Bachelor Thesis

Facial composite production:

Development of a new technique to identify perpetrators in a more reliable way

Franciszka Bayer, s1369040

Faculty of Behavioral and Management Sciences, University of Twente

Bachelor thesis

1st supervisor: Dr. Elze G. Ufkes

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

Specialization: Psychology of Conflict, Risk & Safety

University of Twente

Date: 21 June 2015

Abstract

Eyewitness misidentifications play an important role in wrongful convictions nationwide. Research suggests that this might be due to the mismatch between how faces are represented in memory (holistically) and how current facial composite systems attempt to retrieve this memory (by individual features). This study examined whether the reverse-correlation image classification technique proves to be a suitable tool for face composite production. Furthermore we are suggesting that factors such as prejudice or being the member of an in- or out-group have an impairing effect on the usability of the composite images. In the first part of the study we assessed the participants' level of prejudice by means of an IAT and let them construct a composite sketch with the use of the reverse-correlation image classification technique. Resemblance judgments of participants of the second study show that the composite sketches for offenders of an in-group resemble the actual offender significantly more than those made for offenders of an out-group. Furthermore, whether an eyewitness is implicitly prejudiced does not have any effect on the usability of the composite sketches. Recommendations for future research are suggested.

Development of a new technique to identify perpetrators in a more reliable way

Jennifer Thompson was a 22-year old college student in 1984 when someone broke into her apartment and raped her. During this incident she tried to study the face of her perpetrator as detailed as possible, so that she would be able to reconstruct his face afterwards. After working up a composite sketch of her offender with the police, she still identified the wrong man as her rapist when she was presented a photographic line-up. She testified against him twice, even after seeing the actual perpetrator Bobby Poole, who had admitted to being the true rapist to one of his fellow inmates. Ronald Cotton, the man who was mistakenly sentenced to prison, served 10,5 years of his sentence, until DNA testing conclusively proved that Poole was indeed the rapist.

This example is only one of many incidents that show that the testimonies of eyewitnesses may be prone to bias. Several studies have already proven that the construction of composite sketches and eyewitness identification are often an unreliable way to detect an offender: Eyewitness misidentification alone is the single greatest cause of wrongful convictions nationwide, playing a role in 75% of convictions through DNA testing (Osborne & Davies, 2013). It is necessary to note however that the construction of composite sketches and the subsequent line-up situation are two different processes in eyewitness identification procedures. Whereas the construction of composite sketches is implemented for tracing possible suspects, a line-up is useful as evidence when a possible suspect has already been found (Rennison & Dodge, 2015). Nevertheless those processes are interrelated, as police officers often base their search for possible suspects they want to present in a line-up on the composite sketches previously made by eyewitnesses. Considering these facts it becomes evident that a new system is necessary which helps to reconstruct faces with high resemblance to its perpetrators, so that accurate eyewitness identification in a line-up will be supported.

Usefulness of current face composite systems

An important question that arises is why current face composite systems fail to produce a good replication of faces. Wells and Hasel (2007) pointed out that this might be due to the mismatch between how faces are represented in memory and how composite systems attempt to retrieve this memory. "Faces are generally processed, stored and retrieved at a holistic level rather than at the level of individual facial features" (Wells & Hasel, 2007, p.9). Faces are not processed as sets of different features, but rather as a system that includes the properties between the features, such as distance, sizes and other types of information (Wells & Hryciw, 1984). Cooper and Wojan (2000) were trying a different approach by stating that faces are represented in "a coordinate spatial relations system that includes distances between features, relative sizes of features, and so on that cannot be separated from the features themselves" (p.470). Le Grand, Mondloch, Maurer and Brent (2004) gathered evidence from three months old infants, showing that they are integrating their mother's facial features into a whole rather than perceiving them as individual features. It seems as if "early visual experience naturally sets up a neural substrate for holistic processing of faces" (Le Grand et al., 2004, p.764).

Still, most of the current face composite systems require individuals to recall exactly these individual facial features that are harder to remember. In the study of Wells and Hasel (2007) it became evident that human face processing is designed more for face recognition, which is facilitated by holistic programs, than it is for face recall, which requires individual feature representations. As this mismatch between holistic face processing and detailed retrieval of individual features is one of the most important causes for wrongful eyewitness identifications (Wells & Hasel, 2007), research lately focused on developing a new system approach to face reconstruction. Davies and Christie (1982) already recommended that a consideration should be given to "developing systems of facial recall that allow the witness to reconstruct a face on the basis of groups of features

or to select between alternative whole faces reflecting different feature combinations" (p. 108). Hancock (2000) and Gibson, Pallares-Bejarano, & Solomon (2003) for instance focused on whole-face methods, in which eyewitnesses had to choose from several holistic faces, rather than choosing individual features. Eyewitnesses had to select several faces that were most similar to their memory for the target face, so that eventually a replication of the face of the perpetrator could be composed.

In the current study we are going to assess whether a new approach to face recall emerged by Mangini and Biedermann (2004) is applicable for reconstructing faces. They are introducing the *reverse-correlation image classification technique* which embodies the principle of holistic face recognition by letting participants choose between several sets of different base faces with various levels of noise on them. Eventually a replication of the face can be constructed by combining all these faces into one classification image.

Factors of possible influence on (mis)identification

Still, an important aspect that needs to be assessed is whether the reverse correlation image classification technique is free from possible biases. As stated by van Koppen and Wagenaar (2010), the apperception of offenders can be biased in any one of these moments: 1) during perception, when an individual encodes the physical appearance of another person, 2) during the retention period, when an individual stores the physical appearance of another person and tries to keep it in memory, and 3) during retrieval, when an individual has to reproduce the information stored in memory. As this paper focuses solely on the errors that can occur during retrieval of information, we are mainly interested in biases that may have an effect on this period.

When trying to retrieve information, all available information that seems to be related to the case comes up in an individuals' mind: People tend to use expectations, prior knowledge, and assumptions about what was likely to have happened to fill in their gaps in memory (Valentine, 2002). The human memory does not work like a "tape recorder",

but it is an "active process that is vulnerable to suggestion and biases" (Bijvank, 2014; van Koppen & Wagenaar, 2010). These examples give an idea of the possible biases that may occur when trying to retrieve information from memory.

