UNIVERSIDADE DE LISBOA

FACULDADE DE PSICOLOGIA

# Metamemory for faces: Self-other awareness of typicality and race effects

**Alexandre Emanuel Reis Gonçalves Vieira**

**MESTRADO INTEGRADO EM PSICOLOGIA**

**Área de Especialização em Cognição Social Aplicada**

**2020**

UNIVERSIDADE DE LISBOA

FACULDADE DE PSICOLOGIA

# Metamemory for faces: Self-other awareness of typicality and race effects

**Alexandre Emanuel Reis Gonçalves Vieira**

**Dissertação orientada pelo Doutor Tomás Palma e Professor Doutor André Mata**

## MESTRADO INTEGRADO EM PSICOLOGIA

**Área de Especialização em Cognição Social Aplicada**

**2020**

Dedico este trabalho aos meus pais e irmão.

Dedico a ti, Madalena, e ao nós.

Ao Matias e ao Gaspar.

Dedico ao Tomás, pois sem ti não seria (verdadeiramente) possível.

Dedico ao André, pelo insight e conclusões fantásticas.

A todos que estiveram, estão e, continuarão a estar, um sincero obrigado.

*E se inventássemos o mar de volta,*

*e se inventássemos partir,*

*para regressar.*

*José Mário Branco*

## Abstract

The Own-race bias (ORB) – the tendency to recognize own-race faces better than other-race faces – has been widely scrutinized and replicated across multiple studies. Recently, some have tried to link metamemory to the ORB, realizing that it also occurs when a person attempts to predict their future memory. Specifically, this thesis strives to further investigate how accurate a person is about their future recognition performance of own and other-race faces and whether typicality – a face being atypical or typical – affects the ORB and participants' metamemorial predictions. In addition, we also tried to understand if people are aware of these effects when making predictions for others and themselves. Using a standard recognition paradigm and JOL's (*judgment of learning*) to assess participants' future recognition with Caucasian participants, our results replicated the ORB and an effect of typicality was found, exposing the fact that people tend to better discriminate atypical faces than typical ones, due to the salience of face features. No differences were found in relative metamnemonic accuracy, nevertheless, people predicted their future memory performance above chance level. Importantly, we found that people are aware of these effects and, this awareness is grounded in a well-adjusted naïve theory about the functioning of memory. This awareness was established not only for themselves but for others. Limitations and proposals for future studies are discussed.



**Keywords:** Own-race bias (ORB), judgment of learning, typicality, metamemory, face recognition

## Resumo Alargado

A investigação relativa ao reconhecimento de faces tem mostrado que o sistema visual humano é particularmente proficiente no anterior em diversos contextos e domínios. Apesar desta aparente adequação, a discussão sobre quais os mecanismos e processos que são fundamentais para o reconhecimento de faces tem sido dinâmica. Especificamente, estudos mostram que percecionamos faces de forma holística e que, desta forma, torna-se mais fácil extrair características destas mesmas. Ao mesmo tempo, a existência de uma área cerebral dedicada ao processamento de faces indica que estes estímulos são especiais e, além disso, permitem-nos extrair e inferir uma grande quantidade de informação social (e.g., raça, idade, género) que é importante para o ser humano operar nos contextos em que se insere. Globalmente, os indivíduos afiguram-se como *experts* no reconhecimento de faces; contudo, os efeitos de pertença grupal têm sistematicamente mostrado que existe um "embora" na *expertise* relativa ao reconhecimento de faces.

O *own-race bias* (ORB) é um dos exemplos mais robusto e estudado dos efeitos acima mencionados, tendo sido aplicado em contextos experimentais e ecologicamente válidos. De forma sintética, o ORB revela que os indivíduos são melhores a reconhecer faces da sua própria raça e cometem menos erros em relação a essas mesmas faces. As causas relativas a este fenómeno são inúmeras, podendo ser explicadas através de: *expertise* superior em relação a faces da própria raça; motivação intrínseca dos indivíduos relacionada com a pertença grupal em termos de raça e, ainda, pela interação dos anteriores fatores.

Simultaneamente, as faces diferem tanto em termos gerais como em termos mais específicos (i.e., tipicidade, diferenças mais subtis e graduais) relativamente à categoria a que pertencem (e.g., raça). Desta forma, é expectável que estas diferenças também sejam

importantes e impactantes no que concerne aos indivíduos e ao subsequente reconhecimento de faces da própria e de outra etnia.

Ao mesmo tempo, a existência deste fenómeno tem sido demonstrada não só em termos de memória de reconhecimento como em julgamentos metacognitivos (i.e., em metamemória). Especificamente, os resultados relativos ao ORB em metamemória são contraditórios, sendo que alguns reportam que os indivíduos preveem a sua memória para faces com precisão (i.e., que se irão recordar melhor de faces da sua própria raça do que de outra raça), outros mostram o oposto. Desta forma, as conclusões relativas ao ORB em metamemória mostram que, por vezes, os indivíduos estão conscientes e são precisos nas suas previsões, noutras o padrão oposto sucede.

