clear, perhaps it has something to do with the consequences for the chosen perpetrator. When participants accuse someone, there will be consequences for the accused, so they are more careful in their judgment. In contrast, when participants reject someone, there are no consequences.

Input multiple witnesses

Research has shown that aggregating individual composites might hold better potential in identification evidence. Using statements of multiple eyewitnesses holds some

practical advantages. Often crime investigators are confronted with differing descriptions of the same culprit. A combination of eyewitness memories could retrieve a better impression of the perpetrator as it filters out all inaccurate details since those are uncorrelated between eyewitnesses (Hasel, & Wells, 2007). In a study by Bruce et al. (2002) the aggregated composite, produced by morphing four individual composites was rated as a better likeness than the individual composites on average and as good as the best individual composite. In a lineup task, combined composites produced most correct choices and fewest false positives. Based on our second and third hypothesis we found similar results. We found strong support for the improved effects of multiple witnesses in the Moroccan condition, the aggregated composites score even higher than the best individual composite. We found partial support for the second hypothesis in the Caucasian condition, the aggregated composite of both options ('clearly' and 'probably') scored higher on resemblance than the best individual composite, however the aggregated composite based on 'clearly' scored lower. Critical examination of the data shows that the aggregated Caucasian composite based on 'clearly' is based on only 7.6 % of the total responses in the Caucasian condition. Whereas in the Moroccan condition the aggregated composite based on 'clearly' is based on 14.1% of the total responses in the Moroccan condition. To further test the aggregated composites, we hypothesized that based on the aggregated composite, the correct perpetrator is chosen above chance level by the majority of the participants. We found strong support for our expectations. The majority of participants selected the correct perpetrator from the lineup.

Cross race effect

"They all look alike to me" is a common statement from Caucasians when they are confronted with African American faces. These statements about out-group homogeneity articulate one possible explanation for differences in ability to recognize other-race faces

(Sporer, 2001). Another possible explanation is motivation (Bernstein et al., 2007). Current study investigated whether perpetrator group membership influenced performance on the reverse correlation image classification task. Based on analyses, we found no evidence for a cross-race effect. Our expectations were that the Moroccan perpetrator is recognized more often than the Caucasian perpetrator. This was not the case. We found no significant difference in resemblance scores between the Caucasian and the Moroccan target perpetrator. Our expectations were based on results of studies where targets were categorized as in-group or out-group explicitly (Bernstein et al., 2007). We did not use explicit categorization in this study. Perhaps because the participants were dealing with perpetrators, no categorization process took place. We did find an interesting effect however, that is not in line with the out-group homogeneity bias. Based on the out-group homogeneity bias, we would expect that participants of the first part of the study chose more often for the ‘clearly’ option in the Moroccan condition. However, some participants displayed the reverse effect. Critical examination of the data shows that three participants did not use the ‘clearly’ option for the Moroccan perpetrator. It must be mentioned, as post experiment reactions indicated, participants expected that we measured prejudice, therefore participants could have been more careful with their assessment of the Moroccan perpetrator.

Limitations and recommendations

Even though we confirmed the majority of our hypotheses, some practical and procedural limitations must be mentioned. First of all, the base face we used consisted of a morph of the Caucasian and the Moroccan perpetrator. This study was designed to start with the basics of the usability of the reverse correlation image classification technique for perpetrator composite production, so this was sufficient for this study. However, because the base face only consisted of both perpetrators, we merely showed it is possible to use this task to make a

distinction between these two perpetrators. The majority of participants selected the correct perpetrator in a lineup. The lineup was constructed in a way that no cross-racial identifications could be made. For following research in this field, it is interesting to measure whether the correct perpetrator would still have been chosen if both perpetrators were in the lineup. In addition, the author must mention that both for the pilot study, as for collecting subsequent judgment by independent participants, a convenience sample was used. Often this is an indication of a selection bias. By using a convenience sample, we probably reached more highly educated people. Furthermore, more women than men participated. Participants indicated the task was very challenging and also too boring and monotonous. In analyzing the results, we used all data from all participants even though we acknowledge the possibility that many answers were given out of frustration and boredom. In real life, we expect the victims to be more motivated to produce a good composite. If this task is investigated further, we suggest to shorten the amount of classification images or look critically at the data to omit suspicious answer series. 800 trials seem to be too much for student participants. In addition, critical examination of response data showed that participants barely used the 'clearly' response option. We recommend to use a forced choice design with only two response options. Furthermore we recommend to use the keyboard as an input device for the responses instead of a mouse. This will minimize the cognitive load for the participants.