According to Bernstein, Young and Hugenberg (2007) for example, there is a difference in the recognition accuracy for individuals that belong to the same race as oneself and individuals that belong to another race. This is called the cross-race recognition deficit, known more commonly as the cross-race effect (CRE): This deficit implies that there is "a tendency for recognition accuracy to be better for same-race faces than for cross-race faces" (p. 706). There are two kinds of models that try to explain this phenomenon: perceptual expertise models and social category models.

According to perceptual expertise models, early racial segregation leads to differences in a person's expertise of processing information about same-race or cross-race faces (Bernstein et al., 2007). People are being more exposed to faces of members of their in-group, and this familiarity leads to better recognition performances for targets categorized as in-group members than for targets categorized as out-group members (Bernstein et al.. 2007). Social category models on the other hand emphasize the general tendency for perceivers to think categorically about targets categorized as out-group members (Bernstein et al., 2007). When confronted with faces of out-group members, this has the effect of leading individuals to search for category-specifying features, instead of individuating ones. The results of their study confirmed the hypothesis that face recognition is more accurate for members of the in-group than for members of the out-group.

These theories imply that the influence of stereotypes might only have an effect for members of out-groups. When people encounter faces of in-group members, they look for individuating features and stereotypes are not affecting any judgments. When

encountering faces of out-group members on the other hand, people start to rely on category-specifying features, which are influenced by stereotypes.

As we already mentioned the concept of stereotypes, we are also interested to know how exactly stereotypes are affecting our judgments. As Dotsch, Wigboldus and van Knippenberg (2011) pointed out, " [people] effortlessly and automatically categorize persons into groups to simplify and to make sense of the enormous amount of social information in the world" (p.1). Stereotypes consist of expectations and beliefs about the characteristics of members of groups that are different than your own, which "influence how people attend to, remember and interpret subsequent information" (Charman, Gregory, & Carlucci, 2009, p.2). According to Bodenhausen and Lichtenstein (1987) stereotypes are considered to be "subjective base-rate probabilities". For example, Marin (1984) found out that many Caucasian Americans judge Hispanics to be much more likely to be aggressive than the population at large or than members of other subgroups. These judgments are usually prone to bias and quite unrelated to true base rates. This example illustrates that eyewitnesses who are highly influenced by stereotypes might be considered as of dubious diagnostic value from an objective standpoint. Especially when being confronted with a task that is relatively complex (i.e. judgments of guilt), people start to rely on heuristic judgment strategies that are activated by stereotypes (Ugwuegbu, 1979).

With regard to stereotypes it is also important to take into consideration the problem of the encoding bias: According to Miller and Turnbull (1986), the encoding bias is one of the possible information processing mechanisms by which stereotypes enter into the social perception of others. As stated by this hypothesis, the activation of stereotypic concepts leads to selective attention toward stereotype-consistent information (Miller & Turnbull, 1986). As Coenders, Lubbers, Scheepers, & Verkuyten (2008) pointed out, for Dutch people the Moroccan population represents a highly stigmatized immigrant group, which is strongly associated with the trait criminal.

Dotsch, Wigboldus, Langner and van Knippenberg (2008) wanted to know whether prejudiced people have more negatively stereotyped mental representations of faces of people in the out-group than of faces of people in the in-group. They implemented their study with Dutch participants and used the stereotype of the Moroccan population. Their results confirmed that the more people are prejudiced, the more criminal-looking their prototype of Moroccan faces is. This current research extends the study of Dotsch et al. (2008) by testing whether the same holds true for individual representations of faces. Thus, when being confronted with an individual out-group face (in this case Moroccan), we expect people with stronger stereotypes to produce less accurate individual classification images than individuals with weaker stereotypes. Dotsch et al. (2008) were testing their assumptions with categorical representations of faces of members of out-groups, whereas we are going to make use of individual representations.

The purpose of this study is to assess to what extent the images constructed by the reverse-correlation image classification technique are useful for identifying the offender. Thus, the question is 'To what degree does the constructed image resemble the face of the actual offender?'. Stereotypes and membership of in-groups respectively out-groups may be factors of possible influence on the accuracy of face recognition. We are going to assess this question by testing these two assumptions:

1) The constructed images made by eyewitnesses in the in-group will resemble the actual offender more than constructed images made by eyewitnesses in the out-group.

2) *The more people are influenced by stereotypes, the less the constructed image will resemble the face of the offender.*

2a) This effect of stereotypes is only true for judging faces of out-group members.

We expect to find the same results as Dotsch and his colleagues (2008) and are testing these hypotheses in the following two experiments. Experiment 1 is concerned with

assessing a person's implicit stereotypes (by making use of the Implicit Association Test), and with letting respondents construct an image of an offender by employing the reverse-correlation image classification technique. Dotsch et al. (2008) asked participants to chose the more Moroccan-looking face from two stimulus faces presented side by side, whereas we let them focus on one particular face that we present as the offender. Experiment 2 is interested in estimating the degree of usability of the constructed image. Methods of each study part are going to be discussed separately.

Method

Study 1 - Generating composites using reverse correlation

Participants and design

Participants. In total, 22 students (9 male and 13 female) of the University of Twente participated in this study. Only the data of 21 of the respondents could be used for further analysis, because data storage for one participant did not work. Of the 21 participants left, 9 were male and 12 were female. Four of them were Dutch and the other 17 were German. Ages varied between 18 and 24 years with a mean score of 20,5 years (*SD* = 1.61). 18 of the students were studying Psychology and the other 3 were studying Communication sciences at the University of Twente. In return, they received one course credit via SONA Systems or some confectionery.

Design. After completing the IAT, participants were randomly assigned to one of the following two conditions: an in-group target condition - with the perpetrator being Caucasian – or an out-group target condition – with the perpetrator being Moroccan. Thus, the experiment employed a 1-factor between participants design, with the 2 conditions explained above. The dependent variable was the accuracy of the reconstructed face of the actual perpetrator.

Procedure

On arrival, participants were directed to a small and quiet room in the library with a laptop in it. Before continuing with the experiment, researchers asked the participant for his or her informed consent. Because the experiment was such a time consuming task, participants were instructed that it is important for the usability of the study that they stay focused, and that they are allowed to take a 5-minute break every 20 minutes to ensure this. The experiment was split into two different parts. First they had to complete an Implicit Association Test (IAT), so that afterwards we would be able to measure their prejudices and stereotypes towards Moroccans. Thereupon the reverse-correlation image classification task started, in which participants had to choose between 800 pairs of faces to reconstruct the offender they saw before.