Assente no supramencionado, a presente dissertação tem como objetivo explorar se os indivíduos estão conscientes deste enviesamento para faces da própria raça e dos efeitos de tipicidade, tanto para o próprio como para os outros. Concomitantemente, procurámos perceber se as pessoas são precisas nas suas previsões e se, ao mesmo tempo, capturam as diferenças de tipicidade e as projetam para a sua memória e respetivos julgamentos metacognitivos.

De forma a analisar os objetivos supramencionados, três experiências foram concebidas onde utilizámos um paradigma *standard* de memória de reconhecimento e um plano experimental intraparticipante com 4 condições, manipulando raça (Caucasiana vs. Africana) e tipicidade (atípica vs. típica), de forma transversal, numa amostra de participantes Caucasianos. As diferenças relativamente a cada experiência são as seguintes: na Experiência 1, os participantes, para além de estudarem faces e serem testados sobre esse mesmo estudo, responderam a dois julgamentos: confiança preditiva (i.e., JOLs; *predictive confidence*) e confiança retroativa (i.e., *postdictive confidence*).; na Experiência 2, introduzimos os julgamentos preditivos em relação aos outros; finalmente, na Experiência 3, introduzimos uma condição que consistia na manipulação do Outro (Caucasiano vs. Africano), de forma a perceber em quem é que os participantes pensaram quando fizeram julgamentos de confiança preditivos.

A generalidade dos resultados demonstra que os participantes atribuíram valores maiores de confiança preditiva para faces da própria raça do que para faces de outra raça, e valores maiores para faces atípicas do que típicas. Ao mesmo tempo, os participantes captaram as diferenças de tipicidade, nos julgamentos de confiança preditiva, mais para faces da própria raça do que de outra. Relativamente à precisão das estimações metacognitivas dos participantes, não encontrámos resultados significativos na globalidade das experiências; contudo, os participantes previram acima do acaso. Em termos de memória de reconhecimento, como era expectável, os participantes recordaram melhor faces da própria raça do que faces de outra raça. Concomitantemente, os participantes recordaram melhor faces atípicas do que típicas.

De realçar que os resultados em relação às estimações preditivas para o outro e o próprio mostraram-se correlacionados na Experiência 2, sendo que as estimações para o outro se assemelham às estimações para o próprio. Ao mesmo tempo, os resultados obtidos na Experiência 3 demonstram que, através da manipulação do Outro, os participantes estimam valores superiores para faces Caucasianas quando o outro é caucasiano. O padrão oposto emerge quando o outro é Africano. Finalmente, os participantes estimaram valores maiores para confiança retroativa para maiores proporções de respostas corretas.

Conjuntamente, estes resultados mostram que as pessoas estão conscientes dos efeitos de raça (i.e., ORB) e tipicidade. Por um lado, as pessoas aparentam basear as suas estimações numa crença de que as faces da própria raça serão por si melhor recordadas do que as faces de outra raça. Por outro lado, a fluência na codificação mnésica de faces atípicas em comparação com faces típicas parece informar os indivíduos nas suas estimações de *performance* futura. Desta forma, a conjunção de crenças e fluência aparenta indicar que os indivíduos possuem

uma boa teoria leiga sobre o funcionamento da memória, embora não a apliquem quando são testados para tal. Adicionalmente, os resultados da Experiência 2 e 3 configuram-se como *key findings* nesta dissertação. Especificamente, verificámos que, apesar dos indivíduos apresentarem uma boa teoria leiga do funcionamento da memória, não são capazes de captar as diferenças de tipicidade quando o Outro é Africano, sendo apenas capazes de o fazer quando o outro é Caucasiano (i.e., congruente com a raça do participante). Uma explicação por detrás destas diferenças, pode ser edificada na dificuldade em individuar os indivíduos do *out-group*. De facto, os participantes adotam a noção que os indivíduos do *out-group* são homogéneos (i.e., não os conseguem distinguir de forma fina) e, assim, apenas os membros do próprio grupo é que são capazes de o fazer. Desta forma, as pessoas assumem que quando o Outro é Africano este será igualmente bom para faces Africanas atípicas e típicas (i.e., faces da própria raça). Ao mesmo tempo, conscientes das diferenças de tipicidade do seu próprio grupo, a diferença de tipicidade é assim mais facilmente capturada quando as faces são Caucasianas (i.e., congruente com a raça do participante). Desta forma, os indivíduos possuem uma teoria leiga do funcionamento da memória ajustada à realidade pelo menos em termos gerais.

Como qualquer dissertação, a presente possui limitações que servem também o propósito de projetar estudos futuros. Especificamente, a escolha de uma amostra composta por participantes Caucasianos e Africanos seria importante de forma a conseguir obter uma análise mais transversalmente adequada e robusta aquando da manipulação da Experiência 4.

**Palavras-chave:** Own-race bias (ORB), julgamentos de confiança, tipicidade, metamemória, reconhecimento de faces

# Table of Contents

## Introduction

Extensive research has shown that people's perceptive capabilities are relatively good when it comes to faces. Simultaneously, along the everyday life's course of individuals, these capabilities are developed (Mondloch, Le Grand, & Maurer, 2002) through experience while exhibiting a reliance on various face characteristics, to represent a face as a whole.

