PLACING FACES: RECOLLECTION AND FAMILIARITY IN THE OWN-RACE BIAS FOR FACE RECOGNITION

Ruth Horry

Thesis submitted for the degree of Doctor of Philosophy

University of Sussex

May 2010

Statement

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

……………………………………..

Ruth Horry

May 2010

Acknowledgements

I could not have written this thesis without the support and encouragement of my friend and mentor, Dan Wright. Thank you for being such a fantastic supervisor, for helping me to learn to become a better scientist, for long chats in the Bridge Café, and for your faith in me over the last four years.

Thank you to Colin Tredoux and the University of Cape Town for your hospitality throughout my research visit. Thanks also to the postgraduates in the Department of Psychology at UCT who made me feel welcome during my stay. It was a fantastic experience that I will always feel privileged to have had.

Thank you to the ESRC for funding this research, and to the many reviewers and editors who have provided comments on my work. My experience of the review process has been largely positive, and your suggestions have helped me to improve myself as a cognitive scientist.

To all of the wonderful people who have made my time at Sussex such a great experience: My office mates – Ryan, Louisa, Fran, Mark, Andy, Tina, and Rebecca – and my other friends – Liz, Carina, Ruth, Ali, Zöe, Elin, and Gini. I can't possibly name everybody here, as the postgrad community at Sussex is as big as it is friendly. All of you have helped me get through the last few years with my sanity intact, and we have had some great times! I know that I have made many friendships at Sussex which will stay with me for the rest of my life.

To my oldest and dearest friends – Mandy, Yvonne, Matt, Piers, Keith – just knowing that you were all there any time I needed to talk or to get away from my thesis has been a big help to me. I am fortunate to have such a close group of friends, and I want you all to know that you mean the world to me.

To my family – Mum, Dad, and Rhys – your love and support throughout my entire life has led me to be where I am today. Mum, when I graduated with my degree, you wrote in a card that every day I have made you proud. Well, every day, I have counted my blessings for having such a wonderful family. Thank you all so much for just being you, and being there.

Last but by no means least, thank you to my wonderful husband, Miles. You are my friend, my confidante, my companion, and my inspiration. You know better than anyone that the last three years have been a marathon. You made me laugh when I felt down, and you encouraged me when I lost my confidence. I hope that I never forget how lucky I am to have found you. Thank you.

UNIVERSITY OF SUSSEX

RUTH HORRY

Thesis submitted for the degree of Doctor of Philosophy

PLACING FACES: RECOLLECTION AND FAMILIARITY IN THE OWN-RACE BIAS FOR FACE RECOGNITION

SUMMARY

The research presented in this thesis examined the roles of recollection and familiarity in the own-race bias (ORB) in recognition memory for faces. In Paper 1, Jacoby's (1991) process-dissociation procedure was used to estimate the relative contributions of recollection and familiarity in recognizing own- and other-race faces. Recollection estimates were higher for own-race faces than for other-race faces, although this effect disappeared when deep or shallow encoding strategies were encouraged. In Paper 2, participants were shown to be less accurate at ignoring previously seen other-race distractors than own-race distractors. Papers 3 and 4 examined how accurately participants were able to remember contextual information about correctly recognized faces. In the encoding phase of an old/new recognition test, each target face was paired with one of several different backgrounds. At testing, old judgments were followed by context judgments, in which the participant attempted to identify with which background the face had been paired. The context judgments were consistently more accurate for correctly recognized own-race faces than for correctly recognized other-race faces. This effect was robust to experimental manipulations such as context reinstatement and divided attention. The overall conclusion from this thesis is that recollection is inferior for other-race faces compared to own-race faces. This recollection deficit means that it is more difficult to retrieve specific information about the circumstances in which other-race faces were encountered. The implications of this recollection deficit for real world behaviour are discussed, with particular reference to eyewitness memory.

Table of Contents

List of Tables

List of Figures

General Introduction

1. The Own-Race Bias in Face Recognition

Despite the remarkable capacity for the human memory to recognize limitless numbers of faces, not all types of faces are recognized equally well. People recognize members of their own race more accurately than members of other races. This was first empirically shown by Malpass and Kravitz (1969) with White and Black university students, although it was acknowledged long before. Feingold (1914) remarked that “to the uninitiated American, all Asiatics look alike, while to the Asiatic all white men look alike” (p.50). This effect has been termed the own-race bias (ORB), or cross-race effect.

The ORB has been shown with members of several different races. Most studies have tested White and Black participants (e.g. Ayuk, 1990; Chiroro & Valentine, 1995; Meissner, Brigham, & Butz, 2005), with significant effects found in both groups (Anthony, Copper, & Mullen, 1992; Bothwell, Brigham, & Malpass, 1982), although larger effect sizes are often found in White participants than in Black participants (Meissner & Brigham, 2001). Superior recognition of own-race faces has also been found in Hispanic participants (Evans, Marcon, & Meissner, 2009; Gross, 2009; MacLin, MacLin, & Malpass, 2001; Platz & Hosch, 1988), East Asian participants (Gross, 2009; Hancock & Rhodes, 2008; Michel, Caldara, & Rossion, 2006; Michel, Rossion, Han, Chung, & Caldara, 2006; Ng & Lindsay, 1994), Turkish participants (Sporer, Trinkl, & Guberova, 2007), and Native Canadian participants (Jackiw, Arbuthnott, Pfeifer, Marcon, & Meissner, 2008).

The ORB has been tested with different paradigms. Most commonly, old/new recognition tests have been used (e.g. Ferguson, Rhodes, Lee, & Sriram, 2001; Hills

& Lewis, 2006; Wright, Boyd, & Tredoux, 2003). Participants study a large set of faces, and must then discriminate old faces from new faces in a testing phase. This procedure is statistically powerful, as multiple data points are collected from each participant. Using a large number of faces also reduces stimulus sampling concerns, as results will be less influenced by any peculiarities of the targets (Sporer, 2001a). Old/new recognition tests also allow for tight experimental control over factors such as presentation duration and retention interval. However, this procedure is low in "mundane realism" (Wells & Olson, 2001), and one must therefore be cautious about applying the findings of these studies directly to memory in the real world.

Lineup studies, in which participants identify the target face from a small group of similar faces, provide a more externally valid method for testing face recognition. Most lineup studies of the ORB have used simultaneous lineups (Brigham, Maass, Snyder, & Spaulding, 1982; Evans et al., 2009; Jackiw et al., 2008; Platz & Hosch, 1988; Wright, Boyd, & Tredoux, 2001), although one study used a sequential lineup followed by a simultaneous lineup (Wright et al., 2001). Some of these studies are field studies, in which naïve participants are approached by one or more confederates, and are some time later asked to identify the confederate or confederates (Brigham et al., 1982; Platz & Hosch, 1988; Wright et al., 2001). Other studies have used videotaped staged crimes (Smith, Lindsay, Pryke, & Dysart, 2001; Smith, Stinson, & Prosser, 2004), or photographs (Evans et al., 2009; Jackiw et al., 2008) during encoding. The results of these lineup studies have supported those from old/new recognition tests, with most showing significant ORBs.

Despite their external validity, lineup studies do have some disadvantages. None of the field studies of the ORB have used target absent lineups, which are considered very important by eyewitness researchers (Sporer, 2001a). Only a small number of targets can be tested, which raises stimulus sampling concerns (Wells & Olson, 2001), as the particular targets used in a given study may not be representative of the group from which they are sampled, and the faces from one group may be more similar to each other than faces from the other group. Lineup studies are not statistically powerful, as only a small number of data points (sometimes just one) can be recorded for each participant. Large sample sizes are required, which can be difficult and expensive to recruit. However, despite these limitations, the general consistency of results across these studies suggests that the ORB does generalize to lineup identifications (Meissner & Brigham, 2001).

Several meta-analyses have been conducted on the ORB literature. Anthony et al. (1992) and Bothwell et al. (1982) showed that the ORB is reliable in both White and Black participants, with 79% of the eleven studies included in Bothwell et al.'s (1982) analysis showing significant crossover effects. The most extensive meta-analysis to date involved thirty nine studies with almost five thousand participants (Meissner & Brigham, 2001). Correct identifications, false identifications, and overall discrimination accuracy were shown to be reliably higher for own-race faces than other-race faces. The ORB was strongest in White participants, although it was also present in participants from other races.

2. Why Does the Own-Race Bias Occur?

Theoretical accounts of the ORB have attempted to tackle both the *why* and the *how* which underpin this robust effect. Early theories of *why* the ORB occurs included differences in the amount of physiognomic or perceived variability between groups, and the interference of participants' racial attitudes and stereotypes with their processing and retention of other-race faces. Both of these positions have received limited support, and have largely given way to the two main current theoretical stances: The contact/perceptual expertise account, and the in-group/out-group model. The evidence for each of these explanations for why the ORB occurs will be discussed in turn below, beginning with the earlier accounts of physiognomic variability and stereotyping.

2.1. Physiognomic differences in variability

An early hypothesis was that perhaps some races are simply less variable in their physiognomy than others. For example, whereas Whites show considerable variation in eye colour and hair colour, other groups such as Blacks and Asians tend to have uniformly dark hair and eyes (Chance, Goldstein, & McBride, 1975). However, this hypothesis has not received much empirical support. White participants can make accurate same/different judgments as quickly for Japanese faces as for White faces (Goldstein & Chance, 1976) and can search for a White or Japanese target in an array of same-race faces with similar speed and accuracy (Goldstein & Chance, 1978). In one set of studies, White participants rated Japanese faces equally as similar to each other as White faces (Goldstein & Chance, 1978), while in another study, Japanese faces were rated as more similar than White faces (Goldstein & Chance, 1979). Zebrowitz, Montepare, and Lee (1993) showed that White, Black, and Korean participants could make individuated impression ratings and feature ratings for White, Black, and Korean faces, and that inter-racial agreement on the ratings was high.

More recently, Walker and colleagues (Walker & Hewstone, 2006a, 2006b, 2008; Walker & Tanaka, 2003) have used morphing techniques to investigate whether other-race faces are more perceptually similar to each other than own-race faces. Two faces (referred to as “parent faces”) of different races are morphed together along a continuum, creating a range of faces with varying ratios of information from each parent face. Participants then see a parent face and a morph, and make a rapid same/different judgment. White, East Asian, and South Asian participants make more errors discriminating faces closer to the other-race parent face than the own-race parent face, suggesting that other-race faces may appear more similar to each other than own-race faces after all. However, the crossover nature of this effect in samples from different races suggests that it is not a matter of physiognomic similarity, but of psychological similarity.

2.2. Prejudice and stereotyping

Another early hypothesis was that the ORB is a product of biased attitudes (Chance et al., 1975), and that people with negative attitudes would have more difficulty recognizing other-race faces than people with positive attitudes. Several studies have searched for associations between self-reported attitudes and the ORB, with most showing no relationship between the two (e.g. Brigham & Barkowitz, 1978; Carroo, 1987; Ferguson et al., 2001; Slone, Brigham, & Meissner, 2000). In a meta-analytic review, Meissner and Brigham (2001) found no direct effect of attitudes on the ORB, but suggested that there may be an indirect effect mediated by contact. Brigham, Bennett, Meissner, and Mitchell (2007) suggested that social attitudes may influence both the quantity and quality of contact with members of other races, and that this could lead to poorer recognition of other-race faces.

Other studies have examined the influence of implicit or unconscious attitudes on the ORB, using procedures such as the implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998) and the bona fide pipeline (Fazio, Jackson, Dunton, & Williams, 1995). The results of these studies have been inconsistent, with some showing significant relationships between implicit attitudes and the ORB (Walker & Hewstone, 2008) and others finding no such relationship (Ferguson et al., 2001).

Lebrecht, Pierce, Tarr, and Tanaka (2009) showed that participants who were trained to distinguish between other-race faces showed smaller implicit biases than participants who were not trained. The authors suggested that difficulties in discriminating between members of other races could lead to stereotyping – if people *look* alike, they must therefore *be* alike. Brigham (2007) suggested that lacking confidence in one's ability to accurately recognize people of other races may cause anxiety about, and therefore avoidance of, interactions with those same people.

Overall, the evidence does not seem to support the hypothesis that ORB is a product of prejudiced attitudes or automatic stereotypes. However, the methods used to measure explicit and implicit attitudes may be subject to measurement error. Participants may not be truthful when completing explicit attitude measures, as negative attitudes towards members of other races are not socially desirable. Recent evidence also suggests that tasks designed to measure implicit prejudice may be less reliable than was previously believed. For example, exposing participants to positive exemplars of members of another race reduces bias as measured by the IAT (Dasgupta & Greenwald, 2001). The presence of a Black experimenter is also sufficient to reduce bias in the IAT (Lowery, Hardin, & Sinclair, 2001). Some researchers have argued that apparent racial bias in the IAT may actually be a product of differential familiarity rather than implicit prejudice (Brendl, Markman, & Meissner, 2001; Kinoshita & Peek-O'Leary, 2005).

Stereotyping and prejudice are very complex social phenomena. Garcia-Marques, Santos, and Mackie (2006) argue that stereotypes are not static abstractions, but dynamic knowledge structures which fluctuate over time even within an individual. Stereotypes and implicit attitudes may be successfully controlled under some conditions (Blair, 2002; Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002;

Moskowitz, Gollwitzer, Wasel, & Schaal, 1999), but may be exaggerated in situations of high anxiety (Lambert et al., 2003) or perceived threat (Schaller, Park, & Mueller, 2003). Stereotype activation may also be moderated by how typical of their race a person is perceived to be. Blair, Judd, Sadler, and Jenkins (2002) showed that Black faces with more "Afrocentric" features (e.g. darker skin) are more likely to be associated with negative stereotypes than Black faces with fewer Afrocentric features. These kinds of subtle variations in the activation and expression of stereotypes stand in stark contrast to older theories of prejudice, which viewed attitudes as relatively stable over time. This "modern racism" is much more difficult to define and measure, and may be one reason why studies searching for associations between prejudice and the ORB have not produced significant results (Brigham, 2007).

2.3. Contact and perceptual expertise

A longstanding and inherently plausible account of the ORB is that lifelong contact with members of one's own racial group leads to a specialized sort of expertise for processing and remembering those faces. This expertise is not generalized to the processing of other-race faces, which are therefore processed less efficiently, or are confused in memory. This idea has been around for some time; Feingold (1914) stated that "it is well known that, other things being equal, individuals of a given race are distinguishable from each other in proportion to our familiarity, to our contact with the race as a whole" (p. 50). Sporer (2001b) argued that unless one assumes that people are genetically hardwired to recognize own-race faces but not other-race faces, the ORB must be attributable to differential contact.

Often in the ORB literature, little or no distinction is made between *actual contact* and *opportunity for contact*. Researchers have used a variety of techniques to investigate the contact hypothesis, some of which explore actual contact (by self-report, and by perceptual training), and some of which explore opportunity for contact (e.g. recruiting participants from areas/countries with different demographics, and by examining the development of the ORB throughout childhood). Many cognitive researchers have failed to acknowledge the importance of the distinction between actual contact and opportunity for contact, making the assumption that individuals within multi-racial societies will inevitably interact with members of other groups. While such contact may occur on a superficial level, social psychological research has shown that even within desegregated multicultural environments, there is a strong bias towards interacting with in-group members to the exclusion of out-group members (Dixon, Tredoux, & Clack, 2005; Tredoux & Dixon, 2009).

Furthermore, many researchers have ignored social psychological research on inter-group relations which demonstrates differences between majority and minority groups as a function of their power. The "golden standard" of ORB research is to find a cross-over interaction between the race of the participant and the race of the face, such that the participants from both groups recognize in-group faces more accurately than out-group faces. Researchers have argued that such interactions can rule out the possibility of a priori differences in the discriminability of the two face sets (e.g. Wells & Olson, 2001). However, there are reasons why one might expect members of minority groups – particularly those of low status – to recognize out-group faces as accurately, if not more accurately, than in-group faces. Minority group members may be reliant on majority group members for education, employment, housing, and so on. This might then necessitate a higher amount of meaningful contact with the out-group for minority groups than for majority groups (Islam & Hewstone, 1993a). Research on in-group/out-group attributional biases has also shown that members of high-status majority groups tend to favour the in-group and derogate the out-group; members of low-status minorities, however, do not show the same patterns of out-group derogation, and may even favour the out-group (Hewstone & Ward, 1985; Islam & Hewstone, 1993b).

Below, the evidence for the role of contact in the ORB will be reviewed, distinguishing between actual contact and opportunity for contact. The former will be reviewed first, in studies which have taken measures of self-reported contact, and in studies which have experimentally manipulated experience with other-race faces using training procedures. Opportunity for contact will then be discussed, with evidence from developmental studies, and from experiments conducted with populations differing in their potential exposure to members of other groups.

2.3.1. Actual contact

In adult populations, some studies have investigated the association between self-reported inter-racial contact and the magnitude of the ORB. Some studies have shown significant positive correlations between contact and accuracy (e.g. Brigham et al., 1982; Hancock & Rhodes, 2008), with the most powerful associations with items which assess current friendships (Carroo, 1986; Michel, Caldara, et al., 2006; Slone et al., 2000) and other meaningful individuating contact (Walker & Hewstone, 2006b, 2008). In some cases, contact is associated with accuracy for one race, but not for another (Michel, Caldara, et al., 2006; Platz et al., 1988; Wright et al., 2003), and in many cases, contact is not significantly associated with accuracy (Brigham & Barkowitz, 1978; Carroo, 1987; Corenblum & Meissner, 2006; Jackiw et al., 2008; Malpass & Kravitz, 1969; Ng & Lindsay, 1994; Sporer et al., 2007; Walker & Hewstone, 2006a). Wright et al. (2003) argued that the variability of contact scores within many samples may be insufficient to detect a relationship between contact and accuracy, and this may particularly be true in areas which are racially homogeneous.

Meissner and Brigham (2001) found that across 29 studies, increased self-reported contact was associated with a decrease in the magnitude of the ORB, although the effect was modest, accounting for around 2% of the variance.

Other studies have introduced experimental manipulations of exposure to own- and other-race faces. The logic is this: if the ORB is, at least to some degree, a product of learning experiences, then it should be possible to train participants to become more accurate at recognizing other-race faces. A small number of studies have tested this hypothesis. Elliott, Wills, and Goldstein (1973) and Goldstein and Chance (1985) trained White participants to individuate Asian faces using paired associate tasks, in which participants learnt associations between individual Asian faces and corresponding numeric labels. Trained participants showed a reduced ORB compared to control participants, who received no such training. Lebrecht et al. (2009) used a similar procedure with White and Black faces, and showed a reduced effect in White participants. Hills and Lewis (2006) found no ORB in White participants trained to focus attention on features which are useful for discriminating between Black faces, such as chin shape and mouth shape. However, verbal training, in which participants are trained to accurately describe other-race faces, does not appear to improve recognition (Malpass, Lavigueur, & Weldon, 1973). These studies strongly support the idea that people can learn to accurately distinguish between and recognize other-race faces. Passive exposure is insufficient to reduce the ORB (Goldstein & Chance, 1985; Hills & Lewis, 2006). Rather, participants must engage in a task which allows them to individuate the other-race faces. It is unclear, however, how long lasting these training effects may be as no studies have investigated the effects of training over an extended period of time.

2.3.2. Opportunity for contact

In recent years, a burgeoning literature has emerged on the developmental time course of the ORB, with participants ranging from newborn infants to children and adolescents. Researchers often make the assumption that the infants and children in these studies will have limited exposure to faces from other-race groups. Some developmental studies have included samples from groups with high potential for inter-group contact and groups with low potential for inter-group contact (Goodman et al., 2007).

Infant studies show that the ORB begins to emerge very early in life. Infants' face recognition can be tested with visual preferences paradigms, in which infants are habituated to a target face, and then shown the target along with a novel face. Infants will spend longer looking at novel stimuli than familiar stimuli, so recognition can be inferred when the looking time is longer for the novel face than for the target face. Kelly et al. (2005) found that White newborn infants were able to recognize White, Chinese, Black, and Middle Eastern faces equally well, but by three months old, the infants could only recognize White faces. The ORB at three months old has been confirmed elsewhere (Hayden, Bhatt, Joseph, & Tanaka, 2007; Kelly, Liu, et al., 2007), although Sangrigoli and de Schonen (2004a) showed that this effect can be eliminated with exposure to a small number of other-race faces. Other studies have found that the ORB develops slightly later in infancy. Ferguson, Kulkofsky, Cashon, and Casasola (2009) also found that four month old infants could discriminate between other-race faces, but that this ability was lost by eight months of age. Kelly, Quinn, et al. (2007) and Kelly et al. (2009) have shown that recognition becomes progressively narrower throughout the first year of life, leading to a reliable ORB by nine months old. Thus, despite some inconsistencies in the exact timing of the ORB, it

Table 1: *Age of onset in studies of the ORB in infants (top) and children and adolescents (bottom)*

Age	ORB present	ORB not present
	Infant studies	
Newborn		Kelly et al. (2005)
3 months	Hayden et al. (2007); Kelly et al. (2005); Kelly, Liu, et al. (2007); Sangrigoli & de Schonen (2004a)	Kelly, Quinn, et al. (2007) Kelly et al. (2009)
4-6 months	Kelly, Quinn, et al. (2007); Kelly et al. (2009)	Ferguson et al. (2009)
8-9 months	Ferguson et al. (2009); Kelly, Quinn, et al. (2007); Kelly et al. (2009).	
	Child/adolescent studies	
2-4 years	Sangrigoli & de Schonen (2004b); Shutts & Kinzler (2007)	
5-6 years	Pezdek et al. (2003); Sangrigoli & de Schonen (2004b); Shutts & Kinzler (2007)	Chance et al. (1982); Goodman et al. (2007)
9-10 years	Chance et al. (1982); Goodman et al. (2007); Pezdek et al. (2003); Sporer et al. (2007)	
13-16 years	Sporer et al. (2007); Walker & Hewstone (2006a)	

appears that face recognition becomes narrowed within the first few months of life. However, these own-race effects remain fairly plastic during infancy and can be reduced by a relatively small amount of contact with members of other races (Sangrigoli & de Schonen, 2004a).

Studies with children have produced slightly less clear results. The ORB has been shown in children as young as two (Shutts & Kinzler, 2007) or three years old (Sangrigoli & de Schonen, 2004b). Studies with slightly older samples have found an ORB in five year olds (Pezdek, Blandon-Gitlin, & Moore, 2003) and seven year olds (Corenblum & Meissner, 2006; Walker & Hewstone, 2006a). Other studies have failed to find an ORB until nine (Goodman et al., 2007) or ten years old (Chance, Lockwood-Turner, & Goldstein, 1982). By ten years and onwards, the ORB appears to be stable (Sporer et al., 2007; Walker & Hewstone, 2006b). Some of these inconsistencies in age of onset may be due to the different methodologies used, including perceptual discrimination tasks (Shutts & Kinzler, 2007; Walker & Hewstone, 2006a, 2006b), a delayed match to sample task (Sangrigoli & de Schonen, 2004b), old/new recognition tests (Chance et al., 1982; Corenblum & Meissner, 2006), a matching task (Sporer et al., 2007), and a lineup task (Pezdek et al., 2003). Nevertheless, the evidence suggests that the ORB develops during early childhood, and is firmly established by around ten years of age. The ages of onset of the ORB in studies with infants and with children are summarized in Table 1.

Outside of the developmental domain, researchers have investigated opportunity for contact by comparing participants from populations which are expected to differ in their potential for interacting with members of other groups. Samples from different countries have been compared (Carroo, 1986; Chiroro & Valentine, 1995; Goodman et al., 2007; Ng & Lindsay, 1994; Wright et al., 2001, 2003), as well as samples from more or less integrated areas within one country (Cross, Cross, & Daly, 1971; Sporer et al., 2007). The results of these studies have been mixed, with some showing reduced ORBs in higher contact groups (Carroo, 1986; Sporer et al., 2007), others showing differences only in one group of participants (Chiroro & Valentine, 1995; Cross et al., 1971), and yet others finding no difference in the magnitude of the ORB between high and low contact groups (Goodman et al., 2007; Ng & Lindsay, 1994; Wright et al., 2001, 2003). Weimann, Fishman, and Rattner (1988) found that the ORB was more pronounced when comparing samples from different countries than when comparing different racial groups within one country.

Corenblum and Meissner (2006) compared White Canadian participants' recognition of Whites, who constitute the majority race in the local population, Native Canadians, who represent 40% of the local population, and Blacks, who make up approximately 1% of the local population. The participants showed the strongest deficit for the Black faces, which were the least numerous group in the local populous. Chance et al. (1975) similarly found that White and Black American participants were especially inaccurate at recognizing Japanese faces compared to their more familiar racial out-group. However, these results must be interpreted with some caution, as it is possible that the low contact stimuli (Black faces for Corenblum & Meissner; Japanese faces for Chance et al.) were less discriminable than the other face sets, due to stimulus sampling errors (Wells & Olson, 2001). Li, Dunning, and Malpass (1998) compared recognition for Black faces in White basketball fans and basketball novices. Basketball is a sport in the US with a large number of Black players; in order to follow a game, fans would need to discriminate between different players accurately. In support of the contact hypothesis, fans showed a smaller ORB than novices.

Sangrigoli, Pallier, Argenti, Ventureyra, and de Schonen (2005) studied the recognition of White and Korean faces in White and Korean adults living in France. Some of the Korean participants had been living in France for a few months, while others had been adopted by French families between three and nine years old. While the Korean participants who had been in France for only a few months were more accurate with Korean faces than with White faces, the adopted participants were more accurate with White faces than Korean faces, showing very similar performance to the White participants. Developmental studies suggest that the ORB remains fairly plastic before the age of ten, leading Sangrigoli et al. (2005) to argue that exposure to a large number of other-race faces during childhood can reverse the ORB.

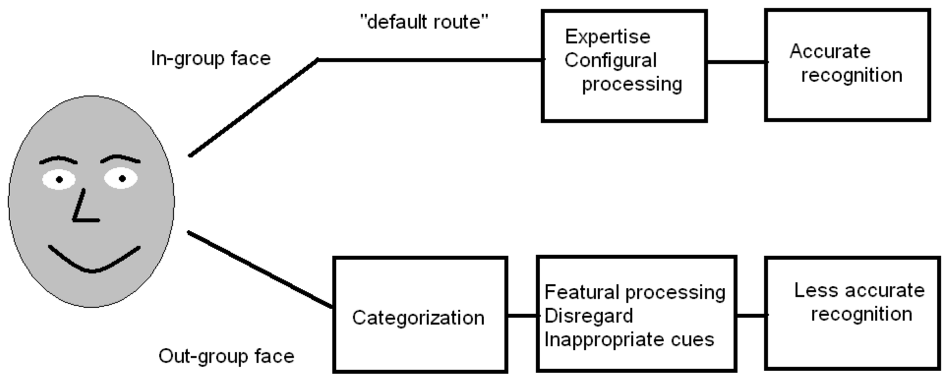
The contact hypothesis, though intuitive and parsimonious, has not been consistently supported. While it does appear that the ORB develops throughout the early months and years of life (e.g. Kelly et al., 2005; Shutts & Kinzler, 2007), effects of contact among adult populations are difficult to predict. To some extent, the success of the contact hypothesis seems to depend on whether researchers measure actual contact (e.g. by self-report), or whether they examine opportunity for contact. Results from studies comparing high and low contact groups across nations or within nations have been somewhat more supportive than studies measuring self-reported contact (Sporer, 2001b), but the results of even these studies have been equivocal. It seems, therefore, that the relationship between contact and the ORB is not straightforward; rather, there may be other mediating factors such as inter-group attitudes, anxiety, and categorization, through which contact exerts its effects on memory.

2.4. The in-group/out-group model

Malpass (1990) provided a utilitarian perspective of the ORB, based on the central idea that humans use their cognitive resources to maximize the rewards, and minimize the losses, of their social interactions. Being able to recognize faces enables humans to predict whether a particular social interaction will be beneficial or harmful, and to what degree. Within an individual's in-group, the value of social interactions is likely to vary quite widely, and there may be no obvious visual cues to denote which members will be more or less valuable to interact with. Recognizing in-group faces is therefore important for maximizing the value of social interactions. Malpass (1990) hypothesized that interactions with out-group members may be less variable in their values than interactions with in-group members. This means that it should be sufficient to categorize an out-group member, and learning to individuate out-group members will be of no substantial benefit. Ackerman et al. (2006) argued that interactions with in-group members are likely to be on an individual level, while interactions with out-group members are likely to be on a group level. Allport (1954) suggested that the categorization of people and objects to oversimplified groups is necessary in order to handle the huge amount of information that we encounter in daily life, and Taylor, Fiske, Etcoff, and Ruderman (1978) showed that categorization leads to errors discriminating between members within a group.

Leading on from this line of reasoning, Sporer (2001b) proposed the in-group/out-group model (IOM) of the ORB. A schematic of this model is shown in Figure 1. According to the IOM, the perception of an out-group face automatically triggers a categorization process. Once a face has been categorized as belonging to an out-group, it will be processed less efficiently or less deeply than a face which has not been categorized as an out-group member. Whereas in-group faces are processed configurally and stored in memory to allow optimal individuation, out-group faces are processed for features which denote their group membership at the expense of information appropriate for individuation.

Figure 1. Schematic diagram of the in-group/out-group model for face recognition.



Note: Adapted from Sporer (2001b).

Several studies have shown that other-race faces are categorized very rapidly. Racial classification is faster for other-race faces than own-race faces (Caldara et al., 2004; Levin, 1996; Zhao & Bentin, 2008). Levin (1996) argues that race is processed as a visual feature which is present in other-race faces but which is absent in own-race faces, and that this creates “pop out” effects when searching for other-race targets among an array of own-race faces (Levin, 2000; Levin & Angelone, 2001). These “pop out” effects for other-race faces have even been found in nine month old infants (Hayden, Bhatt, Zieber, & Kangas, 2009).

Neuroscientific research measuring event-related potentials (ERPs) shows that racial category information is processed very rapidly for other-race faces, even when race is task irrelevant (Ito & Urland, 2003), and when participants are required to make individuating judgments about other-race faces (Ito & Urland, 2005), suggesting that this rapid categorization process is outside of conscious control. Quanty, Keats, and Harkins (1975) suggested that racial categorization may be more accurate among higher prejudiced individuals, as they may become more sensitive to features which distinguish between in-group and out-group members. Skin color is a powerful cue for categorizing faces as in-group or out-group members (Blair et al., 2002). However, featural differences between groups are also very important in producing the ORB. Bar-Haim, Saidel, and Yovel (2009) manipulated the skin color of White and Black faces. Participants recognized own-race faces more accurately than other-race faces, regardless of skin color. However, the largest recognition advantage was for own-race faces with their natural skin colors.

The IOM is not specific to the processing of race – it is a more general model which can be applied to any categorical groups. For example, participants recognize own-age faces (Anastasi & Rhodes, 2005; Chance, Goldstein, & Anderson, 1986; Wiese, Schweinberger, & Hansen, 2008; Wright & Stroud, 2002) and own-gender faces (Sporer, 1993; Wright & Sladden, 2003) more accurately than other-age or other-gender faces. Rule, Ambady, Adams, and Macrae (2007) found that in-group advantages even extend to groups that are visually ambiguous. Heterosexual and homosexual participants recognized own-orientation faces more accurately than other-orientation faces, despite being ostensibly unaware of the group membership of the faces. Sexual orientation of faces may also be processed automatically, in a similar way to race (Rule, Macrae, & Ambady, 2009). Bernstein, Young, and Hugenberg

(2007) showed that participants recognized faces more accurately if they believed them to be students at the same university as themselves, or even if they were members of the same group in a minimal groups paradigm.

Recent years have seen the emergence of a strong body of work in support of the IOM as applied to the ORB. MacLin and Malpass (2001) demonstrated the "ambiguous race face illusion". Black parent faces and Hispanic parent faces were morphed together, creating faces of ambiguous race. When these morphs were given hairstyles typical of Black or Hispanic people, the perceived group membership of the faces changed, and an ORB for recognition emerged among Hispanic participants. Faces categorized as in-group members were recognized more accurately than the very same faces when categorized as out-group members. Categorical racial labels (Pauker & Ambady, 2009), and personal names typical of different groups (Hilliar & Kemp, 2008) have also been used to achieve ambiguous race recognition effects. Ambiguous race faces may become distorted in memory to become closer to the parent face of their perceived race (Corneille, Huart, Becquart, & Brédart, 2004; Eberhardt, Dasgupta, & Banaszynski, 2003).

Racial categorization affects perception as well as recognition memory. Faces categorized as in-group members are encoded more holistically than those categorized as out-group members (Hugenberg & Corneille, 2009; Michel et al., 2007). White participants perceive Black faces as darker than White faces, even when they are matched for luminance (Levin & Banaji, 2006). MacLin and Malpass (2001) showed that ambiguous race faces categorized as Black are perceived as having more Afrocentric features (e.g. fuller lips, darker skin) than the same faces when categorized as Hispanic.

According to the IOM, low motivation to individuate other-race faces is very important for maintaining the ORB. Encouraging participants to include other-race faces in their in-group reduces the magnitude of the bias (Pauker et al., 2009; Shriver, Young, Hugenberg, Bernstein, & Lanter, 2008). Providing participants with instructions to individuate other-race faces also reduces the size of the ORB (Hugenberg, Miller, & Claypool, 2007; Rhodes, Locke, Ewing, & Evangelista, 2009).

All of the above research strongly suggests that the ORB has a large social component. While some degree of expertise may be necessary for accurately discriminating between other-race faces, expertise is not sufficient. There must be a motivation to individuate, and when this condition is met, the ORB can be reduced (Hugenberg et al., 2007; Pauker et al., 2009; Shriver et al., 2008). Unfortunately, the categorization of faces to in-group and out-group status is a rapid and automatic process which may reduce the motivation to individuate other-race faces (Levin, 1996). It seems that the ORB can only be fully understood when the complex socio-cognitive nature of face processing is taken into account; faces are after all, predominantly social stimuli rather than just visual patterns (Sporer, 2001b).

3. What are the Cognitive Mechanisms of the Own-Race Bias?

As well as asking *why* the ORB develops, researchers have also looked for cognitive mechanisms which could explain *how* the ORB operates. The configural/featural hypothesis suggests that own-race faces are processed configurally or holistically, while other-race faces are processed featurally. The multidimensional face space hypothesis (Valentine, 1991) posits that other-race faces are stored in long-term memory in such a way as to make later retrieval more difficult. Both of these hypotheses have received some support, but also some criticisms. The evidence for both hypotheses will be evaluated below.

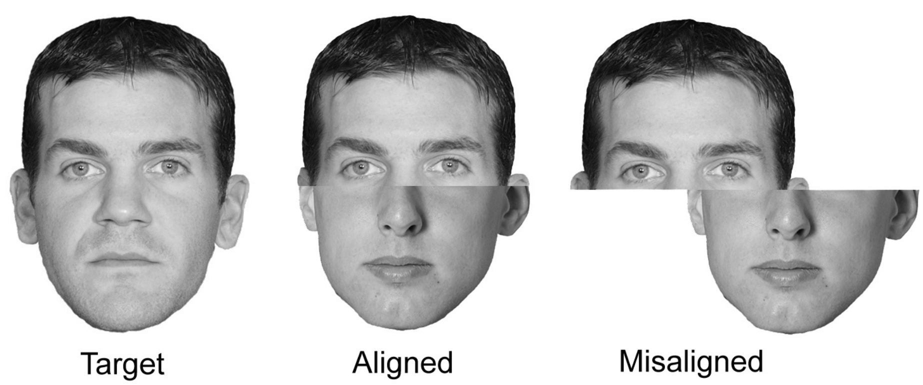
3.1. The configural/featural hypothesis

Visual stimuli can be processed featurally, by decomposing an image into its constituent parts, or configurally, by encoding the position and spacing of the component features in relation to one another. While many visual objects may be processed featurally, faces are believed to be processed configurally. Yin (1969) showed that recognizing inverted faces is much more difficult than recognizing many other inverted objects. This inversion effect is believed to be caused by the configural processing of faces. In support of this hypothesis, Farah, Tanaka, and Drain (1995) found that inverting dot patterns which were encoded holistically was more detrimental to recognition than inverting dot patterns which were encoded featurally. Diamond and Carey (1986) suggested that inversion will disrupt any objects associated with high expertise. Several studies have found larger inversion effects for own-race faces than other-race faces (Hancock & Rhodes, 2008; Rhodes, Brake, Taylor, & Tan, 1989; Sangrigoli & de Schonen, 2004b), suggesting that own-race faces are processed configurally, while other-race faces are processed featurally.