Implicit prejudice. In order to measure prejudice, participants completed an Implicit Association Test (IAT). This test measured indirectly to what extent participants categorized Moroccan or Caucasian faces into the positive or the negative group. We used a shortened version of the IAT, which consisted of a practice, congruent and incongruent block. In the practice block, participants had to classify 10 positive images with one key and 10 negative images with another. After that, participants had to complete 40 trials each in the congruent and in the incongruent block. The order of the last two blocks was randomly distributed across participants. In the congruent block, participants classified Moroccan faces as a negative stimulus and Caucasian faces as a positive one. In the incongruent block, participants did the reverse. Within blocks, stimuli were presented in random order. When there has been a mistake in classification, error feedback was presented to the participant for 1,000 ms. Latencies above 3,000ms were set to 3,000ms. We analysed the data on log-transformed latencies, but untransformed mean latencies are reported. We constructed an IAT score by subtracting the average response latency in the congruent block from the average response latency in the incongruent block. If response

latencies were longer in the incongruent block than in the congruent block, this was assumed to indicate stronger negative than positive associations with Moroccan faces. This difference was interpreted as reflecting higher levels of implicit prejudice.

Reverse-correlation image classification task. Participants were given further instructions about the following task. The system automatically assigned the participants randomly to either the in-group target or the out-group target condition. Two faces were selected from the Radboud Faces Database (Langner, Dotsch, Bijlstra, Wigboldus, Hawk, & van Knippenberg, 2010). The RaFD is a new tool for research using face stimuli, "providing a parametric set of face images varied along important facial characteristics, namely expression, gaze direction, and head orientation" (Langner et al., 2010, p. 1385). From this database we selected one typical Caucasian male face to represent the perpetrator in the in-group target condition, and a typical Moroccan male face for the out-group target condition (Figure 1). At the beginning of the task, participants were exposed with both a three-quarter view from both sides and a full-face view of the perpetrator (Figure 2). They were asked to study the face carefully by using the opportunity to switch between the different viewing angles. Subsequently, they were repeatedly presented with two stimuli side by side. Each pair of stimuli consisted of a base face with two random-noise patterns superimposed over it. The noise patterns were randomly generated at every one of the 800 trials. Participants were instructed to decide to what extent the stimuli resembled the target perpetrator shown in the beginning. They had 4 options to choose from: 1) Clearly A, 2) Probably A, 3) Probably C and 4) Clearly B. Completing the 800 trials took the participants approximately 60 minutes.

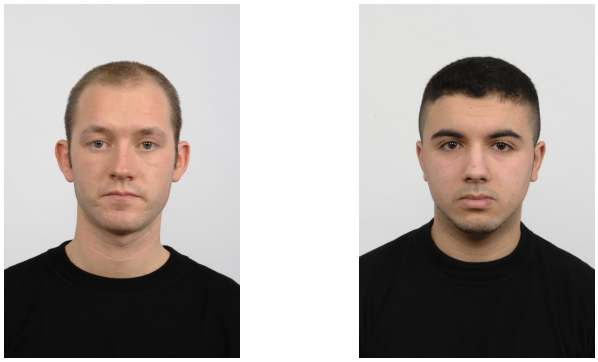

Figure 1. Faces to represent the offender: Caucasian (left) and Moroccan (right).

Figure 2. Initial exposure to participants: Caucasian (above) and Moroccan (below).

Materials

Stimuli. As already mentioned above, participants repeatedly had to choose which of the base face with a random-noise pattern superimposed over it resembled the target more. In order to construct a base face, 12 faces have been selected from the Radboud Faces Database (Langner et al., 2010): Six faces that appeared to be typical Caucasian and six faces that were typical Moroccan. As already explained above, the RaFD offers a

parametric set of face images, which provided us with prototypical faces for both the Caucasian and Moroccan condition (Langner et al., 2010). We then morphed the features of those six faces into one base face for each condition (by using the program PsychoMorph; Figure 3). The noise pattern was generated by randomly calculating one set of parameters for each stimulus. Within a single trial, stimulus A consisted of the base face with a random-noise pattern, while the base face of stimulus B consisted of the inverse pattern. The noise patterns distorted the base face to such an extent, that the pair of faces appeared to be different in every trial.

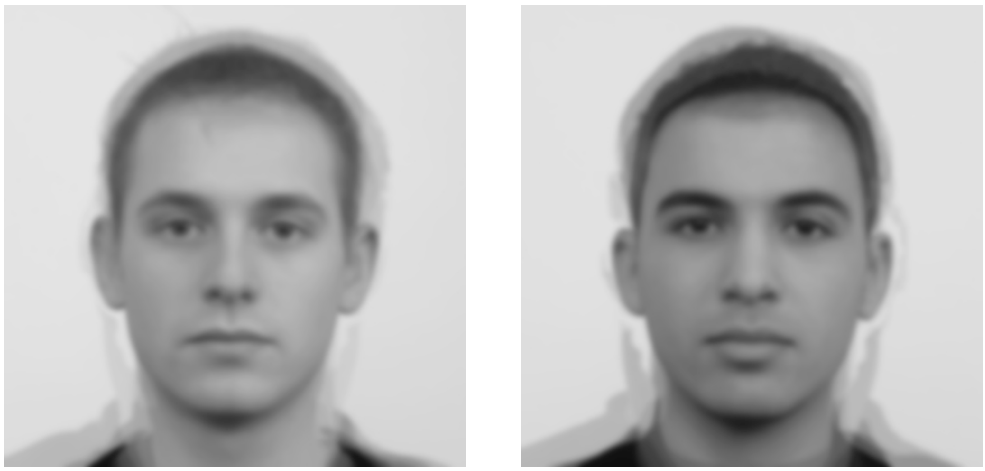

Figure 3. Base face for each condition: Caucasian (left) and Moroccan (right).

Results

Implicit prejudice. Any incorrect trials and the first practice block were omitted from further analysis. Conducting a one-sample t-test revealed that on average participants had stronger negative associations (M= 1135.81, SD= 370.68) with the Moroccan category (out-group) than positive ones, and stronger positive associations with the Caucasian category (in-group) than negative ones (M= 965.82, SD= 265.58), $t(20)$= 2.67, p< 0.001.

By using the median split, one could identify participants that scored below or above the median. An IAT score in the Moroccan condition (out-group) above the median indicated a relatively stronger negative than positive association with Moroccan faces and

was thus an expression of implicit prejudice. Vice versa, an IAT score in the Caucasian condition (in-group) below the median indicated a relatively stronger positive than negative association with Caucasian faces and confirmed that people have better associations with people of their in-group than with people of their out-group.