Indeed, the human face is a significant feature, partially or in its whole, when we reflect on what dazzles us the most in our body. Although may this be considered a bold statement, a face gathers an immense collection of characteristics that we perceive, such as gender, age, ethnicity, etc (e.g., Bruce & Young, 2012).

For example, once it struck me while I was entering a bus, I almost immediately knew who the most attractive person was. Instantly, by looking directly at my fellow bus companions' faces, I was able to extract all the information I needed to make a quick judgment about their attractiveness. Years after that curious encounter, in a conversation with a friend, the same issue arose, since he had faced the same experience a few days before, unaware that it had happened to me as well.

Retracting from this example, it seems that we are particularly tuned to faces since they can draw our attention and resources. However, this apparent expertise seems to be affected by group membership (Anastasi & Rhodes, 2005; Wright & Sladden, 2003).

Importantly, research has been documented the fact that individuals show a deficit when it comes to memory for other-race faces when compared to own-race faces (Meissner & Brigham, 2001). This is a significant finding since it has practical implications (Brigham & Ready, 1985), is robust (Meissner & Brigham, 2001), and generalizes across different conditions (Sporer, 2001a). Interestingly, since people show this bias toward their own-race faces, which can be detrimental to societal relationships, one question arises: Are they aware of it?

A novel trend of research has suggested not only individuals are better at remembering their own-race faces, but they can also predict this effect, showing that they are aware of this deficit (Hourihan, Benjamin, & Liu, 2012). However, people's insight capabilities regarding face recognition appear to be modest (Palermo et al., 2017), and this metacognitive prediction found by Hourihan et al. (2012) has not been found in other related research (Chen & Zhu, 2019).

Reverting to the example given above, people sometimes experience the same effects, still not being aware of others' experiences. Interestingly, to the best of our knowledge, no study has investigated the ability to forecast others' judgments regarding own- and other-race faces. Particularly, it seems that individuals tend to estimate other's predictions by anchoring on their own (Koriat, Bjork, Sheffer, & Bar, 2004), therefore, it is possible the people's estimations regarding the ORB may follow that anchoring effect.

Additionally, since faces are similar, their differences are subtle and gradual (i.e., from more distinctive to more typical exemplars). As such, effects of typicality might be important (Cohen & Carr, 1975; Chiroro & Valentine, 1995; Valentine & Endo, 1992), but they have thus far been somewhat ignored in face recognition research concerning own- and other-race faces. For example, when we consider race as a category, we rely on shared features between faces in order to correctly recognize them, whereas for typicality, we rely on features that distinguish among faces within each category.

In this thesis, we tried to comprehend how metamemory, face recognition, and typicality are linked regarding own- and other-race faces. Specifically, 1) we explored how accurate individuals are in predicting their future own- and other-race memory, 2) tried to understand people's beliefs when predicting own and other people's future own- and other-race memory, and 3) in what manner does typicality affect those same predictions and memory performance for own-and other-race faces.

In the next sections, we briefly review the current theories and research that revolve around these three topics.

**Current approaches to face recognition**

In order to obtain the amount of information a face conveys, a singular form of face processing must be involved. Specifically, Yin (1969) indicated that given the difficulties experienced in recognizing inverted faces, some special factor must be behind face recognition. A fit candidate to assume this position corresponds to a distinctive type of processing – *holistic processing* – that is engaged when a face is encountered. In holistic processing people process a face by perceiving all its constituents in a joint manner (i.e., holistically; Tanaka & Farrah, 1993; Young, Hellawell, & Hay, 1987). Supporting this affirmation are the various outcomes from investigations with normal participants such as face inversion (Yin, 1969), the composite effect (Young et al. 1987), and the part-whole effect (Tanaka & Farrah, 1993) that sustain the experimental basis for a holistic face processing. This unique type of processing that humans apply to faces explains why we distinguish them among a myriad of analogous category exemplars (Rezlescu, Susilo, Wilmer, & Caramazza, 2017). Furthermore, as Wilmer et al. (2010) demonstrated, this specific mechanism tends to be significantly directed to faces and genetically transmissible.

Simultaneously, this apparent domain specificity indicates that our brain possesses a dedicated area for face processing. In fact, the fusiform face area (FFA) located in the fusiform gyrus tends to a have a greater response to faces when compared with other stimuli (Sergent, Ohta, & MacDonald, 1992), showing a greater sensitivity to the mere presence of faces (Kanwisher, 2000).

Concurrent to a view of domain specificity introduced before, a notion that a more general mechanism is at play, not only dedicated to faces but other objects, has been introduced and discussed in the literature (Kanwisher, 2000). For example, contrasting with Yin (1969), Diamond and Carey (1986) retorted the uniqueness of face stimuli by showing that other stimuli (e.g., dogs) provoked a similar inversion effect, provided that the participant was a face or dog expert. This result reveals expertise as a solid characteristic for correct discrimination and, consequently, better face processing.