Event related potential (ERP) studies of the own-race inversion effect have produced conflicting results. The N170 is a waveform which is associated with early encoding of the configural information present in faces, which peaks around 170ms after stimulus onset. Inverting faces causes a delay in the peak of the N170. Gajewski, Schlegel, and Stoerig (2008) found that the delay in the N170 was larger for own-race faces than other-race faces, but only for White participants. Wiese, Stahl, and Schweinberger (2009) found similar delays in the peak of the N170 for own- and other-race faces.

The inversion effect has been criticized as an indirect test of configural processing (Diamond & Carey, 1986). Over recent years, researchers have found other ways to examine configural processing. For example, the composite face effect is the illusion that the top half of a face, when combined with the bottom half of a different face, appears changed (see Figure 2 for example composite faces). This illusion is created by the brain's tendency to process faces as a whole. When the two face halves are misaligned (as on the right of Figure 2), the composite effect disappears. This is because the misaligned composites are not processed holistically, and so the bottom half of the face does not interfere with the processing of the top half. Michel, Rossion, et al. (2006) found that White participants suffered a larger composite face effect for own-race than other-race faces, although this effect was not replicated in Asian participants. Michel, Corneille, and Rossion (2007) showed that this effect extended to morphed ambiguous-race faces when perceived as own-race group members, but not when perceived as other-race group members.

Figure 2: Example composite effect stimuli.



Note: The face on the left is the original target, the centre face is the aligned composite, and the face on the right is the misaligned composite.

Further evidence for the configural/featural hypothesis is that own-race faces show a whole/part advantage, whereas other-race faces do not (Michel, Caldara, et al., 2006; Tanaka, Kiefer, & Bukach, 2004). That is, features from own-race faces are recognized more accurately when presented as part of a whole face than when presented in isolation. Features from other-race faces are recognized as accurately whether presented as part of a whole face or in isolation. Participants are also more sensitive to configural changes made to own-race faces than other-race faces (Rhodes, Hayward, & Winkler, 2006). Configural processing of faces emerges in the early months of life. Ferguson et al. (2009) found that four month old infants process faces featurally. However, by eight months of age, faces are processed configurally. However, this configural processing only emerges for own-race faces, with other-race faces continuing to be processed featurally.

A reasonable amount of evidence has been provided for the configural/featural hypothesis of the ORB. However, there are still some inconsistencies and unanswered questions which need addressing. For example, the amount of configural processing that participants show for own- and other-race faces does not correlate with the magnitude of ORB participants show in a memory test (Michel, Caldara, et al., 2006; Michel, Rossion, et al., 2006). Some studies suggest that own-race faces benefit from improved featural processing as well as configural processing (Hayward, Rhodes, & Schwaniger, 2008; Rhodes et al., 2006). McKone, Brewer, MacPherson, Rhodes, and Hayward (2007) showed that highly familiar other-race faces can be processed configurally, although this is not likely to generalize to unfamiliar other-race faces. Furthermore, there is some evidence that configural processing extends to chimpanzee faces, despite participants' complete lack of expertise in discriminating chimpanzee faces. Taubert (2009) found significant inversion and composite effects for human and chimpanzee faces, but not for faces of other species with less similarity to humans, such as hens, lizards, and sheep. Taubert argued that holistic or configural processing of human-like faces is innate, and that some elements of this broad processing system are retained into adulthood. Thus, it is unclear at this stage whether configural processing is either necessary or sufficient for the successful recognition of other-race faces.

3.2. The multi-dimensional face space model

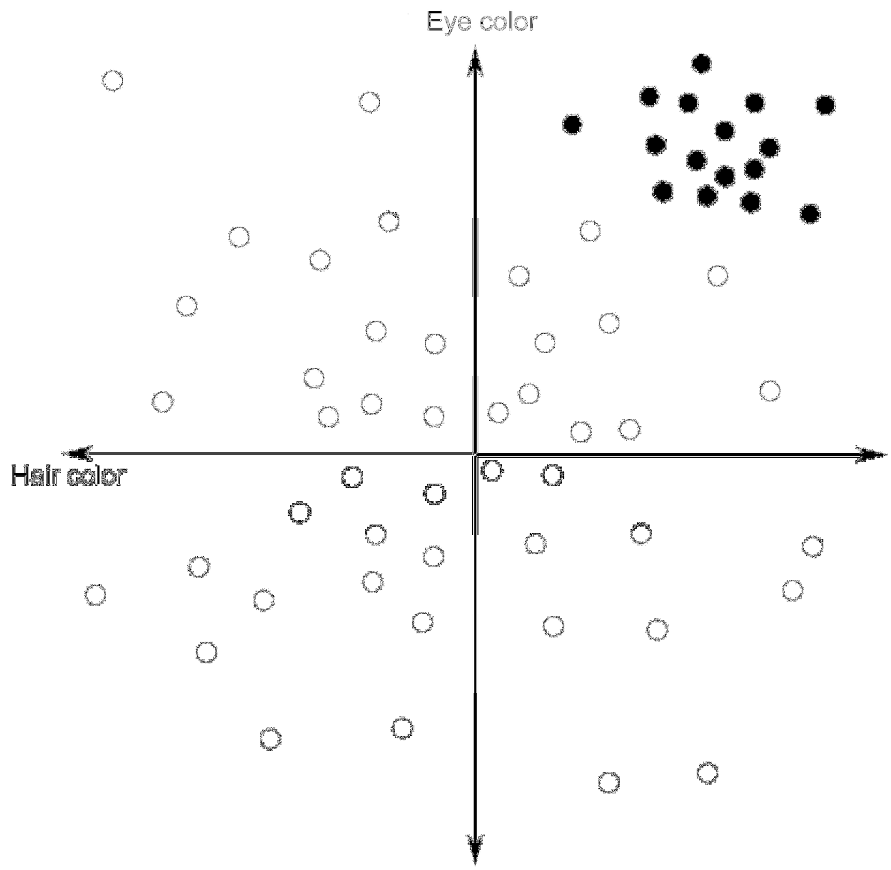
Valentine (1991) proposed that face exemplars are stored in memory in a multi-dimensional “face space”. The number of dimensions is not specified, but could consist of such variations as nose length, eye colour, and eye separation, or could correspond to dimensions that are not easily verbally described. Each face can be plotted as a specific point in this face space, according to its position along each dimension, such that similar faces which share many properties will be clustered closely together. The ease with which a face is recognized will depend on the density of exemplars in the surrounding area of the space. Distinctive faces are therefore remembered relatively easily, while typical faces are less easily remembered. Catz, Kampf, Nachson, and Babkoff (2009) created a six-dimensional face space, in which the locations of 200 face exemplars were plotted. The dimensions of this face space included size (of forehead, chin, whole face), face appearance (eye color, skin tone, distinguishing marks), and eye size and shape. Faces with fewer surrounding exemplars (i.e. those located in less dense areas of face space) were rated as more distinctive and were recognized more accurately than those with more surrounding exemplars (those located in more dense areas of face space).

The dimensions around which the face space is constructed are likely to be shaped by the perceptual experience of the perceiver (Valentine, 1991). Thus, a perceiver who has had extensive contact with White faces will structure their face space around dimensions which are useful for discriminating White faces. These same dimensions may be less appropriate for distinguishing between faces from other racial groups. For example, Ellis, Deregowski, and Shepherd (1975) found that when describing faces, White participants frequently describe variations in eye colour, hair colour, and hair texture. Whereas these characteristics are useful for discriminating between White faces, they will be less appropriate for discriminating between Black or Asian faces. Because the dimensions within the face space are optimal for distinguishing between own-race faces, these faces will be distributed throughout face space, although there is likely to be a tighter grouping around the centre of the face space (Catz et al., 2009). However, other-race faces will be left densely clustered in outlying regions of face space. This dense clustering, coupled with the large distance from the centre of the space, could explain the perceptual homogeneity of other-race faces (Walker & Hewstone, 2008) and the deficit in recognizing other-race faces.

Figure 3 shows a hypothetical two dimensional face space for a White perceiver, with own- and other-race exemplars plotted. The two dimensions within this space are hair color and eye color. These features vary more widely in White faces than in Black or Asian faces, for example. Therefore, while the own-race faces (light circles) are distributed throughout the face space, the other-race faces (dark circles) are clustered within one small region of the space. As a result, there is higher density of exemplars within this region. The other-race exemplars also fall further from the centre of the space than the majority of the own-race exemplars.

Caldara and Abdi (2006) simulated the ORB using a neural network trained on a majority of faces from one race and a minority of faces from another race, and showed that the minority exemplars became clustered in a three dimensional face space. Byatt and Rhodes (2004) asked White participants to rate the similarity of pairs of White and Chinese faces. Using multidimensional scaling analysis, own- and other-race faces were shown to form separate clusters in face space, with other-race faces being more densely clustered than own-race faces. Caldara, Rossion, Bovet, and Hauert (2004) also argue that the face space model can explain the reaction speed advantage for classifying other-race faces by race over own-race faces, as densely clustered faces would produce stronger activation than more evenly spaced faces.

Figure 3: A hypothetical two dimensional face space, with own- and other-race exemplars plotted.



Note: Light circles represent own-race faces, and dark circles represent other-race faces.

However, Levin (1996) argued that the face space model, though intuitive, is an abstraction which relies upon classifications produced by the experimental context. Sporer (2001b) proposed several questions about the face space model, which remain unanswered. How many dimensions would be necessary to adequately represent own- and other-race faces? In order to discriminate between other-race faces more accurately, would one add new dimensions to an existing face space, or would an entirely separate face space be required?

The face space model is an elegant way of conceptualizing the storage of faces in memory, and neatly explains the ORB in face recognition. However, the model does not acknowledge the social nature of the ORB. Faces are stimuli of huge social importance, and the processing of these stimuli may be subject to top-down influences stemming from social interactions with and expectations toward members of other races. A comprehensive account of the ORB should be able to explain the role of social influences such as explicit and implicit attitudes in recognizing own- and other-race faces.

4. Dual-Process Theories of Recognition

Dual-process theories posit that recognition memory can arise from two separable memory processes. Familiarity is the fluid feeling of having encountered a stimulus before, which varies along a continuum. Recollection, on the other hand, is the retrieval of specific information about a particular study event, which is thought to be an all-or-nothing process (Yonelinas, 2002; although see Mickes, Wais, & Wixted, 2009, for an alternative view). There are several different models of how these two processes interact. Atkinson and Juola (1974) argue that items which evoke a sense of very high or very low familiarity will be very quickly categorized as old or new, whereas items with moderate familiarity will activate a search for recollective information. In this conditional search model, the recollection process is only activated when the information gained from familiarity is ambiguous. Other models (e.g. Mandler, 1980) argue that the two processes are activated in parallel, although familiarity is typically faster than recollection. Despite some disagreements among dual-process models, it is commonly assumed that familiarity is faster and less effortful than recollection, and that the two processes function independently at the time of retrieval (Yonelinas, 2002).

4.1. Measuring recollection and familiarity

Several methods have been developed to test dual-process theories of recognition. These can broadly be divided into task dissociation methods, which attempt to isolate the two processes from each other using different tasks such as recall and recognition, and process estimation methods, which attempt to estimate the relative contribution of the two processes in a single task (Yonelinas, 2002). Because familiarity is widely assumed to be faster than recollection, many task dissociation methods have focussed on response time. Some of these studies have examined recollection and familiarity for fast and slow responses (Dewhurst, Holmes, Brandt, & Dean, 2006), while others have imposed response deadlines of varying lengths on participants (Gardiner, Ramponi, & Richardson-Klavehn, 1999; Konstantinou & Gardiner, 2005; McElree, Dolan, & Jacoby, 1999; Mulligan & Hirshman, 1995). The results of these studies have been inconsistent; some studies show that familiarity is the dominant process at short response times (McElree et al., 1999), while others suggest that recollection may also be a very fast process under certain circumstances (Dewhurst et al., 2006; Gardiner et al., 1999; Konstantinou & Gardiner, 2005).

Jacoby's (1991) process-dissociation procedure (PDP) was developed in response to concerns about the assumptions of task dissociation methods. Prior to the introduction of the PDP, researchers had compared accuracy on different memory tests, such as recall and recognition tests, in order to make inferences about the recollection and familiarity processes. However, this methodology relies on the assumption that the memory tasks are process pure – i.e. that recognition relies on familiarity and *only* familiarity, while recall relies on recollection and *only* recollection. Jacoby (1991) argued that this is almost certainly not the case, and so estimates of recollection and familiarity produced by task dissociation methods will be contaminated by the opposing process.

In the PDP, recollection and familiarity are estimated by comparing performance in a single task with two conditions: one in which both processes facilitate recognition, and one in which the two processes are placed in opposition. In the PDP, participants are presented with two stimulus lists, followed by a recognition task with two conditions – inclusion and exclusion trials. In inclusion trials, participants simply decide whether each stimulus is old or new, regardless of list membership. Recollection and familiarity should work in concert for this task, as they would in an ordinary recognition test. In exclusion trials, however, participants are asked only to identify stimuli from one particular presentation list, while treating stimuli from the other list as new. This places recollection and familiarity in opposition, as the participants must recollect information concerning the list membership of the stimulus; responses based solely on familiarity may be incorrect. The proportion of trials in which participants *correctly* identify stimuli as old in the inclusion trials (*P*[Inclusion]), and in which they *incorrectly* identify stimuli as old in the exclusion trials (*P*[Exclusion]) trials are then used to estimate recollection: *P*[Inclusion] – *P*[Exclusion], and familiarity: *P*[Exclusion]/(1-Recollection).

The PDP makes certain assumptions about the underlying memory processes of recollection and familiarity, some of which have drawn criticism. For example, the PDP assumes that recollection and familiarity are independent, and that the response on any given trial will be produced by one of these processes, never both. Curran and Hintzman (1995) argue that violations of this assumption can lead to apparent dissociations between the two processes, which are actually artefacts of the experimental design. For example, dividing attention at encoding has been found to decrease recollection estimates in the PDP, while leaving familiarity estimates unaffected. Curran and Hintzman (1995) argued that if the two parameter estimates are positively correlated at the item level, then increases in the recollection parameter will necessarily lead to decreases in the familiarity parameter. Jacoby, Begg, and Toth (1997) rejected this argument, claiming that the familiarity parameter is robust even to very high item level correlations. However, if performance in the exclusion trials becomes very accurate under deep processing conditions, then ceiling effects can cause familiarity estimates to be reduced (Richardson-Klavehn, Gardiner, & Ramponi, 2002).

One serious criticism of the PDP is that its definitions of recollection and familiarity are very narrow. Recollection is defined as the ability of the participant to accurately recall the list membership of a target item. Familiarity is then defined as the inability to correctly recall this list membership. Gruppuso, Lindsay, and Kelley (1997) argue that recollection as defined in the PDP does not reflect the retrieval of all aspects of the encoding episode, as there are many other details of the encoding event that a participant could recall without identifying the correct list membership. Similarly, familiarity as defined in the PDP is not simply an undifferentiated feeling of oldness, as a participant may be able to recall some other information about a target which is not relevant to list membership. Mulligan and Hirshman (1997) distinguished between *diagnostic* recollection and *nondiagnostic* recollection. Diagnostic recollection allows a participant to accurately discriminate list membership, whereas nondiagnostic recollection does not. The PDP only includes diagnostic recollection in the *R* parameter, which means that the *F* parameter will include nondiagnostic recollection. This becomes more problematic as the similarity of the two study lists increases, and recollection becomes less diagnostic. If the two study lists are very similar, the *R* parameter will be underestimated and the *F* parameter will be overestimated (Gruppuso et al., 1997; Mulligan & Hirshman, 1997).

A major weakness of the PDP is its narrow definition of recollection as the ability to recall the list membership of a target stimulus. The Remember-Know (RK) procedure, developed by Tulving (1985), allows a much broader pool of responses to be included under the heading of recollection. In the RK procedure, participants make experiential judgments about their memories for recognized items, categorizing them as Remembered (R) or Known (K). Participants make R responses when they are able to recollect *any* specific information about the encoding event associated with a target stimulus. This information may include thoughts that the stimulus evoked, associations that the participant made between stimuli, or any other detail of the encoding event. Participants make K responses when they believe that they have seen an item before, but are unable to recall any information about the encoding event. Some studies also include Guess (G) responses (e.g. Gardiner, Java, & Richardson-Klavehn, 1996; Konstantinou & Gardiner, 2005), and this helps to ensure that K responses are uncontaminated by guessing.

Tulving (1985) argued that recollection arises from an episodic memory system, whereas familiarity is a product of a semantic memory system. While the RK procedure does not directly test these two systems, R responses are presumed to accompany old/new judgments based on episodic memory, while K responses are presumed to accompany old/new judgments based on semantic memory. However, the nature of the RK procedure makes it very difficult to know how accurately R and K responses map onto recollection and familiarity, respectively. Some authors have questioned whether participants are able to understand RK instructions, and whether they are then able to introspect on their own memories to apply those labels in a meaningful way (Dunn, 2004). Instead, participants may use RKG responses to reflect differences in confidence, with R responses, K responses, and G responses being applied to decisions made with high, medium, and low confidence respectively (Donaldson, 1996; Dunn, 2004; Rotello, MacMillan, Reeder, & Wong, 2005).

Gardiner, Ramponi, and Richardson-Klavehn (1998) rejected this argument, however. They asked participants to provide explanations for their RKG judgments, and analyzed the content of these explanations. Only R responses were accompanied by reports of recollective experience; K responses were accompanied by reports of just knowing, and by feelings of familiarity. K responses were also sometimes associated with quite high confidence. These transcript data therefore suggest that participants are able to understand and apply instructions meaningfully in the RK paradigm. However, it is also possible that the participants in that particular study were more stringent in their RKG judgments, as they knew that they would have to explain these decisions. Participants in standard RKG experiments may be more likely to use confidence as a heuristic for making RKG judgments (Rotello et al., 2005).

There is some contention around the analysis of RKG responses. The recollection process is easily estimated from R responses, as participants will make an R response whenever they recollect an item. Estimating familiarity is more complex, as participants will only make K responses for items which are familiar but *not* recollected. However, the familiarity process will almost certainly contribute to the recognition of recollected items, and K responses will therefore underestimate the contribution of familiarity to recognition (Gardiner et al., 1996). Yonelinas and Jacoby (1995) proposed transforming K responses using the Independence Remember-Know procedure (IRK). In this procedure, familiarity is estimated by dividing the number of K responses by the number of trials in which R responses are not given: *Familiarity = K/(1-R).* However, this procedure relies on the assumption that the two processes can be subjectively independent – that not only can one experience familiarity in the absence of recollection, but that one can also experience recollection in the absence of familiarity. While this state of awareness may be possible in some cases, it would seem to be an unusual state of awareness, and one that participants may be unable to understand (Gardiner et al., 1996).

An alternative way of estimating familiarity, which would not make the assumption of independence, would be as a total of K and R responses: *Familiarity = K + R*. This would reflect the likelihood that most items which are recollected are also familiar. Many other studies have simply analyzed the proportions of R and K responses for items categorized as old (e.g. Conway & Dewhurst, 1995; Hockley & Consoli, 1999; Mäntylä, 1997). This allows the subjective feelings of remembering and knowing to be examined, but caution must be taken when equating these responses to the recollection and familiarity processes.

4.2. Factors influencing recollection and familiarity

Dual-process theories predict that some variables will influence one process while leaving the other unaffected. Dividing attention during encoding, for example, should reduce recollection more than familiarity. Brandt, Macrae, Schloerscheidt, and

Milne (2003) asked participants to count backwards in threes during the encoding phase of an old/new face recognition test. During the recognition test, participants provided RK responses. Divided attention reduced the number of R responses given to distinctive faces. Parkin, Gardiner, and Rosser (1995) found a similar reduction in R responses when attention was divided by having participants detect auditory tones. Reinitz, Morrissey, and Demb (1994) found that participants made more familiarity-based errors in line drawings of faces when attention was divided than when attention was undivided.

Craik and Lockhart (1972) introduced the levels-of-processing (LOP) framework. According to this framework, objects can be processed at various "depths". Shallow processing may include surface characteristics of the stimulus – the color of ink that a word is printed in, or the size of the nose on a face. Deeper processing involves making inferences or semantic links – for example, thinking of synonyms for a word, or attributing personality characteristics to a face. The LOP approach has become a very popular framework for considering memory, with Craik and Lockhart's (1972) original paper being cited more than 3,200 times to date (Web of Science, 2009). Many of these studies have shown that when task instructions encourage deeper processing, recognition is more accurate than when task instructions encourage shallow processing (e.g. Berman & Cutler, 1998; Bloom & Mudd, 1991; Bower & Karlin, 1974; Patterson & Baddeley, 1977). However, some authors have argued that it is the *amount* of information which is processed that improves recognition, rather than the *depth* of the processed information (Bloom & Mudd, 1991; Winograd, 1981). Sporer (1991) argued that LOP effects are caused by the inferiority of shallow processing instructions over natural encoding strategies, rather than superiority of deep encoding instructions over natural encoding strategies.

Dual-process theories predict that recollection should be enhanced by deeper processing, but that familiarity should be unaffected (Gardiner et al., 1996; Yonelinas, 2002). Mulligan and Hirshman (1995) showed that deep processing was more beneficial to accuracy in unspeeded recognition tests, in which recollection can contribute to recognition, than in speeded recognition tests, which rely much more heavily on familiarity. However, other studies have shown that deep processing benefits familiarity as well as recollection (Brown & Lloyd-Jones, 2006; Gardiner et al., 1999; Konstantinou & Gardiner, 2005). Overall, however, recollection does seem to benefit from deeper processing to a larger extent than familiarity (Yonelinas, 2002).

4.3. Dual-process theory and the own-race bias

Meissner et al. (2005) recently proposed that dual-process theory could be applied to the ORB. The ORB is often characterised by a marked increase in false alarm rates (Doyle, 2001; Meissner & Brigham, 2001; Slone et al., 2000), which could be the result of impoverished recollection for other-race faces. Yonelinas (2002) argued that false alarms can be interpreted as an index of familiarity, as they presumably reflect recognition without recollection. Meissner et al. (2005) showed that participants gave more R responses to own-race faces than to other-race faces in an RK task. Marcon, Susa, and Meissner (2009) used a repetition lag paradigm, in which distractor faces are repeated during the recognition test. Participants mistakenly identified previously seen distractors more frequently for other-race faces than own-race faces. This suggests that for other-race faces, the ability to recollect the details of the encoding event was impaired.

Many of the mechanisms believed to contribute to the ORB could explain why recollection is impaired for other-race faces. For example, faces which are processed configurally may be more likely to be recollected than faces processed featurally (Yonelinas, Kroll, Dobbins, & Soltani, 1999). Faces densely clustered in face space may be difficult to recollect, as other surrounding face exemplars may also be activated by the presentation of a target (Valentine, 1991); Brandt et al. (2003) and Brown and Lloyd-Jones (2006) found that distinctive faces are more likely to be recollected than typical faces. Categorizing faces by race could reduce effortful processing, leading to a weaker memory trace at retrieval (Sporer, 2001b). Mäntylä (1997) showed that faces which are processed for similarities are more likely to be familiar than recollected, while faces processed for differences are more likely to be recollected than familiar.

4.4. Memory for context

In old/new recognition tests, the ability to identify a previously seen target is tested. This type of recognition is item recognition, and it can be distinguished from context memory, which is the ability to remember information concerning the encoding event in which a stimulus was presented. For example, participants may be asked to remember the paired associate of a target word (Hockley & Consoli, 1999; Westerman, 2001), the occupational label associated with a face (Yovel & Pallier, 2004), the color of ink in which a target word was presented (Mickes et al., 2009), or the voice in which a target word was spoken (Bornstein & LeCompte, 1995).

The distinction between item recognition and context memory can be intuitively understood when considering face recognition. For example, upon encountering a face, we may initially be struck with a sense of knowing that person, yet the source of that familiarity may be outside of our grasp. This will activate a search in memory for plausible contexts in which we may have encountered that person. If the recollective search is successful, we will attribute our familiarity to a particular encounter; if unsuccessful, we may decide that we were mistaken in our feeling of familiarity. Brown, Deffenbacher, and Sturgill (1977) showed participants two groups of faces, in two different rooms, and later tested recognition for those faces in a third room. As well as old/new judgments for each face, participants were asked to identify the room in which faces judged as old had been seen. While recognition of the faces was very high, participants could often not recollect in which room the faces had been presented.

Dual-process theory predicts that recollection, but not familiarity, should support context memory (Yonelinas, 2002). For example, Mandler (1980, p. 253) argues that the "specific identification of an event is not possible on the basis of familiarity alone", and memory for context is associated with a higher proportion of R responses than K responses using the RK procedure (Hockley & Consoli, 1999). However, familiarity may contribute to context memory when decisions are made very rapidly (Westerman, 2001) or when the target stimulus and the contextual information are perceived holistically as one item (Yonelinas et al., 1999).

4.5. Context reinstatement

Reinstating the context in which a stimulus was originally seen improves the ability to recall that item. Godden and Baddeley (1975) famously demonstrated that words which were learned on land or underwater were remembered more accurately if recalled in the same environment. Other research has shown that words are recalled more accurately by participants in the same mood state (Balch, Myers, & Papotto, 1999) or physical state (Schramke & Bauer, 1997) at encoding and testing. Marian and Neisser (2000) found that bilingual participants recalled autobiographical memories more easily if the language used at recall was the same as that being spoken during the remembered event.

The effects of context reinstatement on recognition are less clear, however. Some studies have found that context reinstatement improves accuracy in old/new recognition tasks (Evans et al., 2009; Gruppuso, Lindsay, & Masson, 2007; Hockley, 2008; Sporer, 1993), while others have found no improvement when context is reinstated (Baddeley & Woodhead, 1982; Macken, 2002). Smith (1988) argued that during free recall, relatively few cues are available to the participant, so context reinstatement provides cues that increase accuracy. However, in recognition, there are very strong cues available to the participant (the stimulus itself), so context reinstatement can provide only a very small benefit at best. In line with this argument, context reinstatement seems more beneficial when recognizing unfamiliar stimuli such as faces and non-words, than when recognizing familiar stimuli such as known words (Russo, Ward, Geurts, & Scheres, 1999).

The success of context reinstatement in improving recognition accuracy seems to depend on the way in which context is manipulated. For example, local context is associated with one or few stimulus items, whereas global context is associated with many stimulus items. Dalton (1993) found that reinstating local context improved recognition accuracy for both familiar and unfamiliar stimuli, whereas reinstating global context improved recognition accuracy for unfamiliar stimuli but not for familiar stimuli. Many of the studies which have shown context reinstatement effects on recognition have manipulated local context using unique visual or verbal information for each stimulus at encoding (Evans et al., 2009; Gruppuso et al., 2007; Hockley, 2008).However, not all studies have shown improved recognition accuracy following reinstatement of local context (Baddeley & Woodhead, 1982).

Contexts can also be interactive or independent (Baddeley & Woodhead, 1982). Interactive contexts are always processed and interfere with the encoding of a stimulus. Independent contexts, on the other hand, are not always processed and do not interfere with encoding. Sporer (1993) suggested that clothing cues function as interactive contexts when recognizing faces. Participants saw a series of faces with clothing cues present or hidden during encoding, and present or hidden during testing. Participants were more accurate when the clothing cues were congruent (either present or hidden) at both times. Interactive contexts may be more beneficial to recognition than independent contexts.

From a dual-process perspective, context reinstatement can enhance recollection (Gruppuso et al., 2001; Macken, 2002), and to a lesser degree, familiarity (Hockley, 2008). Context reinstatement also improves recognition more for stimuli which are less familiar (Russo et al., 2009). Applied to the ORB, one might predict that context reinstatement would reduce the magnitude of the ORB by boosting recollection for other-race faces. However, Evans et al. (2009) found that context reinstatement only benefited own-race faces, and did not improve recognition of other-race faces. The contextual information used by Evans et al. (2009) was verbal information, which may not have acted as an interactive context. Further research into context reinstatement at the ORB therefore seems warranted.

5. Eyewitness Memory

As well as informing theories about how faces are processed and stored in memory, the ORB has direct implications for real world behavior, including eyewitness identification. Over the last few decades, a wealth of research has shown that eyewitnesses make errors when identifying suspects, which can and do lead to false convictions (Brigham, 2007; Doyle, 2001). Eyewitness memory was one of the first applied topics to be studied by psychologists (Loftus, 1993) and has continued to receive a great deal of attention due to the grave consequences of mistaken identifications within the legal system. Inaccurate identifications create two distinct problems: on the one hand, an innocent person is imprisoned, and on the other hand, the guilty party eludes justice (Malpass & Devine, 1981).

To date (1st February, 2010), 249 falsely convicted people have been exonerated from US prisons by the Innocence Project, using DNA testing. The average sentence served by these individuals was 12 years; 17 of these people had been sentenced to death. These cases are thought to be just the “tip of the iceberg” (Naughton, 2003), as most crimes will not leave behind any DNA marker. Scheck, Neufeld, and Dwyer (2003) argued that inaccurate eyewitness identifications are the leading cause of wrongful imprisonment. Gross, Jacoby, Matheson, Montgomery, and Patil (2004) examined 340 exonerations in the US between 1989 and 2003, and found that in 64% of cases, at least one eyewitness erroneously identified the defendant; this increased to 88% in rape cases.

Eyewitness identifications often form key evidence in criminal trials. Although many eyewitnesses will provide accurate testimony, jurors may find it very difficult to discriminate between accurate and inaccurate identifications (Penrod & Cutler, 1995; Lindsay, Wells, & Rumpel, 1981). Earlier in the investigative process, investigating officers may also be “over-influenced” by eyewitness decisions in response to lineups (Dahl, Lindsay, & Brimacombe, 2006). This may stem from a tendency to equate confidence with accuracy, leading to the belief that confident witnesses will make correct identifications. However, research has consistently shown that the correlation between a witness’ confidence and the accuracy of their testimony is small (e.g. Penrod & Cutler, 1995; Perfect, Watson, & Wagstaff, 1993; Smith, Kassin, & Ellsworth, 1989; although see Odinot & Wolters, 2006), although the correlation is stronger for witnesses who make identifications than for witnesses who do not make identifications (Sporer, Penrod, Read, & Cutler, 1995). Further, jurors and judges may underestimate the influence of many factors on the reliability of a witnesses' memories (Benton, Ross, Bradshaw, Thomas, & Bradshaw, 2006; Wise & Safer, 2004), despite high consensus on these factors among researchers (Kassin, Tubb, Hosch, & Memon, 2001).

Laboratory research on eyewitness memory has been criticized by some as inapplicable to real world behavior (Bartolomey, 2001; Egeth, 1993). For example, ethical constraints limit the stress under which a participant can be placed, and practical constraints limit the number of factors that can be taken into account in any single study (Memon, Mastroberardino, & Fraser, 2008). However, eyewitness research has successfully informed policy on at least some issues (Wells et al., 2000). In 1999, *Eyewitness Evidence: A Guide for Law Enforcement* (Technical Working Group for Eyewitness Evidence, 1999) was published in the USA. This guide contained information from the psychological literature on factors such as lineup instructions, simultaneous vs. sequential lineups, and selection of fillers for lineups. This collaboration between researchers and policy makers within the legal system can at least provide some optimism that eyewitness research can be applied to real world behavior, and can make a difference to the collection of eyewitness evidence.

5.1. Eyewitness memory and the ORB

Wells (1978) distinguished between system variables and estimator variables in eyewitness memory research. System variables are controllable within the legal system, and include such factors as lineup format and selection of fillers. Estimator variables, however, refer to aspects of the witness or crime, and are not controllable within the system. These include such factors as stress, viewing conditions, and the ORB. Exoneration statistics reveal the potential impact of the ORB on the reliability of eyewitness memory. Gross et al. (2004) found that 39 of 80 exonerations for rape involved a Black defendant and a White victim/witness, yet less than 10% of all rapes involve Black assailants and White victims. Of the first 180 exonerations brought about by the Innocence Project, 61% of the defendants were Black and 78% of the witnesses were White.

Most studies of the ORB have used old/new recognition tests (Sporer, 2001a), and some authors have expressed concern about the applicability of these studies to eyewitness identifications, which will usually be made from lineups (Bartolomey, 2001). The consistency of findings across field studies (Brigham et al., 1982; Platz & Hosch, 1988; Wright et al., 2001) and other lineup studies (Evans et al., 2009; Jackiw et al., 2008) suggests that the ORB does generalize to eyewitness identification. A few archival studies, in which patterns across large numbers of real cases are analyzed, have provided mix results. Wright and McDaid (1996) examined lineup identifications of 616 suspects by 1561 witnesses in the UK. Although the race of the witnesses was not recorded, White suspects were less likely to be identified than suspects from other races. Valentine, Pickering, and Darling (2003) conversely found that Black witnesses in UK lineups were less likely to be identified than White suspects, and the race of the witness did not interact with the race of the suspect. Behrman and Davey (2001) analyzed 159 own-race and 72 other-race lineups in California, and found higher suspect identification rates in same-race than cross-race lineups. Suspect identification rates in field show-ups (which are similar to old/new recognition judgments) were similar in other-race and own-race cases. However, causal conclusions are difficult to draw from archival studies. Witnesses are not randomly allocated to groups (Wright & McDaid, 1996); there is no control over factors such as retention interval, crime seriousness, or selection of fillers; and there is no guarantee that the suspect is the culprit (Behrman & Davey, 2001).

Ninety percent of eyewitness researchers surveyed by Kassin et al. (2001) believed that the ORB is a reliable effect, and 72% would testify about the ORB in a criminal trial. While it is clear that more research is needed to establish the reliability of the ORB in the eyewitness world, it is also clear from exoneration statistics that the legal system is failing members of minority races. The extent of this failure extends well beyond eyewitness misidentification. For example, jurors may perceive members of minority races as more culpable (Lindholm & Christianson, 1998) and less credible (Abshire & Bornstein, 2003) than members of majority races. Black juvenile offenders are more likely to be prosecuted as adults than White juvenile offenders (Gross et al., 2004), and Black motorists are more likely to be stopped by traffic officers than White motorists (Buerger & Farrell, 2002).

5.2. Memory for context in eyewitness identifications

For a witness attempting to identify a suspect, it is not enough that the witness finds a face familiar. The witness must be sure that they saw the suspect in a particular time and place, committing a particular act. This involves discriminating between the sources of one's own memories, and has been termed *source monitoring* (Johnson, 2006). Source monitoring failures in eyewitness identifications could have severe consequences, as innocent individuals with some familiarity to the witness could be identified (Lindsay, 2007). Loftus (1976) suggested that these kinds of errors are a "by-product of the integrative malleable nature of human memory" (p. 97). Errors may occur when an innocent bystander to a crime is misidentified (Loftus, 1976), or when a witness is exposed to intervening mugshots or lineups between witnessing the crime and making the final identification.