Reverse-correlation image classification task. We used the script from Dotsch and Todorov (2012) to construct the classification images of the responses of the 21 participants for the second part of the study. The analysis yielded 11 classification images for the Caucasian condition, and 10 classification images for the Moroccan condition (Figure 4). Furthermore, one final classification image was constructed for each condition by combining the features of all classification images into one average classification image (Figure 5).

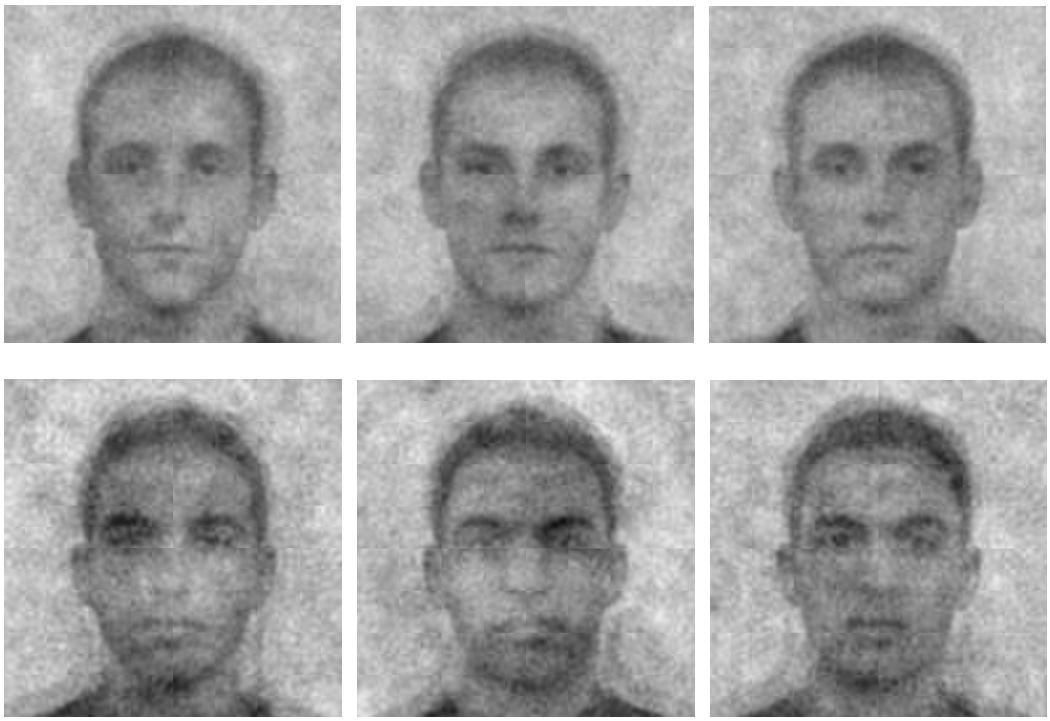

Figure 4. Three types of classification images with low (left), moderate (center) and high levels of implicit prejudice. Caucasian (above) and Moroccan (below).


Figure 5. Average classification images: Caucasian (left) and Moroccan (right).

Study 2 - Rating the generated composites

Participants and design

Participants. Participants were approached by social network websites such as Facebook and by email. This left us with a convenience sample, because only those with a Facebook- or Email account could be reached. 198 of the approached participants started the survey, but only 116 (58.59%) completed it. Consequently, these 82 uncompleted surveys were omitted from further analyses. Of the 116 participants who completed the survey, 50 were men and 65 were women. One participant did not indicate his/her gender. Their age varied from 17 to 57 (M= 25.94, SD= 8,92).

Design. The second study employed a 2 (group membership: Moroccan vs. Caucasian face) x 2 (level of prejudice: high vs. low) mixed design, with the two different conditions of group membership as a between-subjects variable and the level of prejudice as a within-subject variable. Participants were randomly assigned to one of the following two conditions: either the in-group target condition (n= 57), in which participants were exposed to a Caucasian perpetrator, or the out-group target condition (n= 59), in which participants were exposed to a Moroccan perpetrator. It took the participants approximately 20 minutes to complete the survey.

Procedure

As this survey was an online study, we were not able to have any influence on the location where participants completed the study. At the beginning of the survey participants were given a short introduction and were asked to give their informed consent. The survey consisted of three parts.

In the first part, participants were told that there had been a robbery at the local night store and that the police, with the support of several eyewitnesses, had been able to compose a sketch of one of the offenders (Figure 4). The sketch that we used was the average classification image constructed for each condition in study 1. Participants were asked to take a close look at the sketch in order to memorize it as good as possible. Furthermore they were told that noise was added to the sketch in order to make it more difficult. Subsequently they saw 6 faces in a line up from which they had to choose the possible offender. At the beginning of the second part participants were presented with the photo of one of the possible offenders in the line-up. We told participants that the experiment randomly selected a face. However, in reality we always presented the photo of the true offender. Participants had the opportunity to take a close look by scrolling back and forth between different viewing angles of the photo (Figure 2). It lasted at least 20 seconds until the "continue" button to get to the next page appeared. Subsequently participants were shown the composite sketches of the offender that were constructed in the first study (Figure 4). For each of these sketches, we asked participants to indicate the resemblance between the sketch and the offender they just saw. In the last part participants had to indicate the general impression they had for each of the sketches. They were provided with 10 adjectives for each sketch in order to rate how aggressive the sketch appeared to the participants.

At the end of the survey participants had to provide their demographics and indicate to what extent they were in touch with Caucasian or Moroccan persons.

Materials

The twenty-one eyewitnesses of Study 1 provided us with 21 classification images: 10 classification images for the Moroccan offender and 11 classification images for the Caucasian offender. Furthermore we constructed one final classification image for each condition by combining the classification images into one average classification image.

Dependent on the condition participants were assigned to, they rated 11 (12) classification images, thus the individual classification images as well as the average classification image for each condition. Unfortunately, due to an unforeseen problem with the survey portal only eight of the individual classification images were uploaded in the survey in the Moroccan condition. In order to judge the resemblance between the actual perpetrator and the classification images, participants were confronted with the question "To what extent does the sketch resemble the offender?". They could respond on a Likert Scale from 1 (not at all) to 7 (very much). To get a clear picture of the general impression of the offender, we asked participants to indicate to what extent they thought the sketch embodied several characteristics (five positively and five negatively valued adjectives). Here participants could respond on a 5-point Likert Scale from 1 (not at all) to 5 (very much).