Some empirical evidence has supported this view of a more general mechanism for face recognition instead of a more domain-specific hypothesis where expertise concerning holistic processing is more dominant than a neural substrate especially attuned and developed for face processing. Specifically, Gauthier, Skudlarski, Gore, and Anderson (2000) showed that expertise with cars and birds seems to be a sufficient trait for FFA recruiting and activation. Concurrently, one investigation carried out by Gauthier, Williams, Tarr, and Tanaka (1998) described a similar result while using a different type of stimuli resemblant to a face. They used *Greebles,* which had a structural association and were sorted in different categories such as gender or family. By training participants to achieve expertise in these unique objects, Gauthier and colleagues found that a configural type of processing was also engaged upon Greebles recognition

Thus, an immense debate has been taking place in the face processing literature. On the one hand, some researchers endorse a more specific mechanism that encompasses the face and its processing as a whole. At the same time, empirical evidence based on neuroimaging studies displays a dedicated brain area (i.e., FFA) that is particularly important and developed explicitly for face processing. On the other hand, some researchers acknowledge that a general mechanism which is grounded on expertise acquired during human development is at play, while sharing the same mental machinery between faces and other objects. Nevertheless, it is apparent that we are somewhat competent at processing faces; moreover, our perceptual system displays a predilection for this type of stimuli.

For example, some researchers brought forward the concept of *pareidolia*, which can be described as a tendency to identify faces in everyday life stimulus, which have no direct relation to the latter, therefore, implying a more general tuned mechanism for face percepts (Omer, Sapir, Hatuka, & Yovel, 2019). In addition, eye-tracking studies show that our visual system is primarily driven by faces (e.g., Crouzet, Kirchner, & Torpe, 2010), since we detect them very rapidly (around 110 ms).

A glimpse at a face is enough to extract and infer a wealth of social evidence, such as identity (e.g., Young et al., 1987), age (Anastasi & Rhodes, 2005; Rhodes & Anastasi, 2012), or race (Michel, Corneille, & Roisson, 2007). For example, attractiveness by way of facial beauty emerges as rather noticeable to people, in non-optimal conditions (13 ms stimulus presentation; Olson & Marshuetz, 2005). At the same time, Willis and Todorov (2006) demonstrated that people do not need much time, in terms of stimulus presentation, to make an inference about a particular face. These fast interpretations and evaluations do not improve with increased time, having only an effect on expressed confidence judgments.

To sum up, our perceptual (visual) system is particularly tuned to human faces in various domains and situations. The discussion around which type of mechanism, their connections and symbioses in relation to face processing is still an unfinished issue (Rossion, 2018; Sunday & Gauthier, 2018; Young & Burton, 2018). Nonetheless, our preferences remain clear. Humans are particularly *attracted* to faces and their constituents since it is a source of great information. Still, this apparent expertise does not translate to all face categories.

**Effects of group membership on face processing**

Relevant to this work, in the past few decades, several face-related biases have emerged from research. For example, Anastasi and Rhodes (2005) found that individuals tend to show better recognition memory for faces of their own age group, a phenomenon known as *own-age bias*. One explanation proposed considers that a superior proficiency must be in play and initiated from a significant and recurrent contact with those specific faces from their own-age groups.

In fact, the own-age bias has some implications to real life contexts. For instance, one must consider that in a crime where an eyewitness is present, the age of the latter and of the criminal must be taken in consideration, since it can downplay the testimony's acuity (Anastasi & Rhodes, 2005; Wright & Stroud, 2002). Moreover, in a recent meta-analysis, Rhodes and Anastasi (2012) demonstrated the robustness of the own-age bias, indicating a clear interplay of the perceiver's memory and their age group when there is a mismatch between participants' and target's age. The same pattern of results occurs in the context of gender, where there is a greater recognition memory competence for own-gender faces (Wright & Sladden, 2003).

Central to this thesis is another, but most well-know, bias that focuses on race and it is known by different names such as: *cross-race effect, other-race effect, own-race bias* and *in-group face recognition* (for an extensive review, see Young, Hugenberg, Bernstein, & Sacco, 2012). The seminal work of Malpass and Kravitz (1969) showed that participants had superior recognition memory for "old" Caucasian versus "new" items when compared with a poorer "old" Black versus "new" items memory. This was the first demonstration of the ORB in recognition memory. In 2001, Meissner and Brigham conducted a meta-analysis and found that across different samples, in terms of discrimination accuracy, a moderate ORB was present and accounted for a portion of variance related to the phenomenon. At the same time, the response criterion of participants showed that for own-race faces a more stringent criterion was used instead of a more liberal criterion for other-race faces. Moreover, Meissner and Brigham (2001) found a “mirror” effect, implicating that the hit rate was larger for own-race faces versus a higher of false alarm for other-race faces.