The effects of mugshot exposure on eyewitness memory have been studied extensively. Searching mugshots for the suspect decreases correct identifications (Davies, Shepherd, & Ellis, 1979; Deffenbacher, Bornstein, & Penrod, 2006) and increases mistaken identifications (Brown et al., 1977; Deffenbacher et al., 2006; Deffenbacher, Carr, & Leu, 1981) in later lineups. Suspects who are identified by the witness during a mugshot search are much more likely to be identified in a later lineup, even if innocent (Dysart, Lindsay, Hammond, & Dupuis, 2001). Behrman and Davey's (2001) archival study included 66 lineups in which the lineup identification was preceded by an earlier mugshot or photographic lineup identification. Suspect identification rates in these lineups were higher than for lineups where no previous identification had been made. Furthermore, 45% of the witnesses who had not identified the suspect at the first opportunity went on to identify the suspect in the later lineup. Although mugshot searches can be useful investigative tools (Lindsay, Nosworthy, Martin, & Martynuck, 1994), witnesses who view mugshots should not be later asked to make a lineup identification, as the risk of error is too high (Dysart et al., 2001).

The cognitive interview (CI) is a technique used to increase the reliability of eyewitnesses' reported memories. A core element of the CI is to mentally reinstate contextual information about the crime scene. Malpass and Devine (1981) used a guided memory procedure to enhance participants' memories of a culprit of a staged crime, five months after the event. The participants were asked to recall details about the room in which the incident took place, their thoughts and feelings at the time, and the immediate reactions to the incident. This guided memory technique increased the accuracy of lineup identifications. Krafka and Penrod (1985) showed that mentally reinstating an encounter with a confederate while viewing physical evidence from the encounter increased hit rates in a lineup, without increasing error rates. Context reinstatement therefore appears to be a useful tool for increasing correct identifications from lineups, while leaving misidentifications unaffected.

6. Introduction to Papers

Throughout this thesis, the own-race bias in face recognition will be examined from a dual-process perspective. The relative contributions of recollection and familiarity in recognizing own- and other-race faces will be explored, as well as the ability to retrieve contextual information about those faces. The aim of this thesis is to provide support for the hypothesis that recollection is impaired when identifying faces of other races, and to show that participants therefore have difficulty in remembering specific details of the previous occurrence of other-race faces.

In Paper 1, Jacoby's (1991) process-dissociation procedure is used to estimate the contributions of recollection and familiarity to White participants' recognition of White and Black faces. Level of processing effects on recollection and familiarity are also tested. The main predictions of Paper 1 are that recollection estimates will be higher for the White faces than for the Black faces, while familiarity estimates may be similar for the two groups of faces. There is some evidence that participants process other-race faces more deeply than own-race faces in recognition tasks (Chance & Goldstein, 1981; Sporer, 2001b). Encouraging participants to adopt deep or shallow processing strategies for all faces may therefore reduce the difference in recollection for the White and the Black faces.

In Paper 2, a procedure similar to the exclusion trials of the PDP is used to investigate White participants' abilities to discriminate between familiar White and Black faces from two intermingled study lists. Distinguishing between familiar stimuli from the two lists requires a recollective experience of the encoding encounter. Therefore, the participants should be more accurate with the White faces than with the Black faces. Paper 2 also investigates whether accuracy can be improved if the participants are encouraged to consider their responses more carefully.

Papers 3 and 4 both explore memory for context in own- and other-race identifications. Faces are shown on one of several backgrounds at encoding. At testing, these faces are seen again on blank background, along with a series of new faces. Participants make old/new recognition decisions for each faces. Following an old judgment, the participant must attempt to identify upon which background the face had originally appeared. In Paper 3, White participants are recruited; in Paper 4, both White and Black participants are recruited. The main prediction for these papers is that participants will be more accurate at identifying the contexts in which correctly recognized own-race faces were seen than other-race faces.

In eight experiments across four papers, the importance of recollection in the own-race bias for face recognition is explored. Process estimation methods such as Jacoby's (1991) process-dissociation procedure and Tulving's (1985) Remember-Know procedure are used to estimate the contribution of recollection to recognizing own- and other-race faces. Factors which influence recollection, such as depth of processing and context reinstatement are manipulated. Memory for specific contextual information about correctly recognized own- and other-faces is tested. The overall aim of this thesis is to show that recollection is impaired for other-race faces, and that there is therefore a reduction in the ability to recall specific information about previously seen other-race faces.

Paper 1: Recollection and Familiarity in Recognizing Own- and Other-Race Faces

Abstract

Recognition accuracy is higher for faces of one's own race than for faces from other races. Recent evidence from a dual-process perspective suggests that people may have difficulty retrieving specific information about the previous occurrence of other-race faces. In two studies, White participants completed a process dissociation task with White and Black faces, in which recollection and familiarity were estimated based on performance in inclusion and exclusion trials. In Experiment 1, the recollection estimate was higher for White faces than for Black faces, and participants made more errors with Black faces than with White faces. In Experiment 2, depth of processing was manipulated. Deep processing was associated with higher recollection estimates than shallow processing for both White and Black faces. Recollection estimates for White and Black faces were very similar in the shallow and deep conditions.

Introduction

For many years, a widespread belief has persisted that people have difficulty discriminating between and recognizing people from other races. Feingold, in 1914, stated that "to the uninitiated American, all Asiatics look alike, while to the Asiatic all White men look alike" (p. 50). This hypothesis was first tested empirically by Malpass and Kravitz (1969) with White and Black Americans. Since then, dozens of papers have been published on the topic, including several meta-analyses. The most recent meta-analysis, conducted by Meissner and Brigham (2001), concluded that the own-race bias (ORB) is very robust, and generalizes to participants from many ethnic and racial groups.

Most studies of the ORB have used old/new recognition tests, in which participants study a set of faces, and later attempt to identify the studied faces from a larger set including some new distractor faces (e.g. Cross, Cross, & Daly, 1971; Hills & Lewis, 2006; Wright, Boyd, & Tredoux, 2003). Participants frequently make more false alarms for other-race faces than own-race faces (e.g. Chiroro & Valentine, 1995; Slone, Brigham, & Meissner, 2000; Teitelbaum & Geiselman, 1997). Often, these higher false alarm rates for other-race faces are accompanied by lower hit rates, creating a “mirror effect” (Glanzer & Adams, 1990; Meissner, Brigham, & Butz, 2005; Teitelbaum & Geiselman, 1997).

Over several decades, theoretical models of recognition memory have been developing. The number of processes involved in recognition has been debated, with some models proposing that recognition is the product of a single process, and others proposing the existence of two separate processes, recollection and familiarity. Single-process models, such as traditional signal detection theory, posit that recognition is based upon the strength of activation produced by a stimulus, which is compared to some decision threshold (Swets, 1964). Dual-process theories, on the other hand, posit that in addition to this continuous activation strength process (familiarity), there is a second independent route to recognition (Yonelinas, 2002). This second route, recollection, involves the retrieval of episodic information related to the encoding of the stimulus. For example, an observer may recollect some thought or feeling that the stimulus evoked at encoding, or they may remember some details about the context in which a stimulus was presented.

In support of dual-process models, neuropsychological and behavioral dissociations have been found between recollection and familiarity. Separate event related potentials (ERPs) have been identified for recollected stimuli and familiar stimuli (Curran, 2000; Wilding, Doyle, & Rugg, 1995). Amnesic patients who are unable to recognize items using recollection may still be able to recognize items based on familiarity (Yonelinas, 2002). And some experimental manipulations affect one process while leaving the other relatively unaffected. For example, dividing attention at encoding reduces recollection but does not affect familiarity (Brandt, Macrae, Schloerscheidt, & Milne, 2003; Parkin, Gardiner, & Rosser, 1995), and reinstating elements of the encoding context benefits recollection to a much greater degree than familiarity (Macken, 2002).

Meissner et al. (2005) recently considered the ORB in a dual-process framework. People may have difficulty retrieving episodic information about the previous occurrence of other-race faces, which then reduces the ability to reject faces which appear somewhat familiar but which have never actually been seen. Meissner et al. (2005) provided some support for this hypothesis using a Remember-Know task (Tulving, 1985) in which participants describe the subjective experiences associated with their old judgments in an old/new task using Remember and Know labels. The participants made more Remember responses for own-race faces than for other-race faces, but made a similar proportion of Know responses for both groups of faces. Marcon, Susa, and Meissner (2009) used a repetition lag paradigm to examine recollection and familiarity, in which some of the distractors were repeated during the testing phase of an old/new recognition task. The participants made more repetition errors for other-race faces than for own-race faces. Marcon et al. (2009) concluded that participants were unable to retrieve information about the previous encounter with the repeated distractors, and so misattributed their familiarity to them having appeared during the encoding phase.

Many methods have been developed for testing dual-process theories of recognition memory. One method is the process dissociation procedure (PDP), created by Jacoby (1991). The PDP was introduced as a way of estimating the relative contribution of recollection and familiarity to a given task. Previously, researchers had compared performance across two different tasks, such recall and recognition. The conclusions from these studies were based on the assumption that these tasks were process pure (i.e. each task involved either recollection or familiarity, not both), which Jacoby (1991) argued was an invalid assumption. The PDP uses a single task with two different trial types to derive estimated parameters for the contribution of recollection and familiarity to recognition performance.

During the encoding phase of the PDP, participants see two stimulus lists. At testing, the stimuli from these two lists are seen again, along with some distractors. Participants take part in inclusion trials and exclusion trials. In inclusion trials, participants are simply asked to decide whether each stimulus is old or new, regardless of list membership. The inclusion trials therefore simply consist of targets and distractors, as in a standard old/new recognition test. In exclusion trials, participants are asked only to identify the stimuli which appeared in one of the two encoding lists, treating stimuli from the other list as new. In addition to targets and distractors, the exclusion trials also contain *lures*, previously seen stimuli which are to be ignored (from the non-identified list).

In order to estimate recollection (R) and familiarity (F), the proportion of old responses to targets in the inclusion trials and lures in the exclusion trials must first be calculated:

P(Inclusion) = hits / (hits + misses)

P(Exclusion) = lure IDs / (lure IDs + lure rejections)

Old responses to targets in the inclusion trials (P(Inclusion)) will consist of some recollection-based responses, and also some familiarity based responses: $R + F(1 - R)$. Old responses to lures (P(Exclusion)) will consist only of familiarity-based responses: $F(1 - R)$. To estimate recollection (R), the proportion of lures incorrectly identified as old in the exclusion trials is then subtracted from the proportion of targets correctly identified as old in the inclusion trials:

$R = P(\text{Inclusion}) - P(\text{Exclusion})$.

Recollection and familiarity are assumed to be independent in the PDP; therefore, familiarity (F) is estimated with the following formula:

$F = P(\text{Exclusion})/(1 - R)$.

Despite the widespread use of the PDP in memory research, the procedure has received some criticism. Some of these criticisms have concerned the statistical assumptions made by the PDP, whereas others have concerned the way in which memory is conceptualized within the PDP framework. For example, a central assumption of the PDP is that recollection and familiarity are independent. That is, in any given trial, either recollection or familiarity will be used, but never both. This "independence assumption" has been widely criticized (McBride, 2007), with opponents of the PDP claiming that violations of this assumption can create artifactual dissociations between recollection and familiarity. Curran and Hintzman (1995) showed that if the estimated R and F parameters are highly correlated at the item level, then increases in R will lead to decreases in F. This causes an underestimation of the contribution of familiarity to recognition. Jacoby, Begg, and Toth (1997) refuted this claim, showing that F estimates are robust even to very high item-level correlations. However, Jacoby et al. (1997) did concede that the independence assumption could be violated if performance is near ceiling, with very few errors in the exclusion trials. This is because an error rate of 0 in the exclusion trials will produce a parameter estimation of 0 for familiarity (because $F = P(\text{Exclusion})/(1 - R)$).

Conceptually, the PDP has been criticized for its narrow definition of recollection. Within the PDP, recollection is defined as the ability to correctly identify the list membership of a previously seen stimulus. However, there may be many other aspects of the encoding event that a participant is able to retrieve, whether or not they are able to determine the list membership of a given face (Buchner, Erdfelder, Steffens, & Martensen, 1997). Mulligan and Hirshman (1997) distinguished between *diagnostic* recollection and *nondiagnostic* recollection. Diagnostic recollection includes the retrieval of information pertinent to the task. In the PDP, this would be the list membership of a stimulus. Nondiagnostic recollection could include any other information about the encoding event, which is not relevant to the task. This could include such information as thoughts, feelings, or associations evoked by the stimulus at encoding. The PDP only includes diagnostic recollection in the *R* estimate, meaning that recollection will be underestimated, and the *F* estimate will be contaminated by nondiagnostic recollection. Gruppuso, Lindsay, and Kelley (1997) also argued that the PDP only includes a narrow range of responses in the *R* parameter, meaning that the *F* estimate represents more than just a “completely undifferentiated feeling of oldness” (p. 259). Gruppuso et al. (1997) showed that the *R* and *F* estimates in the PDP are heavily influenced by the similarity of the two encoding lists. As similarity increases, making list discrimination more difficult, *R* decreases and *F* increases. Yet it does not seem logical to suggest that a participant's ability to recollect information about a stimulus decreases as list similarity increases.

Dodson and Johnson (1996) argued that the PDP views recollection as a controlled process which will always produce a correct response. Familiarity is viewed as an automatic process, which will produce either a correct response or an incorrect response, depending on task conditions. However, recollection may not always be the controlled all-or-nothing process it is represented as in the PDP (Mickes, Wais, & Wixted, 2009), and familiarity may not always be automatic and uncontrollable. Dodson and Johnson (1996) suggested that the source monitoring framework provides a much more flexible and inclusive way of thinking about memory processes than the PDP, and that source monitoring tasks produce data which is easier to interpret and to apply to real world behavior.

Despite these criticisms, the PDP can still provide a useful insight into the memory processes involved in recognition. The parameter estimates themselves should perhaps be treated with some caution, as they may not reflect the "true" contributions of recollection and familiarity to recognition, due to measurement error and biases in the data. However, the patterns of R and F estimates for different types of stimuli, and across different conditions, still provide valuable information on the ways in which memory processes are used in recognition.

Dual-process theorists have identified several factors with dissociable effects on recollection and familiarity. One such factor is depth of processing. The levels-of-processing (LOP) framework was introduced by Craik and Lockhart (1972), who suggested that perception involves a series of complex stages which become progressively "deeper". Upon perceiving a visual stimulus, for example, early visual areas of the brain detect properties of the object such as color, contrast, angles, and lines. Further along the perceptual pathway, this input will be compared to real world knowledge in order to interpret the stimulus in a meaningful way. At deeper stages, associations may be formed between the stimulus and other items stored in memory. Experimentally, depth of processing is usually manipulated by giving participants instructions which encourage either shallow processing or deep processing. In face recognition studies, "shallow" instructions often ask participants to make judgments about facial features, and "deep" instructions ask participants to make personality judgments about the people depicted (e.g. Patterson & Baddeley, 1977; Winograd, 1981).

Deep processing instructions often produce higher recognition accuracy than shallow processing instructions (e.g. Berman & Cutler, 1998; Bower & Karlin, 1974). Dual-process theory predicts that LOP instructions should greatly influence recollection but should influence familiarity to a much smaller degree. Deeper or more elaborate processing provides participants with more cues for recollection; for example, they may be able to retrieve particular thoughts and associations which were elicited by the stimulus at encoding. Familiarity, on the other hand, should not be affected by depth of processing, as this more automatic process stems from the perceptual fluency associated with a stimulus (Yonelinas, 2002). Some studies have supported this hypothesis (e.g. Gardiner, Java, & Richardson-Klavehn, 1996; Mulligan & Hirshman, 1995), while others have found that deep processing improves familiarity as well as recollection (Brown & Lloyd-Jones, 2006; Konstantinou & Gardiner, 2005).

Sporer (1991) compared the recognition accuracy of participants who were given deep or shallow encoding instructions with control participants who were given no instructions about encoding strategy. A typical LOP effect was found, with higher accuracy in the deep condition than in the shallow condition. However, the deep instructions provided no extra benefit over the control condition. Sporer (1991) therefore argued that LOP effects are caused by the inferiority of shallow instructions rather than the superiority of deep instructions. Including a control condition in which no processing instructions are given can therefore provide some insight into how participants "naturally" process stimuli within the confines of a recognition task. This is something that few LOP studies have done, with most including only shallow and deep conditions (e.g. Bower & Karlin, 1974; Bloom & Mudd, 1991; Patterson & Baddeley, 1977).

Socio-cognitive explanations of the ORB propose that when an out-group face is encountered, an automatic categorization process is triggered (Ito & Urland, 2005; Sporer, 2001b; Zhao & Bentin, 2008). Faces categorized as out-group members may then be processed using a shallow encoding strategy, as motivation to individuate may be low (Ackerman et al., 2006; Hugenberg, Miller, & Claypool, 2007). Chance and Goldstein (1981) asked White participants to write down the first thought they had upon seeing a series of White, Black, and Japanese faces. The content of the participants' responses suggested that White faces were spontaneously processed more deeply than the Black or Japanese faces.

In this paper, two experiments are reported which use the PDP to investigate the roles of recollection and familiarity in the ORB. White participants' memories of White and Black faces are tested in inclusion and exclusion trials, from which *R* and *F* parameters are estimated. The main predictions are that participants will falsely identify more Black lures than White lures in the exclusion trials, and that the estimated *R* parameter will be higher for White faces than for Black faces. In Experiment 2, depth of processing is manipulated. Participants are randomly allocated to a shallow processing condition, a deep processing condition, or a control condition, with no encoding instructions. As depth of processing increases, the estimated R parameter should also increase; the estimated F parameter should not be influenced by depth of processing.

When deep or shallow encoding strategies are encouraged, performance may be very similar for White and Black faces. In the control condition, participants are free to adopt any encoding strategy they wish. Previous research (e.g. Chance & Goldstein, 1981; Hugenberg et al., 2007) suggests that under such conditions, participants adopt more shallow processing strategies for other-race faces than for own-race faces. However, when participants are more constrained in their processing strategies as a result of the task instructions, these “natural” processing differences might be overridden. That is to say, in the shallow condition, participants’ natural tendencies to process own-race faces deeply will be inhibited; and in the deep condition, participants’ tendencies to categorize and process other-race faces in a shallow manner will also be prevented. Thus, whether LOP effects occur through deep-superiority processes or through shallow-inferiority processes, one would predict a smaller ORB in the shallow and deep conditions than in the control condition.

Experiment 1

Method

Participants and Design

Fifty two White undergraduates participated for credit on an introductory psychology course. The mean age of the participants was 20 years; forty (77%) were female.

The experiment had a 2 x 2 repeated measures design. The first factor was race of face, with two levels, White and Black. The second factor, trial type, had two levels, inclusion and exclusion. The dependent measures were the proportions of "old" responses to targets, distractors, and lures; and the Recollection (*R*) and Familiarity (*F*) estimates.

Materials and Apparatus

Color photographs of one hundred and sixty male faces (80 White, 80 Black) were cropped in Adobe Photoshop to remove necks, shoulders, and backgrounds. These photographs were taken from two databases, which were of comparable high quality. All of the Black faces and approximately half of the White faces were taken from a South African database. However, the South African database did not contain a sufficient number of White males, so the remaining White faces were taken from an Australian database. Both of the databases had been set up specifically for research purposes, and all of the individuals had given permission for their images to be used in psychological research.[1]

None of the faces had glasses, jewellery, or distinguishing marks, and all had a neutral facial expression. For each person depicted, two separate photographs were prepared – one in a frontal pose (shown in the learning phase), and the other in a three quarter profile pose (shown in the testing phase). The experiment was created and run using E-Prime software.

---

[1] Due to copyright reasons, the face images cannot be reproduced here.

Procedure

The participants were tested in groups of up to five by a White experimenter, with each participant working individually at their own PC. The experimenter remained in the room throughout. The participants were told that they would be taking part in a face recognition experiment, and that they should pay close attention to all of the faces so that they could answer questions about them later on. They were not given any information about inclusion or exclusion trials, nor were they told that they would see faces from different racial backgrounds.

The experiment was divided into four blocks, each of which had a learning phase followed by a testing phase. For each participant, two blocks consisted of inclusion trials, and two consisted of exclusion trials. The order of the blocks was randomized for each participant. In each learning phase, participants saw two groups of 10 target faces – Group A and Group B. Faces in Group A appeared first on a screen with a red border. After a thirty second delay, the Group B faces were shown on a screen with a blue border. The colored borders were added to decrease the list similarity, as high similarity can lead to an underestimation of recollection (Gruppuso et al., 1997). The faces were presented sequentially for 10 seconds each, with a 1 second inter-stimulus interval. The order of the faces within each list was randomized for each participant.

The learning phase was followed by a thirty second delay, after which the testing phase began. In the inclusion trials, participants were told to say that any face that they recognized was old, regardless of whether the face had appeared in Group A or Group B, while unfamiliar faces should be classed as new. In the exclusion trials, participants were told only to identify faces from one particular group (e.g. Group A) as old, and to identify faces from the other group (e.g. Group B) as well as unfamiliar faces as new. The participants responded by pressing 1 for old or 0 for new. In each block, 20 distractors appeared along with the 20 faces from the learning phase. The faces were shown sequentially in a random order and remained on the screen until the participant had made a response.

The experiment was blocked by race of face, so that participants couldn't use race as a cue to list membership. Each participant took part in one inclusion and one exclusion block for both White and Black faces. In the exclusion blocks, the target group (Group A or B) and the lure group (Group A or B) were counterbalanced between participants. Which faces appeared as targets and which appeared as distractors was also counterbalanced between participants.

Results and Discussion

The mean proportions of trials in which participants said old are shown in Table 2, broken down by trial type (inclusion or exclusion), stimulus type (target, foil, or lure), and race of face. The table indicates that participants made correct old responses to White targets more frequently than Black targets. Conversely, they made more incorrect old responses to Black distractors and lures than to White distractors and lures. Repeated measures t tests were conducted to confirm these observations. The effect size Cohen's d is reported for all comparisons (the formulae for calculating d in repeated measures and between groups designs are given in Appendix 1).

The differences were significant in all cases, and the effect sizes were large. Participants said old more often to White targets than Black targets in both the inclusion trials, $t(51) = 4.87$, $p < .001$, $d = .75$, and the exclusion trials, $t(51) = 4.19$, $p < .001$, $d = .62$. Conversely, participants falsely accepted Black distractors as old more often than White distractors in both the inclusion trials, $t(51) = 8.80$, $p < .001$, $d = 1.18$, and in the exclusion trials, $t(51) = 4.83$, $p < .001$, $d = .79$. Participants also identified Black lures as old more often than White lures, $t(51) = 3.69$, $p = .001$, $d = .64$.

Table 2: *Mean proportion of old responses to White and Black faces in inclusion trials (top) and exclusion trials (bottom) in Experiment 1.*

	Mean		Standard Deviation	
Stimulus type	White	Black	White	Black
	Inclusion			
Targets	.68	.57	.15	.13
Distractors	.26	.41	.14	.13
	Exclusion			
Targets	.58	.48	.15	.18
Distractors	.22	.33	.14	.13
Lures	.32	.42	.13	.18

These results support previous findings in the ORB literature. The proportions of old responses to targets are akin to hit rates in standard old/new recognition tasks, which are often higher for own-race faces than for other-race faces (e.g. Meissner et al., 2005; Teitelbaum & Geiselman, 1997). Old responses to distractors are akin to false alarm rates in old/new recognition tests, which are often higher for other-race faces than for own-race faces (e.g. Chiroro & Valentine, 1995; Meissner et al., 2005; Slone et al., 2000). The lures in the PDP form a category of stimuli which are not included in standard recognition tests, as they are previously seen faces which are to be ignored. The results of this study suggest that participants have greater difficulty in recalling the list membership of Black faces than White faces, and so they mistakenly identify a larger number of Black lures than White lures.

The proportion of targets correctly identified as old in the inclusion trials (P(Inclusion)) and the proportion of lures incorrectly identified as old in the exclusion trials (P(Exclusion)) were used to derive estimates for the Recollection (R) and Familiarity (F) parameters using the following formulae: $R = P(\text{Inclusion}) - P(\text{Exclusion})$; $F = P(\text{Exclusion})/(1 - R)$. The mean estimates are shown in Table 3, broken down by race of face. The table shows that the estimated R parameter for White faces was more than twice the estimated R parameter for Black faces. The F parameter, however, is very similar for White and Black faces. These observations were confirmed in repeated measures t tests. The R estimate was significantly higher for White faces than for Black faces, $t(51) = 6.62$, $p < .001$, $d = 1.05$, whereas the F estimate was not significantly different for White and Black faces, $t(51) = 0.78$, $p = .44$, $d = .13$.

Table 3: *Mean estimated recollection (R) and familiarity (F) parameters for White and Black faces in Experiment 1.*

	Mean		Standard deviation	
Parameter	White	Black	White	Black
R	.36	.15	.19	.20
F	.50	.48	.18	.15

As predicted, participants showed a specific impairment in recollecting other-race faces. Participants made more errors with Black lures than with White lures, and also correctly identified fewer Black targets than White targets, which lead to substantially higher *R* estimates for White faces than for Black faces. However, participants were no more or less likely to use familiarity to recognize own-race faces compared to other-race faces. These results support those of Meissner et al. (2005), who found a significant difference in Remember responses but not Know responses in a Remember-Know task. Similarly, Marcon et al. (2009) found that Hispanic participants mistakenly identified repeated Black distractors more frequently than repeated Hispanic distractors. Using equations from the PDP, Marcon et al. (2009) estimated that participants were specifically impaired at recollecting other-race faces compared to own-race faces.

Experiment 2

Experiment 1 showed that White participants are less likely to use recollection to recognize Black faces than White faces. Experiment 2 introduces an LOP manipulation to the process-dissociation procedure. Participants are allocated to one of three conditions, with instructions designed to encourage shallow processing, deep processing, or with no instructions about processing strategy. LOP instructions influence recollection more than familiarity (Gardiner et al., 1996; Yonelinas, 2002). Therefore, the largest effects of the LOP instructions should be on the lure response rates and the *R* estimates. Participants in the shallow condition should identify more lures than participants in the deep condition, leading to lower *R* estimates. The participants in the control condition are likely to be using a range of different encoding strategies, and so their performance may fall somewhere between the two experimental conditions. The effect size for the difference in recollection between White and Black faces should be smaller in the shallow and deep conditions than in the control condition, as participants' natural tendencies to process own-race faces deeply but to categorize other-race faces will be constrained.

Method

Participants and Design

A power analysis was conducted to determine the appropriate sample size for Experiment 2. In order to replicate the difference in the estimated R parameter for White and Black faces, the effect size (d = 1.05) of this difference in Experiment 1 was used. The power analysis showed that, in order to achieve 95% power with an α level of .05, 17 participants were required per condition. Sixty White undergraduates took part for credit on an introductory psychology course or for £4. The mean age of the participants was 21 years; forty six (77%) were female.

The experiment had a 2 x 2 x 3 mixed design. The first factor, race of face, had two levels, White and Black. The second factor, trial type, had two levels, inclusion and exclusion. The final factor, LOP condition, had three levels, shallow, deep, and control (n = 20 in each condition). The first two factors were manipulated within participants, and the third factor was manipulated between groups. As in Experiment 1, the dependent measures were the proportions of "old" responses to targets, distractors, and lures; and the Recollection (R) and Familiarity (F) estimates.

Materials and Apparatus

The stimuli were identical to those used in Experiment 1.

Procedure

The procedure was similar to that in Experiment 1, but with the addition of new instructions in the shallow and deep conditions. Participants in these conditions answered one question about each face as it appeared during the encoding phase. A total of three different questions were used in each condition, and they were randomly assigned to the target faces. The participant could make their response at any time while the face was on screen. The face remained on screen for the full 10 seconds, regardless of when the participant responded to the question. All of the questions were answered on a 1-5 scale, which appeared on the screen below each face. Participants in the control condition were just told to pay attention to all of the faces, as they would be asked questions about them later. These instructions were identical to those given in Experiment 1.

The questions in the shallow condition were selected to avoid drawing attention to physiognomic differences between White and Black faces, such as nose width and lip fullness. Some characteristics, such as eye and hair color, vary more widely in White faces than in Black faces. To avoid tuning participants' attention to racial differences, the shallow questions only ask about features which do not systematically vary between White and Black faces, and which are equally likely to vary within both racial groups. The three questions used were: "How round is the chin?", "How far apart are the eyes?", and "How thick are the eyebrows?".

In the deep condition, care was taken to avoid questions which might make participants feel uncomfortable. Personality characteristics with very obvious polar positive and negative opposites (such as intelligent vs. unintelligent) were avoided, as participants may have objected to making such judgments about minority race faces. Instead, personality traits with less obvious positive and negative associations were selected. The three questions used were: "How outgoing is this person?", "How spontaneous is this person?", and "How competitive is this person?".

Results and Discussion

The mean proportions of old responses for targets, distractors, and lures were analyzed in separate 2 (race of face) x 3 (LOP instructions) mixed ANOVAs. Table 4 shows the group means, collapsed across LOP instructions, but broken down by race of face. The pattern of means for targets and distractors appear similar to those found in Experiment 1, with more old responses to White targets than to Black targets, but more old responses to Black distractors than to White distractors.

Table 4: *Mean proportion of "old" responses to White and Black faces in inclusion trials (top) and exclusion trials (bottom) in Experiment 2.*

Trial type	Mean		Standard Deviation	
	White	Black	White	Black
		Inclusion		
Targets	.61	.57	.16	.14
Distractors	.28	.40	.16	.11
		Exclusion		
Targets	.54	.51	.16	.18
Distractors	.24	.29	.12	.13
Lures	.44	.41	.18	.17

The ANOVAs confirmed these observations. Participants made slightly more old responses to White targets than Black targets in the inclusion trials, $F(1, 57) = 3.79$, $p = .06$, $d = .27$, but not in the exclusion trials, $F(1, 57) = 1.24$, $p = .27$, $d = .18$. Black distractors were incorrectly identified as old more often than White distractors in both the inclusion trials, $F(1, 57) = 29.88$, $p < .001$, $d = .87$, and the exclusion trials, $F(1, 57) = 7.28$, $p = .01$, $d = .40$.

Table 5 shows the mean proportions of old responses to lures, broken down by race of participant and LOP condition. The ANOVA confirmed that the lure responses differed across the LOP conditions, $F(2, 57) = 4.60$, $p = .01$, $\eta_p^2 = .14$. Participants in the shallow condition identified a higher proportion of lures (*M* = .49, *SD* = .12) than participants in the control condition (*M* = .39, *SD* = .12), *t*(38) = 2.78, *p* = .008, *d* = .28, or participants in the deep condition (*M* = .40, *SD* = .12), *t*(38) = 2.50, *p* = .02, *d* = .25. For both White and Black faces, the mean number of errors was smaller in the deep condition than in the shallow condition; however, the magnitude of this difference appears to be greater for Black faces than for White faces. Despite these apparent differences between White and Black faces, neither the main effect of race nor the interaction term were significant (main effect of race: *F*(1, 57) = 1.00, *p* = .32, *d* = .17; interaction: *F*(2, 57) = 0.48, *p* = .62, η_p^2 = .02).

Table5: *Mean proportion of old responses to lures, broken down by race of face and LOP condition in Experiment 2.*

	Mean			Standard deviation		
Race	Shallow	Control	Deep	Shallow	Control	Deep
White	.49	.40	.43	.16	.16	.21
Black	.49	.38	.36	.14	.18	.16

The LOP effects were confined to the identification of lures. More lures were identified in the shallow condition than in the control or deep conditions, suggesting that shallow encoding instructions impair participants' abilities to recollect information concerning the list membership of familiar faces at testing. The LOP instructions did not influence responding for targets or for distractors (maximum *F*(2, 57) = 2.28).

Using the same formulae as in Experiment 1, the *R* and *F* parameters were estimated for White and Black faces in the three LOP conditions. The mean estimates are shown in Table 6, broken down by race of face and LOP instructions. The table shows that *R* appears to increase as the LOP instructions become deeper, whereas *F* remains relatively constant across the LOP conditions. Further, the pattern of means for the *R* estimate appears to be different for White and Black faces. For White faces, there is a sharp increase in *R* between the shallow and control conditions, with the mean *R* estimate in the control and deep conditions almost identical. For Black faces, the increase in *R* seems to be more gradual, with the parameter estimate in the control condition falling between the estimates in the shallow and control conditions. To test these patterns, *R* estimates across the three LOP conditions were compared separately for White and Black faces, in one way ANOVAs.

Table 6: *Mean estimated R (top) and F (bottom) parameters for White and Black faces, broken down by LOP condition in Experiment 2.*

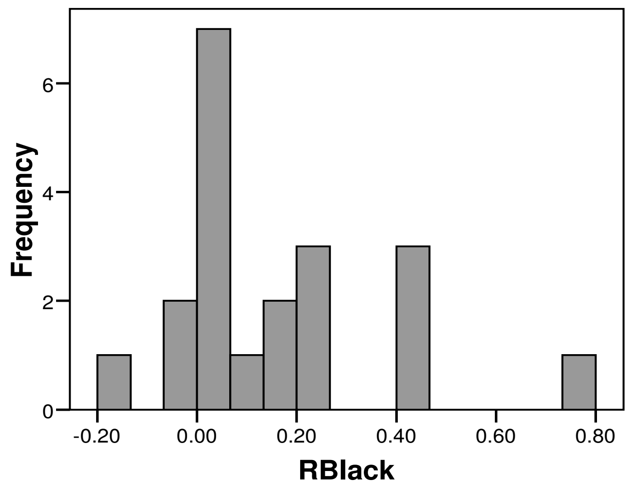
	Mean			Standard deviation		
Race	Shallow	Control	Deep	Shallow	Control	Deep
			Recollection (*R*)			
White	.06	.22	.22	.19	.17	.19
Black	.08	.15	.25	.12	.22	.20
			Familiarity (*F*)			
White	.53	.51	.52	.16	.15	.18
Black	.53	.43	.47	.12	.17	.13

For White faces, the ANOVA confirmed that the mean estimated *R* parameter varied across the LOP conditions, $F(2, 57) = 4.98$, $p < .001$. Post hoc *t* tests (with a Bonferroni corrected α of .02) showed that the mean *R* estimate in the shallow condition (M = .06, SD = .19) was lower than the mean estimate in the control condition (M = .22, SD = .17), $t(38) = -2.81$, $p = .008$, $d = .28$, and the deep condition (M = .22, SD = .19), $t(38) = -2.61$, $p = .01$, $d = .26$. Participants in the control condition appear to be using deep encoding strategies. When participants were prevented from using a deep processing strategy in the shallow condition, their ability to recollect White faces was impaired.

The R parameter also varied across the LOP conditions for Black faces, $F(2, 57) = 4.24$, $p = .02$. The R estimate was lower in the shallow condition (M = .08, SD = .12) than in the deep condition (M = .25, SD = .20), $t(38) = -3.26$, $p = .002$, $d = .33$. The estimate in the control condition (M = .15, SD = .22) did not significantly differ from the estimate in the shallow condition, $t(38) = 1.34$, $p = .19$, $d = .13$, or in the deep condition, $t(38) = -1.43$, $p = .16$, $d = .14$. This makes it difficult to draw any inferences about encoding strategies for Black faces in the control condition. Participants may have been using a variety of different processing styles, or they may have been using some medium-depth processing strategy. Figure 4 shows the distribution of R estimates for Black faces for participants in the control condition. If participants were using two or more distinct processing styles, then one might expect to see a bimodal or multimodal distribution. However, there is quite clearly a single peak in the distribution, which suggests that this is unlikely.

The ANOVAs confirmed that the LOP instructions did not influence the estimated F parameter for either White faces, $F(2, 57) = 0.13$, $p = .88$, or Black faces, $F(2, 57) = 2.43$, $p = .10$. Previous research has shown that LOP instructions influence R estimates but not F estimates in dual-process investigations (Yonelinas, 2002). Familiarity stems from the fluency with which a stimulus is perceived, which should not be affected by encoding strategy. These results therefore support the existing literature, that depth of processing is less important for familiarity than for recollection.