Results

The hypotheses that we wanted to test with this study were the following:

1) The constructed images made by eyewitnesses in the in-group will resemble the actual offender more than constructed images made by eyewitnesses in the out-group.

2) *The more people are influenced by stereotypes, the less the constructed image will resemble the face of the offender.*

2a) This effect of stereotypes is only true for judging faces of out-group members.

Before finding out whether the data support our assumptions, we wanted to know whether the reverse-correlation image classification technique in general provides a good tool for face reconstruction of eyewitnesses. On average, the mean resemblance score for all faces revealed that this technique does not acquire classification images with high resemblance to the offender ($M = 3.54$, $SD = .90$). Conducting a one sample t-test revealed that the mean resemblance score is not significantly different from 3.5, the scale midpoint: $t(115) = .48$, $p = .32$). As the scores do not differ from the scale midpoint we cannot reject the assumption that scores are based on neutral answers. This indicates that the individual constructed classification images are not sufficiently similar to the actual offender. The range for the resemblance score of all individual composite sketches goes from 2.25 to 4.89. Furthermore, an analysis of the line-up revealed that only 7 participants (11.9%) of those in the Moroccan condition ($n = 59$) in study 2 identified the actual offender after seeing the average classification image. The number of times a suspect should be identified as the offender based on chance is 10. This demonstrates that the actual offender was identified even less often than someone who would have been identified by chance. Two other potential offenders were picked more often (54.2% and 20.6%; Figure 6). In the Caucasian condition however ($n = 57$), the majority of participants ($n = 40$, 70.2%) picked the accurate face as the offender (Figure 6). Considering all these results led to the assumption that eyewitnesses in general failed to construct classification images with satisfactory resemblance scores to the actual offender seen at the beginning of the study.

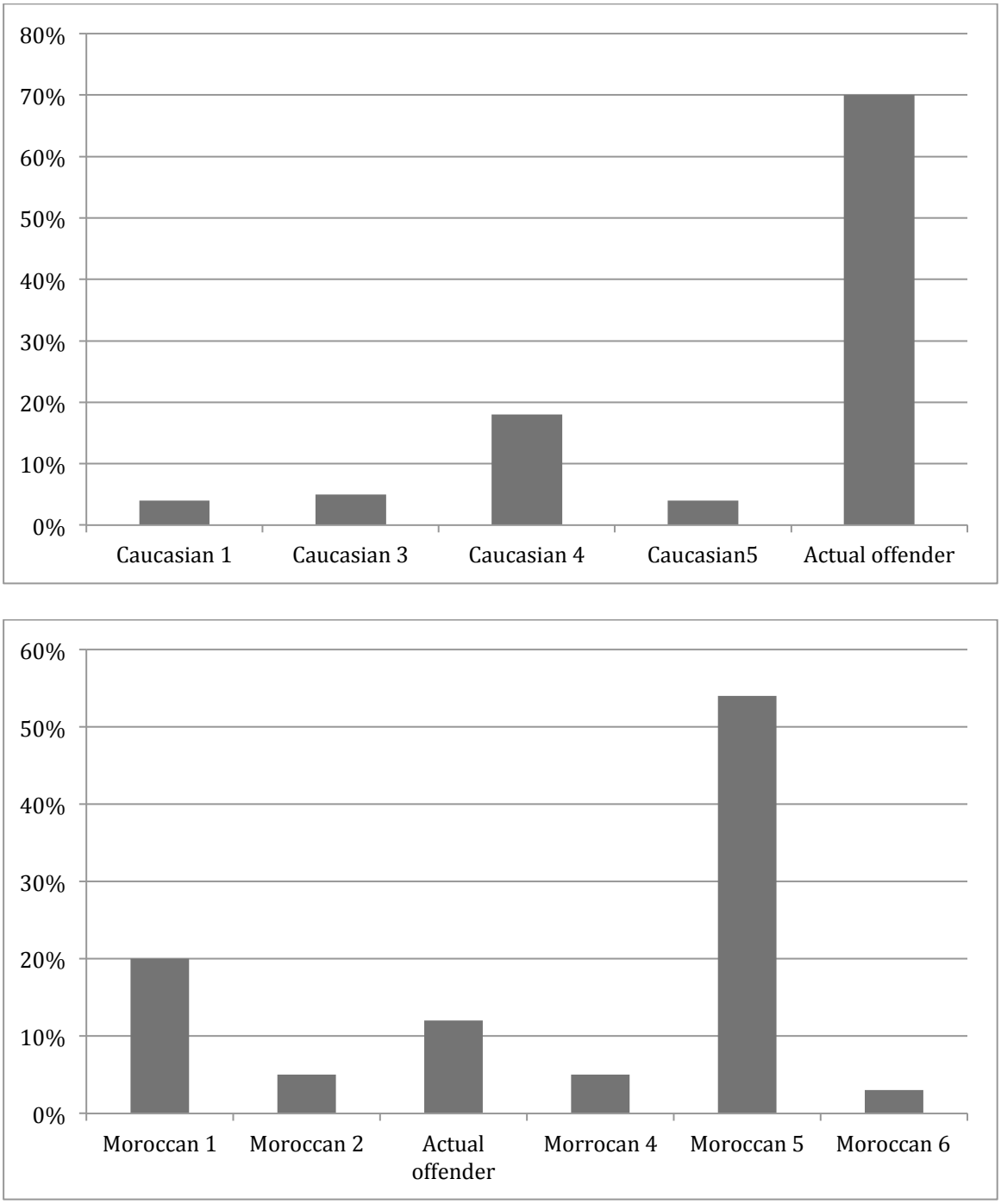

Figure 6. Analysis of the line-up: Percentage of face identification in the Caucasian (above) and in the Moroccan condition (below).

With the aim to test whether our assumptions stated above prove to be true, we examined the data for the existence of three different effects: First of all, we examined whether there is a main effect for Group Membership of eyewitnesses on the resemblance scores, thus whether there is any difference between those assigned to the in- or out-group.

Furthermore, we tested whether there is a main effect for the Level of Prejudice that eyewitnesses in the first study had on the resemblance of their composite sketches. Finally, we checked whether there is an interaction effect between Level of Prejudice and Group Membership of eyewitnesses.