At this point, a clear definition for the own-race bias (ORB) must be made. Specifically, the ORB is a tendency to recognize more easily and consistently faces of own-race members when compared with faces of other-race members. In other words, individuals seem to be better at remembering and, consequently, recognizing faces of their own-race. This reveals a sensitivity to own-race faces reflected in a better discrimination and strict criterion in terms of face memory recognition (Wilson, Bernstein, & Hugenberg, 2016).

In fact, the ORB has deep social implications (Young et al., 2012) and it is a very robust and replicable effect (Meissner & Brigham, 2001). For example, Bothwell, Brigham, and Malpass (1989) executed a meta-analysis and found that the ORB was present in 79% of the samples they examined. Moreover, this phenomenon is not only observed in experimental context and “inside the laboratory”. For instance, Sporer (2001a) described the reliability of the ORB in more ecologically valid studies.

As an example, research revolving line-up identification and eyewitness testimony has shown a clear bias in these situations (Brigham & Ready, 1985; Wilson, Hugenberg, & Bernstein, 2013). For example, it has been noted that lineup identification and its contruction tends to suffer in terms of fairness (Brigham, Meissner, & Wasserman, 1999; Malpass and Lindsay, 1999). Thus, we should discuss the mechanisms, factors and what is behind the ORB and the theoretical accounts proposed to explain and classify the latter.

**What causes the ORB?.** In general, three kinds of theoretical accounts must be considered: perceptual expertise, social cognitive, and hybrid (i.e., conjunction of the two first models) accounts (Young et al., 2012).

Firstly, regarding perceptual expertise accounts, one must pinpoint the face recognition processes discussed at the onset of this discussion. Specifically, since greater expertise leads to better face recognition (Diamond & Carey, 1986) and experience does affect it (Tanaka & Gauthier, 1997), a separated processing style must occur when observing own and other-race faces. Simultaneously, a clear reliance on more holistic processing is a hallmark of face recognition (Young et al.,1987).

On the one hand, research has shown that own-race faces are perceived holistically, whereas other-race faces are perceived in a more featural manner (Michel, Caldara, & Rossion, 2006; Mondloch et al., 2010; Rhodes, Brake, Taylor, & Tan, 1989; Tanaka, Kiefer, & Bukach, 2004). For example, Rhodes et al. (1989), utilizing inverted stimuli (i.e., European and Chinese faces shown in an upright or inverted position), revealed that own-race faces triggered a larger effect than other-race faces. This result indicates that individuals rely on a configural type of processing when observing own-race faces. Similarly, Tanaka et al. (2004) showed the same pattern of results while utilizing a different paradigm (whole/part) and that experience with own-race face tends to stimulate its processing in a White and Asian sample. In addition, using a similar procedure as Tanaka et al. (2004), Michel et al. (2006) exhibited that same holistic processing preference for own-race faces with the same sample characteristics.

On the other hand, experience can have an effect not only on the recognition of own-race faces but also on recognition of other-race faces. Particularly, Tanaka et al. (2004) and Michel et al. (2006) found that the Asian participants tended to process both White and Asian faces in the same configural manner. Thereby, these Asian participants were equally experienced on own- and other-race faces, highlighting the role of experience in these situations.

In line with the aforementioned, Lavarkas, Buri, and Mayzner (1976) found that the manner in which a participant (in this case, White) contacts with an another-race face is a crucial feature of experience when determining this expertise advantage.

In this sense, whether you have other-race friends, or you live in a densely other race populated area, this type of contact is influential on how you perceive other-race faces. This can be seen in the investigation of Hancock and Rhodes (2008), where they measured the quantity and quality of experience that both Caucasian and Chinese participants had with each other. They showed that any type of contact has its toll on some of the ORB effects, whether altering how holistic was the processing of both faces or the differentiated recognition accuracy for both faces. The quantity of contact, albeit its importance, does not necessarily reduce the abovementioned effects (Hancock & Rhodes, 2008).

Therefore, it appears that quality of contact is more decisive when it comes to own- and other-race to face processing and recognition memory. Prior to developing on other perceptual-expertise accounts, a discussion should be made regarding differences between perceptual-based tasks and memory-based tasks.

Specifically, Horry, Cheong, and Brewer (2015) found that perceptual tasks are not correlated with recognition performance, arguing against the notion that holistic processing drives the ORB effect. Similarly, it seems that behind it, a relationship between facial features exists and compels the ORB (Lewis & Hill, 2018). Consequently, perceptual effects on own- and other-race faces can be limited, and other accounts must be considered in order to explain the ORB.

An alternative account is the multidimensional face-space framework (MDS) postulated by Valentine (1991, 2001), which states that the different types of processing that govern own- and other-race faces are rooted in a memory disparity of stored exemplars of the latter. Indeed, it is assumed that own and other-race faces are depicted in a continuum, along different dimensions as if distributed in a Euclidean space. This means that each face characteristic corresponds to a specific dimension (e.g., nose size, distance between the eyes). At the same time, the previously seen faces are represented accordingly, varying in terms of characteristics that, in turn, are dimensions.