Figure 4: Histogram showing the distribution of R estimates for Black faces for participants in the control condition in Experiment 2.



Repeated measures *t* tests were run to compare the estimated *R* parameters for White and Black faces in each LOP condition. The prediction was that *R* would be significantly higher for White faces than for Black faces in the control condition, but that the differences in the shallow and deep conditions would be smaller and non-significant. The *t* tests confirmed that the *R* estimates were not significantly different for White and Black faces in either the shallow condition, $t(19) = 0.27$, $p = .79$, $d = .13$, or the deep condition, $t(19) = 0.40$, $p = .70$, $d = .14$. However, contrary to predictions, the difference in the *R* estimates in the control condition was also non-significant, $t(19) = 1.29$, $p = .21$, $d = .36$. However, the means in the control condition were in the predicted direction (*M* White = .22, *SD* = .17; *M* Black = .15, *SD* = .22), and the effect size was larger than those found in the shallow or deep conditions.

There may therefore have been a problem with statistical power in this study, with too few participants in each condition to detect a significant effect. In support of Experiment 1, the estimated *F* parameters did not differ for White and Black faces in any of the LOP conditions (for all comparisons, $t(38) \leq 1.63$, $p \geq .20$, $d \leq .50$).

The results of Experiment 2 support the hypothesis that depth of processing influences recollection to a greater degree than familiarity. Participants in the shallow condition identified more lures than participants in the deep condition, which led to a lower mean estimated *R* parameter. Participants in the control condition performed very similarly to participants in the deep condition with White faces. This suggests that the control participants were using deep encoding strategies for own-race faces. The mean performance of control participants for Black faces fell somewhere between the means for the shallow and deep conditions, making it difficult to make any solid claims about processing strategies. However, the unimodal distribution of *R* estimates suggests that there were not two or more distinct strategies being used.

General Discussion

Summary of race effects

In two experiments, the roles of recollection and familiarity in recognizing own- and other-race faces were explored using Jacoby's (1991) process-dissociation procedure (PDP). White participants studied White and Black faces, presented in two separate study lists, and took part in inclusion trials and exclusion trials. The mean proportion of old responses to targets, distractors, and lures (exclusion trials only) were compared for White and Black faces. In both studies, participants correctly identified more White targets as old than Black targets. Black distractors were also identified as old more often than White distractors. This pattern of results replicates the mirror effect often found in ORB studies, with lower hit rates and higher false alarm rates for other-race faces compared to own-race faces (Meissner et al., 2005; Teitelbaum & Geiselman, 1997).

In Experiment 1, participants misidentified more Black lures as old than White lures. These errors are similar to the repetition errors seen in Marcon et al.'s (2009) study, and they suggest that participants have difficulty retrieving episodic information in order to discriminate the list membership of familiar other-race faces. This effect was not replicated in Experiment 2, however, with no overall difference in lure identification rates between White and Black faces. As a result, the large difference in the estimated R paramaters found for White and Black faces in Experiment 1 was not found in Experiment 2.

In both experiments, however, the estimated *F* parameters were similar for White and Black faces. This supports previous findings by Meissner and colleagues (Marcon et al., 2009; Meissner et al., 2005), who found a dissociation between recollection and familiarity for own- and other-race faces. These results indicate that people are able to recognize other-race faces on the basis of familiarity. Reduced recollection and preserved familiarity in recognizing other-race faces could well lead to the pattern of results often found in ORB studies, with lower hit rates and higher false alarm rates for other-race faces than for own-race faces (Teitelbaum & Geiselman, 1997).

The inconsistencies between Experiments 1 and 2 may have been produced by the LOP manipulation in Experiment 2. Although there were no significant interactions between race of face and LOP condition for any outcome measure, the decreased accuracy rates in the shallow condition may have obscured any differences between White and Black faces. Table 6 shows that the mean *R* estimates in the shallow and deep conditions were very close for both groups of faces. It is in the control condition that the difference is beginning to emerge. With a larger sample size, this difference may have reached significance. However, the power analysis conducted prior to the data collection showed that this difference should have been detected with greater than 95% power with the current sample size.

Summary of LOP effects

The LOP instructions in Experiment 2 influenced the mean proportions of lure misidentifications, with more errors in the shallow condition than in either the control or deep conditions. The means in the control and deep conditions were almost identical, suggesting that there was a shallow-inferiority effect present rather than a deep-superiority effect. This supports Sporer's (1991) argument that within intentional learning paradigms, faces are processed deeply, and that shallow encoding instructions reduce accuracy. There were no LOP effects for targets or distractors, however. Depth of processing instructions did not influence participants' abilities to correctly identify previously seen faces. Nor did they influence the rates at which participants falsely identified previously unseen faces. LOP effects therefore seem to be confined to responses which require the rejection of a familiar stimulus using specific information about the encoding event (list membership).

This was confirmed in the analysis of the estimated *R* and *F* parameters. In line with previous research (e.g. Gardiner et al., 1996; Mulligan & Hirshman, 1995), LOP instructions had dissociable effects on recollection and familiarity. Whereas *R* increased as the depth of processing increased, *F* remained constant across the LOP conditions. As predicted, *R* estimates for White and Black faces were very similar in the shallow and deep conditions, with a difference beginning to show in the neutral condition. With a larger sample size, this difference may have reached significance.

The pattern of mean *R* estimates was different for White faces and Black faces. For White faces, there was a clear shallow-inferiority effect, with a lower *R* in the shallow condition than in either the control or deep conditions. However, the *R* estimate for Black faces in the control condition fell somewhere between the estimates in the shallow and deep conditions. It seems, therefore, that although racial categorization does reduce accuracy for out-group compared to in-group faces, some degree of accuracy is maintained over and above that found when featural processing strategies are encouraged. Some authors (e.g. Levin, 2000; Sporer, 2001) have argued that certain features act as “racial markers” (e.g. skin tone), which automatically cause the face to be categorized as an out-group member, and to therefore be processed in a shallow manner. In line with this account, participants have been found to rely more on featural processing than on configural processing for other-race faces (e.g. Michel, Corneille, & Rossion, 2006; Tanaka, Kiefer, & Bukach, 2004). However, in Experiment 2, participants who were explicitly told to attend to specific facial features performed less accurately with other-race faces than the control participants, who were left to use more natural encoding strategies. The difference between these two groups (though statistically non-significant) does seem to suggest that featural processing per se may not be the cause of the ORB in the control participants. However, the features which were used in the shallow task here (chin shape, eye separation, eyebrow thickness) were chosen specifically to avoid artificially inducing categorization effects. Perhaps if features which are more indicative of group membership (e.g. skin tone, lip fullness, nose width) a different pattern of results would have emerged.

Limitations

The studies presented here contained samples of White participants only. Ideally, similarly sized samples of Black participants would also have been recruited. Testing participants from only one race can raise concerns about stimulus sampling (Wells & Olson, 2001). For example, if the sample of Black faces were less variable and discriminable than the sample of White faces, then apparent cross-race differences may have been artifacts of the stimulus selection. However, as the size of the stimulus set increases, stimulus sampling concerns should decrease, as any idiosyncrasies of individual faces will have less influence on the overall results (Sporer, 2001a). With eighty faces from each group, the likelihood of stimulus sampling errors posing major concerns for these data seems small.

Furthermore, full crossover effects are not always found in ORB research (e.g. Ferguson, Rhodes, Lee, & Sriram, 2001; Wright et al., 2003). When crossover interactions are found, they are often asymmetric, with White participants typically showing larger effects than participants of other races (e.g. Jackiw, Arbuthnott, Pfeifer, Marcon, & Meissner, 2008; Meissner et al., 2005; Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005). There are several reasons why these asymmetries could occur, including a priori differences in the variability of the two stimulus sets. However, social psychological research on intergroup relations can also inform our understanding of these recognition effects. Minority groups are often lacking in social status and in power, and so are reliant upon the powerful majority group in their everyday life (Islam & Hewstone, 1993a). As a result, they may come to perceive the out-group as heterogeneous, avoiding many of the attributional biases and stereotypes subscribed to by members of majority groups (Hewstone & Ward, 1993; Islam & Hewstone, 1993b). Anthony, Copper, and Mullen (1992) argued that the numerical infrequency of out-group members in the environment of a majority group member encourages prototypical representation, which leads to perceptions of homogeneity. For minority group members, the numerical frequency of out-group members in the environment is high, encouraging exemplar-based representations, and reducing perceptions of homogeneity. Therefore, while one might expect the finding of Paper 1 to replicate with members of other groups, such effects may well be smaller and more elusive.

The results of these experiments should be interpreted with the criticisms of the PDP in mind. The validity of the independence assumption made in the PDP has been called into question, and violations of this assumption can create artifactual distinctions between the *R* and *F* estimates. In particular, the *F* parameter can be decreased as the *R* parameter increases. However, this did not occur in Experiment 2. Depth of processing increased *R* estimates, but *F* estimates remained constant.

The PDP includes only a narrow range of responses in the *R* estimate, as recollection is defined as the ability to remember the list membership of a previously seen stimulus. This leads to an underestimation of *R* and an overestimation of *F*. However, the interest in these studies lies in the pattern of mean estimates across different stimulus types and conditions, rather than in the absolute values of the estimates. If the degree of under- and overestimation is similar for both groups of faces, then the overall patterns of means will not be affected.

Evidence from the PDP should be combined with convergent evidence using different paradigms, such as source monitoring tasks, and Tulving's (1985) Remember-Know procedure. Using multiple methods will provide a much clearer picture of how memory is really operating when recognizing faces from other racial groups. Paper 2 will use a task designed to test participants' memories of contextual information associated with own- and other-race faces, in search of further support for the dissociation between recollection and familiarity in the ORB.

Conclusions

In support of previous research, White participants make greater use of recollection when recognizing White faces than when recognizing Black faces. Own-race faces produced higher hit rates and lower false alarm rates than other-race faces. The participants were more accurate at recalling the list membership of familiar White faces than familiar Black faces, at least in some situations. When deep encoding strategies were encouraged, the number of lure identifications for Black faces decreased, leading to an increase in the estimated R parameter. Conversely, encouraging shallow processing reduced the R estimate for both groups of faces. These results suggest that the encoding strategies adopted by participants in old/new recognition tasks might be different for own- and other-race faces. When the processing strategies are controlled, cross-race effects become much smaller.

Paper 2: Recollection Impairment for Other-Race Faces: A Question of Mental Effort at Retrieval?

Abstract

In two experiments, White participants studied White and Black faces, each of which was assigned one of two verbal labels. At testing, participants saw these faces again along with some novel faces. Participants were required to identify the faces which had appeared with one of the encoding labels (targets), and to reject the faces which had appeared with the other label (distractors). The results were consistent with a dual-process approach to the own-race bias in face recognition. In Experiment 3, the participants were unable to discriminate between Black targets and Black distractors, suggesting they did not retrieve recollective information about the Black faces. In Experiment 4, when the task was framed to encourage more careful consideration at testing, participants could successfully discriminate between Black targets and distractors. These results suggest that participants are generally less willing, rather than less able, to search for recollective evidence when identifying other-race faces.

Introduction

Dual-process theories posit that there are two independent routes to recognition – recollection and familiarity (see Yonelinas, 2002, for a review). Recollection involves the retrieval of specific information about the previous occurrence of an item. Familiarity, on the other hand, refers to the strength of memory activation created by an item. Although there are several variations of dual-process theory (e.g. Atkinson & Juola, 1974; Mandler, 1980; Mickes, Wais, & Wixted, 2009), the dominant view is that recollection is an all-or-nothing process, whereas familiarity is a continuous process, to which some decision criteria must be applied. Recollection is also believed to be a relatively slow and effortful process, which involves actively searching episodic memory, whereas familiarity is a fast and automatic process (Yonelinas, 2002).

In standard old/new face recognition tests, participants can be correct in two ways: They can identify a target (a hit) and they can reject a novel face (a correct rejection). They can also be incorrect in two ways: They can identify a novel face (a false alarm) and they can reject a target (a miss). In these tasks, recollection and familiarity act in concert, providing two sources of evidence upon which decisions can be based. However, Jacoby (1991) created a task that places the two processes in opposition, the process-dissociation procedure (PDP). Participants study two separate study lists, and then take part in inclusion trials and exclusion trials. Inclusion trials are similar to old/new recognition tests – participants must simply discriminate old items from new items, regardless of list membership. In exclusion trials, however, participants are required only to identify stimuli from one of the two study lists. One study list becomes the target list, and the other becomes the distractor list. This creates an additional type of correct response (rejection of distractors) and an additional type of incorrect response (identification of distractors).

When completing an exclusion task, participants can choose to use an effortful recollective search strategy, or they can choose not to use this effortful strategy, and they can instead rely on the relatively effortless familiarity process. The successful discrimination of targets from distractors requires a recollective search strategy, as one must recall-to-reject (i.e. one must recall seeing the stimulus in the task-irrelevant study list). Identifications of distractors therefore reflect familiarity without recollection.

Let us make the assumption that upon seeing a face during the testing phase, the participant will always attempt to search for any episodic information about a previous encounter with that face. If the face is a target, then a successful search will lead the participant to identify the face. If the face is a distractor, then a successful search will lead the participant to reject the face. If searches are unsuccessful (as they will always be with novel faces, excepting false recollection), then the participant could either decide to reject the face or to use an alternative decision strategy, such as familiarity.

Figure 5: The hypothetical familiarity distributions for targets, distractors, and novel faces.

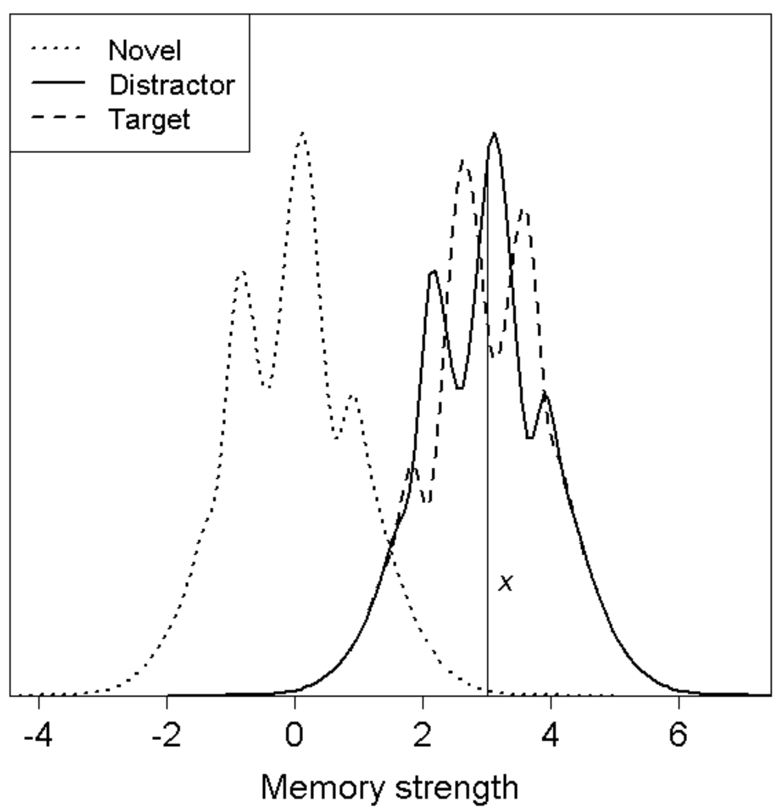


Figure 5 shows the hypothetical distributions of familiarity for targets, distractors, and novel faces. As in a standard recognition test, targets should be more familiar than novel faces, with some overlap in the distributions. Targets and distractors, however, should be approximately equal in their familiarity distributions, as all of the faces have been seen before. The two distributions will therefore overlap almost completely. No matter where a participant places their response criterion (for example, in position *x* on Figure 5), they will be unable to accurately distinguish between targets and distractors.

One can then predict that successful recollective searches will increase target identifications and decrease distractor identifications. As the success rate of these searches increases, the difference in identification rates of targets and distractors should also increase. If recollective searches are consistently unsuccessful, or are not attempted, then there should be no difference in target and distractor identifications. The size of the difference can therefore indicate how successfully participants are able to recollect studied items.

A robust finding in the face recognition literature is that participants are more accurate at recognizing faces of their own race than faces of other races (Meissner & Brigham, 2001). This effect has been termed the own-race bias (ORB). Most studies of the ORB have used old/new recognition tests (e.g. Corneille, Hugenberg, & Potter, 2007; Hugenberg, Miller, & Claypool, 2007; Malpass & Kravitz, 1969). In these tests, the ORB is often seen as a "mirror effect", with lower hit rates and higher false alarm rates for other-race faces. However, the effect size is often larger for false alarms than for hits. Signal detection measures of response bias show that participants respond more leniently to other-race faces than to own-race faces (Meissner & Brigham, 2001).

Recently, Meissner and colleagues (Evans, Marcon, & Meissner, 2009; Marcon, Susa, & Meissner, 2009; Meissner, Brigham, & Butz, 2005) have examined the ORB from a dual-process perspective. Using a Remember-Know task, Meissner et al. (2005) showed that participants make more Remember responses to own-race faces than to other-race faces. Participants are also more likely to misidentify repeated distractor faces of a different race than of their own race (Marcon et al., 2009). In Paper 1, recollection estimates from the PDP were higher for own-race faces than for other-race faces.

Here, two studies are reported, in which White participants see White and Black faces associated with one of two verbal labels. Participants then see both groups of faces again, along with a number of novel faces. Using a paradigm similar to Jacoby's (1991) exclusion trials, participants are asked to identify faces which were associated with one of the two verbal labels (targets), while rejecting faces which had been seen with the other verbal label (distractors), as well as novel faces.

The main prediction is that participants will successfully retrieve recollective information more frequently for the White faces than for the Black faces. As a result, the difference between identification rates for targets and distractors will be larger for the White faces than for the Black faces.

Experiment 3

Early studies of the ORB investigated racial attitudes as potential moderators of recognition accuracy (e.g. Brigham & Barkowitz, 1978; Carroo, 1987). Overall, little evidence was found linking participants' explicit attitudes to the magnitude of the ORB (Meissner & Brigham, 2001). More recent studies have used more sensitive techniques to measure attitudes and stereotypes. Findings have been inconsistent, with some but not others finding significant relationships between attitudes and recognition accuracy (Ferguson, Rhodes, Lee, & Sriram, 2001; Lebrecht, Pierce, Tarr, & Tanaka, 2009; Walker & Hewstone, 2008). Here, we explore whether “criminal” and “non-criminal” labels influence identification rates for White and Black faces. Participants assigned to one condition were asked to identify criminals and to reject non-criminals, while participants assigned to the other condition had the opposite task.

Previous research has shown that the information presented with a face can influence the ways in which the face is perceived, and the subsequent memory for that face. For example, Shepherd, Eliis, McMurran, and Davies (1978) presented participants with the photograph of a target, who they were told was either a murderer or a lifeboat captain. The participants rated the target more favourably along a range of physical and personality dimensions, including attractiveness and intelligence. After a short delay, the participants reconstructed the target face using Photofit. The resulting composites were rated by an independent group of judges, who rated the “murderers” to be less attractive and intelligent than the “lifeboat captains”. Shepherd et al. argued that the information distorted the participants’ perceptions of the faces, moving them towards their stereotypic norms for the categories of “murderer” and “lifeboat captain”.

Further evidence comes from Sporer (1989, Experiment 1), who presented participants with a group of faces, and asked them to decide which were murderers, and which were not criminals. In a later old/new recognition test, recognition accuracy was higher for the faces that participants had labelled as murderers than those they had labelled as innocent. Within a cross-race context, the work of Oliver, Jackson, Moses, and Dangerfield (2004) has shown that Black faces presented in the context of news reports concerning crime become distorted in memory to contain more Afrocentric features. Based on the aforementioned research, and the widely acknowledged stereotype of young Black males as criminals (Eberhardt, Goff, Purdie, & Davies, 2004), it is expected that participants will identify more criminals than non-criminals, and that this effect may be larger for the Black faces than for the White faces.

Method

Participants and Design

Fifty two White participants took part in this experiment, for either £3 or 30 minutes course credit. Forty three (83%) of the participants were female; ages ranged from 18 to 41 years, with a mean age of 21 years.

The experiment had a 2 x 2 x 3 mixed design. The first factor, condition, was manipulated between groups (n = 26 in each condition). In the "criminal" condition, participants were asked only to recognize faces that had been labeled as criminals. In the "non-criminal" condition, participants were asked only to recognize faces that had been labeled as non-criminals. The second factor, race of face (White or Black), varied within subjects. The third factor, stimulus type (target, distracter, or novel face), also varied within subjects.

The dependent variable was the mean proportion of faces identified.

Materials and Apparatus

Photographs of 36 White and 36 Black males were used in this experiment. All of the photographs were taken from the same high quality database (used also in Papers 1 and 4). The faces all had neutral expressions, and had no distinguishing marks, glasses, or jewellery. The photographs were cropped in Adobe Photoshop to remove necks and shoulders. All faces were then pasted onto a blank white background. For each face, two photographs in different poses were used: a frontal shot, and a three quarter profile shot. This ensures that participants make their recognition decisions based on the faces, rather than on any low-level features of the photographs (Bruce, 1982).

Procedure

Participants were tested in groups of up to three by a White experimenter. Each participant sat at an individual PC. The experimenter remained in the room throughout the experiment.

The experiment was split into two blocks, each with a learning phase followed by a testing phase. Pilot testing showed that two blocks were necessary for accuracy to be above chance (there were no significant block effects in any of the analyses). In each learning phase, participants saw twenty four faces presented in a random order. Half of these faces were White, and half Black. For each race of face, half were labeled as criminals and half as non-criminals. This was denoted by the word "criminal" or "non-criminal" underneath the face. The face and the associated label remained on screen for five seconds. Participants were informed that they would be performing a recognition test, but were not given any instructions about whether they would be required to remember criminals or non-criminals. All participants were in the same condition across both blocks, and so had to identify either criminals or non-criminals both times. However, they were told that the second block may be different from the first block, and so they should pay attention to all faces in both encoding phases.

In order to encourage deeper processing of the faces (Craik & Lockhart, 1972), participants were asked to make a judgment for each face. For criminal faces, participants were asked to decide what crime that individual may have committed. For non-criminal faces, participants were asked to decide what occupation that individual may do. After the face disappeared, participants were given five seconds to write down their decision on a numbered piece of paper. Participants were asked to attempt a response for every face. However, on trials where they could not make a response, they were asked to leave a blank space on the paper.

Immediately following the learning phase, the testing phase began. Thirty six faces were shown sequentially in a random order. These included the twenty four faces from the learning phase plus twelve foils. The participants were not given any information on the ratio of old to new faces. Research by Sporer (1989, Experiment 1) suggests that faces labeled as criminals may be more memorable than faces labeled as non-criminals. To test for this effect, half of the participants were asked only to recognize criminals, and the other half were asked only to recognize non-criminals. So, for example, in the criminal condition, participants were asked "Was this person a criminal? Yes or no". They had to respond "yes" if they recognized a face as a criminal (target), and "no" if they either recognized a face as a non-criminal (distractor) or if they thought a face was new (novel face). In this way, three types of responses could be examined: identifications of targets (correct), identifications of distractors (incorrect), and identifications of novel faces (incorrect). There was no time limit on responding. Participants were told to take as long as they needed, and the face remained on screen until they had made their decision.

After completing the learning and testing phase of block 1, participants moved on to block 2. The order of the blocks was counterbalanced between participants. Which faces appeared as criminals, non-criminals, and novel faces were also counterbalanced. After completing the experiment, the participants were debriefed, and were informed that the labels had been randomly allocated to the faces.

It was stressed that the labels in no way indicated that the people depicted in the study were actually criminals.

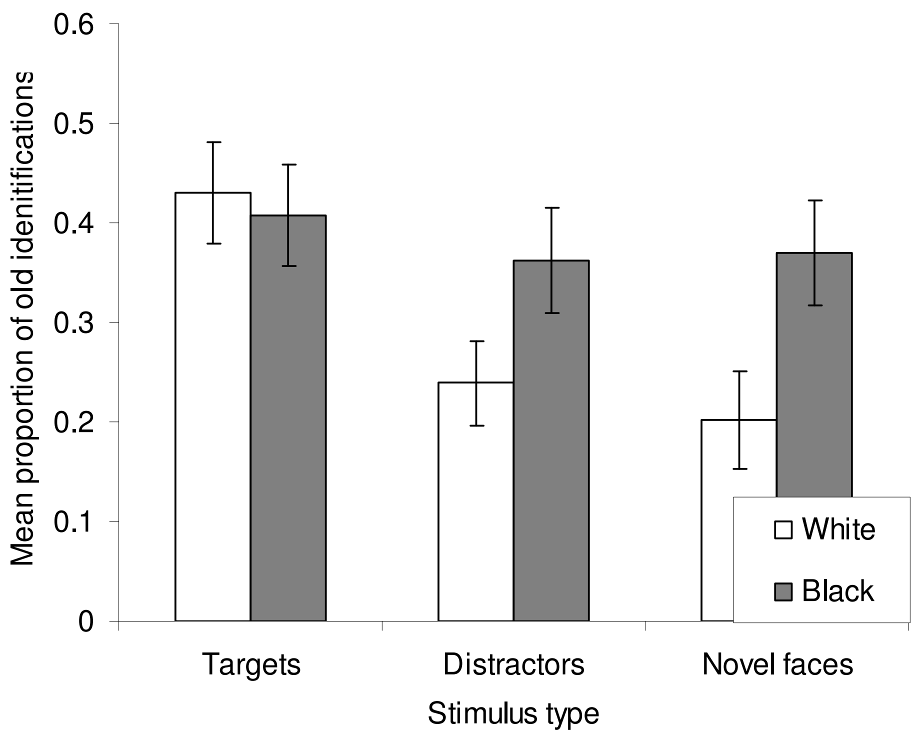
Results and Discussion

For each participant, the mean proportion of old identifications for White and Black targets, distractors, and novel faces was calculated. These proportions were analyzed in a 2 (race of face – White or Black) x 3 (stimulus type – target, distractor, or novel face) x 2 (condition – identify criminals or identify non-criminals) mixed ANOVA. There were no significant effects of condition in the analysis (maximum $F(2, 100) = 1.53$, $p = .22$, $\eta_p^2 = .03$). Therefore, the results will be discussed without reference to condition.

Figure 6 shows that the pattern of responses is different for White and Black faces. This observation was confirmed by a significant interaction term between race of face and stimulus type, $F(2, 100) = 14.50$, $p < .001$, $\eta_p^2 = .23$. The main prediction was that the difference in identification rates between targets and distractors would be larger for White faces than for Black faces. This hypothesis was confirmed. White targets ($M = .43$, $SD = .18$) were identified significantly more often than White distractors ($M = .24$, $SD = .15$), $t(51) = 6.69$, $p < .001$, $d = .1.13$, whereas Black targets ($M = .41$, $SD = .18$) and Black distractors ($M = .36$, $SD = .19$) were identified at similar rates, $t(51) = 1.63$, $p = .11$, $d = .24$. These results suggest that the participants made more successful recollective searches for White faces than for Black faces.

Identifications of novel faces were then compared to identifications of targets and distractors. For the White faces, novel faces were identified less often ($M = .20$, $SD = .18$) than targets, $t(51) = 7.14$, $p < .001$, $d = 1.26$, and distractors, $t(51) = 2.16$, $p = .04$, $d = .22$. For the Black faces, however, novel faces were identified at a similar rate (*M* = .37, *SD* = .19) to targets, $t(51) = 1.30$, $p = .20$, $d = .20$, and to distractors, $t(51) = 0.26$, $p = .80$, $d = .04$. These results suggest that the participants were successfully recollecting White faces on many occasions, increasing target identification rates and reducing distractor and novel face identification rates. Black faces were not successfully recollected, producing similar identification rates across response types.

Figure 6: Mean proportion of old identifications of White and Black targets, distractors, and novel faces in Experiment 3. Error bars represent 95% confidence intervals.



There were also significant main effects of race of face and stimulus type in the analysis. Overall, more identifications were made for Black faces (*M* = .38, *SD* = .14) than for White faces (*M* = .29, *SD* = .13), $F(1, 50) = 14.21$, $p < .001$, $d = .65$.

Identification rates varied across the different stimulus types, $F(2, 100) = 24.07$, $p < .001$, $\eta_p^2 = .33$. Targets were identified more often ($M = .42$, $SD = .14$) than distractors ($M = .30$, $SD = .13$), $t(51) = 5.98$, $p < .001$, $d = .87$, and novel faces ($M = .29$, $SD = .14$), $t(51) = 5.51$, $p < .001$, $d = .92$. Identification rates did not significantly differ for distractors and novel faces, $t(51) = 0.79$, $p = .43$, $d = .11$.

The results of Experiment 3 are consistent with the hypothesis that participants encode other-race faces in a manner which is less conducive for later recollection than own-race faces. White targets were identified significantly more often than White distractors, suggesting that the participants were using a recollective strategy to discriminate between the two stimulus types. Black targets, distractors, and novel faces were identified at similar rates, suggesting that the participants were not successfully recollecting the Black faces. The high identification rates of novel faces also suggest that participants adopted a lenient response criterion for non-recollected Black faces. There was no evidence of any stereotyping effects in Experiment 3, as there were no significant effects of condition.

Experiment 4

Experiment 3 showed that participants were less likely to recollect other-race faces than own-race faces. A question which arises from these data is whether participants were less *able* to recollect other-race faces, or whether they were less *willing* to attempt a recollection. The dominant view of recollection is that it is an effortful and controlled process (Yonelinas, 2002). Participants may have conserved cognitive effect by only attempting recollective searches for White faces, and not for Black faces.

In Experiment 4, participants were allocated to one of two conditions, with strict and lenient response instructions. The strict instructions encouraged participants to be concerned with the accuracy of their target identifications compared to distractor identifications; the lenient instructions encouraged participants to be concerned with the quantity of their target identifications. One can make different predictions for the influence of response instructions on identifications of Black faces, depending on whether participants are less *able* or less *willing* to use recollection. If the participants in Experiment 3 were attempting recollective searches for Black faces, but these searches were frequently unsuccessful, then the instructions should not influence the difference in Black target and distractor identification rates. If the participants in Experiment 3 were not attempting recollective searches for the Black faces, then the strict instructions should encourage the participants to make these attempts. Assuming that at least some of these searches will be successful, target identifications should increase, and distractor identifications should decrease.

In order to manipulate response bias, Experiment 4 was framed as a stop and search exercise. Racial disparities in police stop and search statistics have attracted political attention over recent years. The Race Equality Duty in the UK states that stop and searches based solely upon race, ethnicity, or religion are unlawful (Home Office Stop and Search Manual, 2005, p. 17). Similarly in the USA, the Attorney General's Office from the US Department of Justice issued a set of guidelines in 2003 which expressly prohibit the use of race in spontaneous law enforcement activities such as stop and search. Recent empirical research examining real world police practices has produced mixed results. For example, Petrocelli, Piquero, and Smith (2003) analysed traffic stop data in one US city, which showed that the number of traffic stops resulting in a search could be predicted by the percentage of Black people living in the neighborhood. Smith and Petrocelli (2001) analysed similar traffic stop data and found that Black and minority drivers were more likely to be stopped than

White drivers. However, once stopped, White drivers were more likely to be searched than Black drivers. Waddington, Stenson, and Don (2004) argue that stop and search statistics are in proportion with the population available to be searched, as young men of ethnic minorities are more likely to frequent public places in certain areas than Whites. Some of the inconsistencies in these studies may be due to methodological constraints, such as relying on self-reports of police officers, and missing data. Despite the ambiguities and controversies present in real-world stop and search data, the perception of unfairness towards minority groups persists in the public consciousness (Buerger & Farrell, 2002). Experiment 4 was couched in "stop and search" terms in order to manipulate response bias in a way which would hopefully be meaningful to participants.

The encoding phase was identical to that in Experiment 3. However, in the testing phase, participants were told to imagine that they were police officers, and that they would need to stop any individuals that they recognized as criminals from the encoding phase. In the strict condition, it was stressed that they should try not to stop many non-criminals; in the lenient condition, participants were told that it was very important to catch the criminals, and that stopping a few non-criminals did not matter.

Method

Participants and Design

Seventy two White participants took part in the experiment for course credit or £3. Sixty five (90%) of the participants were female; ages ranged from 18 to 30 years, with a mean age of 21 years.

The experiment had a 2 x 2 x 3 mixed design. The first factor, condition (strict or lenient), was manipulated between groups (n = 36 in each group). The second factor, race of face (White or Black), varied within subjects. The third factor, stimulus type (target, distractor, or novel face), also varied within subjects.

The dependent variable was the mean proportion of faces identified.

Materials and Apparatus

The stimuli and apparatus were identical to those used in Experiment 3.

Procedure

The encoding phase was identical to that in Experiment 3. In the testing phase, participants had to imagine that they were police officers "on the beat" and that they would need to "stop and search" anybody that they remembered as a criminal. Participants responded to each face by pressing 1 if they decided to stop and search, or 0 if they decided not to stop and search.

Across two conditions, response criterion was manipulated. In the lenient condition, participants were given the following instructions:

> "You should imagine that you are a police officer on the beat. There has been a spate of crimes in this particular area over the last few hours. You need to catch as many of the criminals from the learning phase as you can. For each face, you will be asked if you would like to stop and search the individual. Due to the recent crimes, catching the criminals is your top priority. Stopping a few non-criminals by mistake is a small price to pay for ensuring the safety of the area."

In the strict condition, participants received these instructions:

> "You should imagine that you are a police officer on the beat. It is a standard day, with crime levels no higher than usual. You need to catch as many of the criminals from the learning phase as you can. For each face, you will be asked if you would like to stop and search the individual. Although you want to catch

the criminals, you should also strive not to inconvenience any of the non-criminals by mistakenly stopping and searching them."

The experiment had two blocks, each with a learning phase and a testing phase.[2] Before the experiment began, participants were told that the two blocks may be the same or they may be different, so they would not be able to predict what they would need to do in the second block from what they did in the first block. This was to reduce the possibility that participants would only pay attention to the criminal faces in the second block. In reality, both blocks followed the same procedure for all participants.

Results and Discussion

The proportion of identifications of White and Black targets, distractors, and novel faces were entered into a 2 (race of face) x 3 (stimulus type) x 2 (condition) mixed ANOVA. The mean proportions of identifications are shown in Table 7. Overall, more identifications were made in the lenient condition (M = .45, SD = .09) than in the strict condition (M = .39, SD = .12), F(1, 70) = 5.69, p = .02, d = .57. This indicates that the instructions were successful in encouraging participants to shift their response criteria, although the increase in identifications was relatively small, at approximately six per cent. None of the interaction terms including condition were statistically significant (maximum F(2, 140) = 1.96, p = .15, η_p^2 = .03). The shifts in response criteria between the strict and lenient conditions therefore appear to be of a

---

[2] There was a small main effect of block number, F(1, 70) = 4.46, p = .04, d = .23, with slightly fewer identifications in block 2 (M = .40, SD = .12) than in block 1 (M = .43, SD = .12). None of the interaction terms including block number were significant (maximum F(2, 140) = 2.26, p = .11, η_p^2 = .03).

similar magnitude for White and Black faces, and for targets, distractors, and novel faces.

Table 7: *Mean proportions of identifications of White and Black targets, distractors, and novel faces in the strict (top) and lenient (bottom) conditions in Experiment 4.*

	Percentage			Standard deviation		
Race	Target	Distractor	Novel	Target	Distractor	Novel
			Strict			
White	.51	.35	.28	.18	.17	.18
Black	.50	.32	.35	.22	.17	.22
			Lenient			
White	.56	.37	.36	.17	.19	.20
Black	.53	.44	.42	.14	.15	.16

The interaction between race and stimulus type was marginally significant, $F(2, 140) = 2.59$, $p = .08$, $\eta_p^2 = .04$. White targets ($M = .53$, $SD = .18$) were identified significantly more often than White distractors ($M = .36$, $SD = .18$), $t(71) = 6.52$, $p < .001$, $d = .95$. This difference is similar to that found in Experiment 3, and again suggests that participants can use recollection to successfully distinguish between two groups of familiar own-race faces. Contrary to Experiment 3, the difference in target and distractor identification rates was also significant for Black faces (targets: $M = .52$, $SD = .19$; distractors: $M = .38$, $SD = .17$), $t(71) = 5.32$, $p < .001$, $d = .76$. Thus in Experiment 4, participants do seem to be showing some recollective ability for the Black faces.