Implementing a repeated measures ANOVA revealed a significant main effect for Group Membership of eyewitnesses on the resemblance scores: $F(1,114) = 29.12$, $p < .001$ This means that the resemblance scores of classification images made by eyewitnesses in the in-group (Caucasian) condition ($M = 3.95$, $SD = .76$) are significantly higher than those made by eyewitnesses in the out-group (Moroccan) condition ($M = 3.14$, $SD = .86$). In the Caucasian condition, the range for the resemblance scores of the individual composite sketches goes from 2.98 to 4.89, whereas in the Moroccan condition the range only goes from 2.25 to 3.68. Also the average resemblance score for the aggregated classification images for each condition is significantly higher in the in-group ($M = 4.56$, $SD = 1.60$) than in the out-group condition ($M = 3.22$, $SD = 1.43$): $t(115) = 1.65$, $p = .04$. These results confirmed our first assumption that constructed images made by eyewitnesses in the in-group resemble the actual offender more than the constructed images made by eyewitnesses in the out-group.

Furthermore, the repeated measures ANOVA revealed that there is no significant main effect for the Level of Prejudice of eyewitnesses on the resemblance scores: $F(1,114) = .15$, $p = .70$. Whether eyewitnesses in the first study were highly affected by implicit stereotypes ($M = 3.55$, $SD = 1.00$) did not seem to have any influence on the usability of the constructed classification images compared to eyewitnesses with low levels of implicit prejudice ($M = 3.53$, $SD = 1.00$). This means that neither one of the two groups (with low and high levels of prejudice) constructed classification images with high levels of accuracy. These results were not consistent with our second assumption that the

more people are influenced by stereotypes, the less the constructed image will resemble the face of the offender.

Finally, no significant Level of Prejudice x Group Membership interaction was found: $F(1,114) = .30$, $p = .56$. The third assumption that stereotypes influence the usability of classification images made by eyewitnesses is only true with regard to faces of out-group members could not be confirmed.

Discussion

Recent research emphasized that eyewitness identification is often unreliable. In fact, it is the single greatest cause of wrongful convictions nationwide, playing a role in 75% of exonerations through DNA testing (Osborne & Davies, 2013). One of the possible explanations may be the mismatch between how faces are represented in memory (by means of holistic representations) and how composite systems attempt to retrieve this memory (by means of individual feature representations; Wells & Hasel, 2007).

This led researchers (see Wells & Hasel, 2007) to conclude that face composite systems should focus more on face recognition, which is better accomplished by holistic programs, than on face recall, which relies on individual feature representations. The reverse-correlation image classification technique, a new approach to face composite systems by Mangini and Biedermann (2004), does focus on face recognition and holistic face representation. Consequently, this study aimed to assess whether this technique is a reliable approach to eyewitness identification. Furthermore, we did not only want to assess the usability of this new method, but also whether factors such as being a member of the in- or out-group and prejudices could be of possible influence on the accuracy of eyewitness identifications. We did this by implementing two experiments: In the first experiment, we measured the implicit prejudice of participants and let them construct a composite sketch of an offender by means of the reverse-correlation image classification

technique. In the second experiment participants were asked to rate the resemblance of the composite individual sketches.

A positive finding was that performances of participants in composing a composite sketch of a member of an in-group were generally satisfactory. The resemblance scores for composite sketches of the in-group were above the average, and more than half of the participants in the second experiment (n = 40, 70,2%) identified the right photo as the actual offender. These results imply that at least in the context of a crime within an in-group, the reverse-correlation image classification technique seems to provide a suitable tool for facial composite production.

As stated by the cross-race effect (CRE) there is a general tendency for recognition accuracy to be better for same-race faces than for cross-race faces (Bernstein et al., 2007). This led us to the assumption that composite sketches constructed by eyewitnesses in the in-group resemble the actual offender more than composite sketches constructed by eyewitnesses in the out-group. Although the study of Bijvank (2014) could not confirm this assumption we found indeed that the group to which one has the feeling to belong to did have an effect on the usability of the constructed classification images. Eyewitnesses were better at constructing faces of offenders that were members of their in-group than of offenders that were considered to be members of their out-group. A study by Eysenck and Keane (2013) found similar results: By examining more than 200 real court cases they showed that eyewitnesses correctly identified 65% of possible suspects of their own race, but only 45% of offenders for other-race members. This means that when being confronted with a crime in which the offender does not belong to an individuals' in-group, this difference in group-membership has a serious impairing effect on the resemblance of the classification image that one is about to compose.

There are several interpretations that try to explain an effect like that. The term infrahumanization for example refers to the perception of the in-group as more defined

than the out-group by uniquely human features (Capozza, Boccato, Andrighetto, & Falvo, 2009). Furthermore a study of Harris and Fiske (2006) using functional magnetic resonance imaging suggests that out-groups may be processed as objects, and not as human beings. In a study by Capozza et al. (2009) they found that participants did protect the human integrity of their in-group by avoiding animal contamination. The same principle of protecting one's in-group might apply to this context: Because eyewitnesses want to protect members of their in-group, they are acting more carefully in constructing and rating composite sketches of offenders.

Due to the low range of resemblance scores for eyewitnesses in the out-group condition, another likely explanation for the bad resemblance scores in this condition is that we encountered a floor effect. According to Hessling, Schmidt and Traxel (2004), "a floor effect occurs when a measure possesses a distinct lower limit for potential responses and a large concentration of participants score at or near this limit" (p.393). In this manner, any possible variance in the results of those that already performed badly at constructing composite sketches is restricted.

Calculating the average resemblance score of the constructed classification images for all participants however puts the overall usability of the reverse-correlation image classification technique for face reconstruction of eyewitnesses into question. The resemblance scores for both groups lie in the middle range of all options, and the overall general resemblance thus appears to be unsatisfactory. The fact that less than the half of all participants in the Moroccan condition identified the right offender in a photographic line-up supports this assumption. In a recent study of Dotsch and Todorov (2012) on the other hand they found that the reverse-correlation image classification technique "provides an excellent tool to extract psychologically meaningful images that map onto social perception" (p.562).