Therefore, faces are encoded as typical in the point of origin and more distinctive as we progress along the continuum. Thus, own-race faces that are often experienced tend to concentrate in the middle and are seen as more typical according to its various dimensions. Likewise, other-race faces, which are less experienced, tend to locate in the extremes and close together, since the insufficient contact with the latter does not provide differentiated characteristics and exemplars (See Figure 1). Therefore, as Young et al. (2012) wrote:

> the dimensions of face-space represent variations from the "typical" or average features of a face, such that the center of the space would represent a prototypical face exemplar and spaces farther from the center would represent faces deviating greatly from the appearance and variability of faces usually encountered. (p.120)

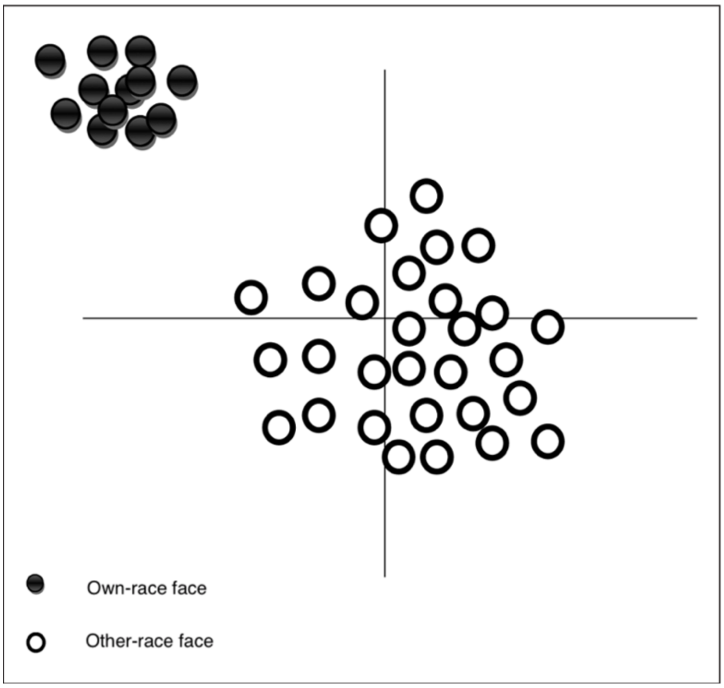


*Figure 1*. A graphic representation of the MDS distribution of own- and other-race faces, showing the own-race face dispersion and other-race clustering. Adapted from Young et al. (2012).

As an example, Rhodes, Locke, Ewing, and Evangelista (2009) found that there is a lack of expertise with other-race faces even when giving instructions to encode particular characteristics that compose these faces. According to the MDS, this indicates that, since individuals do not possess a great range of exemplars for other-race faces, the mere solicitation to perceive determinate characteristics is not sufficient to accomplish a better recognition of those faces. At the same time, Byatt and Rhodes (2004) showed that participants indeed organize their perceptual experience with own and other-race faces within dimensions. As a result, individuals demonstrated a poorer distribution of exemplars from other-race faces, as they were not coded in all of the dimensions. Therefore, the clustering of other-race faces occurred while own-race faces were evenly distributed. In fact, Papesh and Goldinger (2010) constructed an experimental face-space model of own- and other-race faces dispersion, showing that other-race faces are densely clustered and own-race faces are more disperse. Therefore, own-race faces are more well-organized and distributed along the MDS space.

Even though the perceptual expertise accounts are capable of giving some explanations for how we perceive and recognize own- and other-race faces, some caveats still linger. In fact, the experience acquired throughout the years with other-race faces – the basis for the contact hypothesis – only accounts for a 2% portion of the variance regarding this phenomenon (Meissner & Brigham, 2001). Additionally, Ng and Lindsay (1994) found no evidence to support the contact hypothesis, since experience with the other-race did not diminish the ORB. Other factors like the repeated presentation of stimulus seem to play a detrimental role in the ORB as well. In this case, Palma and Garcia-Marques (under revision) manipulated the repetition of own- and other-race faces in a series of experiences and uncovered an incremental effect of repetition, in the sense that, with repeated presentation of faces, the ORB increased. In turn, this means that having experience or being given the opportunity does not necessarily translate into expertise.

Thus, the quality and quantity of contact that an individual has with other-race faces seem insufficient as a factor to explain why some have a better recognition performance than others. Hence, another type of model that goes beyond the limitations of perceptual expertise theoretical accounts must enter this discussion, since other factors can be at play.

The social-cognitive perspective is based on the premise that individuals perceive the world in terms of categories, discerning between those who belong to their group or another group, as well as individuating in-group members while categorizing out-group members (e.g., Allport, 1954; Fiske & Neuberg, 1990; Hewstone, Rubin, & Willis, 2002; Macrae & Bodenhausen, 2000; Tajfel & Turner, 1979). Specifically, the *cognitive disregard* model is one of the hallmarks concerning the abovementioned perspective. Proposed by Rodin (1987), it states that individuals tend to use cues to disentangle which stimulus is irrelevant, consequently, “disregarding” it and economizing cognitive resources for those to which they are motivated.