For both White and Black faces, target identification rates were higher than novel face identification rates: White faces, $t(71) = 8.09$, $p < .001$, $d = 1.14$; Black faces, $t(71) = 6.61$, $p < .001$, $d = .72$. Distractor identification rates and novel face identification rates did not significantly differ for White faces, $t(71) = 1.70$, $p = .09$, $d = .22$, or for Black faces, $t(71) = .00$, $p = 1.00$, $d = .00$.

The main effect of stimulus type was significant in the analysis. Targets were identified more frequently ($M = .53$, $SD = .14$) than distractors ($M = .37$, $SD = .13$), $t(71) = 8.10$, $p < .001$, $d = 1.14$, and novel faces ($M = .35$, $SD = .15$), $t(71) = 10.28$, $p < .001$, $d = 1.19$. Identification rates for distractors and novel faces were not significantly different, $t(71) = 1.10$, $p = .28$, $d = .14$. This pattern of results is very similar to that seen in Experiment 3. However, unlike in Experiment 3, the main effect of race of face was not significant, $F(1, 70) = 1.50$, $p = .22$, $d = .18$, although the mean proportion of identifications was numerically higher for Black faces ($M = .43$, $SD = .14$) than for White faces ($M = .41$, $SD = .13$).

The results of Experiment 4 are generally consistent with the view that participants are able to use recollection when recognizing other-race faces, but that they tend to do so less often than for own-race faces. Although there were no significant interactions between condition and race of face, participants in Experiment 4 were able to discriminate between Black targets and Black distractors. It is possible that the instructions in both the strict and lenient conditions drew participants' attention to their decision making strategies, encouraging more effortful processing at testing. It is also possible that the "stop and search" framework influenced participants' retrieval strategies. These instructions may have encouraged participants to consider the real world consequences that identification decisions can sometimes have, therefore encouraging more careful and effortful responding.

General Discussion

In two experiments, White participants' abilities to discriminate between previously seen White and Black targets and distractors were examined. Participants studied a series of faces, each of which was seen with one of two verbal labels. These faces were later seen again, along with a group of novel faces. Using a procedure similar to the exclusion trials of Jacoby's (1991) process-dissociation procedure, the participants were asked only to identify the faces which had been studied with one of the two labels (targets). Faces which had been seen with the other label (distractors), and novel faces, were to be rejected. In order to accurately reject a distractor, participants must recollect seeing the face associated with the task-irrelevant label. Thus recollection and familiarity are placed in opposition.

Previous research has shown that recollection is impaired for recognizing other-race faces (Evans et al., 2009; Marcon et al., 2009; Meissner et al., 2005; Paper 1). Based on this research, it was predicted that the participants would have greater difficulty in distinguishing Black targets from Black distractors than in distinguishing White targets from White distractors. This hypothesis was confirmed in Experiment 3. The participants identified more White targets than White distractors, but identified Black targets and Black distractors at a similar rate. It also seemed that participants were adopting more lenient response criteria for Black faces than for White faces, an effect that is commonly found in cross-race recognition studies (Meissner & Brigham, 2001).

In Experiment 4, participants were allocated to one of two conditions designed to encourage strict or lenient responding. The experiment was framed as a "stop and search" exercise, in which the participants had to catch the criminals from the study phase. Participants in the strict condition were told to avoid stopping non-criminals, while participants in the lenient condition were told that catching the criminals was of the utmost importance, and that a few identifications of non-criminals were unavoidable. The aim of Experiment 4 was to find out whether participants could, in some circumstances, recollect the Black faces, or whether they were simply unable to recollect the Black faces.

The results of Experiment 4 were consistent with the hypothesis that participants can, when encouraged to, use recollection in their identifications of other-race faces. Participants in both the strict and lenient conditions were able to discriminate Black targets from Black distractors. Overall, identification rates were higher in Experiment 4 than in Experiment 3. The "stop and search" instructions may have given the participants in Experiment 4 a different view of the task compared to the participants in Experiment 3, who received more standard recognition test instructions. This "stop and search" framework might have made the participants more aware of the real world consequences that identification decisions can have, which then encouraged more careful and controlled responding. Further research using this method seems needed in order to establish why even participants in the strict condition made more identifications than participants in Experiment 3, and why the increase in identifications seemed to be largest for the target identifications.

In Experiment 3, half of the participants were asked only to identify criminals in the testing phase, while half were asked only to identify non-criminals. Previous research has shown that memory can be enhanced for face stimuli presented with crime-related information (Sporer, 1989), and that target faces can also become distorted in memory to become closer to stereotypical representations of "criminals" (Shepherd et al., 1978; Oliver et al., 2004). Based on this research, it was predicted that the participants who were asked to identify the criminals would perform more accurately than those who were asked to identify the non-criminals. However, no significant differences were found between the two groups.

There are many differences between the method used here and the methods used in previous research which could account for this failure of replication. Here, a large number of targets were presented from each group, creating a high demand on memory resources. In contrast, previous studies used a very small number of target faces – between one (Shepherd et al., 1978) and eight in total (Sporer, 1989). Furthermore, Sporer (1989) and Oliver et al. (2004) used quite strong manipulations of criminality. The criminal and non-criminal labels in Sporer's study were self-generated, while the contextual information in Oliver's study consisted of an entire news report. In comparison, brief verbal labels were used in Experiment 3, which also included a large lexical overlap, with both labels containing the label "criminal". As a result, the two stimulus lists were quite similar. High list similarity can decrease the accuracy of list discrimination (Gruppuso, Lindsay, & Kelley, 1997).

In conclusion, the results reported here provide evidence consistent with a dual-process approach to the ORB in face recognition. White participants had difficulty discriminating between familiar Black targets and distractors, but had no such difficulty discriminating between familiar White targets and distractors. However, when participants were encouraged to consider their response strategies, they were able to successfully distinguish between Black distractors and targets. This suggests that participants are not less able to use recollection when recognizing other-race faces, but that they are less willing to do so.

Paper 3: Memory for Context in Own- and Other-Race Identifications[3]

Abstract

People recognize faces from their own race more accurately than faces from other races. In three experiments, memory for context in own- and other-race identifications was investigated. White participants saw White and Black faces presented with different visual contexts at encoding. In an old/new recognition test, participants saw these target faces again along with some distractor faces. Following "old" decisions, participants were also asked to identify on which context the face had appeared during the encoding phase. In all three experiments, face recognition and context memory were more accurate for White faces than for Black faces. Remembering context is important in real world identifications, and failures in context memory can have serious consequences in eyewitness identifications. The results are discussed in a dual-process framework.

Introduction

People are more accurate at recognizing faces from their own race than faces from other races (Meissner & Brigham, 2001). This own-race bias (ORB) has been shown in participants from several races (Chiroro & Valentine, 1995; Evans, Marcon, & Meissner, 2009; Hancock & Rhodes, 2008; Jackiw, Arbuthnott, Pfeifer, Marcon, & Meissner, 2008; Sporer, Trinkl, & Guberova, 2007), and meta-analyses have shown the effect to be robust and reliable (Anthony, Copper, & Mullen, 1992; Bothwell

---

[3] Experiment 5 published in: Horry, R. & Wright, D. B. (2008). I know your face but not where I saw you: Context memory is impaired for other-race faces. *Psychonomic Bulletin & Review,* **15**, 610-614.

Brigham, & Malpass, 1989; Meissner & Brigham, 2001). The ORB has worrying implications when applied to eyewitness memory. Mistaken identifications by eyewitness can and do lead to miscarriages of justice (Scheck, Neufeld, & Dwyer, 2003), which may be more likely when the witness and the suspect are of different races (Doyle, 2001).

In day to day life, we encounter many individuals in different environmental contexts. Sometimes upon recognizing a person, we may be unable to recall any details of the situation in which we encountered that person (Mandler, 1980). The ability to recall such contextual information is known as source monitoring (Conway & Dewhurst, 1995); failures in source monitoring can lead to false memories being accepted as true memories (Johnson, 2006) and to familiar yet irrelevant items being incorrectly identified as targets (Jacoby, 1991). Source monitoring becomes particularly important in the eyewitness world, as a witness making an identification must be sure that the suspect was seen in a specific place (the crime scene) at a specific time (during the crime), carrying out a specific act (committing the crime). If the witness cannot recollect this contextual information about a familiar yet innocent suspect, they may mistakenly identify a bystander to the crime (Read, Tollestrup, Hammersley, McFazden, & Christensen, 1990; Ross, Ceci, Dunning, & Toglia, 1994), or an individual seen during a mugshot search (Deffenbacher, Bornstein, & Penrod, 2006).

Dual-process theories assume that recollection and familiarity arise from separate and functionally independent memory processes (Yonelinas, 2002). Recollection is the product of an episodic memory system, which allows an individual to remember contextual information about a recognized item. Familiarity is a more subjective sense of knowing that an item has been seen before based on the strength of activation produced by a presented item. Recollection is therefore an all-or-nothing process in which episodic information will either be retrieved or not, whereas familiarity is a continuous index of memory strength. Familiarity is believed to be a very fast and automatic process, whereas recollection is slower and more effortful (McElree, Dolan, & Jacoby, 1999).

Meissner, Brigham, and Butz (2005) suggested that the ORB could be caused by a specific impairment in the recollection process for other-race faces. Participants took part in an old/new recognition test with own- and other-race faces, and were also asked to report on the experiential basis of their memories, using the Remember-Know procedure (Tulving, 1985). Participants made more Remember responses to own-race faces than other-race faces, but Know responses were made at a similar rate to faces from both groups. Marcon, Susa, and Meissner (2009) also found that participants make more familiarity-based errors with other-race faces than with own-race faces. Here, these results were supported in Paper 1, as White participants made more exclusion errors for Black faces than White faces in a process-dissociation task, and the estimated recollection parameter was higher for White faces than Black faces.

Dual-process theorists argue that recollection is necessary for contextual information to be accurately remembered (Mandler, 1980). The impairment in the recollection process seen in other-race faces should therefore lead to deficits in remembering context. Across three studies, this hypothesis is tested in samples of White participants. At encoding, a series of White and Black target faces are shown, and each face appears on one of several backgrounds. At testing, the targets are seen again, along with a series of distractor faces. Participants make old/new decisions for the faces, followed by confidence judgments. Following an old judgment, the participants are also asked to identify on which background the face had appeared at encoding. In Experiments 5 and 6, the contexts used are complex visual scenes; Experiment 7 explores memory for simple contexts colored contexts.

In Experiment 6, the effects of context reinstatement on face recognition and context memory are explored. During the testing phase, the target faces are seen on the same background as at encoding, or on a different background. Some studies have found that reinstating encoding context improves recognition accuracy (Gruppuso, Lindsay, & Masson, 2007) while others have found no such improvement (Macken, 2002). When these effects are found, they are generally stronger for classes of stimuli which are less familiar (Russo, Ward, Geurts, & Scheres, 1999). Other-race faces form a less familiar class of stimuli than own-race faces, and so reinstating the encoding context may particularly benefit these faces. Context reinstatement appears to enhance recollective memory to a greater degree than familiarity-based memory (Gruppuso et al., 2007; Macken, 2002). This may reduce the size of the ORB in both face recognition and context memory, by increasing participants' abilities to recollect other-race faces.

Experiment 5

The aim of Experiment 5 is to explore how accurately White participants can remember contextual information about correctly identified faces from their own race and from a different race. White participants were previously shown to have difficulties in using recollection to identify Black faces compared to White faces (Paper 1, Paper 2). As recollection is believed to play a central role in memory for context (Mandler, 1980; Yonelinas, 2002), White participants may make more context errors for correctly recognized Black faces than for correctly recognized White faces.

Previous research has shown that stereotypic attitudes towards out-group faces can be modified by the context in which they are presented (e.g. Barden, Maddux, Petty, & Brewer, 2004; Wittenbrink, Judd, & Park, 2001). Wittenbrink et al. showed that automatic racial bias, as measured by the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) was reduced after watching a video showing Black actors in a positive context (a family barbecue). No such reduction occurred for participants who watched a video of Black actors in a negative context (a gang incident). Using static images as contexts, Barden et al. also showed that context can moderate racial bias. Faces were presented on backgrounds which were congruent either with positive aspects of racial stereotypes (e.g. Black faces on a basketball court, White and Asian faces in a classroom) or with negative aspects of racial stereotypes (Black faces in a prison context, White faces in a factory context). As predicted, when the contexts were positive-congruent, positive racial bias was found; when the contexts were negative-congruent, negative racial bias was found.

Experiment 5 uses four backgrounds similar to those used by Barden et al. (2004) – the basketball court, the factory, the classroom, and the prison. An additional gray screen background is used as a baseline control. The results of Barden et al. suggest that the basketball and prison contexts are congruent with positive and negative aspects of Black stereotypes, respectively, while the classroom and factory contexts are congruent with positive and negative aspects of White stereotypes. Congruency between a context and the stimulus in terms of stereotypical representations could then influence memory, such that memory is enhanced when the information is congruent.

Method

Participants and Design.

Forty six White participants took part in this experiment for £4. The mean age of the participants was 27 years (range 18 to 56), and thirty five (76%) of the participants were female. Data from one participant were lost due to a computer error.

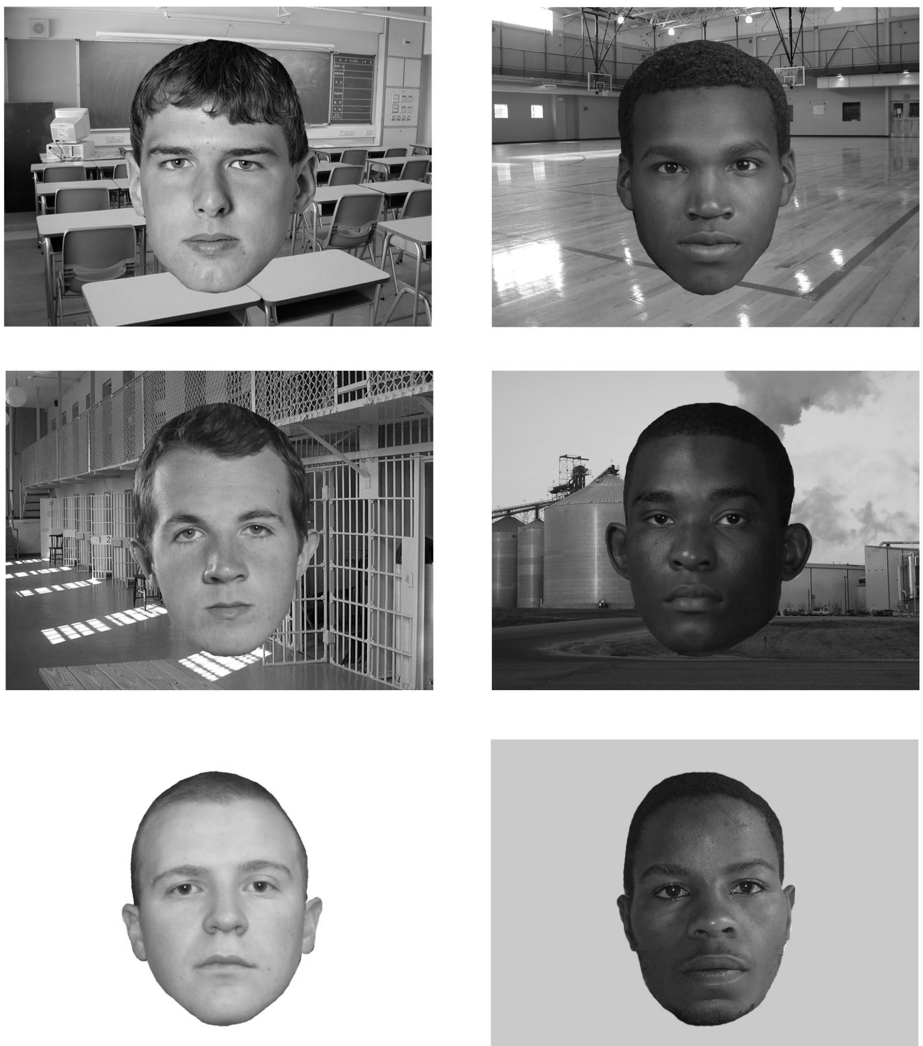
A 2 x 5 repeated measures design was used. The two independent variables were race of face (White or Black) and context (basketball court, classroom, factory, jail cell, and gray screen). Several measures were used to analyze face recognition: hit rates, false alarm rates, and the signal detection measures d' and C. Hit rates, d', and C were analyzed in separate repeated measures ANOVAs; false alarm rates were analyzed in a repeated measures t test. Context memory was analyzed in a multilevel logistic regression. The outcome variable was whether, on any given hit trial, the context judgment was correct or incorrect.

Materials and Apparatus

Eighty White faces and eighty Black faces were collected from the internet. The faces were looking straight ahead with neutral expressions, and had no distinguishing marks, glasses, or jewellery. The photographs were cropped in Adobe Photoshop, to remove necks, shoulders, and backgrounds. The four visual scenes used as contexts (basketball court, classroom, factory, and jail cell) were also collected from the internet; the gray screen context was created in Microsoft Paint. The faces were superimposed onto the different contexts in Adobe Photoshop, with the size and position of the faces relative to the contexts held constant. Examples of the stimuli are shown in Figure 7.

The software E-Prime was used to create and run the study, and to collect the data.

Figure 7. Example stimuli from Experiment 5. White faces (left) and Black faces (right) on all five encoding contexts plus the blank screen used at testing.



Procedure

Participants were tested in groups of up to three by a White experimenter. Each participant worked individually at their own PC. The experimenter remained in the room throughout the experiment to ensure that there was no talking among participants. Participants were told that they would be performing a face recognition task, and that they would need to pay close attention to both the faces and the backgrounds on which they were presented.

The experiment was split into four blocks, each with an encoding phase and a testing phase. In each encoding phase, 25 targets were presented for 5s, with a 1s inter-stimulus interval (ISI). Within each encoding phase, all possible combinations of face race and context were shown at least once. Each encoding phase was immediately followed by a testing phase, in which the 25 targets were shown along with 15 foils. In total, participants saw 50 targets (10 in each context) and 30 foils from each race.

During the testing phases, the stimuli were shown against a blank background. For each face, participants made an old/new judgment by pressing 1 on the keyboard for yes, or 0 for no, followed by a confidence rating on a 9-point scale (1 being not confident at all, and 9 being very confident). For faces judged as old, a context judgment was required. Participants made this judgment by pressing the numbered key which corresponded to the context that they believed was correct (1 for basketball, 2 for classroom, 3 for factory, 4 for jail cell, and 5 for gray screen). This was a forced choice task, in which participants had to make a response. The face remained on screen throughout these questions. Participants were asked to be as accurate as possible, and to take as much time as they needed over their answers.

Targets and distractors were counterbalanced across participants, as were the contexts associated with each target. The order of the blocks was randomized for each participant, as was the presentation order of the stimuli within each encoding phase and testing phase.

Results and Discussion

Face recognition

Previous research has shown that the ORB is often characterized by a marked increase in false alarm rates for other-race faces compared to own-race faces (Meissner & Brigham, 2001; Slone, Brigham, & Meissner, 2000). Hit rates and false alarm rates (FA rates) were therefore analyzed separately, in addition to the signal detection measures d' and C. The formulae used to calculate these measures are given in Appendix 2. The hit rates, d', and C scores were analyzed in 2 (race of face) x 5 (context) repeated measures ANOVAs. The FA rates for White and Black faces were compared in a repeated measures t test. For all comparisons with two levels, the effect size Cohen's d is reported. For interaction terms and comparisons with more than two levels, the effect size η_p^2 is reported.

Table 8 shows the mean hit rates, FA rates, d', and C for White and Black faces, collapsed across context. Whereas hit rates were similar for White and Black faces, $F(1,44) = 0.77$, $p = .39$, $d = .13$, false alarm rates were higher for Black faces than White faces, $t(44) = 5.60$, $p < .001$, $d = .88$. When hit rates and false alarm rates were combined in the d' measure, accuracy was found to be higher for White faces than Black faces, $F(1, 44) = 8.53$, $p = .005$, $d = .52$. Participants were also more conservative when responding to White faces than Black faces, $F(1, 44) = 9.00$, $p = .004$, $d = .60$. Thus, in line with previous research, participants were more accurate at recognizing own-race faces than other-race faces (e.g. Malpass & Kravitz, 1969; Meissner & Brigham, 2001). However, this effect was found only in false alarm rates, with similar hit rates for both groups of faces. This is a pattern of results which has been found elsewhere (e.g. Slone et al., 2000). In this study, the ORB was characterized by a criterion shift for other-race faces, with less conservative responding than for own-race faces.

Table 8: *Mean hit rates, false alarm (FA) rates, d', and C for White and Black faces in Experiment 5.*

	Mean		Standard deviation	
Measure	White	Black	White	Black
Hit rate	.75	.77	.11	.11
FA rate	.14	.23	.08	.12
d'	1.94	1.64	.55	.62
C	-.19	-.01	.28	.33

There were no observed differences between the encoding contexts for any measure: for hit rates, $F(4,176) = 0.22$, $p = .93$, $\eta_p^2 = .01$; for *d'*, $F(4, 176) = 0.43$, $p = .78$, $\eta_p^2 = .01$; and for *C*, $F(4, 176) = 0.43$, $p = .79$, $\eta_p^2 = .01$. The interaction term between race and context was not significant in any of the analyses: for hit rates, $F(4,176) = 0.54$, $p = .70$, $\eta_p^2 = .01$; for *d'*, $F(4, 176) = 1.18$, $p = .32$, $\eta_p^2 = .03$; and for *C*, $F(4, 176) = 1.18$, $p = .32$, $\eta_p^2 = .03$.

It was predicted, on the basis of previous research (Barden et al., 2004; Wittenbrink et al., 2001), that stereotypic congruency between the faces and the context might enhance memory. However, there were no significant differences in accuracy across any of the contexts. The contexts which were used were similar to those used by Barden et al. One possible reason for the lack of significant effects was that the content of racial stereotypes differs between the United States (where the work of Barden et al. was conducted) and the United Kingdom. For example, basketball is not a particularly popular sport in the UK, and so basketball might not form part of the stereotype of Black males here. No ratings were taken from participants to confirm or disconfirm the applicability of these contexts to UK stereotypes, so it is possible that the manipulation was unsuccessful in activating stereotypic attitudes in this study.

Furthermore, although Barden et al. make the claim that the basketball court and classroom contexts form positively valenced contexts, while the prison and factory contexts form negative contexts, these assumptions seem somewhat objective. An individual's attitudes towards classrooms might vary with their enjoyment of studying; attitudes towards factories might vary with employment history and socio-economic status; and attitudes towards the basketball court might vary with enjoyment of sport in general, and with basketball in particular. Overall, the different contexts used here may have provided only weak manipulations of valence, therefore leading to small and unreliable effects. It is possible that backgrounds which are more strongly negative and positive (e.g. the scene of a violent crime, or a pristine beach) would have elicited larger effects.

Memory for context

Context memory was analyzed in a regression model. Participants only made context judgments after "old" decisions for faces, so data was only available on hit and false alarm trials. As all context judgments for false alarm trials are in error, only hit trials are included in the analysis. Because the number of hits varies by person, a multilevel model was used, so that inappropriate weight was not given to people with different numbers of hits (Wright & London, 2009). The outcome measure, whether the context judgment was accurate or inaccurate, was binary, and so a logistic regression was used (Wright, 1997).

Multilevel statistics provide a powerful method of analysis at the level of individual trials rather than averages calculated across trials for each participant (Wright, Horry, & Skagerberg, 2009). Memory data can be thought of as clustered or nested within participants, giving the data a hierarchical structure (Wright, 1998). In a multilevel regression, participant variation can be treated as a random effect in which the regression slopes are allowed to vary for each participant. In a similar way, variation around individual stimuli can be treated as a random effect (Clark, 1973). This may be particularly important for naturalistic stimuli such as faces, which will vary along many dimensions.

The regression model was built in two steps, shown in Table 9. In the first step, confidence was entered as a predictor of context accuracy. Participants may be more accurate at identifying the contexts for faces recognized confidently than for faces recognized less confidently. Accuracy ranged from 37.25% for faces identified with very low confidence (confidence ratings of 1) to 68.26% for faces identified with the highest confidence (confidence ratings of 9). The regression model confirmed that confidence significantly predicted accuracy, with an estimated regression parameter of .35, $p < .001$. The second model included race of face as a predictor. Context judgments were accurate for 58.72% of White face trials, compared to 52.05% of Black face trials. This difference was statistically significant, with an estimated regression parameter of .22, $p = .02$. Including race of face improved the fit of the model to the data, $\chi^2(1) = 4.96$, $p = .03$.

According to dual-process theory, recollection is necessary to support memory for context. Recent evidence suggests that people use recollection less frequently when recognizing other-race faces than when recognizing own-race faces (Marcon et al., 2009; Meissner et al., 2005). Experiment 5 showed that White participants were more accurate at identifying the contexts in which they saw White faces than the contexts in which they saw Black faces.

Table 9: *Estimated regression parameters and error terms for confidence and race of face in Experiment 5.*

		Model 1: Context and confidence			Model 2: Race of face included		
Parameter		Estimate	SE	Sig.	Estimate	SE	Sig.
Fixed effects:							
Intercept	β_0	-2.41	.21	<.001	-2.50	.21	<.001
Confidence	β_1	.35	.02	<.001	.35	.02	<.001
Race of face	β_2	------	----	------	.22	.10	.02
Random effects:							
Participant level	σ^2_u	.18			.17		
Stimulus level	σ^2_e	.44			.44		
Difference in model fit:		$\chi^2(1) = 4.96$, $p = .03$					

Experiment 6

Experiment 6 investigates the effects of context reinstatement on face recognition and memory for context. During the testing phase, the target faces are either shown in a congruent context or an incongruent context. Faces should be recognized more accurately when presented on a congruent context than on an incongruent context (Gruppuso et al., 2007). This improvement may be larger for Black faces than White faces, as they form a relatively less familiar class of stimuli (Russo et al., 1999). If context reinstatement specifically benefits recollection, then memory for context should be more accurate when the contexts are congruent than when they are incongruent. This may reduce the difference in context accuracy for White and Black faces.

Method

Participants and Design

Fifty four White participants took part in the experiment for credit on an introductory psychology course or for £5. The mean age of the participants was 21 years (range 18 to 36), and forty three (80%) of the participants were female.

A 2 x 2 repeated measures design was used. The two independent variables were race of face (White or Black) and congruency (congruent or incongruent). The face recognition measures were hit rates, false alarm rates, d', and C. For context memory, the outcome measure was whether, on any given trial, the context judgment was accurate or inaccurate; these data were analyzed in a multilevel logistic regression.

Materials and Apparatus

The faces used in this experiment were the same as those used in Experiment 5. Only three of the encoding contexts from Experiment 5 were used: the classroom, the jail cell, and the gray screen. This was to reduce the number of possible encoding and testing context combinations. Once again, the experiment was created and run using E-Prime software.

Procedure

The procedure was the same as Experiment 5, with the following exceptions. A total of 54 targets from each race were seen; 18 of these targets appeared on congruent contexts, and the remaining 36 appeared on incongruent contexts. One third of the distractor faces appeared on each context. The presentation order was randomized for each participant. Which faces were targets and which were foils was counterbalanced between participants, as was the pairings of faces to the different contexts.

Results and Discussion

Face recognition

Table 10 shows the mean hit rates, *d′*, and *C* scores broken down by congruency and race of face. These data were analyzed in separate 2 (race of face) x 2 (congruency) repeated measures ANOVAs. The false alarm rates for White and Black faces were compared in a repeated measures *t* test. Overall accuracy, shown in the *d′* scores, was higher for White faces (*M d′* = 1.55, *SD* = .59) than for Black faces (*M d′* = 1.35, *SD* = .55), $F(1, 53) = 8.56$, $p = .005$, $d = .35$. Participants made more false alarms for Black faces (*M* = .29, *SD* = .16) than for White faces (*M* = .20, *SD* = .12), $t(53) = 5.33$, $p < .001$, $d = .63$. This replicates the ORB found in Experiment 5, and elsewhere in the face recognition literature (e.g. Malpass & Kravitz, 1969; Meissner & Brigham, 2001). However, hit rates were somewhat higher for Black faces (*M* = .75, *SD* = .12) than for White faces (*M* = .72, *SD* = .11), $F(1, 53) = 4.22$, $p = .05$, $d = .25$.

Table 10: *Mean hit rates, d', and C for White and Black faces shown on congruent (top) and incongruent (bottom) contexts in Experiment 6.*

Measure	Mean		Standard deviation	
	White	Black	White	Black
		Congruent		
Hit rate	.72	.77	.14	.15
d'	1.57	1.44	.69	.68
C	-.14	.11	.34	.41
		Incongruent		
Hit rate	.72	.73	.12	.12
d'	1.53	1.26	.57	.52
C	-.15	.02	.30	.40

Overall accuracy was higher for faces presented on congruent contexts at testing (*M d'* = 1.50, *SD* = .59) than for faces presented on incongruent contexts at testing (*M d'* = 1.40, *SD* = .49), $F(1, 53) = 5.12$, $p = .03$, $d = .20$. Congruency increased hit rates slightly (*M* congruent = .74, *SD* = .12; *M* incongruent = .72, *SD* = .10), but the difference was not significant, $F(1, 53) = 2.31$, $p = .14$, $d = .18$. The interaction term between race of face and congruency did not reach significance for hit rates, $F(1, 53) = 2.58$, $p = .11$, $\eta_p^2 = .05$, or for *d'*, $F(1, 53) = 2.46$, $p = .12$, $\eta_p^2 = .04$. The hypothesis that context reinstatement would increase face recognition accuracy was therefore supported, although the effect size was quite small. The prediction that Black faces might benefit most from context reinstatement, however, was not supported, as the interaction was not significant. However, the mean differences in hit rates and *d'* were numerically larger for the Black faces than for the

White faces. With a larger sample size, these differences may have become statistically significant.

Participants were more conservative with White faces (*M C* = -.14, *SD* = .30) than with Black faces (*M C* = .07, *SD* = .38), $F(1, 53) = 25.07$, $p < .001$, $d = .61$. This shift in response criterion could explain the higher hit rates found for Black faces, as participants were making more “old” decisions overall for Black faces than for White faces. Participants were also somewhat more conservative with faces presented on incongruent contexts than with faces presented on congruent contexts, $F(1, 53) = 5.12$, $p = .03$, $d = .16$, although this effect was quite small. The interaction term was non-significant, $F(1, 53) = 2.46$, $p = .12$, $\eta_p^2 = .04$.

Memory for context

Context memory was again analyzed in a multilevel logistic regression. The predictor variables were confidence, congruency, and race of face. Confidence and congruency were entered in model 1, and race of face was added in model 2. Table 11 shows the estimated regression parameters for each of the main effects in the two models.

Participants were more likely to be accurate for faces which had been identified with high confidence than for faces identified with low confidence. Accuracy ranged from 40.43% for faces identified with the lowest confidence (ratings of 1) to 66.12% for faces identified with the highest confidence (ratings of 9). The estimated regression parameter for confidence was .20, $p < .001$. Participants were no more accurate when the contexts were congruent (54.97%) than when they were incongruent (54.55%). The estimated regression parameter for congruency was .01, $p = .88$. The hypothesis that context memory would be more accurate when context was reinstated was therefore not supported. This is somewhat surprising, as context reinstatement has been shown to enhance recollective memory (Gruppuso et al., 2007; Macken, 2002).

Table 11: *Estimated regression parameters and error terms for confidence, congruency, and race of face in Experiment 6.*

		Model 1: Congruency and confidence			Model 2: Race of face included		
Parameter		Estimate	SE	Sig.	Estimate	SE	Sig.
Fixed effects:							
Intercept	β_0	-1.25	.14	<.001	-1.37	.15	<.001
Congruency	β_1	.01	.07	.88	.01	.07	.86
Confidence	β_2	.20	.02	<.001	.20	.02	<.001
Race of face	β_3	------	----	------	.27	.08	.001
Random effects:							
Participant level	σ^2_u	.14			.14		
Stimulus level	σ^2_e	.12			.10		
Difference in model fit:		$\chi^2(1) = 10.12$, $p = .001$					

Accuracy was higher for White faces (58.24%) than for Black faces (51.24%). The estimated regression parameter for race was .27, $p = .001$. This supports the findings of Experiment 5, that memory for context is impaired for correctly identified other-race faces. The interaction term between race of face and congruency was non-significant, with an estimated regression parameter of .03, $p = .80$. There was therefore no evidence that the ORB for context memory was reduced when context was reinstated.

Experiment 7

Experiments 5 and 6 showed that White participants were less accurate at identifying the contexts in which Black faces were seen than the contexts in which White faces were seen. In both of these studies, complex visual scenes were used as the contexts. These contexts may have carried different associations for different participants, some positive and some negative. Although no differences in recognition accuracy between the different contexts were found, it nevertheless seems important to show that this cross-race effect is not an artefact of the particular contexts used in Experiments 5 and 6. In Experiment 7, simple backgrounds of different colors are used as encoding contexts.

Method

Participants and Design

Fifty four White participants took part in the experiment for credit on an introductory psychology course or for £4. The mean age of the participants was 21 years (range 18 to 32), and forty seven (87%) were female.

The experiment had a 2 x 3 repeated measures design. The two independent variables were race of face (White and Black) and encoding context (red, green, or blue screen). The face recognition measures were hit rates, false alarm rates, d', and C. Context memory was analyzed in a multilevel logistic regression.

Materials and Apparatus

The faces were the same as those used in Experiments 5 and 6. The encoding contexts were a red screen, a green screen, and a blue screen. At testing, all faces were presented on a blank screen.

Procedure

The procedure was similar to that of Experiment 5, with the following changes. As there were only three encoding contexts, the number of White and Black targets shown on each background was increased to 18. This gave a total of 54 targets and 26 distractors of each race. Pilot studies showed that the context judgment task was more difficult with colored backgrounds than with visual scenes. In order to help participants build associations between each face and its corresponding background, participants were asked to press a key on the keyboard which had been labelled with a red, green or blue sticker during the encoding phase. This did not affect the presentation time of the stimulus. Participants also used the colored keys to make their context memory judgements.

Results

Face recognition

Table 12 shows the mean hit rates, false alarm rates, d', and C for White and Black faces. Overall face recognition accuracy, shown by d', was higher for White faces (M = 1.59, SD = .53) than for Black faces, (M = 1.26, SD = .58), $F(1, 52)$ = 24.89, $p < .001$, d = .59. Participants made more false alarms with Black faces (M = .28, SD = .17) than with White faces (M = .17, SD = .12), $t(52) = 7.26$, $p < .001$, d = .77. However, hit rates were higher for Black faces (M = .71, SD = .11) than for White faces (M = .68, SD = .11), $F(1, 52) = 6.68$, $p = .01$, d = .25. This pattern of results is very similar to that found in Experiment 6. A significant ORB was found, but in false alarm rates and not in hit rates.