There are two main distinctions in the approaches of Dotsch et al. and this study that are possible explanations for this difference in results. First of all, Dotsch et al. (2008) were working with categorical and not with individual faces. Furthermore they did not conceptualize the context of a crime like we did: Instead of confronting eyewitnesses in the first study with an individual image of an offender, they just asked participants to think of a typical Moroccan face and then let them complete the reverse-correlation image classification technique. The influence of this distinction in design is going to be assessed in the next section. Secondly, they used a different base face for the reverse-correlation image classification technique: We only used the gray scale of the aggregated features of the Moroccan and the Caucasian perpetrator as the base face, whereas Dotsch and Todorov used a gray scale average of all male faces in the Karolinska Face Database. It is possible that the base face used by Dotsch and Todorov made it easier for eyewitnesses to make a distinction between the two stimuli presented. Also the study of Bijvank (2014) suggests that the reverse-correlation image classification technique can be used to create meaningful composites: The majority of participants selected the correct perpetrator from the line-up, and analysis of the resemblance scores revealed satisfactory results.

Besides the race or group that someone has the feeling to belong to, there are other factors that may influence the accuracy of face composite systems: Prejudices, which consist of expectations and beliefs about the characteristics of members of groups that are different to your own, influence how people attend to, remember and interpret subsequent information (Charman et al., 2009). As Fiske (1998) already pointed out, prejudice biases cognition, affect, and behavior toward ethnic out-groups. Furthermore, Dotsch et al. (2008) found that the classification images of participants with high levels of prejudice were rated as more criminal and less trustworthy than the classification images of participants with moderate and low levels of prejudice. Thus, in the same way as our results confirm that being a member of an out-group influences the way people remember

things about others, we assumed that prejudices may have the similar effect of impairing the accuracy of face recognition for certain groups.

Contrary to our expectations there is no difference in the usability of classification images between those participants with high or low levels of prejudice. No matter to what extent an individual is implicitly influenced by stereotypes and judges over members of other groups accordingly, this does not have any effect on the accuracy and usability of the classification image. As already mentioned above, Dotsch et al. (2008) on the other hand found the opposite effect, which we were expecting to find in our study as well. We need to take into consideration though the main differences in the approach of Dotsch et al. and this current study that were previously stated. Although our results did not confirm the hypothesis that prejudices influence the resemblance of composite sketches, this does not necessarily mean that there is no influence at all.

As Dotsch et al. (2008) did not examine the reverse-correlation image classification technique in the context of a crime like we did, a possible explanation for the difference in results might be that being confronted with a perpetrator alone already influenced eyewitnesses to such an extent, that race of the perpetrator did not have that much of an impact anymore. The term confirmation bias describes how people selectively look for only expectation-consistent information and interpret incoming information accordingly. This in turn leads to the effect that "a pre-existing belief in the suspect's guilt influences the evaluators subjective similarity judgement" (Charman et al., 2009, p. 86). Thus, the effect of the confirmation bias may override any effect of an individuals' prejudice in this task.

Furthermore, Dotsch et al. (2008) were working with categorical and not with individual faces: They asked participants to think of a typical Moroccan face and then let them complete the reverse-correlation image classification technique, instead of confronting them with an individual photo like we did. Thus, in the study of Dotsch et al.

(2008) the direct effect of a group stereotype might have been evoked more strongly than in our study, which employed a rather indirect way of stereotype activation. "A key feature of group stereotypes is that they are rich cognitive structures whose various traits are linked in interconnected associative networks" (Berinsky & Mendelberg, 2005, p.846). As soon as one particular element of the stereotype is activated, a process of spreading activation across the stereotype's associative network is started by which all other connected elements of this stereotype will become available as well (Berinsky & Mendelberg, 2005). As a result of merely thinking about members of an out-group, the spreading activation generated in the study of Dotsch et al. (2008) might have made stereotypes and prejudices more easily available. It is unclear whether being presented with the photo of a member of an out-group is sufficient to produce the same effect. Due to the difference in the set-up of the experiments it is difficult to compare the results of Dotsch et al. (2008) with this current study.

One positive aspect that is important to underline here in general is the finding that the reverse-correlation image classification technique in our particular context does not seem to be influenced by the level of prejudices that an eyewitness has. The fact that there is no difference in the resemblance of composite sketches constructed by eyewitnesses with either high or low levels of implicit prejudice towards Moroccans indicates that the reverse-correlation image classification technique seems to be a suitable tool for eyewitness identification.

There is one last observation that we would like to mention: With regard to the reverse-correlation image classification technique, it feels like we have encountered two propositions that create a paradox situation. First, there is a general tendency for humans to process faces of out-group members by category-specifying features. A category is "a class or group of things or people, possessing some quality or qualities in common" (Dictionary.com Unabridged, n.d). It does not refer to a single characteristic, but to a

quality that is rather more all embracing. This leads to the suggestion that faces of members of out-groups are processed by rather holistic mechanisms than by individual features. Second, the reverse-correlation image classification technique is specifically designed for holistic face recognition. As a result, one might assume that this technique works especially well for faces of out-group members. The in-group/out-group model (IOM) by Sporer (2001) however, which explains the CRE by integrating social-categorization and perceptual-expertise models, invalidates the assumption of this paradox: Because of greater expertise with faces of members of in-groups, they are processed in a default, automatic manner, which results in holistic processing and superior recognition. In the case of out-group faces on the other hand, social categorization is a cue to disrupt this default and automatic processing, which in turn leads to disregarding the stimulus and poor recognition. This explanation may also account for the results in this current study.

Limitations of this study

Although part of our assumptions could be confirmed, we have to mention some practical and procedural limitations. In our first study, in which we conducted the IAT and the reverse-correlation image classification technique, the majority of participants were German (17 out of 21). Because we already expected to have lots of German participants, we used two versions of the program (one with the explanations in Dutch and the other in German), but we still used the same stimuli and faces in both versions. According to Coenders et al. (2008), for Dutch people the Moroccan population represents a highly stigmatized immigrant group that is strongly associated with the trait criminal. That is why we decided to select Moroccan stimuli for the IAT and a Moroccan perpetrator for the out-group condition. In the German population however there may be other immigrant groups that are considered to be more stigmatized and evoke stronger negative associations than the Moroccan population. It is possible that the German population does not have any

strongly negative associations with Moroccans, which would explain why there is no difference in the usability of classification images constructed by participants with high level of prejudice and those with low levels of prejudice. Due to the small number of Dutch participants in our study ($n = 4$) we did not consider any supplementary analysis as being representative enough to further investigate this supposition. Using another, more stigmatized population for the out-group in the German version might have produced a main effect for Level of Prejudice or an interaction effect for Level of Prejudice and Group Membership, and would thereby have supported our hypothesis. This is something that needs to be assessed in future research.