Thus, the cognitive disregard model rests on the expectation that categorization will lead to inferior recognition performance for other-race faces since perceivers are not attuned to individuate and differentiate those same faces as they are for own-race faces.

One investigation that builds evidence and is coherent with the assumptions of this model was presented by Bernstein, Young, and Hugenberg (2007). The results obtained by the researchers went beyond the ORB. Participants in this experiment were White as were the faces provided as the stimulus, varying only in college affiliation. Bernstein and colleagues showed that this difference in in-group vs. out-group was sufficient to cause a deficit in recognition. Specifically, participants had poorer recognition of faces that pertain to the other university, whereas those faces belong to the same race category (i.e., White).

According to the cognitive disregard model, these participants were not necessarily motivated to individuate those out-group faces. Since differences in group membership are sufficient to determine recognition performance, perceptual expertise models cannot account for these results, as the experience with those faces was constant (Young et al., 2012).

At the same time, Hugenberg and Corneille (2009) found similar results to the abovementioned and demonstrated that maintaining the effect of perceptual expertise fixed (i.e., participants’ race matches stimulus’ race), the holistic processing of own-race faces suffers since participants were better at recognizing in-group faces than out-group faces.

Another account that pertains to the social cognitive perspective corresponds to Levin’s (1996, 2000) feature-selection model, which lays on the notions of individuation and categorization. In this model, it is assumed that individuals give preponderance to specific features that allow them to process better in-group members by individuating them, which, in turn, leads to superior recognition of own-race faces. On the contrary, categorizing out-group members leads to a lack of attention to specific individuating face characteristics that results in poor recognition of other-race faces. Evidence supporting this idea of racial categorization specifies that other-race faces are perceptually harder to discriminate and, therefore, recognize (Susa, Meissner, & De Heer, 2010). Finally, hybrid accounts try to pursue explanations for the ORB while integrating both types of the abovementioned effects (Young et al., 2012).

The *dual-process framework* proposed by Meissner, Brigham, and Butz (2005) postulates that face recognition of own- and other-race faces are an intertwined process between memorial processes and social cognitive views. Specifically, dual-process theories of memory (e.g., Yonelinas, 1994) are based on the notion that two processes occur in recognition memory. While some recognition judgments rest upon an evaluation of familiarity with the stimulus, others are driven by a recollection-based assessment (Yonelinas, 1994; Yonelinas, 2002).

On the one hand, Meissner et al. (2005) propose that other-race recognition is prone to errors since the familiarity-based process associated with them tends to be undemanding and rests on the availability of stored material. Drawing from social cognitive accounts like Rodin's cognitive disregard (1987), this lenient process is interconnected with the subjective neglecti of other-race faces, since individuals are not motivated to attend to them. On the other hand, behind the ORB is a more effortful and strenuous recollection-based process where participants effectively encode information and retrieve them. At the same time, the subjective significance of own-race faces drives participants to be attuned to them (Rodin, 1987).

To sum up, there is not a single, most suitable explanation for the ORB and its mechanisms. Overall, individuals are clearly biased in terms of face memory and recognition, whether it's driven by social cognitive influence or by perceptual experience.

As we shall see next, it turns out that the ORB also exists in metacognitive judgments, specifically metamemory.

**Metamemory predictions**

Before entering the more specific literature concerning *metamemory*, we should address a broader topic - *metacognition.* The latter can be described as simply thinking about thinking or being aware of your own thoughts (e.g., Koriat, 2016). Specifically, Nelson and Narens (1990) devised a framework to explain this interplay of cognitive processes and elucidate what processes undergo metacognition. These two levels correspond to the *object-level* and *meta-level*, which have two specific ways to interact – control and monitoring. The former refers to a mechanism that enables modifications by the meta-level in the object level. The latter denotes an online process of analysis of the processes occurring in the object-level, informing the meta-level (See Figure 2).

This assumes that individuals are rather active in knowing and assessing what they perceive in a feedback and feedforward loop of information (Koriat, 2016). At the same time, it is well-defined that these relations occur in terms of our memorial processes and should be an integral part of how our memory is and scrutinizes its performance (Goldsmith & Koriat, 2007). Hence, we should define and discuss this particular segment of metacognitive functioning related to memory.

Firstly, we can explain metamemory as a process embedded in meta-analytic judgments about our actual memory. It can be accessible in a prospective (e.g., judgments of learning) or retrospective (e.g., confidence judgments) manner in order to assess our memory's content. Briefly, metamemory refers to actively reflecting on our own memory. (e.g., Metcalfe & Dunlosky, 2008; Schwartz & Metcalfe, 2017).