Table 12: *Mean hit rates, false alarm (FA) rates, d', and C for White and Black faces in Experiment 7.*

	Mean		Standard deviation	
Measure	White	Black	White	Black
Hit rate	.68	.71	.11	.11
FA rate	.17	.28	.12	.17
d'	1.59	1.26	.53	.58
C	-.27	-.01	.34	.37

Hit rates varied across the different encoding contexts, $F(2, 104) = 3.18$, $p = .05$, $\eta_p^2 = .06$, ranging from .69 ($SD = .13$) for the red context to .72 ($SD = .11$) for the green context. The interaction terms between context and race of face were not significant in either the hit rate or the *d'* analyses: hits, $F(2, 104) = 1.72$, $p = .18$, $\eta_p^2 = .03$; *d'*, $F(2, 104) = 2.05$, $p = .13$, $\eta_p^2 = .04$.

Participants were more conservative with White faces ($M = -.27$, $SD = .34$) than with Black faces ($M = -.01$, $SD = .37$), $F(1, 52) = 54.34$, $p < .001$, $d = 1.07$. This means that participants were making more "old" decisions in response to Black faces than White faces, which could explain the increase in hit rates as well as the increase in false alarm rates. Response criteria did not vary across encoding contexts, $F(2, 104) = 2.30$, $p = .11$, $\eta_p^2 = .04$. The interaction term between race of face and context in the *C* analysis was not statistically significant, $F(2, 104) = 2.05$, $p = .13$, $\eta_p^2 = .04$.

Memory for context

Context memory was analyzed in a multilevel logistic regression. The outcome variable was whether, on a given trial, the context judgment was accurate or inaccurate. The predictor variables were confidence (entered in model 1) and race of face (entered in model 2). The estimated regression parameters are shown in Table 13.

Table 13: *Estimated regression parameters and error terms for confidence and race of face in Experiment 7.*

		Model 1: Context and confidence			Model 2: Race of face included		
Parameter		Estimate	SE	Sig.	Estimate	SE	Sig.
Fixed effects:							
Intercept	β_0	-1.02	.14	<.001	-1.15	.14	<.001
Confidence	β_1	.13	.02	<.001	.13	.02	<.001
Race of face	β_2	------	----	------	.27	.07	<.001
Random effects:							
Participant level	σ^2_u	.07			.07		
Stimulus level	σ^2_e	.03			.02		
Difference in model fit:		$\chi^2(1) = 15.02$, $p < .001$					

Confidence ratings significantly predicted accuracy; the estimated regression parameter for confidence was .13, $p < .001$. Accuracy ranged from 50.00% for faces identified with very low confidence (ratings of 1) to 56.68% for faces identified with very high accuracy (ratings of 9). This relationship is somewhat weaker than those found in Experiments 5 and 6. This may have been because remembering associations between faces and simple, colored contexts was more difficult than remembering associations between faces and complex visual scenes.

Participants were more accurate at identifying the contexts on which White faces appeared (51.29%) than the contexts on which Black faces appeared (44.55%). The estimated regression parameter for race of face was .27, $p < .001$. This supports the findings from Experiments 5 and 6, and shows that impaired memory for context in other-race identifications extends beyond the specific visual scenes used previously.

General Discussion

A large body of work over the last forty years has shown that people are more accurate at recognizing faces from their own race than faces from other races (Malpass & Kravitz, 1969; Meissner & Brigham, 2001). Here, across three studies, White participants recognized White faces more accurately than Black faces. Participants' response criteria were consistently more lenient for Black faces than for White faces, leading to a marked increase in false alarm rates, as well as a smaller increase in hit rates. However, even when a face is correctly identified, an observer may be unable to remember information about the context in which that face was originally encountered. These studies show, for the first time, that such failures in context memory are more likely when identifying faces from a different race. In all three experiments, context judgments were more often accurate for White faces than for Black faces.

Recently, dual-process theories of memory have been applied to the ORB. Meissner et al. (2005) found that subjective reports of remembering in a Remember-Know task were more frequent in own-race than other-race identifications, and

Marcon et al. (2009) showed that participants made more familiarity-based errors for other-race faces than for own-race faces. Recollection is the product of an episodic memory system, which allows individuals to retrieve contextual information about the encoding of a stimulus (Yonelinas, 2002). Deficits in recollection should therefore lead to failures in context memory. The three experiments presented here support the dual-process account of the ORB by showing this predicted impairment in context memory for other-race faces.

Research from the source monitoring literature has shown the importance of being able to retrieve information about the source of an item's familiarity. This type of contextual memory is particularly important for eyewitnesses (Loftus, 1976), who need to be sure that the suspect that they identify is the culprit, and is not simply familiar from some other source. For example, the risk of a misidentification by a witness is increased if the witness is exposed to mugshots between witnessing the crime and viewing a lineup. In such a case, the witness may identify a previously seen innocent suspect on the basis of familiarity, as he or she is unable to recall the source of that familiarity (Deffenbacher et al., 2006). The results presented here raise the worrying possibility that these types of errors may be more likely if the suspect and the witness are of different races.

Experiment 6 explored the potential of context reinstatement in improving both face recognition and context memory. While a modest improvement was found in face recognition accuracy, no improvement was found in context memory. Previously, context reinstatement has been shown to benefit recognition of unfamiliar classes of stimuli more than familiar classes of stimuli (Russo et al., 1999). It was therefore predicted that context reinstatement may be more beneficial for Black faces than for White faces. The difference in face recognition accuracy between stimuli shown on congruent and incongruent backgrounds was numerically larger for Black faces than for White faces, although the interaction term was not statistically significant.

Many studies which have found significant increases in recognition accuracy following context reinstatement have manipulated *local* context. Local context is associated with just one or very few stimulus items (e.g. Evans et al., 2009; Gruppuso et al., 2007; Hockley, 2008). In contrast, global context is associated with a larger subset of items. For example, if each stimulus was paired with a unique piece of contextual information, such as a unique image, that would be local context. If all items on one study list were associated with one context (e.g. a particular testing room), and all items on another study list were associated with another context (a different study room), then the room would constitute the global context. In Experiment 6, all of the stimuli were associated with one of three background images. The backgrounds would therefore have acted as a global context rather than a local context, perhaps reducing the likelihood that context reinstatement would provide benefits to recognition. Future studies could examine the impact of local context reinstatement on the ORB, by pairing each face with a unique background image.

There are two major limitations to these studies. First, the same photograph of each face was used during the encoding and testing phases. Bruce (1982) argued that participants could use low-level features of the photographs to make their recognition judgments. Rather than recognizing the faces, they may simply be recognizing the pictures. Using different images of the same individual during encoding and testing ensures that *face* recognition is really being tested. This is a limitation which will be addressed in Paper 4, by using a different stimulus set with two images of each face. Second, only White participants took part in this experiment. This is known as a half-design within the cross-race recognition literature, and raises the possibility of stimulus sampling errors (Wells & Olson, 2001). If the stimuli from the two groups differ in their homogeneity, then apparent cross-race effects may actually be artefacts of the stimulus selection. Testing participants from both racial groups reduces the likelihood of this possibility if a cross-over interaction can be found in recognition accuracy. In Paper 4, samples of White and Black participants will be tested to address this limitation.

Across three experiments, White participants showed advantages in recognizing own-race faces, and in retrieving contextual information about those faces from memory. These results support the dual-process theory of the ORB (Meissner et al., 2005), which posits that own-race faces are more likely to be recollected than other-race faces. Although low in external validity, one can speculate on the implications of these results outside of the laboratory. In order to successfully interact with those around us, we need not only recognize whether we have seen an individual before, but we need to recall in what context we have seen somebody before. Remembering such information will inform our decision to approach or to avoid, and how to behave in a given encounter (Mandler, 1980).

Perhaps the most extreme situation in which this sort of contextual memory may prove crucial is when an eyewitness is making an identification. In a lineup, an eyewitness is presented with a series of individuals, which will evoke varying feelings on familiarity in the witness. One way to make the identification decision would be for the witness to pick the individual in the lineup who seems most familiar. In many cases, the most familiar individual may well be the culprit. However, the familiarity associated with a given individual could arise from many different sources – perhaps the witness has seen the individual at some other time and place, or perhaps the individual was present at the crime scene, but was not the culprit. Eyewitness identification therefore requires person recognition, rather than the simple target recognition often tested in old/new recognition studies (Sporer, 2001b).

Source monitoring errors have been demonstrated in lab experiments using eyewitness paradigms. For example, innocent bystanders might be misidentified as the culprit (Loftus, 1976), or innocent suspects seen in mugshot searches might be identified in subsequent lineups (Deffenbacher, Bornstein, & Penrod, 2006). The studies in this paper provide the first empirical evidence that source monitoring errors are more prevalent in cross-race identifications than in own-race identifications. Further research could explore such an effect using more externally valid paradigms with mock crimes and lineups. If the source monitoring errors observed here generalise to real world identifications, then mugshot exposure effects and bystander effects might be larger when the suspect and the witness are from different ethnic groups than when they are from the same ethnic group.

Paper 4: The Role of Recollection in the Own-Race Bias in Item Recognition and Context Memory[4]

Abstract

People recognize own-race faces more accurately than other-race faces. This own-race bias (ORB) in recognition may be the result of a deficit in recollective memory for other-race faces. In a single experiment, White and Black participants saw White and Black faces presented with different visual contexts. Recognition memory was tested using an old/new recognition paradigm. For faces categorized as old, participants also produced context judgments. Participants were more accurate at identifying own-race faces than other-race faces. They were also more accurate at remembering the contexts in which correctly recognized own-race faces were seen. Using a Remember-Know task, participants gave more Remember responses to own-race faces than to other-race faces.

Introduction

Humans have a remarkable capacity for recognizing faces. As a class of visual stimuli, faces are highly similar to one another, each consisting of the same basic features in the same general configuration. Yet people are able to detect even very small changes in a familiar face (Brooks & Kemp, 2007), and are able to recognize familiar faces despite drastic changes (Hole, George, Eaves, & Rasek, 2002). However, people are not able to remember all types of faces equally well. Out-group

---

[4] Experiment 8 in press: Horry, R., Wright, D. B., & Tredoux, C. G. (In press). Recognition and context memory for faces from own and other ethnic groups: A Remember-Know investigation. *Memory & Cognition.*

recognition deficits are well documented in the literature (Sporer, 2001b). For example, people are impaired at recognizing faces from other age groups (Anastasi & Rhodes, 2005; Wright & Stroud, 2002) or faces of the opposite gender (Wright & Sladden, 2003). Perhaps the most robust of these out-group recognition effects is the own-race bias (ORB), the tendency for people to recognize faces of their own race more accurately than faces of other races (Malpass & Kravitz, 1969; Meissner & Brigham, 2001).

Several theories have been put forward to explain the ORB including cognitive disregard following categorization (Rodin, 1987), attention to inappropriate features during encoding (Ellis, Deregowski, & Shepherd, 1975), reliance on featural processing for other-race faces (Hancock & Rhodes, 2008), and storage in memory along inappropriate dimensions in “face space” (Chiroro & Valentine, 1995). Many of these theories flow from the premise that people generally have more contact, and therefore acquire more perceptual expertise, with members of their own race than with members of other races. Sporer (2001b) argued that unless the brain is hardwired from birth to selectively recognize own-race faces, contact must play an important role in the ORB. Sangrigoli and de Schonen (2004) showed that three-month old infants could recognize own-race faces more accurately than other-race faces, although this effect could be eliminated by exposure to a few exemplars of other-race faces.

Dual-process theorists argue that that there are two separable routes to recognition – recollection and familiarity. Recollection is the explicit episodic memory of having encountered a stimulus before, which includes some information about the encoding event. Familiarity is a more fluid sense of knowing that a stimulus has been encountered before, in the absence of recalling any specific details about that encounter (for a review of dual-process theory, see Yonelinas, 2002). Recently, Meissner and colleagues (Evans, Marcon, & Meissner, 2009; Marcon, Susa, & Meissner, 2009; Meissner, Brigham, & Butz, 2005) suggested that the ORB could be produced by a specific impairment in recollection for other-race faces. This could account for the reduced hit rates and increased false alarm rates for other-race faces which are commonly found in old/new recognition tests (Meissner & Brigham, 2001).

Tulving's (1985) Remember-Know (RK) procedure is a widely used paradigm for investigating participants' subjective experiences of their recognition decisions. When an item is recognized as old, the participant makes a second judgment, concerning his or her reasons for making a positive identification. If the participant is able to recall any aspects about the encoding event in which a stimulus was encountered, then he or she makes a Remember (R) response. If a participant feels that an item is familiar, but is unable to recall any information about the encoding event, then he or she makes a Know (K) response. Many studies also include a Guess (G) option, for when recognition decisions are made with no memorial basis (e.g. Konstantinou & Gardiner, 2005). The advantages and disadvantages of the RK procedure are discussed by Yonelinas (2002) and Gardiner, Ramponi, and Richardson-Klavehn (2002). While some authors have equated Remember and Know responses to recollection and familiarity respectively, other authors have cautioned against these interpretations of the RK procedure (e.g. Dunn, 2004; Donaldson, 1996). This issue remains contentious, and will be returned to in the General Discussion.

Recollection and familiarity can both produce accurate item recognition. Dual-process theorists argue that memory for context, however, relies much more heavily on recollection than familiarity. Remember responses are often associated with higher context memory accuracy than Know responses, although context memory accuracy may still be above chance for Know responses (Meiser & Sattler, 2007). Meiser, Sattler, and Weisser (2008) showed that the binding of multiple elements of an encoding event into a coherent memory representation is evident in Remember, but not in Know, responses. Therefore, while familiarity may allow the retrieval of some residual context information, recollection is required for retrieving more complex and specific context information.

In Paper 3, White participants were found to be less accurate at remembering contextual information about other-race faces than about own-race faces. At encoding, faces of White and Black males were presented on one of five different backgrounds. At testing, the faces were all presented on a blank screen. Participants made old/new recognition decisions, and also made context judgments for faces that they categorized as old. As well as the expected impairment in item recognition for Black faces compared to White faces, context memory was impaired for Black faces.

Paper 3 included samples of White participants only, limiting the generalizabilty of the results. Wells and Olson (2001) argued that a priori differences in the discriminability of the two stimulus groups can create apparent cross-race effects, which are actually artifacts of the stimulus selection. Recruiting participants from both groups greatly reduces the likelihood of this if a crossover interaction can be shown. These studies present data from participants living and studying in South Africa. A small number of ORB investigations have been conducted in Africa (Chiroro, Tredoux, Radaelli, & Meissner, 2008; Chiroro & Valentine, 1995; Wright, Boyd, & Tredoux, 2001, 2003), with some finding full crossover effects and others finding asymmetric effects. South Africa provides a unique societal context for exploring the ORB, due to its recent history of enforced segregation, and its ethnic multiplicity. The data were collected at a large university with a diverse student population. In this particular setting, there is considerable opportunity for students to interact with people from different ethnic and racial backgrounds, so one might expect smaller cross-race effects. However, even within multicultural societies or organisations, people are much more likely to have meaningful and intimate relationships with people from their own race than with people from other races (Dixon, Tredoux, & Clack, 2005; Tredoux & Dixon, 2009). Participants from such areas could provide some new insights into the subtleties and complexities of cross-race effects, which would strengthen the generalizabilty of the results from the United States and Europe.

This paper presents a single experiment which will investigate the role of recollection in recognizing own- and other-race faces, and in remembering contextual information about those faces. This study uses a similar procedure to that used in Paper 3. Faces are shown at encoding on one of several backgrounds. Participants then take part in an old/new recognition test for those faces, and make context judgments for faces categorized as old. Remember-Know judgments are also provided when old decisions are made.

This paper has three main aims. First, to add to the emerging evidence that the ORB is characterized by a marked decrease in recollection for other-race faces compared to own-race faces, using Tulving's (1985) RK procedure. Second, to show that memory for context is impaired in other-race identifications, and that this generalizes beyond White participants. And finally, to collect cross-race recognition data from a population outside of North America or Europe.

Method

Participants and Design

Forty one White and forty four Black undergraduates at the University of Cape Town took part in the study for credit on an introductory psychology course or for payment (approximately £3). Nine of the participants were male (2 White, 7 Black). The mean age was 21 years; ages ranged from 17 to 43.

The experiment had a 2 (race of participant – White or Black) by 2 (race of face – White or Black) by 4 (context – basketball court, classroom, factory, jail cell) mixed design. The first factor was between groups, and the latter two factors were repeated measures. The dependent variables were face recognition measures (hits, false alarms, d' and C), decision strategy (Remember, Know, Guess judgments), and context memory accuracy.

Materials and Apparatus

Eighty White faces and eighty Black faces of young males were used as stimuli. All of the Black males and approximately half of the White males were taken from a South African database. However, this database did not contain a sufficient number of White faces, and so the remaining White faces were taken from an Australian database of comparable quality. Chiroro et al. (2008) suggested that local stimuli should be used wherever possible when testing the ORB. However, this can be practically difficult to do, especially when large numbers of faces are required. Even when local stimuli are used, the people depicted may nonetheless come from a range of different backgrounds due to the multicultural nature of many societies, including South Africa. Therefore, both groups of faces would have had some variation in the ethnic backgrounds of the people shown.

Two photographs of each face were used in the experiment – one with a full frontal pose, and one with a three quarter profile pose. A different view of each face was presented at encoding and testing to ensure that participants were recognizing the faces rather than any low level features of the particular photograph (Bruce, 1982). Each photograph was cropped so that the necks and shoulders were removed. The cropped photographs were then pasted onto the four backgrounds using Adobe Photoshop. Which faces appeared on which backgrounds was counterbalanced between participants. Which faces were targets and which were foils was also counterbalanced. The experiment was designed and run with the E-Prime software system.

Procedure

Participants were tested in groups of up to five. Each participant sat at an individual PC and worked independently. Participants were told that they would see a series of faces presented on different backgrounds, and that they would later be asked to remember the faces and their associated backgrounds.

The experiment was broken down into ten short blocks, each with a learning phase followed by a testing phase. During each learning phase, eight targets appeared (one target for each combination of race and context). Each target and its associated background was shown for 8s, with a 1s inter-stimulus interval. In each testing phase, the eight targets were seen again, along with eight new faces (four White and four Black). The recognition test was self-paced, and each face remained on screen until the participant had responded to all of the questions. In total, throughout the experiment, ten targets of each race appeared on each background, giving a total of forty White and forty Black targets. Forty distractors from each group were also seen. Participants were not told the proportion of old to new faces. The order of the blocks was randomized for each participant, and the order in which the faces were presented within each block was also randomized.

Participants made an old/new recognition decision for each face. Following an old judgment, participants made an RKG judgment followed by a context memory decision. Instructions for the RKG judgments were provided on the screen, and were also printed on a sheet of paper so that the participants could refer to them at any time. The instructions were adapted from Conway and Dewhurst (1995), to make them suitable for use in a face recognition experiment:

> Sometimes when we recognize a face we have seen before, we can consciously REMEMBER specific details about the previous occurrence of the face. At other times, we simply KNOW that we have seen a face before, even though we cannot recall specific details about the event.
>
> For each face, you will have 3 options:
>
> 1. REMEMBER: When you see a face in this test that you recognize, you may be able to remember specific details about seeing that face before. You may for example recollect the thoughts or feelings that the face evoked when you saw it, or an association that you formed with another face, or some aspect of the face's physical appearance. In short, any additional detail which supports your belief that the face appeared before.
>
> 2. KNOW: For other faces, you may know that they appeared before, even though you cannot remember any specific details of their occurrence.
>
> 3. GUESS: For some faces, you may have simply guessed, without any reason to believe that the face appeared before.

For the context memory judgments, participants were required to press 1 for the basketball court, 2 for the classroom, 3 for the factory, or 4 for the jail. This was a forced choice with task, with no option to respond “don’t know”.

Results

Face recognition accuracy and response bias

Table 14 shows the mean hit rates and false alarm rates, and the mean *d'* and *C* for White and Black participants with White and Black stimuli. Separate 2 (race of participant) x 2 (race of face) mixed ANOVAs were conducted for each outcome measure[5]. Across all four measures, the interaction term between race of participant and race of face was significant and large: for hit rates, $F(1, 83) = 60.83$, $p < .001$, $\eta_p^2 = .42$; for false alarm rates, $F(1, 83) = 45.72$, $p < .001$, $\eta_p^2 = .36$; for *d'*, $F(1, 83) = 55.69$, $p < .001$, $\eta_p^2 = .40$; for *C*, $F(1, 83) = 55.69$, $p < .001$, $\eta_p^2 = .40$. White participants had higher hit rates for in-group (White) faces than for out-group (Black) faces, $t(40) = 8.00$, $p <.001$, $d = .98$, and higher false alarm rates for out-group (Black) faces than for in-group (White) faces, $t(40) = 6.43$, $p < .001$, $d = .63$. Black participants showed the same pattern of results, with higher hit rates for in-group (Black) faces than for out-group (White) faces, $t(43) = 3.24$, $p = .002$, $d = .50$, and higher false alarm rates for out-group (White) faces than for in-group (Black) faces, $t(43) = 3.04$, $p = .004$, $d = .33$.

---

5 Analyses including encoding context revealed no significant main effects or interactions on any outcome measure (minimum $p = .16$). Therefore, context will not be discussed further.

Table 14: ***Mean hit rates and false alarm (FA) rates, and mean d' and C for White participants (top) and Black participants (bottom) with White and Black faces in Experiment 8.***

	Mean		Standard deviation	
Measure	White faces	Black faces	White faces	Black faces
	White participants			
Hit rates	.76	.65	.11	.12
FA rates	.12	.20	.11	.14
d'	1.77	1.36	.59	.52
C	-.07	-.27	.37	.37
	Black participants			
Hit rates	.63	.70	.15	.13
FA rates	.15	.12	.12	.10
d'	1.71	1.87	.59	.59
C	-.46	-.38	.36	.33

The signal detection measure *d'* confirmed that White participants were more accurate with White faces than with Black faces, $t(40) = 7.25$, $p < .001$, $d = .74$, whereas Black participants were more accurate with Black faces than with White faces, $t(43) = 3.12$, $p = .003$, $d = .27$. The criterion measure *C* showed that White participants were more conservative with White faces than with Black faces, $t(40) = 7.25$, $p < .001$, $d = .53$, whereas Black participants were more conservative with Black faces than with White faces, $t(43) = 3.12$, $p = .003$, $d = .24$.

RKG responses

The way in which RKG responses are analyzed is a contentious issue, and one's position will depend upon the assumptions made about how recollection and familiarity operate at the trial level. For example, one can assume that the two processes are exclusive, i.e. that on any given trial, *either* recollection *or* familiarity can influence the recognition judgment. Alternatively, one can assume that the two processes are inclusive, i.e. that items which are recollected will also be familiar. Authors who favour the exclusivity assumption argue that the Independence Remember-Know (IRK) procedure, developed by Yonelinas and Jacoby (1995) should be applied to K responses. In the IRK procedure, familiarity is estimated as follows: $F = K/(1-R)$. Authors who favour the inclusivity assumption (e.g. Gardiner, Java, & Richardson-Klavehn, 1996) argue that the proportions of R and K responses should be analyzed as they are, as they reflect the participant's actual state of awareness. Here, we present data for both the proportion of K responses, and for familiarity estimates from the IRK procedure. Table 15 shows the mean proportions of hit trials producing R, K, and G responses, and the IRK familiarity estimates, broken down by race of participant and race of face. As G responses are dependent upon R and K responses, G responses were not analyzed. All of the other measures were analyzed in separate 2 (race of participant) x 2 (race of face) mixed ANOVAs.[6]

Table 15 shows that both groups of participants made more R responses for own-race faces than for other-race faces, although the mean difference appears larger for White participants than for Black participants. This observation was confirmed in

---

[6] The same analyses were conducted on R and K responses for false alarm trials. No significant effects were found. For main effects: Maximum $F(1, 83) = 2.40$, $p = .13$, $d = .34$. For interaction terms: Maximum $F(1, 83) = 1.78$, $p = .18$, $\eta_p^2 = .02$.

the ANOVA, as the interaction term was significant, $F(1, 83) = 34.40$, $p < .001$, $\eta_p^2 = .29$. White participants made significantly more R responses to White faces than Black faces $t(40) = 6.25$, $p < .001$, $d = .74$. In contrast, Black participants made slightly more R responses to Black faces than White faces, although this difference was not statistically significant, $t(43) = 1.72$, $p = .09$, $d = .15$.

Table15: ***Mean proportions of Remember, Know, and Guess responses in hit trials, along with transformed Independence Remember-Know (IRK) responses, broken down by race of participant and race of face in Experiment 8.***

	Mean		Standard deviation	
Measure	White faces	Black faces	White faces	Black faces
	White Participants			
Remember	.70	.54	.20	.23
Know	.21	.27	.17	.16
IRK	.71	.63	.39	.29
Guess	.09	.18	.10	.19
	Black participants			
Remember	.57	.61	.25	.24
Know	.28	.28	.23	.24
IRK	.58	.66	.29	.25
Guess	.14	.11	.11	.09

Table 15 shows that White participants made more K responses for Black faces than for White faces, whereas Black participants were equally likely to make K responses for White and Black faces. The ANOVA confirmed these observations. The interaction term was significant, $F(1, 83) = 4.43$, $p = .04$, $\eta_p^2 = .05$, but the difference was significant for White participants $t(40) = 3.00$, $p = .005$, $d = .40$, and not Black participants, $t(43) = 0.34$, $p = .74$, $d = .01$.

A different pattern seems to have emerged when the IRK transformation was applied to the RKG data, with higher means for own-race faces than other-race faces within both groups of participants. Again, the interaction between race of participant and race of face was significant, $F(1, 83) = 4.73$, $p = .03$, $\eta_p^2 = .05$. However, the differences between White and Black faces were not statistically significant for either the White participants, $t(40) = 1.36$, $p = .18$, $d = .22$, or the Black participants, $t(43) = 1.74$, $p = .09$, $d = .29$.

Context memory

As in Paper 3, context memory was first analyzed in a multilevel logistic regression. Participant-level variance and stimulus-level variance were included as random effects, and the predictor variables were race of face, race of participant, and RKG response. The regression model was built in two steps, shown in Table 16. The first step (shown in the left hand columns), included the main effects of race of participant and race of face. White participants were more likely to be accurate (67.84%) than Black participants (60.80%), $\beta_1 = .38$, $p = .01$. Overall, context judgments for White faces (66.79%) were more accurate than context judgments for Black faces (61.72%), $\beta_2 = .22$, $p = .02$.

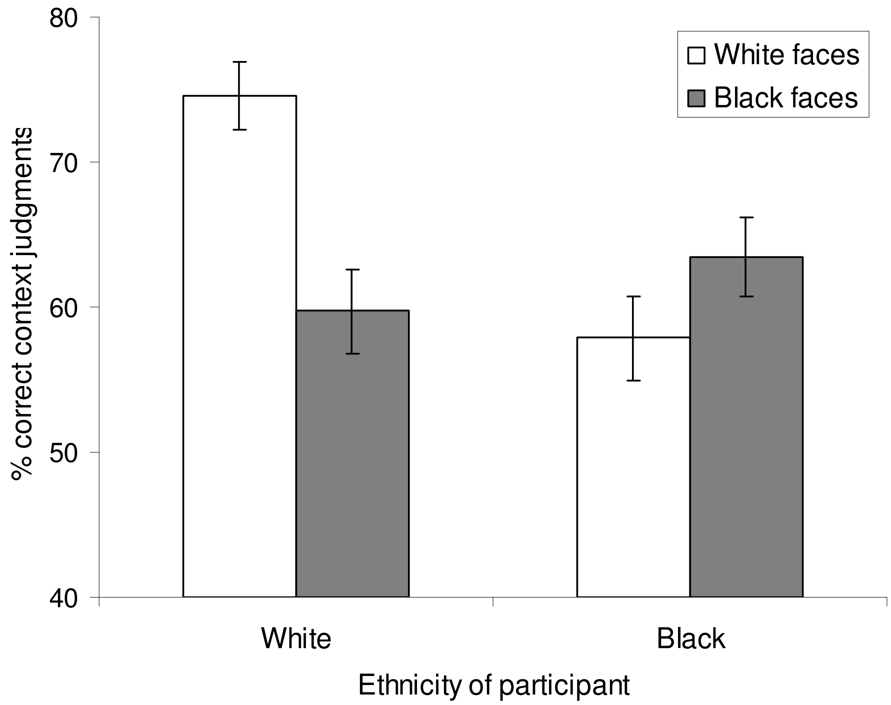
In the second step of the model, the interaction between race of participant and race of face was entered. The interaction term was significant, $\beta_3 = .86$, $p < .001$. This interaction is shown in Figure 8. The figure shows that White participants were more accurate with White faces than with Black faces. For Black participants, there seems to be a smaller difference in the opposite direction, with higher accuracy for Black faces than for White faces. To confirm these observations, separate regressions were run on the data from White and Black participants. In both cases, race of face was a significant predictor of accuracy: for White participants, $\beta = .80$, $p < .001$, and for Black participants, $\beta = -.29$, $p = .01$.

Table 16: ***Comparison of regression models predicting context accuracy from race of participant, race of face, and Remember-Know-Guess response in Experiment 8.***

		Model 1: Without ORB			Model 2: With ORB		
Parameter		Estimate	SE	Sig.	Estimate	SE	Sig.
Fixed effects:							
Intercept	β_0	.32	.12	.007	.99	.12	<.001
Race of participant	β_1	.38	.15	.01	.05	.17	.75
Race of face	β_2	.22	.10	.02	.25	.11	.02
ORB interaction	β_3	------	----	------	.86	.13	<.001
Remember-Know	$\beta_4^{(a)}$	------	----	------	.86	.08	<.001
Remember-Guess	$\beta_4^{(b)}$	------	----	------	1.45	.11	<.001
Guess-Know	$\beta_4^{(c)}$	------	----	------	.59	.11	<.001
Random effects:							
Participant level	σ^2_u	.39			.39		
Stimulus level	σ^2_e	.20			.13		

Difference in model fit: $\chi^2(4) = 298.66$, $p < .001$

***Figure** 8*: The percentage of correct context recognition judgments on hit trials, broken down by race of face and race of participant in Experiment 8. Error bars represent 95% confidence intervals.



RKG responses were also added to the second step of the regression model. Context judgments were more likely to be accurate in Remember trials (73.67%) than in Know trials (54.11%), $\beta_4^{(a)} = .86$, $p < .001$, or Guess trials (39.83%), $\beta_4^{(b)} = 1.45$, $p < .001$. Know trials were also more accurate than Guess trials, $\beta_4^{(c)} = .59$, $p < .001$.

Context responses always followed old decisions and RKG judgments. A possible confound could have occurred if the RKG responses influenced participants' later abilities to remember contextual information. For example, providing a Remember judgment may have generated additional retrieval cues, allowing easier access to the representation of the encoding context. A further regression was run, using only data from Remember trials (*N* = 2909 trials). Again, the interaction between race of participant and race of face was a significant, $\beta_3 = .88$, $p < .001$. White participants were more accurate with White faces (82.13%) than with Black faces (70.02%), $\beta = .71$, $p < .001$. Although Black participants were slightly more accurate with Black faces (70.92%) than with White faces (67.68%), this difference was non-significant, $\beta = .18$, $p = .19$.

To confirm the validity of these analyses, the context data were used to calculate a measure of source identification which is independent of item recognition, the *average conditional source identification measure* (ACSIM; see Murnane & Bayen, 1996). The calculation of this measure is based upon the ratio of correct to incorrect context judgments for faces correctly judged as old. ACSIM values can vary between 0 and 1, with higher scores indicating more accurate memory for context. The procedure for calculating these values is described in Appendix 3. A 2 (race of face) x 2 (race of participant) mixed ANOVA was run on the ACSIM data. The mean ACSIMs are shown in Table 17, broken down by race of face and race of participant.

The main effects of Race of participant and Race of face were significant. Overall, accuracy was higher for White faces ($M = .66$, $SD = .17$) than Black faces ($M = .61$, $SD = .18$), $F(1, 82) = 8.71$, $p = .004$, $d = .28$. White participants were more accurate ($M = .66$, $SD = .15$) than Black participants ($M = .60$, $SD = .14$), $F(1, 82) = 3.79$, $p = .06$, $d = .43$. However, these main effects were qualified by a significant interaction, $F(1, 82) = 36.57$, $p < .001$, $\eta_p^2 = .31$. White participants were significantly more accurate with White faces than with Black faces, $t(40) = 6.13$, $p < .001$, $d = .92$, while Black participants were more accurate with Black faces than with White faces, $t(43) = 2.27$, $p = .03$, $d = .34$. The ACSIM analysis confirms the findings from the regression analysis, that memory for context is impaired in other-race identifications compared to own-race identifications.

Table 17: ***Mean ACSIM values for White and Black faces, broken down by race of participant in Experiment 8.***

Race of participant	Mean		Standard deviation	
	White faces	Black faces	White faces	Black faces
White	.74	.59	.15	.19
Black	.57	.63	.14	.18

Discussion

This paper presented a single experiment, in which White and Black participants living in South Africa viewed White and Black faces presented on different contexts. In a recognition test, participants attempted to discriminate old faces from new faces, and also made RKG judgments and context judgments for faces identified as old. The three main aims of this paper were as follows. First, to show that recognition of other-race faces is characterized by a marked decrease in Remember responses when Tulving's (1985) Remember-Know procedure is used. Second, to replicate the context impairment found for other-race faces in Paper 4 with a crossover sample. And third, to investigate the ORB in a population outside of North America or Europe.

The analyses of the RK judgments clearly showed that the ORB has dissociable effects on remembering and knowing. White participants made significantly more R judgments for White faces than for Black faces, while Black participants showed a non-significant trend in the opposite direction. Consistent with the findings of Meissner et al. (2005), differences in knowing were smaller than differences in remembering. However, K responses are difficult to interpret within the RK paradigm. The analysis of RK responses is a contentious issue, as there are still debates concerning the relationship between recollection and familiarity. The decision to use the IRK procedure (Yonelinas & Jacoby, 1995) or to analyze the proportion of K responses (e.g. Gardiner et al., 1996) can influence the pattern of results. This was the case in Experiment 8. When the proportion of K responses was analyzed, White participants made significantly more K responses to Black faces than White faces. When familiarity was estimated using the IRK procedure, both groups of participants had slightly higher familiarity estimates for own-race faces than for other-race faces. These two analyses could lead to quite different conclusions. Therefore, one should always be very careful when discussing RK responses as if they map directly on to underlying memory processes. Gardiner et al. (1996) argue that RK responses are best treated as what they are – a participant's subjective reports of their state of awareness associated with a recognition decision.

Participants from both groups were less accurate at identifying the contexts associated with other-race faces than with own-race faces. Analyzing memory for context can be difficult, as context recognition can be confounded with item recognition. To confirm the validity of the ORB for context memory, the analysis was conducted in two ways. First, a multilevel regression was run on hit trial data only; second, ACSIM values were analyzed in a traditional ANOVA (Murnane & Bayen, 1996). Using both of these methods, the results confirmed those of Paper 3, while expanding this finding to Black participants as well as White participants.

Context memory was much more accurate on R trials than on K trials. This is not surprising, as reports of Remembering are often associated with increased memory for source information than reports of Knowing (Meiser & Sattler, 2007). Yet K and G responses also produced correct context judgments at a higher rate than chance. It may be that the K and G responses were not “process pure”, and that they included some identifications in which there was some degree of recollection present. It is also possible that some context memory can be supported by familiarity. This position is supported in the implicit learning literature (Scott & Dienes, 2008).

Most ORB studies have been carried out in North America or Europe. A handful of studies carried out in Africa have produced inconsistent results, with some but not others finding crossover effects (Chiroro et al., 2008; Chiroro & Valentine, 1995; Wright et al., 2001, 2003). Here, participants from two racial groups living and studying in the same city in South Africa were recruited. The face recognition data showed the typical ORB, with higher accuracy for own-race faces than for other-race faces. The ORB is often particularly strong for false alarm rates (e.g. Chiroro & Valentine, 1995; Ng & Lindsay, 1994; Slone, Brigham, & Meissner, 2000), with accompanying hit rate differences also sometimes found (Meissner et al., 2005; Teitelbaum & Geiselman, 1997). Here, crossover interactions were found for both hit rates and false alarm rates, although the effects were larger the false alarm rates than for the hit rates.