Furthermore, the majority of respondents in both studies were psychology students. It is generally questionable whether psychology students form a representative population for conducting such a study. Psychology students are by nature keen to question the purpose of something and are likely to act different than other students would do. Besides they might have encountered the Implicit Association Test before and could thereby have acted accordingly to protect their reputation. Considering this it is questionable whether the results of this study are representative for the whole population. Due to the time limit and the circumstances given in our situation, it was not possible to reach further to a population that is more representative than our respondents. It is therefore recommendable for future research to try to conduct this research in another setting with a wider range of respondents.

Another aspect that needs to be considered is that the majority of participants in the first study mentioned that the task was too long, mentally exhausting and boring. The performance of participants while completing the reverse-correlation image classification task might have suffered due to the low motivation and boredom that comes along with this task. This leads to the suggestion that the number of trials in the reverse-correlation image classification should be shortened. In real life however, eyewitnesses are expect to

be more willing and intrinsically motivated to produce a good composite sketch of the offender.

Conclusion

The purpose of this study was to assess to what extent the images constructed by the reverse-correlation image classification technique are useful for identifying an offender. Furthermore we wanted to know whether factors such as being the member of an in- or out-group and prejudices might have any effect on the usability of this technique. The results of this study let us draw mixed conclusions: Whereas the reverse correlation image classification technique holds potential for facial composite production of in-group members, it is still wide open for improvement in the case of members of out-groups. One positive finding though is that the technique does not seem to be influenced by the prejudices an eyewitness possesses. As participants in our study mentioned that this task was mentally exhausting and boring, we would recommend to shorten the number of classification images in the future. We assume however that eyewitnesses in the real world are highly motivated to find the perpetrator and are willing to act more conscientious when completing this task.

As eyewitness misidentification is such an important problem in wrongful convictions nationwide, we strongly recommend future research to continue in this field.

Literature

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*(8), 706-712.

Berinsky, A. J., & Mendelberg, T. (2005). The indirect effects of discredited stereotypes in judgments of Jewish leaders. *American Journal of Political Science*, *49*(4), 845-864.

Bijvank, M. (2014). Facial composite production: the usability of the reverse correlation image classification technique for perpetrator recognition.

Bodenhausen, G. V., & Lichtenstein, M. (1987). Social stereotypes and information-processing strategies: The impact of task complexity. *Journal of Personality and Social Psychology*, *52*(5), 871.

Capozza, D., Boccato, G., Andrighetto, L., & Falvo, R. (2009). Categorization of ambiguous human/ape faces: Protection of ingroup but not outgroup humanity. *Group processes & intergroup relations*, *12*(6), 777-787.

category. (n.d.). *Dictionary.com Unabridged.* Retrieved June 19, 2015, from Dictionary.com website: http://dictionary.reference.com/browse/category

Charman, S. D., Gregory, A. H., & Carlucci, M. (2009). Exploring the diagnostic utility of facial composites: Beliefs of guilt can bias perceived similarity between composite and suspect. *Journal of Experimental Psychology: Applied, 15*(1), 76.

Coenders, M., Lubbers, M., Scheepers, P., & Verkuyten, M. (2008). More than two decades of changing ethnic attitudes in the Netherlands. *Journal of Social Issues*, *64*(2), 269-285.

Cooper, E. E., & Wojan, T. J. (2000). Differences in the coding of spatial relations in face identification and basic-level object recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 470

Davies, G., & Christie, D. (1982). Face recall: An examination of some factors limiting composite production accuracy. *Journal of Applied Psychology*, *67*(1), 103.

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, *3*(5), 562-571.

Dotsch, R., Wigboldus, D. H., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of Personality and Social Psychology*, *100*(6), 999.

Dotsch, R., Wigboldus, D. H., Langner, O., & van Knippenberg, A. (2008). Ethnic out group faces are biased in the prejudiced mind. *Psychological Science*, *19*(10), 978 980.

Eysneck, M., & Kaene, M.T. (2010). *Cognitive Psychology: A Student's Handbook* (6th ed.), Florence, Kentucky: Psychology Press.

Fiske S.T. (1998). Stereotyping, prejudice and discrimination. In Gilbert D.T., Fiske S.T., Lindzey G. (Eds.), *The handbook of social psychology* (pp. 357–411). Boston: McGraw-Hill

Gibson, S. J., Solomon, C. J., & Pallares-Bejarano, A. P. (2003). Synthesis of Photographic Quality Facial Composites using Evolutionary Algorithms. In *BMVC* (pp. 1-10).

Hancock, P. J. (2000). Evolving faces from principal components. *Behavior Research Methods, Instruments, & Computers*, *32*(2), 327-333.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low neuroimaging responses to extreme out-groups. *Psychological science*, *17*(10), 847-853.

Hessling, R., Schmidt, T., & Traxel, N. (2004). Floor Effect. In Michael S. Lewis-Beck, A. Bryman, & Tim Futing Liao (Eds.), *The SAGE Encyclopedia of Social Science Research Methods.* (pp. 391-392). Thousand Oaks, CA: Sage Publications, Inc. doi: http://dx.doi.org/10.4135/9781412950589.n344

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377-1388.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychological Science*, *15*(11), 762-768.

Mangini, M. C., & Biedermann, 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

Marín, G. (1984). Stereotyping Hispanics: The differential impact of research method, label, and degree of contact. *International Journal of Intercultural Relations*, *8*, 17– 27.

Miller, D. T., & Turnbull, W. (1986). Expectancies and interpersonal processes. *Annual Review of Psychology*, *37*, 233– 256.

Rennison, C. M., & Dodge, M. (2015). *Introduction to Criminal Justice: Systems, Diversity, and Change*. SAGE Publications.

Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law*, *7*(1), 36.

Ugwuegbu, D. C. E. (1979). Racial and evidential factors in juror attribution of legal responsibility. *Journal of Experimental Social Psychology*, *15*, 133– 146.

Valentine, T. (Ed.). (2002). *Cognitive and computational aspects of face recognition: Explorations in face space*. Routledge.

Van Koppen, P. J., & Wagenaar, W. A. (2010). Herkennen van gezichten. *Reizen met mijn rechter.* Deventer: Kluwer.

Wells, G. L., & Hasel, L. E. (2007). Facial composite production by eyewitnesses. *Current Directions in Psychological Science, 16*(1), 6-10.

Wells, G. L., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory & Cognition*, *12*(4), 338-344.