Conclusion

This first step showed that a distinction between two persons can be made. All participants successfully created the target perpetrator as indicated by resemblance ratings of independent participants. The results of the present study indicate that the reverse correlation image classification task holds potential for facial composite production. Provided that more

research is conducted to improve this task. In laboratory conditions, implementers must be aware of the cognitive highly demanding nature of the task, as it can affect the willingness to participate and concentrate on the task. It may therefore be especially important to find extrinsic ways to motivate participants. We assume that real victims are more intrinsically motivated than student participants to create a good composite of the perpetrator. We are not there yet, more research has to be done to justify implementing this task in real police investigations. The researchers strongly recommend to continue in this line of research as the consequences of misidentifications continue to put innocent persons behind bars.

Acknowledgements

I would like to thank Dr. Ron Dotsch for his contributions in this research. He calculated the classification images and their subsequent aggregated composites. Furthermore I would like to thank my supervisors Dr. Elze Ufkes and Dr. Ir. Peter de Vries, for their input and critical remarks.

References

- Ackerman, J.M., Shapiro, J.R., Neuberg, S.L., Kenrick, D.T., Becker, D.V., Griskevicius, V., Maner, J.K., & Schaller, M. (2006). They all look the same to me (unless they're angry) from out-group homogeneity to out-group heterogeneity. *Psychological Science*, 17, 836-840. doi: 10.1111/j.1467-9280.2006.01790.x
- Bernstein, M.J., Young, S.G., & Hugenberg, K. (2007). The cross-category effect mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, 18, 706-712. doi: 10.1111/j.1467-9280.2007.01964.x
- Brigham, J.C., & Wolfskeil, M.P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identifications. *Law and Human Behavior*, 7, 337-349. doi: 10.1007/BF01044736
- Bruce, V., Ness, H., Hancock, P.J.B., Newman, C., & Rarity, J. (2002). Four heads are better than one: combining face composites yields improvements in face likeness. *Journal of Applied Psychology*, 87, 894. doi: 10.1037/0021-9010.87.5.894
- Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *the Quarterly Journal of Experimental Psychology*, 48, 879-894. doi: 10.1080/14640749508401421
- Clifford, B.R., & Scott, J. (1978). Individual and situational factors in eyewitness testimony. *Journal of Applied Psychology*, 63, 352. doi: 10.1037/0021-9010.63.3.352
- Cutler, B.L., Penrod, S.D., & Dexter, H.R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior*, 14, 185-191. doi: 10.1007/BF01062972
- Cutler, B.L., Penrod, S.D., & Stuve, T.E. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior*, 12, 41-55. doi: 10.1007/BF01064273

- Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, 3, 562-571. doi: 10.1177/1948550611430272
- Dysart, J.E., Lindsay, R.C.L., Hammond, R., & Dupuis, P. (2001). Mug shot exposure prior to lineup identification: Interference, transference, and commitment effects. *Journal of Applied Psychology*, 86, 1280-1284. doi: 10.1037//0021-9010.86.6.1280
- Egeth, H.E. (1993). What do we not know about eyewitness identification? *American Psychologist*, 48, 577-580. doi: 10.1037/0003-066X.48.5.577
- Farah, M.J., Wilson, K.D., Drain, M., & Tanaka, J.N. (1998). What is “special” about face perception? *Psychological review New York*, 105, 482-498. doi: 10.1037/0033-295X.105.3.482
- Hasel, L. E., & Wells, G. L. (2007). Catching the bad guy: Morphing composite faces helps. *Law and Human Behavior*, 31, 193–207. doi:10.1007/s10979-006-9007-2
- Hosch, H.M., Leippe, M.R., Marchioni, P.M., & Cooper, D. (1984). Victimization, self-monitoring, and eyewitness identification. *Journal of Applied Psychology*, 69, 280-288. doi:10.1037/0021-9010.69.2.280
- Innocence Project. (n.d.-a). Know the Cases, Derrick Williams. Retrieved May, 8, 2013, from http://www.innocenceproject.org/Content/Derrick_Williams.php
- Innocence Project. (n.d.-b). Understand the causes, eyewitness misidentification. Retrieved May, 8, 2013 from <http://www.innocenceproject.org/understand/Eyewitness-Misidentification.php>
- Innocence Project. (n.d.-c). Reevaluating Lineups: Why witnesses make mistakes an how to reduce the chance of a misidentification. Retieved June, 30, 2013, from http://www.innocenceproject.org/Content/Reevaluating_Lineups_Why_Witnesses_Make_Mistakes_and_How_to_Reduce_the_Chance_of_a_Misidentification.php

- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, 24, 1377-1388. doi: 10.1080/02699930903485076
- Loftus, E.F. (1979). The malleability of human memory: Information introduced after we view an incident can transform memory. *American Scientist*, 67, 312-320.
- Mangini, M.C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*, 28, 209-226. doi: 10.1016/j.cogsci.2003.11.004
- Ross, D.R., Ceci, S.J., Dunning, D., & Toglia, M.P. (1994). Unconscious transference and mistake identity: When a witness misidentifies a familiar with innocent person. *Journal of Applied Psychology*, 79, 918-930. doi: 10.1037/0021-9010.79.6.918
- Schmidt, H.C., & Tredoux, C.G. (2006). Utilisation and usefulness of face composites in the South African police service- an evaluation study. *South African Journal of Criminal Justice*, 19, 303-314.
- Schooler, J.W., & Engstler-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology*, 22, 36-71.
- Sporer, S. (1992). Postdicting eyewitness accuracy: confidence, decision-times and person descriptions of choosers and non-choosers. *Journal of Social Psychology*, 22, 157-180.
- Sporer, S. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy and Law*. 7, 36-97. doi:10.1037/1076-8971.7.1.36
- Tanaka, J.W., & Farah M.J. (1993). Parts and wholes in face recognition. *the Quarterly Journal of Experimental Psychology*, 46, 225-245. doi: 10.1080/14640749308401045

- Tanaka, J.W., Kiefer, M., & Bukach, C.M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition*, 93, B1-B9. doi: 10.1016/j.cognition.2003.09.011
- Valentine, T. (2012). Identification evidence. *Forensic Psychology, crime, justice, law, interventions*. Chichester. John Wiley & Sons Ltd.
- van Koppen, P.J., & Wagenaar, W.A. (2010). Herkennen van gezichten. *Reizen met mijn rechter*. Deventer: Kluwer.
- Wells, G.L. (1993). What do we know about eyewitness identification? *American Psychologist*, 48, 553-571. doi: 10.1037/0003-066X.48.5.553
- Wells, G.L., & Hasel, L.E. (2007). Facial composite production by eyewitnesses. *Current Directions in Psychological Science*, 16, 6-10. doi: 10.1111/j.1467-8721.2007.00465.x
- Wells, G.L., & Hryciw B. (1984). Memory for faces: Encoding and retrieval operations. *Memory & Cognition*, 12, 338-344. doi: 10.3758/BF03198293
- Wells, G.L., Olson, E.A., & Charman, S.D. (2002). The confidence of eyewitnesses in their identifications from lineups. *Current Directions in Psychological Science*, 11, 151-154. doi: 10.1111/1467-8721.00189
- Wright, D.B, Memon, A., Skagerberg, E.M., & Gabbert, F. (2009). When eyewitnesses talk. *Current Directions in Psychological Science*, 18, 174-178. doi: 10.1111/j.1467-8721.2009.01631.x