For every measure reported, the effect size was substantially larger for the White participants than for the Black participants. Asymmetric effects are quite common in the ORB literature (e.g. Ferguson, Rhodes, Lee, & Sriram, 2001; Wright et al., 2003), with larger effects often being found in participants from the majority group than for participants in the minority group. Members of minority groups are likely to have more opportunity for contact with members of majority groups than vice versa, due to the numerical frequency of out-group members in their environment (Anthony, Copper, & Mullen, 1992). Minority groups are also usually associated with low status, and are reliant on members of the high-status majority group for employment and education (Islam & Hewstone, 1993a).

South Africa provides a particularly interesting location in which to study cross-race effects, due to its relatively recent history of segregation. Uncommonly, the numerical minority group (Whites) in South Africa had high status and power, while the numerical majority group (Blacks) had low status and power. Cape Town is a particularly diverse city, and the University of Cape Town, from which the participants were all recruited, enrols approximately 50% Black students and 50% White students. Members of both groups should therefore have had ample opportunity for contact with individuals from the out-group. However, opportunity for contact does not necessarily translate into actual meaningful contact. A recent study conducted on desegregated beaches in Cape Town showed that people still almost exclusively associate with members of their own race (Dixon et al., 2005). Self-reported contact was not recorded in this study, so one can only speculate on why the effect sizes were larger for the White participants than for the Black participants. However, as members of a group with a history of low status, the Black participants may have found interacting with White authority figures necessary throughout their lives. Being able to accurately discriminate between out-group members may have therefore been more important for the Black participants than for the White participants.

The results of this paper have been interpreted within a dual-process framework. Supporters of dual-process models argue that Remember and Know responses affect recollection and familiarity, respectively (e.g. Gardiner et al., 1996, 2002), and that memory for contextual information relies heavily on recollection (Meiser & Sattler, 2007; Meiser et al., 2008). The current study, when related to dual-process theory, suggests that recollection is specifically impaired for other-race faces. However, these results of could also be consistent with single-process theories of memory. Dunn (2004) and Donaldson (1996) argued that Remember responses are made to items recognized with high confidence, and that apparent dissociations between R and K responses can be explained by a simple memory strength model. DeCarlo (2003) also argued that source memory can be explained by an extended multivariate signal detection model, which does not assume the existence of two separate memory processes. Glanzer, Hilford, and Kim (2004) showed that Receiver Operating Characteristics (ROCs) for item recognition and source recognition share similar properties (e.g. the ROCs are convex, the *z*-ROCs are linear), and that items which influence item recognition have very similar effects on source recognition. A memory strength interpretation of the current results would suggest that own-race faces produce stronger activation that other-race faces. This leads to decreased item recognition, fewer reports of Remembering, and impaired memory for contextual details of other-race faces.

In conclusion, across two experiments, an own-race bias in face recognition and in memory for context was shown. The effects were larger for White participants than for Black participants. Own-race faces were more likely to be given Remember responses than other-race faces. Dividing attention at encoding reduced overall accuracy for both face recognition and context memory, but did so at a similar rate for own- and other-race faces. These results are generally consistent with the dual-process approach to the ORB proposed by Meissner et al. (2005).

Thesis Discussion

The own-race bias (ORB) is the well-established finding that people recognize faces of their own race more accurately than faces of other races (Meissner & Brigham, 2001). Several theories have been proposed as to why and how this bias develops. For example, lifelong contact with members of one's own group creates perceptual expertise which is lacking for other-race faces (Chiroro & Valentine, 1995; Hancock & Rhodes, 2008; Sporer 2001b). Other-race faces are then processed less holistically than own-race faces (Michel, Caldara, & Rossion, 2006; Michel, Rossion, Han, Chung, & Caldara, 2006; Hancock & Rhodes, 2008), or inappropriate features are encoded (Ellis, Deregowski, & Shepherd, 1975; Lindsay, Jack, & Christian, 1991). These faces will be represented in long-term memory in a way that is detrimental for later recognition (Valentine, 1991).

Meissner and colleagues (Evans, Marcon, & Meissner, 2009; Marcon, Susa, & Meissner, 2009; Meissner, Brigham, & Butz, 2005) recently approached the ORB from a dual-process perspective. Dual-process theories of memory posit that there are two separate routes to recognition: through an episodic process of recollection, and through an automatic process of familiarity (Yonelinas, 2002). Although both processes can lead to correct recognition judgments, recollection more often produces correct decisions than familiarity, and is necessary for the retrieval of contextual information about the encoding event in which a stimulus was seen (Mandler, 1980). Meissner and colleagues have shown that recollection is impaired for other-race faces, but familiarity is preserved.

The main aims of this thesis were to expand upon the work of Meissner et al. (2005) by providing further evidence for the dual-process account of the ORB, and to show that this recollection impairment for other-race faces leads to inferior contextual memory. Eight experiments were reported across four papers, using methods such as Jacoby's (1991) process-dissociation procedure (PDP) and Tulving's (1985) Remember-Know (RK) task to estimate the relative contributions of recollection and familiarity to recognizing own- and other-race faces. Factors believed to influence recollection, such as depth of processing and context reinstatement were also manipulated. The results of these studies will be summarized below. This summary will be followed by more in-depth discussions of the theoretical and applied implications of the data, as well as a discussion of the limitations of these studies, and future directions for research.

1. Summary of Results

In Paper 1, Jacoby's (1991) PDP was used to estimate the contributions of recollection (*R*) and familiarity (*F*) to White participants' recognition of White and Black faces. Based on previous findings of Meissner et al. (2005) and Marcon et al. (2009), it was predicted that the estimated *R* parameter would be higher for the White faces than for the Black faces. The estimated *F* parameter was predicted to be similar for White and Black faces. In Experiment 1, these hypotheses were supported. The participants made more identifications of Black lures than White lures in the exclusion trials, which produced a significantly higher *R* estimate for the White faces than for the Black faces. The effect size for this difference was large (d = 1.05). The estimated *F* parameter, on the other hand, was not significantly different for the White and Black faces (d = .13).

Level of processing (LOP) was manipulated in Experiment 2. The depth of encoding strategy influences recollection to a much larger extent than familiarity (Gardiner, Java, & Richardson-Klavehn, 1996; Yonelinas, 2002). Some authors have argued that people process other-race faces less deeply than own-race faces (e.g.

Chance & Goldstein, 1981). The predictions for Experiment 2 were that as the depth of processing increased, the estimated *R* parameter would increase, while the estimated *F* parameter would remain relatively constant. Control participants who were given no LOP instructions would perform similarly to those in Experiment 1, with a higher *R* estimate for the White faces than for the Black faces. Participants in the shallow and neutral conditions, on the other hand, would not show such a difference.

The results were generally supportive of these hypotheses. For both White and Black faces, the *R* estimate was higher in the deep condition than in the shallow condition. The *F* estimate did not vary across the LOP conditions. Participants in the shallow and deep conditions showed little difference in recollection for the White and Black faces ($d = .13$ and $d = .14$ respectively). The control participants had somewhat higher *R* estimates for the White faces than for the Black faces ($d = .36$), although this difference was not statistically significant. The lack of significance could have been due to low power, with 20 participants in each LOP condition. Nevertheless, the results were in the predicted direction. Paper 1 therefore supported the hypothesis that other-race faces are less likely to be recollected than own-race faces, and suggested that this could be due to less elaborate processing at encoding.

Paper 2 used a procedure similar to the exclusion trials of the PDP. White participants studied White and Black faces, presented with one of two verbal labels – "criminal" or "non-criminal". At testing, the participants were asked only to identify faces that had appeared with one of the two labels (target faces), while rejecting faces associated with the other label (distractor faces). As both groups of faces are equally familiar, recollection is required in order to discriminate between targets and distractors. The main prediction was that participants would be less able to make this discrimination for the Black faces than for the White faces, and that they would therefore identify similar numbers of Black targets and Black distractors. This hypothesis was supported in Experiment 3. Participants identified significantly more White targets than White distractors (*d* = 1.13), but identified Black targets and Black distractors at a similar rate (*d* = .24). There were at least two explanations for these results. First, that the participants were unable to successfully retrieve recollective information about the Black faces. And second, that the participants were not attempting to retrieve recollective information about the Black faces.

Experiment 4 attempted to tease these two explanations apart, by manipulating the instructions given to participants at testing. The task was framed as a "stop and search" exercise, and participants were allocated to conditions with strict or lenient response instructions at testing. If participants are *unable* to recollect other-race faces, then these instructions should not have influenced the difference between target and distractor identification rates. If participants are simply *unwilling* to attempt to recollect other-race faces, then the instructions should have influenced the difference between target and distractor identification rates. The results showed that the participants in both the strict and the lenient conditions were able to discriminate between Black targets and Black distractors. It is possible that the stop and search framework encouraged participants to respond more carefully, and that the strict and lenient instructions therefore had only a small effect. Overall, these results were consistent with the hypothesis that, under some conditions, participants are able to recollect other-race faces as accurately as own-race faces. The recollection impairment seen in Papers 1 and 2 may therefore be a result of reduced cognitive effort when responding to Black faces than when responding to White faces.

Paper 3 investigated White participants' memories for contextual information associated with correctly recognized White and Black faces. At encoding, each target face appeared on one of several backgrounds. At testing, the target faces were seen again, along with a number of new faces. Participants made an old/new decision and a confidence judgment for each face. For faces identified as old, context judgments were also required. A successful context judgment requires a recollective memory of the target stimulus as it appeared during the original encoding event. Based on the findings of Papers 1 and 2, the main prediction for Paper 3 was that context judgments would be less accurate for Black faces than for White faces.

Across three experiments, this hypothesis was confirmed. The participants consistently recalled the contexts in which White faces had appeared more accurately than Black faces. This effect remained significant even after participants' confidence ratings had been accounted for. Context accuracy was above chance in all studies, but also relatively modest. This confirms previous research, which has shown that recalling contextual information about familiar stimuli is reasonably difficult (Brown, Deffenbacher, & Sturgill, 1977). Experiments 5 and 6 used meaningful and complex visual scenes as contexts, while Experiment 7 extended this effect to simple coloured backgrounds. The replication of this effect across different backgrounds suggests that the effect was not a product of the particular scenes which were chosen.

Experiment 6 investigated the influence of context reinstatement on face and context recognition accuracy. Reinstating elements of the encoding context at testing has larger benefits for recollection than for familiarity (Gruppuso, Lindsay, & Masson, 2007; Macken, 2002). Recognition improvements following context reinstatement are also larger for relatively unfamiliar classes of stimuli (Russo, Ward, Geurts, & Scheres, 1999). Therefore, it was predicted that congruency between encoding and testing contexts would increase both face and context recognition accuracy, and that the effect would be larger for the Black faces than for the White faces. There was a small yet significant improvement in face recognition accuracy when the contexts were congruent compared to when they were incongruent ($d = .20$). However, this improvement was of a similar size for the White and the Black faces. Context reinstatement did not influence the accuracy of the context judgments. The small effects found are perhaps not surprising. Context reinstatement effects are often larger in recall studies than in recognition studies, with some recognition studies finding no benefit of reinstating contextual information (Macken, 2002). The presence of a target item itself is such a powerful cue for recognition that the context may have no further real value. This may have been the case here, with only small differences in accuracy between congruent and incongruent trials.

Paper 4 extended the findings of Paper 3 to samples of Black participants as well as White participants. Both groups of participants lived in a cosmopolitan and racially diverse city, Cape Town. Both groups of participants showed own-race biases in their item recognition and in their context memory accuracy. The effect sizes were larger for the White participants than for the Black participants. Although contact was not measured in any of the samples, it is possible that the Black participants had had extended meaningful contact with White authority figures throughout their lifetimes, thus reducing the magnitude of their ORB compared to the White participants. Asymmetric effects are common when groups differ in social status (e.g. Islam & Hewstone, 1993a, 1993b), and when one group is reliant on another for many aspects of their daily lives (education, employment, etc.).

In Experiment 8, Tulving's (1985) RK procedure was used to investigate participants' subjective reports of remembering and knowing. Participants were expected to make more Remember (*R*) responses to own-race faces than to other-race faces, while proportions of Know (*K*) responses were expected to be more similar for own- and other-race faces. RK responses were also expected to be strong predictors of context memory accuracy. The results confirmed these hypotheses. Both groups of participants made more *R* responses to own-race faces than to other-race faces, although the difference was not significant for the Black participants. White participants made more *K* responses for Black faces than for White faces, while Black participants made approximately the same number of *K* responses for White and Black faces. RK responses strongly predicted accuracy. However, even after controlling for RK responses, the ORB for context memory was still present. These results suggest that remembering and knowing play some role in producing other-race context impairment, but that there is some variance left unexplained. RK responses are unlikely to be "process pure" (Jacoby, 1991), such that they will not map directly onto the processes of recollection and familiarity. Some of the unexplained variance may be a result of measurement error within the RK procedure.

Across eight studies, recollection was shown to be impaired for other-race faces compared to own-race faces. As a result of this, participants had difficulty in retrieving contextual information associated with correctly recognized other-race faces. The implications of these results for the theoretical accounts of the own-race bias will be discussed below, followed by a discussion of the practical implications of these data.

2. Theoretical Implications

2.1. Theories of the own-race bias

Explanations of the ORB can be broadly split into two main camps. Perceptual expertise accounts suggest that extensive lifelong contact with people from our own group finely tunes our cognitive systems to be optimal for processing and recognizing own-race faces (e.g. Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005; Valentine, 1991). Socio-cognitive accounts suggest that processes such as stereotyping and out-group categorization interfere with the efficient processing of other-race faces, which impairs later attempts at recognition (e.g. MacLin & Malpass, 2001; Shriver, Young, Hugenberg, Bernstein, & Lanter, 2008; Sporer, 2001b). Both of these positions predict that sub-optimal encoding strategies are used when processing other-race faces. This may involve the use of featural processing rather than holistic processing (Hancock & Rhodes, 2008; Tanaka, Kiefer, & Bukach, 2004), the processing of inappropriate facial features (Ellis et al., 1975; Hills & Lewis, 2006), or the use of less elaborate encoding for other-race faces than for own-race faces (Chance & Goldstein, 1981).

Any of these social and cognitive mechanisms would lead to other-race faces being represented in long-term memory more poorly than own-race faces. The memorial representations of other-race faces could be generally weaker than the memorial representations of own-race faces, with less associative information and semantic links attached. Or other-race exemplars could be stored in densely clustered regions of a multi-dimensional face-space (Valentine, 1991), rendering later attempts at recognition more difficult.

Whichever theoretical stance one takes concerning the underlying mechanisms of the ORB, one would ultimately predict that other-race faces are encoded less effectively than own-race faces, and are therefore poorly represented in memory. The dual-process perspective of the ORB (Meissner et al., 2005) therefore fits reasonably well with the varying accounts of the effect. Insufficient processing at encoding and weak memory representations of other-race faces should create difficulties in retrieving recollective information about those faces. For example, if less elaborative encoding strategies are used for other-race faces than for own-race faces, then there will be fewer semantic links associated with those faces, producing fewer cues for recollection. Or if other-race faces are densely clustered in face-space, then the increased subjective similarity of a target face to its surrounding exemplars will reduce the likelihood of a successful recollective search.

The dual-process account of the ORB has several advantages. First, it gives some insight into the subjective experiences of participants when trying to recognize other-race faces, which can lead to new ways of attempting to reduce the magnitude of the ORB. Second, by separating the recollection component of recognition memory from the familiarity component, one can make more subtle predictions about the effects of some variables upon the ORB, such as divided attention and levels of processing. And third, it allows new questions to be asked about the ORB, such as whether participants will be able to accurately recall contextual information associated with other-race faces.

Throughout this thesis, several factors were manipulated which were expected to influence the magnitude of the ORB, either by reducing accuracy for own-race faces or by increasing accuracy for other-race faces. The ORB remained robust to manipulations of response criterion and context reinstatement, despite significant main effects of these factors upon accuracy. In order to detect significant interactions, higher statistical power may have been required. However, the magnitude of the ORB *was* influenced by a levels-of-processing (LOP) manipulation. When participants were given no encoding instructions, they were significantly more accurate with own-race faces than with other-race faces. However, when the participants were encouraged to use deep or shallow encoding strategies, this difference disappeared. These results suggest that participants who are given no encoding instructions tend to process own-race faces more deeply than other-race faces, and that this contributes to the ORB in recognition.

Overall, the results presented in this thesis are consistent with the dual-process account of the ORB proposed by Meissner et al. (2005). Recollection does appear to be impaired for other-race faces, while familiarity is preserved. As a result of this recollection deficit, participants make more errors when trying to distinguish between familiar other-race faces from different sources. Participants are also less accurate at recalling specific contextual information associated with other-race faces. The implications of these results for dual-process theories of memory will be discussed below, followed by a discussion of the applied importance of these data.

2.2. Dual-process theories of recognition memory

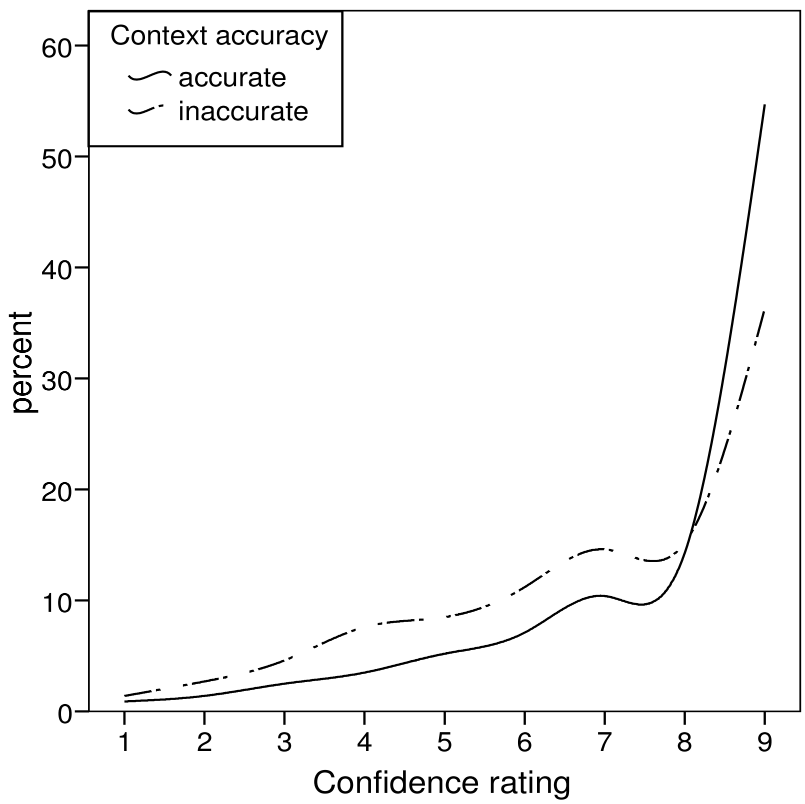
There is considerable debate among memory theorists concerning the number of processes involved in recognition memory. Dual-process theories compete with single-process models of memory, which argue against the existence of a separate recollection process, and for a single signal detection-type process (Swets, 1964). Supporters of these models argue that many of the apparent dissociations which have been found between recollection and familiarity can also be explained by the more parsimonious single-process models.

Consider the RK procedure (Tulving, 1985). Dual-process theorists argue that the consistency of RK responses across studies, coupled with the dissociable effects of many factors on *R* and *K* responses, confirm the validity of the RK procedure (Yonelinas, 2002). However, single-process theorists argue that participants simply make *R* responses to items recognized with high confidence, and *K* responses to items recognized with lower confidence (Donaldson, 1996; Dunn, 1994). Donaldson suggests that participants have two decision thresholds which are applied to one underlying process. The first threshold will govern whether an item is judged as old or new. The second threshold, placed higher on the memory strength curve, will govern whether *R* or *K* responses are given. Items in between the two thresholds will be given *K* responses, whereas items above the second threshold will be given *R* responses.

In Paper 3, confidence ratings were collected for each identification decision before a context judgment was made. Correct context judgments are more likely to be made for recollected items than are incorrect judgments (although the mapping between context accuracy and recollection is unlikely to be exact). Donaldson's (1996) single-process theory would predict that inaccurate context trials should be associated with a range of different confidence ratings, whereas accurate context trials should be associated almost exclusively with very high confidence ratings.

Figure 10 shows the percentages of accurate and inaccurate context trials given confidence ratings of 1 to 9. The confidence ratings are shown on the x axis, and the percentages are shown on the y axis. So, for example, approximately 5% of accurate context trials and approximately 10% of inaccurate context trials were rated with a confidence of 5. It is immediately obvious that the shapes of the curves are quite similar. For both accurate and inaccurate trials, there is a shallow increase in confidence ratings from 1 to 8, followed by a sharp increase in confidence ratings of 9. Approximately 55% of accurate context trials were given confidence ratings of 9, as were approximately 35% of inaccurate context trials. This suggests that there are situations is which a person can very confidently recognize a face without being able to retrieve recollective information about its previous occurrence.

Figure 9: The percentages of accurate and inaccurate context recognition trials given confidence ratings of 1 to 9 in Experiments 5, 6, and 7.



There is also considerable variation in confidence ratings for both accurate and inaccurate context trials. Roughly 45% of accurate context trials were given confidence ratings between 1 and 8, and roughly 65% of inaccurate context trials were given confidence ratings between 1 and 8. This variation suggests that recollective information about a face's previous occurrence can be retrieved even when the face itself is recognized with medium or low confidence. These patterns of results are inconsistent with Donaldson's (1996) argument that recollection is nothing more than high confidence recognition.

Throughout this thesis, several factors were manipulated which have been shown to have dissociable effects on recollection and familiarity. Of all of the factors which were tested, the most successful one was the LOP manipulation in Experiment 2. LOP effects are stronger for estimates of recollection than familiarity (Gardiner et al., 1996; Yonelinas, 2002), as deeper or more elaborate processing provides more cues for later recollection. In line with previous research, recollection estimates from the PDP increased as depth of processing increased; familiarity estimates were unaffected. However, the other manipulations tested in this thesis were not so successful at dissociating recollection and familiarity. While context reinstatement influenced overall accuracy, it did so at similar rates for own- and other-race faces. As own-race faces are more likely to be recollected than other-race faces, this manipulation should have influenced the size of the ORB. However, in order to detect such effects, greater statistical power may have been required.

3. Practical Applications

When recognizing faces in an old/new recognition test, it is usually sufficient to decide whether faces are familiar or not. The familiarity associated with a test stimulus can likely be attributed to having encountered that stimulus during the encoding phase. However, when recognizing faces in the real world, one often wishes to know more than whether a face is familiar or not. Retrieving information about the source of a face's familiarity will be crucial for guiding social interactions. The information that one is able to retrieve about a face will influence decisions about whether to interact with that person, and what form that interaction should take. In the eyewitness world, the retrieval of accurate source information by a witness regarding a familiar suspect will be essential in ensuring that the correct person is identified. In order to illustrate this point, some real life cases from the Innocence Project (www.innocenceproject.org) will be discussed.

Timothy Cole, an African-American man, was convicted of raping a Caucasian woman. The witness was shown a biased photographic lineup. In this lineup, Cole's photograph was the only color photograph, presented in a lineup of black and white mugshots. Cole was facing the camera, whereas all of the lineup fillers were facing to one side. The witness identified Cole as her attacker. Some time later, the witness was shown a live lineup, and identified Cole for the second time. On the basis of this evidence, Cole served thirteen years before dying in prison. He was posthumously exonerated on the basis of DNA evidence.

Gilbert Alejandro served three and a half years in prison for rape, based on an inaccurate identification by the victim. The witness got only a brief glimpse of her attacker during the crime, as her face was covered by a pillow. The witness was repeatedly shown photographic lineups containing Alejandro. She did not identify him as her attacker. She later identified Alejandro from a sketch lineup and a live lineup. The real perpetrator is still unknown.

A. B. Butler, an African-American man, was convicted of kidnapping and raping a Caucasian woman, for which he served 16 years in prison. The victim initially identified Butler during a mugshot search. Some time later, she identified Butler again in a live lineup. Butler was exonerated based on an analysis of DNA evidence from the crime scene. The real perpetrator has not yet been found.

Above are just three cases of mistaken eyewitness identifications which have resulted in miscarriages of justice. In all of the above cases, it seems likely that failures to identify the source of the suspect's familiarity played at least some role in the eyewitness's decision to ultimately identify the suspect. In the cases of Cole and

Alejandro, the witnesses were given multiple lineups including the suspect. The witness in Alejandro's case did not identify him until after seeing his image in several photographic lineups. In Cole's case, the original photographic lineup was biased, indicating very clearly to the witness which lineup member was the suspect. In Butler's case, the identification was made following a mugshot search, during which the witness picked out Butler.

Laboratory studies have consistently shown that viewing an innocent suspect more than once, in multiple lineups or in mugshot searches, greatly increases the likelihood of mistaken identifications (see Deffenbacher, Bornstein, & Penrod, 2006, for a review). In these situations, the witness may unconsciously transfer the identity of the suspect onto the identity of the attacker (Loftus, 1976). The results of this thesis show that recollection, which is crucial for retrieving source information about familiar stimuli, is impaired for other-race faces. As a result of this, people struggle to discriminate between other-race faces from two different sources (Papers 1 and 2), and have difficulty in recalling specific details about contextual details associated with other-race faces (Papers 3 and 4). These studies therefore indicate that unconscious transference errors in eyewitness identifications may be more common when the suspect and the victim/witness are of different races. Of the three cases presented above, two involved cross-racial identifications. This will be discussed further in the following section, on limitations and future directions.

4. Limitations and Future Directions

Faces form a very variable class of objects, which differ in many ways. Whenever a set of faces is selected for use as stimuli in an experiment, those faces are a small sample of a much lager population of faces. Just as when recruiting participants, one hopes that the stimuli are representative of the wider population of faces from which they were sampled. Stimulus sampling errors occur when the stimulus set is not representative of the population (Wells & Windschitl, 1999). Such errors can lead to unjustified conclusions, which are an artefact of the particular stimuli used in the experiment. Stimulus sampling errors are of particular concern in studies of the ORB, which involve comparing recognition accuracy for two different stimulus sets against one another. If the two stimulus sets differ in their representiveness of the populations from which they are sampled, or in the amount of variability that they show, then the results will be unreliable and may not be replicable with other stimulus sets.

In order to increase the likelihood that the results from ORB studies are valid, it is ideal to recruit participants from both racial groups represented by the stimulus sets (Wells & Olson, 2001). If both groups of participants are more accurate with own-race faces than with other-race faces, then stimulus sampling explanations become implausible. Increasing the number of stimuli used also decreases the likelihood that the results will be artefacts of the idiosyncrasies of the particular faces sampled (Sporer, 2001a). Of the nine studies presented in this thesis, seven recruited participants from only one racial group, and two recruited cross-race samples. This could potentially limit the generalizabilty and reliability of these results. However, the main findings of Papers 1 to 3 were supported when a cross-race sample was recruited in Paper 4, indicating that the stimuli were not producing artifactual effects. The results were also replicated across two different stimulus sets (a different stimulus set was used in Paper 3). Furthermore, large numbers of stimuli were used in these studies (80 faces from each racial group in Papers 1, 3, and 4), reducing the impact of any particularly distinctive individuals on the overall results. Thus, while cross-race samples could not be recruited for every study due to constraints on time and resources, I am confident that these results are both reliable and generalizable.

While considering the meaning of these studies, it is important to bear in mind that the dual-process theory of memory is still a topic of considerable debate. In particular, some of the methods used in this thesis, such as the RK procedure and the PDP, have themselves received some criticism. For example, some authors argue that participants do not understand the distinction between remembering and knowing, and so use *R* and *K* responses as proxies of high and low confidence judgments respectively (Donaldson, 1996; Dunn, 2004). And within the PDP, the underlying assumptions of independence between the recollection and familiarity processes have been attacked (Curran & Hintzman, 1995; McBride, 2007). However, the consistency of findings presented in this thesis using a number of different methods speaks to the reliability of the results found.

A final major limitation of the studies in this thesis is that they are low in ecological validity. One must therefore exercise caution when considering the potential applications of these data to real-world situations such as eyewitness identifications. Old/new recognition tasks place very different demands on participants than a lineup places on an eyewitness. The quality of view in all of these studies was optimal, and ranged between 5 and 10 seconds in duration. A witness could view a culprit under sub-optimal conditions, and could view them from seconds to minutes or possibly even hours. Many other factors which could influence a witness's accuracy, such as stress, alcohol intoxication, and weapon focus, were not taken into account in these studies. The participants in these studies were also aware that there were no consequences for the individuals who were identified – this would obviously not be true for an eyewitness making an identification.

However, far from being discouraged by these limitations, one should view them as exciting new directions for future research. Will these results replicate to lineup paradigms? What would be the effect of factors such as stress, alcohol, and weapon focus on recollection and familiarity? How are the two processes influenced by changes in viewing quality? Will mugshot exposure be more detrimental to cross-race lineup identification than same-race lineup identification? And would participants be more likely to misidentify innocent bystanders of a different race than of their own race? The results of this thesis certainly suggest that this would be the case, and so these questions offer promising lines of research.

As well as these more "applied" lines of research, future studies could also continue to explore the role of recollection within the ORB. An interesting idea which has emerged very recently is that of "graded recollection" (Mickes, Wais, & Wixted, 2009; Palmer, Brewer, McKinnon, & Weber, in press). Traditionally, recollection has been characterized as an all-or-nothing process, in which an item is either recollected or not recollected (Yonelinas, 2002). However, Mickes et al. suggested that recollection may be a continuous process, in which any number of elements of an encoding event could be recalled. For example, a participant could recollect a thought that they had upon seeing one target, but they may recall thoughts, associations, and contextual information about another target. While both of these targets are recollected, the second target elicits more detail than the first target. Mickes et al. argued that participants will be more confident for items which are recollected in greater detail than items which are recollected in less detail. Confidence and recollection should therefore be associated in a positive linear manner. This seems to be the case with the data from Paper 3 (see Figure 10, p. 160). Palmer et al. asked participants to report their memories of identified targets following *R* and *K* responses, and found evidence consistent with this graded recollection process. Future research could similarly analyze the contents of self-reported memories for own- and other-race faces to discover whether participants report more details for recollected own-race faces than for recollected other-race faces.

5. Conclusions

This thesis aimed to investigate the roles of recollection and familiarity in the own-race bias for face recognition. Based on the work of Meissner and colleagues (Evans et al., 2009; Marcon et al., 2009; Meissner et al., 2005), it was predicted that recollection would be impaired for other-race faces, but that familiarity would be preserved. These predictions were supported across nine studies using several different methods, including Jacoby's (1991) PDP and Tulving's (1985) RK procedure. Because of this deficit in recollection, participants make more errors when attempting to distinguish between equally familiar other-race faces from two different study lists. They also struggle to recall specific contextual information associated with other-race faces at encoding. These effects appear to be influenced by cognitive processing strategies at encoding and at testing, suggesting that there may be scope to improve recollection for other-race faces with carefully worded instructions.

The results of this thesis provide important advances in both our theoretical and our applied understanding of the ORB. The dual-process approach to the ORB opens up new doors for future research, allowing new questions to be asked. If this line of research is continued, it may well be possible to find ways of reducing the ORB. Future research should also focus on source monitoring errors in cross-race identifications, and what can be done to reduce the likelihood of such errors.

In conclusion, the ORB can be viewed as a specific deficit in recollection for own-race faces compared to other-race faces. Recollective memory plays important roles in the real world, helping us to regulate our social interactions, and helping us to recall in which situations we have previously encountered others. Deficits in recollection can have far-reaching consequences, particularly within the realm of the legal system. This thesis provides a stepping stone to new lines of research into both the theoretical basis of, and the practical implications of the own-race bias in face recognition.

References

Abshire, J. & Bornstein, B. H. (2003). Juror sensitivity to the cross-race effect. *Law & Human Behavior,* **27**, 471-480.

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.

Allport, G. W. (1954). *The nature of prejudice.* Cambridge, MA: Addison-Wesley.

Anastasi, J. S. & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin and Review,* **12**, 1043-1047.

Anthony, T., Copper, C., & Mullen, B. (1992). Cross-racial facial identification - a social cognitive integration. *Personality and Social Psychology Bulletin,* **18**, 296-301.

Atkinson, R. C. & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.). *Fourth international symposium on attention and performance*. (pp. 583-611). New York: Academic Press.

Atkinson, R. C. & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Vol. 1. Learning, Memory, & Thinking.* San Francisco: Freeman.

Ayuk, R. E. (1990). Cross-racial identification of transformed, untransformed, and mixed race faces. *International Journal of Psychology,* **25**, 509-527.

Baddeley, A. D. & Woodhead, M. (1982). Depth of processing, context, and face recognition. *Canadian Journal of Psychology,* **36**, 148-164.

Balch, W. R., Myers, D. M., & Papotto, C. (1999). Dimensions of mood in mood-dependent memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **25**, 70-83.

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality & Social Psychology,* **87**, 5-22.

Bar-Haim, Y., Saidel, T., & Yovel, G. (2009). The role of skin colour in face recognition. *Perception,* **38**, 145-148.

Bartolomey, D. (2001). Cross-racial identification testimony and what not to do about it: A comment on the cross-racial jury charge and cross-racial expert identification testimony. *Psychology, Public Policy, & Law,* **7**, 247-252.

Behrman, B. W. & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law & Human Behavior,* **25**, 475-491.

Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common sense: Comparing jurors, judges, and law enforcement to eyewitness experts. *Applied Cognitive Psychology,* **20**, 115-129.

Berman, G. L. & Cutler, B. L. (1998). The influence of processing instructions at encoding and retrieval on face recognition accuracy. *Psychology, Crime, & Law,* **4**, 89-106.

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.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality & Social Psychology Review,* **6**, 242-261.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of afrocentric features in person perception: Judging by features and categories. *Journal of Personality & Social Psychology,* **83**, 5-25.

Bloom, L. C. & Mudd, S. A. (1991). Depth of processing approach to face recognition: A test of two theories. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **17**, 556-565.

Bornstein, B. H. & LeCompte, D. C. (1995). A comparison of item and source forgetting. *Psychonomic Bulletin & Review,* **2**, 254-259.

Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-racial identification. *Personality and Social Psychology Bulletin,* **15**, 19-25.

Bower, G. H. & Karlin, M. B. (1974). Depth of processing pictures of faces in recognition memory. *Journal of Experimental Psychology,* **103**, 751-757.

Brandt, K. R., Macrae, C. N., Schloerscheidt, A. M., & Milne, A. B. (2003). Remembering or knowing others? Person recognition and recollective experience. *Memory*, **11**, 89-100.

Brendl, C. M., Markman, A. B., & Meissner, C. A. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the implicit association test. *Journal of Personality & Social Psychology,* **81**, 760-773.

Brigham, J. C. (2007). The role of race and racial prejudice in recognizing other people. In C. Willis-Esqueda, *Motivational Aspects of Prejudice and Racism.* New York: Springer-Verlag.

Brigham, J. C., & Barkowitz, P. (1978). Do they all look alike - effects of race, sex, experience, and attitudes on the ability to recognize faces. *Journal of Applied Social Psychology,* **8**, 306-318.

Brigham, J. C., Bennett, L. B., & Meissner, C. A. (2007). The influence of race on eyewitness memory. In R. Lindsay, D. Ross, J. Read, & M. Toglia, (Eds.) *Handbook of Eyewitness Psychology.* Mahwah, NJ: Lawrence Erlbaum & Associates.

Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of eyewitness identifications in a field setting. *Journal of Personality & Social Psychology,* **42**, 673-681.

Brooks, K. R., & Kemp, R. I. (2007). Sensitivity to feature displacement in familiar and unfamiliar faces: Beyond the internal/external feature distinction. *Perception,* **36**, 1646-1659.

Brown, C. & Lloyd-Jones, T. J. (2006). Beneficial effects of verbalization and visual distinctiveness on remembering and knowing faces. *Memory & Cognition,* **34**, 277-286.

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology,* **62**, 311-318.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology,* **73**, 105-116.

Buchner, A., Erdfelder, E., Steffens, M. C., & Martensen, H. (1997). The nature of memory processes underlying recognition judgments in the process dissociation procedure. *Memory & Cognition,* **25**, 508-517.

Buerger, M. E., & Farrell, A. (2002). The evidence of racial profiling: Interpreting documented and unofficial sources. *Police Quarterly,* **5**, 359-379.

Byatt, G. & Rhodes, G. (2004). Identification of own-race and other-race faces: Implication for the representation of race in face space. *Psychonomic Bulletin & Review,* **11**, 735-741.

Caldara, R. & Abdi, H. (2006). Simulating the "other-race effect" with autoassociative neural networks: Further evidence in favor of the face-space model. *Perception,* **35**, 659-670.

Caldara, R., Rossion, B., Bovet, P., & Hauert, C. A. (2004). Event-related potentials and time course of the "other-race" face classification advantage. *Neuroreport,* **15**, 905-910.

Carroo, A. W. (1986). Other race recognition - a comparison of Black-American and African subjects. *Perceptual and Motor Skills,* **62**, 135-138.

Carroo, A. W. (1987). Recognition of faces as a function of race, attitudes, and reported cross-racial friendships. *Perceptual & Motor Skills,* **64**, 319-325.

Catz, O., Kampf, M., Nachson, I., & Babkoff, H. (2009). From theory to implementation: Building a multidimensional space for face recognition. *Acta Psychologica,* **131**, 143-152.

Chance, J. E. & Goldstein, A. G. (1981). Depth of processing in response to own- and other-race faces. *Personality & Social Psychology Bulletin,* **7**, 465-480.

Chance, J. E., Goldstein, A. G., & Anderson, B. (1986). Recognition memory for infant faces: An analog of the other-race effect. *Bulletin of the Psychonomic Society*, **24**, 257-260.

Chance, J., Goldstein, V., & McBride, L. (1975). Differential experience and recognition memory for faces. *Journal of Social Psychology,* **97**, 243-253.

Chance, J., Lockwood-Turner, A., & Goldstein, A. G. (1982). Development of differential recognition for own- and other-race faces. *The Journal of Psychology,* **112**, 29-37.

Chiroro, P., Tredoux, C. G., Radaelli, S., & Meissner, C. A. (2008). Recognizing faces across continents: The effect of within-race variations on the own-race bias in face recognition. *Psychonomic Bulletin & Review,* **15**, 1089-1092.

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,* **48A**, 879-894.

Clark, H. H. (1973). The language-as-a-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior,* **`12**, 335-359.

Conway, M. A. & Dewhurst, S. A. (1995). The self and recollective experience. *Applied Cognitive Psychology,* **9**, 1-19.

Corenblum, B. & Meissner, C. A. (2006). Recognition of faces of ingroup and outgroup children and adults. *Journal of Experimental Child Psychology,* **93**, 187-206.

Corneille, O., Huart, J., Becquart, E., & Brédart, S. (2004). When memory shifts towards more typical memory exemplars: Accentuation effects in the recollection of ethnically ambiguous faces. *Journal of Personality & Social Psychology,* **86**, 236-250.

Corneille, O., Hugenberg, K., & Potter, T. (2007). Applying the attractor field model to social cognition: Perceptual discrimination is facilitated, but memory is impaired for faces displaying evaluatively congruent expressions. *Journal of Personality & Social Psychology,* **93**, 335-352.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing - framework for memory research. *Journal of Verbal Learning and Verbal Behavior,* **11**, 671-684.

Cross, J. F., Cross, J., & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception and Psychophysics,* **10**, 393-396.

Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition,* **28**, 923-938.

Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia,* **42**, 1088-1106.

Curran, T., & Hintzman, D. L., (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **21**, 531-547.

Dahl, L. C., Lindsay, D. S., & Brimacombe, C. A. E. (2006). Investigating investigators: Examining witnesses' influence on investigators. *Law & Human Behavior,* **30**, 707-732.

Dalton, P. (1993). The role of stimulus familiarity in context-dependent recognition. *Memory & Cognition,* **21**, 223-234.

Dasgupta, N. & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality & Social Psychology,* **69**, 1013-1027.

Davies, G., Shepherd, J. & Ellis, H. (1979). Effects of interpolated mugshot exposure on accuracy of eyewitness identification. *Journal of Applied Psychology,* **64**, 232-237.

DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology,* **47**, 292-303.

Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law and Human Behavior,* **30**, 287-307.

Deffenbacher, K. A., Carr, T. H., & Leu, J. R. (1981). Memory for words, pictures, and faces: Retroactive interference, forgetting, and reminiscence. *Journal of Experimental Psychology: Human Learning & Memory,* **7**, 299-305.

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality & Social Psychology,* **77**, 167-184.

Dewhurst, S. A., Holmes, S. J., Brandt, K. R., & Dean, G. M. (2006). Measuring the speed of the conscious components of recognition memory: Remembering is faster than knowing. *Consciousness & Cognition,* **15**, 147-162.

Diamond, R. & Carey, S. (1986). Why faces are and are not special - An effect of expertise. *Journal of Experimental Psychology: General,* **115**, 107-117.

Dixon, J., Tredoux, C. G., & Clack, B. (2005). On the micro-ecology of racial division: A neglected dimension of segregation. *South African Journal of Psychology,* **35**, 395-411.

Dodson, C. S. & Johnson, M. K. (1996). Some problems with the process-dissociation approach to memory. *Journal of Experimental Psychology: General,* **125**, 181-194.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition,* **24**, 523-533.

Doyle, J. M. (2001). Discounting the error costs: Cross-racial false alarms in the culture of contemporary criminal justice. *Psychology, Public Policy, & Law,* **7**, 253-262.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods,* **1**, 170-177.

Dunn, J. C. (2004). Remember-Know: A matter of confidence. *Psychological Review,* **111**, 524-542.

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.

Eberhardt, J. L., Dasgupta, N., & Banaszynski, T. L. (2003). Believing is seeing: The effects of racial labels on face perception. *Personality & Social Psychology,* **29**, 360-370.

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing Black: Race, crime, and visual processing. *Journal of Personality & Social Psychology,* **87**, 876-893.

Egeth, H. E. (1993). What do we not know about eyewitness identification? *American Psychologist,* **48**, 577-580.

Elliott, E. S., Wills, E. J., & Goldstein, A. G. (1973). Effects of discrimination training on recognition of White and Oriental faces. *Bulletin of the Psychonomic Society*, **2**, 71-73.

Ellis, H. D., Deregowski, J. B., & Shepherd, J. W. (1975). Descriptions of white and black faces by white and black subjects. *International Journal of Psychology,* **10**, 119-123.

Evans, J. R., Marcon, J., & Meissner, C. A. (2009). Cross-racial lineup identification: assessing the potential benefits of context reinstatement. *Psychology*, *Crime, & Law,* **15**, 19-28.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance,* **21**, 141-145.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes - a bona-fide pipeline. *Journal of Personality & Social Psychology,* **69**, 1013-1027.

Feingold, G. A. (1914). The influence of environment on identification of persons and things. *Journal of the American Institute of Criminal Law and Criminology,* **5**, 39-51.

Ferguson, D. P., Rhodes, G., Lee, K., & Sriram, N. (2001). "They all look alike to me": Prejudice and cross-race face recognition. *British Journal of Psychology,* **92**, 567-577.

Ferguson, K. T., Kulkofsky, S., Cashon, C. H., & Casasola, M. (2009). The development of specialized processing of own-race faces in infancy. *Infancy,* **14**, 263-284.

Gajewski, P. D., Schlegel, K., & Stoerig, P. (2008). Effects of human race and face inversion on the N170: A cross-race study. *Journal of Psychophysiology,* **22**, 157-165.

Garcia-Marques, L., Santos, A. S. C., & Mackie, D. M. (2006). Stereotypes: Static abstractions or dynamic knowledge structures? *Journal of Personality & Social Psychology,* **91**, 814-831.

Gardiner, J. M., Java, R. I., & Richardson-Klavehn, A. (1996). How level of processing really influences awareness in recognition memory. *Canadian Journal of Experimental Psychology,* **50**, 114-122.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of Remembering, Knowing, and Guessing. *Consciousness & Cognition,* **7**, 1-26.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1999). Response deadline and subjective awareness in recognition memory. *Consciousness & Cognition,* **8**, 484-496.

Glanzer, M. & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **16**, 5-16.

Glanzer, M., Hilford, A., & Kim, K. (2004). Six regularities of source *recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition,* **30**, 1176-1195.

Godden, D. R. & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and under water. *British Journal of Psychology,* **66**, 325-331.

Goldstein, A. G. & Chance, J. (1976). Measuring psychological similarity of faces. *Bulletin of the Psychonomic Society,* **7**, 407-408.

Goldstein, A. G., & Chance, J. (1978). Judging face similarity in own and other races. *Journal of Psychology,* **98**, 185-193.

Goldstein, A. G. & Chance, J. (1979). Do foreign faces really look alike? *Bulletin of the Psychonomic Society,* **13**, 111-113.

Goldstein, A. G. & Chance, J. (1985). Effects of training on Japanese face recognition - Reduction of the other-race effect. *Bulletin of the Psychonomic Society,* **23**, 211-214.

Goodman, G. S., Sayfan, L., Lee, J. S., Sandhei, M., Walle-Olsen, A., Magnussen, S., Pezdek, K., & Arrendondo, P. (2007). The development of memory for own- and other-race faces. *Journal of Experimental Child Psychology,* **88**, 760-763.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality & Social Psychology,* **74**, 1464-1480.

Gross, S. R., Jacoby, K., Matheson, D. J., Montgomery, N., & Patil, S. (2004). Exonerations in the United States: 1989 through 2003. *The Journal of Criminal Law & Criminology,* **95**, 523-560.

Gross, T. F. (2009). Own-ethnicity bias in the recognition of Black, East Asian, Hispanic, and White faces. *Basic and Applied Social Psychology,* **31**, 128-135.

Gruppuso, V., Lindsay, D. S., & Kelley, C. M. (1997). The process-dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **23**, 259-278.

Gruppuso, V., Lindsay, D. S., & Masson, M. E. J. (2007). I'd know that face anywhere! *Psychonomic Bulletin and Review,* **14**, 1085-1089.

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding, and the other-race effect in face recognition. *British Journal of Psychology,* **99**, 45-56.

Hayden, A., Bhatt, R. S., Joseph, J. E., & Tanaka, J. W., (2007). The other-race effect in infancy: Evidence using a morphing technique. *Infancy,* **12**, 95-104.

Hayden, A., Bhatt, R. S., Zieber, N., & Kangas, A. (2009). Races-based perceptual asymmetries underlying face processing in infancy. *Psychonomic Bulletin & Review,* **16**, 270-275.

Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition,* **106**, 1017-1027.

Hewstone, M. & Ward, C. (1985). Ethnocentrism and causal attributions in Southeast Asia. *Journal of Personality & Social Psychology,* **48**, 614-623.

Hilliar, K. F., & Kemp, R. I. (2008). Barack Obama or Barry Dunham? The appearance of multiracial faces is affected by the names assigned to them. *Perception,* **37**, 1605-1608.

Hills, P. J., & Lewis, M. B. (2006). Reducing the own-race bias in face recognition by shifting attention. *The Quarterly Journal of Experimental Psychology,* **59**, 996-1002.

Hockley, W. E. (2008). The effects of environmental context on recognition memory and claims of remembering. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **34**, 1412-1429.

Hockley, W. E. & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition,* **27**, 657-664.

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception,* **31**, 1221-1240.

Home Office (2005). *Stop & Search Manual.* Accessed online at police.homeoffice.gov.uk/publications/operational-policing/stopandsearch-intermanual1.pdf?view=Binary. Accessed on August 12, 2008.

Hugenberg, K. & Corneille, O. (2009). Holistic processing is tuned for in-group faces. *Cognitive Science,* **33**, 1173-1181.

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology,* **43**, 336-340.

Islam, M. R. & Hewstone, M. (1993a). Dimensions of contact as predictors of inter-group anxiety, perceived out-group variability, and out-group attitude: An integrative model. *Personality & Social Psychology Bulletin,* **19**, 700-710.

Islam, M. R. & Hewstone, M. (1993b). Intergroup attributions and affective consequences in majority and minority groups. *Journal of Personality & Social Psychology,* **64**, 936-950.

Ito, T. A. & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality & Social Psychology,* **85**, 616-626.

Ito, T. A. & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, & Behavioral Neuroscience,* **5**, 21-36.

Jackiw, L. B., Arbuthnott, K. D., Pfeifer, J. E., Marcon, J. L., & Meissner, C. A. (2008). Examining the cross-race effect in lineup identification using Caucasian and First Nations samples. *Canadian Journal of Behavioural Science,* **40**, 52-57.

Jacoby, L. L. (1991). A process-dissociation framework - separating automatic from intentional uses of memory. *Journal of Memory and Language,* **30**, 513-541.

Jacoby, L. L., Begg,. I. M., & Toth, J. P. (1997). In defense of functional independence: Violations of assumptions underlying the process-dissociation procedure? *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **23**, 484-495.

Johnson, M. K. (2006). Memory and reality. *American Psychologist,* **61**, 760-771.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research. *American Psychologist, **56***, 405-416.

Kelly, D. J., Liu, S. Y., Ge, L. Z., Quinn, P. C., Slater, A. M., Lee, K., Liu, Q. Y., & Pascalis, O. (2007). Cross-race preferences for same-race faces extend beyond the African versus Caucasian contrast in 3-month-old infants. *Infancy,* **11**, 87-95.

Kelly, D. J., Liu, S. Y., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge., L. Z. (2009). Development of the other-race effect during infancy: Evidence toward universality? *Journal of Experimental Child Psychology*, **104**, 105-114.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L. Z., Pascalis, O. (2007). The other-race effect develops during infancy - Evidence of perceptual narrowing. *Psychological Science,* **18**, 1084-1089.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., Ge. L. Z., & Pascalis, O. (2005). Three-month olds, but not newborns, prefer own-race faces. *Developmental Science,* **8**, F31-F36.

Kinoshita, S. & Peek-O'Leary, M. (2001). Does the compatibility effect in the race Implicit Association Test reflect familiarity or affect? *Psychonomic Bulletin & Review,* **12**, 442-452.

Knowles, J., Persico, N., & Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy,* **109**, 203-229.

Konstantinou, I. & Gardiner, J. M. (2005). Conscious control and memory awareness when recognising famous faces. *Memory,* **13**, 449-457.

Krafka, C. & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology*, **49***,* 58-69.

Lambert, A. J., Payne, B. K., Jacoby, L. L., Shaffer, L. M., Chasteen, A. L., & Khan, S. R. (2003). Stereotypes as dominant responses: On the "social facilitation" of prejudice in anticipated public contexts. *Journal of Personality & Social Psychology,* **84**, 277-295.

Lebrecht, S., Pierce, L. J., Tarr, M. J., & Tanaka, J. W. (2009). Perceptual other-race training reduces implicit racial bias. *PLoS ONE,* **4**, e4215.

Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **22**, 1364-1382.

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General,* **129**, 559-574.

Levin, D. T. & Angelone, B. L. (2001). Visual search for a socially defined feature: What causes the search asymmetry favoring cross-race faces? *Perception & Psychophysics,* **63**, 423-435.

Levin, D. T. & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General,* **135**, 501-512.

Lindholm, T. & Christianson, S.-A. (1998). Intergroup biases and eyewitness testimony. *The Journal of Social Psychology,* **138**, 710-723.

Lindsay, D. S. (2007). Autobiographical memory, eyewitness reports, and public policy. *Canadian Psychology*, **48**, 57-66.

Lindsay, D. S., Jack, P. C., & Christian, M. A. (1991). Other-race face perception. *Journal of Applied Psychology,* **76**, 587-589.

Lindsay, R. C. L., Nosworthy, G. J., Martin, R., & Martynuck, C. (1994). Using mug shots to find suspects. *Journal of Applied Psychology,* **79**, 121-130.

Lindsay, R. C. L., Wells, G. L., & Rumpel, C. M. (1981). Can people detect eyewitness-identification accuracy within and across situations? *Journal of Applied Psychology,* **66**, 79-89.

Li, J. C., Dunning, D., & Malpass, R. S. (1998). Cross-racial identification among European-Americans: Basketball fandom and the contact hypothesis. Cited by Brigham, J. C., Bennett, L. B., Meissner, C. A., & Mitchell, T. L. (2007). The influence of race on eyewitness memory. In R. Lindsay, D. Ross, J. Read, & M. Toglia, (Eds.) *Handbook of Eyewitness Psychology.* Mahwah, NJ: Lawrence Erlbaum & Associates.

Loftus, E. F. (1976). Unconscious transference in eyewitness identification. *Law and Psychology Review,* **94**, 93-98.

Loftus, E. F. (1993). Psychologists in the eyewitness world. *American Psychologist,* **48**, 550-552.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality & Social Psychology,* **81**, 842-855.

Macken, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **28**, 153-161.

MacLin, O. H., MacLin, M. K., & Malpass, R. S. (2001). Race, arousal, attention, exposure, and delay: An examination of factors moderating face recognition. *Psychology, Public Policy, & Law,* **7**, 134-152.

MacLin, O. H. & Malpass, R. S. (2001). Racial categorization of faces - the ambiguous race effect. *Psychology, Public Policy, & Law,* **7**, 98-118.

Malpass, R. S. (1990). An excursion into utilitarian analysis. *Behavior Science Research,* **24**, 1-14.

Malpass, R. S. & Devine, P. G. (1981). Guided memory in eyewitness identification. *Journal of Applied Psychology,* **64**, 343-350.

Malpass, R. S. & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality & Social Psychology,* **13**, 330 – 334.

Malpass, R. S., Lavigueur, H., & Weldon, D. E. (1973). Verbal and visual training in face recognition. *Perception & Psychophysics,* **14**, 285-292.

Mandler, G. (1980). Recognizing - the judgement of previous *occurrence. Psychological Review,* **87**, 252-271.

Mäntylä, Y. (1997). Recollections of faces: Remembering differences and knowing similarities. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **23**, 1203-1216.

Marcon, J. L., Susa, K. J., & Meissner, C. A. (2009). Assessing the influence of recollection and familiarity in memory for own- and other-race faces. *Psychonomic Bulletin & Review,* **16**, 99-103.

Marian, V. & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General,* **129**, 361-368.

McBride, D. M. (2007). Methods for measuring conscious and automatic memory. *Journal of Consciousness Studies,* **14**, 198-215.

McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **25**, 563-582.

McKone, E., Brewer, J. L., MacPherson, S., Rhodes, G., & Hayward, W. G. (2007). Familiar other-race faces show normal holistic processing and are robust to perceptual stress. *Perception,* **43**, 161-204.

Meiser, T. & Sattler, C. (2007). Boundaries of the relation between conscious recollection and source memory for perceptual details. *Consciousness and Cognition,* **16**, 189-210.

Meiser, T., Sattler, C., & Weisser, K. (2008) Binding of multidimensional context information as a distinctive characteristic of Remember judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **34**, 32-49.

Meissner, C. A. & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, & Law,* 7, 3-35.

Meissner, C. A., Brigham, J. C., & Butz, D. A. (2005). Memory for own- and other-race faces: A dual-process approach. *Applied Cognitive Psychology,* **19**, 545-567.

Memon, A., Mastroberardino, S., & Fraser, J. (2008). Munsterberg's legacy: What does eyewitness research tell us about the reliability of eyewitness testimony? *Applied Cognitive Psychology,* **22**, 841-851.

Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition,* **14**, 55-73.

Michel, C., Corneille, O., & Rossion, B. (2007). Race categorization modulates holistic face processing. *Cognitive Science,* **31**, 911-924.

Michel, C., Rossion, B., Han, J., Chung, C. S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science,* **17**, 608-615.

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychological Science,* **20**, 509-515.

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality & Social Psychology,* **77**, 167-184.

Mulligan, N. & Hirshman, E. (1995). Speed-accuracy trade-offs and the dual process model of recognition memory. *Journal of Memory & Language,* **34**, 1-18.

Mulligan, N. & Hirshman, E. (1997). Measuring the bases of recognition memory: An investigation of the process-dissociation framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **23**, 280-304.

Murnane, K. & Bayen, U. J. (1996). An evaluation of empirical measures of source identification. *Memory & Cognition,* **24**, 417-428.

Naughton, M. (2003). How big is the "iceberg"? – a zemiological approach to quantifying miscarriages of justice. *Radical Statistics,* **81**, 5-17.

Ng, W. J. & Lindsay, R. C. L. (1994). Cross-racial facial identification - failure of the contact hypothesis. *Journal of Cross-Cultural Psychology,* **25**, 217-232.

Odinot, G. & Wolters, G. (2006). Repeated recall, retention interval and the accuracy-confidence relation in eyewitness memory. *Applied Cognitive Psychology,* **20**, 973-985.

Office of the Attorney General (2003). *Guidance regarding the use of race by federal law enforcement agencies.* U.S. Department of Justice, Civil Rights Division: Washington D.C. http://www.usdoj.gov. Accessed June 23, 2008.

Oliver, M. B., Jackson, R. L., Moses, N. N., & Dangerfield, C. L. (2004). The face of crime: Viewers' memory of race-related facial features of individuals pictured in the news. *Journal of Communication,*

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (in press). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica.*

Parkin, A. J., Gardiner, J. M., & Rosser, R. (1995). Functional aspects of recollective experience in face recognition. *Consciousness & Cognition,* **5**, 378-398.

Patterson, K. E. & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning & Memory,* **3**, 406-417.

Pauker, K., Weisbuch, M., Ambady, N., Sommers, S. R., Adams, R. B., & Ivcevic, Z. (2009). Not so Black and White: Memory for ambiguous group members. *Journal of Personality & Social Psychology,* **96**, 795-810.

Penrod, S. & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, & Law,* **1**, 817-845.

Perfect, T. J., Watson, E. L., & Wagstaff, G. F. (1993). Accuracy of confidence ratings associated with general knowledge and eyewitness memory. *Journal of Applied Psychology*, **78**, 144-147.

Petrocelli, M., Piquero, A. R., & Smith, M. R. (2003). Conflict theory and racial profiling: An empirical analysis of police traffic stop data. *Journal of Criminal Justice,* **31**, 1-11.

Pezdek, K., Blandon-Gitlin, I., & Moore, C. (2003). Children's face recognition memory: More evidence for the cross-race effect. *Journal of Applied Psychology,* **88**, 760-763.

Platz, S. J. & Hosch, H. M. (1988). Cross-racial ethnic eyewitness identification - a field study. *Journal of Applied Social Psychology*, **18**, 972-984.

Quanty, M. B., Keats, J. A., & Harkins, S. G. (1975). Prejudice and criteria for identification of ethnic photographs. *Journal of Personality & Social Psychology,* **32**, 449-454.

Read, J. D., Tollestrup, P., Hammersley, R., McFazden, E., & Christensen, A. (1990). The unconscious transference effect: Are innocent bystanders ever misidentified? *Applied Cognitive Psychology,* **4**, 3-31.

Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **20**, 161-168.

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology,* **80**, 313-331.

Rhodes, G., Hayward, W. G., & Winkler, C. (2006). Expert face coding: configural and component coding of own- and other-race faces. *Psychonomic Bulletin & Review,* **13**, 499-505.

Rhodes, G., Locke, V., Ewing, L., & Evangelista, E. (2009). Race coding and the other-race effect in face recognition. *Perception*, **38**, 232-241.

Richardson-Klavehn, A., Gardiner, J. M., & Ramponi, C. (2002). Levels of processing and the process-dissociation procedure: Elusiveness of null effects on automatic retrieval. *Memory,* **10**, 349-364.

Rodin, M. J. (1987). Who is memorable to whom – a study of cognitive disregard. *Social Cognition,* **5**, 144-165.

Ross, D. F., Ceci, D. J., Dunning, D., & Toglia, M. P. (1994). Unconscious transference and mistaken identity: When a witness misidentifies a familiar yet innocent person. *Journal of Applied Psychology,* **79**, 918-930.

Rotello, C. M., MacMillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review,* **12**, 865-873.

Rule, N. O., Ambady, N., & Adams, A. B. (2007). Us and them: Memory advantages in perceptually ambiguous groups. *Psychonomic Bulletin & Review,* **14**, 687-692.

Rule, N. O., Macrae, C. N., & Ambady, A. B. (2009). Ambiguous group membership is extracted automatically from faces. *Psychological Science,* **20**, 441-443.

Russo, R., Ward, G., Geurts, H., & Scheres, A. (1999). When unfamiliarity matters: Changing environmental context between study and test affects recognition memory for unfamiliar stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **25**, 488-499.

Sangrigoli, S. & de Schonen, S. (2004a). Recognition of own-race and other-race faces by three-month-old infants. *Journal of Child Psychology and Psychiatry,* **45**, 1219-1227.

Sangrigoli, S. & de Schonen, S. (2004b). Effects of visual experience on face processing: a developmental study of inversion and non-native effects. *Developmental Science,* **7**, 74-87.

Sangrigoli, S., Pallier, C., Argenti, A-M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in recognition during childhood. *Psychological Science,* **16**, 440-444.

Schaller, M., Park, J. H., & Mueller, A. (2003). Fear of the dark: Interactive effects of beliefs about danger and ambient darkness on ethnic stereotypes. *Personality & Social Psychology Bulletin,* **29**, 637-649.

Scheck, B., Neufeld, P., & Dwyer, J. (2003). *Actual Innocence: When justice goes wrong and how to put it right.* New York: New American Library.

Schramke, C. J. & Bauer, R. M. (1997). State-dependent learning in older and younger adults. *Psychology and Aging,* **12**, 255-262.

Scott, R. B. & Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **34**, 1264-1288.

Shepherd, J. W., Ellis, H. D., McMurran, M., & Davies, G. M. (1978). Effect of character attribution on Photofit construction. *European Journal of Social Psychology,* **8**, 263-268.

Shriver, E. R., Young, S. G., Hugenberg, K., Bernstein, M. J., & Lanter, J. R. (2008). Class, race, and the face: Social context modulates the cross-race effect in face recognition. *Personality and Social Psychology Bulletin,* **34**, 260-274.

Shutts, K. & Kinzler, K. D. (2007). An ambiguous-race illusion in children's face memory. *Psychological Science,* **18**, 763-767.

Skinner, E. I. & Fernandes, M. A. (2009). Illusory recollection in older adults and younger adults under divided attention. *Psychology & Aging,* **24**, 211-216.

Slone, A. E., Brigham, J. C., & Meissner, C. A. (2000). Social and cognitive factors affecting the own-race bias in Whites. *Basic and Applied Social Psychology,* **22**, 71-84.

Smith, M. R. & Petrocelli, M. (2001). Racial profiling? A multivariate analysis of police traffic stop data. *Police Quarterly,* **4,** 4-27.

Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies & D. M. Thomson (Eds.), *Memory in Context: Context in Memory.* New York: Wiley.

Smith, S. M., Lindsay, R. C. L., Pryke, S., & Dysart, J. E. (2001). Postdictors of eyewitness errors - can false identifications be diagnosed in the cross-race situation? *Psychology, Public Policy, and Law,* **7**, 153-169.

Smith, S. M., Stinson, V., & Prosser, M. A. (2004). Do they all look alike? An exploration of decision-making strategies in cross-race facial identifications. *Canadian Journal of Behavioural Science,* **36**, 146-154.

Smith, V. L., Kassin, S. M., & Ellsworth, P. C. (1989). Eyewitness accuracy and confidence: Within- versus between-subjects correlations. *Journal of Applied Psychology,* **74**, 356-359.

Sporer, S. L. (1989). Verbal and visual processes in person identification. In H. Weneger, F. Losel, & J. Haisch (Eds.). *Criminal Justice and the Justice System: Psychological Perspectives.* New York: Springer-Verlag. Pp 303-324.

Sporer, S. L. (1991). Deep-deeper-deepest? Encoding strategies and the recognition of human faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **17**, 323-333.

Sporer, S. L. (1993). Clothing as a contextual cue in facial recognition. *The German Journal of Psychology,* **17**, 183-199.

Sporer, S. L. (2001a). The cross-race effect: Beyond recognition of faces in the laboratory. *Psychology, Public Policy, and Law,* **7**, 170-200.

Sporer, S. L. (2001b). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law,* **7**, 36-97.

Sporer, S. L., Penrod, S. D., Read, D., & Cutler, B. (1995). Choosing, confidence and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin,* **118**, 315-327.

Sporer, S. L., Trinkl, B., & Guberova, E. (2007). Matching faces - Differences in processing speed of out-group faces by different ethnic groups. *Journal of Cross-Cultural Psychology,* **38**, 398-412.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers,* **31**, 137-149.

Swets, J. A. (1964). *Signal detection and recognition by human observers.* New York: Wiley.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the other-race effect in face recognition: evidence from a cross-cultural study. *Cognition,* **93**, B1-B9.

Taubert, J. (2009). Chimpanzee faces are "special" to humans. *Perception,* **38**, 343-356.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality & Social Psychology,* **36**, 778-793.

Teitelbaum, S. & Geiselman, R. E. (1997). Observer mood and cross-racial recognition of faces. *Journal of Cross-Cultural Psychology,* **28**, 93-106.

Tredoux, C. G. & Dixon, J. A. (2009). Mapping the multiple contexts of racial isolation: The case of Long Street, Cape Town. *Urban Studies,* **46**, 761-777.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology,* **26**, 1-12.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology,* **43**, 161-204.

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology,* **17**, 969-993.

Waddington, P. A. J., Stenson, K., & Don, D. (2004). In proportion: Race, and police stop and search. *British Journal of Criminology,* **44**, 889-914.

Walker, P. M. & Hewstone, M. (2006a). A perceptual discrimination investigation of the own-race effect and intergroup experience. *Applied Cognitive Psychology,* **20**, 461-475.

Walker, P. M. & Hewstone, M. (2006b). A developmental investigation of other-race contact and the own-race effect. *British Journal of Developmental Psychology,* **24**, 451-463.

Walker, P. M. & Hewstone, M. (2008). The influence of social factors and implicit racial bias in a generalized own-race effect. *Applied Cognitive Psychology,* **22**, 441-453.

Walker, P. M. & Tanaka, J. W. (2003). An encoding advantage for own-race versus other-race faces. *Perception*, **32**, 1117-1125.

Weimann, G., Fishman, G., & Rattner, A. (1988). Social distance and misidentification. *International Journal of Comparative Sociology,* **27**, 217-225.

Weise, H., Schweinberger, S. R., & Hansen, K. (2008). The age of the beholder: ERP evidence of an own-age bias in face memory. *Neuropsychologia,* **46**, 2973-2985.

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality & Social Psychology,* **36**, 1546-1557.

Wells, G. L., Malpass, R. S., Lindsay, R. C. L., Fisher, R. P., Turtle, J. W., & Fulero, S. M. (2000). From the lab to the police station: A successful application of eyewitness research. *American Psychologist,* **55**, 581-598.

Wells, G. L. & Olson, E. A. (2001). The other-race effect in eyewitness identification: What do we do about it? *Psychology, Public Policy, and Law,* **7**, 230-246.

Wells, G. L. & Windschitl, P. D. (1999). Stimulus sampling and social psychology experimentation. *Personality & Social Psychology Bulletin,* **25**, 1115-1125.

Westerman, D. L. (2001). The role of familiarity in item recognition, associative recognition, and plurality recognition on self-paced and speeded tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **27**, 723-732.

Wiese, H., Stahl, J., & Schweinberger, S. R. (2009). Configural processing of other-race faces is delayed but not decreased. *Biological Psychology,* **81**, 103-109.

Wilding, E. L., Doyle, M. C., & Rugg, M. D. (1995). Recognition memory with and without retrieval of context: An event-related potential study. *Neuropsychologia,* **33**, 743-767.

Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning & Memory,* **7**, 181-190.

Wise, R. A. & Safer, M. A. (2004). What US judges know and believe about eyewitness testimony. *Applied Cognitive Psychology,* **18**, 427-443.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality & Social Psychology,* **81**, 815-827.

Wright, D. B., (1997). Extra-binomial variation in multilevel logistic models with sparse structures. *British Journal of Mathematical & Statistical Psychology,* **50**, 21-29.

Wright, D. B. (1998). Modelling clustered data in autobiographical memory research: The multilevel approach. *Applied Cognitive Psychology,* **12**, 339-357.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology, Public Policy, and Law,* **7**, 119-133.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own-race bias for face recognition in South Africa and England. *Applied Cognitive Psychology,* **17**, 365-373.

Wright, D. B., Horry, R., & Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods,* **41**, 257-267.

Wright, D. B. & London, K. (2009). Multilevel modeling: Beyond the basic applications. *British Journal of Mathematical & Statistical Psychology,* **62**, 439-456.

Wright, D. B. & McDaid, A. T. (1996). Comparing system and estimator variables using data from real line-ups. *Applied Cognitive Psychology,* **10**, 75-84.

Wright, D. B. & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica,* **114**, 101-114.

Wright, D. B. & Stroud, J. N. (2002). Age differences in lineup identification accuracy: People are better with their own age. *Law and Human Behavior,* **6**, 641-654.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology,* **81**, 141-145.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language,* **46**, 441-517.

Yonelinas, A. P. & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory & Language,* **34**, 622-643.

Yonelinas, A. P., Kroll, E. A., Dobbins, I. G., & Soltani, M. (1999). Recognition memory for faces: When familiarity supports associative recognition judgments. *Psychonomic Bulletin & Review,* **6**, 654-661.

Yovel, G. & Pallier, K. A. (2004). The neural basis of the butcher-on-the-bus phenomenon: When a face seems familiar but is not remembered. *NeuroImage,* **21**, 789-800.

Zebrowitz, L. A., Montepare, J. M., & Lee, H. K. (1993). They don't all look alike - individuated impressions of other racial groups. *Journal of Personality and Social Psychology,* 65, 85-101.

Zhao, L. & Bentin, S. (2008). Own- and other-race categorization of faces by race, gender, and age. *Psychonomic Bulletin & Review,* **15**, 1093-1099.

Appendices

Appendix 1: Formulae for Calculating the Effect Size Cohen's *d* for Repeated Measures and Between Groups Designs

Repeated measures:

$$d = \frac{(M1 - M2)}{\sqrt{(SD1 + SD2)/2}}$$

Between groups:

$$d = 2 * \frac{t}{\sqrt{df}}$$

Formulae taken from Dunlap, Cortina, Vaslow, and Burke (1996).

Appendix 2: Formulae for Calculating the Signal Detection Measures *d′* and *C*

$$d' = z(HR) - z(FAR)$$

$$C = .5 * (z(HR) + z(FAR))$$

Where:

$$HR = (Hits + .5)/(Hits + Misses + 1)$$

$$FAR = (FAs + .5)/(FAs + CRs + 1)$$

Note: *HR* = Hit rate, *FAR* = False alarm rate, *FAs* = False alarms, *CRs* = Correct rejections.

Formulae from Stanislaw and Todorov (1999).

Appendix 3: Calculation of Average Condition Source Identification Measure (ACSIM).

Formula below is for an ACSIM calculated with three sources:

$$\text{ACSIM} = \frac{\left(\frac{Yaa}{Yaa+Yab+Yac}\right)+\left(\frac{Ybb}{Yba+Ybb+Ybc}\right)+\left(\frac{Ycc}{Yca+Ycb+Ycc}\right)}{N}$$

Where:

Y = old identifications

aa = items from context a correctly assigned to context a

ab = items from context b incorrectly assigned to context a

N = number of sources/contexts.