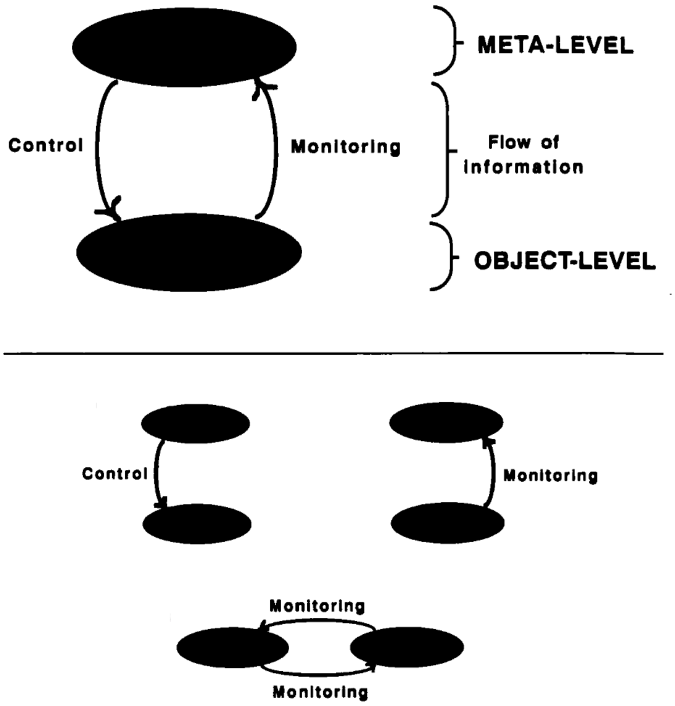


*Figure 2.* Representation of the hierarchical relations between the meta-level and object-level. Adapted from Nelson and Narens (1990)

One of the forefathers of this movement in terms of research interest can be recognized with a seminal work on the feeling-of-knowing experience. Specifically, Hart (1965) devised a couple of experiences to address this retrieval problem and found that this feeling tended to be an accurate predictor of subsequent memory performance. Thus, although participants could not recall some materials when allowed to recognize them, their *a priori* prediction was correlated with performance afterwards. Concomitantly, research has shown similar results on the feeling-of-knowing experience where participants accurately foresee their memory, functioning like a metaphorical window for probing memory for stored elements (Blake, 1973, Nelson, Gerler, & Narens, 1984).

Indeed, this work serves as an introduction for considering how accurate people are at predicting their future memory and the judgments that mainly concern this thesis. For example, Koriat and Goldsmith (1996) found a strong relationship between performance and metacognitive judgments, where correct answers were highly correlated with prior judgments in terms of scale. Additionally, Schwartz and Metcalfe (2017) pointed out that judgments of learning are generally suitable in predicting future memory performance. However, some contradictory evidence has been put forth by Rhodes and Castel (2009), showing that JOLs are prone to illusions and are affected by different types of cues.

Specifically, JOLs are estimates of the likelihood that a certain study item will be remembered in the future (Koriat, 2016; Nelson & Leonesio, 1988). In general, this type of judgment tends to be required after a single or series of stimulus presentations where individuals have to study that material. In addition, JOLs can be given according to different types of scales (e.g., probability (i.e., 0-100), dichotomic (yes/no), or Likert (1-9) (Schwartz & Metcalfe, 2017).

Some theories have been proposed to explain what is behind a JOL. According to a prominent approach, the cue-utilization approach (Koriat, 1997), when making JOLs, individuals rely on their beliefs and other cues as they monitor their learned information stored in memory.

For instance, Mueller, Tauber, and Dunlosky (2013) showed that beliefs are influential when making a JOL. Specifically, these researchers found that participants used their relatedness belief to give JOLs since they gave higher recall ratings to pairs that were more related in comparison to pairs that were not. Contrastingly to the effect that beliefs have on self-predictions, Tauber, Whiterby, and Dunlosky (2019) showed that specific beliefs about aging do not contribute as importantly as previously thought, since individuals' belief about how aging tolls memory did not affect participants' JOLs. At the same time, other cues such as experience seem to play a role in this process of making JOLs.

In fact, other investigations have shown that fluency (e.g., in encoding and retrieval) serves as a cue to direct individuals (Koriat, 2016). For example, Palma, Santos, and Garcia-Marques (2018) found that by manipulating the repetition of stimulus, JOLs were affected through how easily the stimulus was perceived, being an indication of future rememberability. Additionally, manipulations concerning font size have been proven to influence how JOLs typically are given, by facilitation in terms of processing (Mueller, Dunlosky, Tauber, & Rhodes, 2014; Yang, Huang, & Shanks, 2018). Conversely, Yang et al. (2018) demonstrated that while beliefs are significant, fluency can be considered as being side by side in terms of importance when an individual reflects on given a JOL.

Thus, although different aspects are prominent in these prospective types of judgments, it seems clear that people actively monitor their judgments. However, in spite of these different implicit or explicit strategies, individuals are still affected, and these prospective judgments do not always reflect accurate memory foreseeing.

Nevertheless, some examples can be found in the literature where specific task manipulations are able to improve relative accuracy in JOLs. Comparably to Palma et al. (2018) work, where they employed stimulus repetition, researchers have been using the latter manipulation combined with practice in order to improve JOL accuracy. For example, Koriat, Sheffer, and Ma'ayan (2002) presented an analysis of different investigations samples that showed that while practice impaired overall accuracy, it actually improved relative accuracy.

Another technique corresponds to soliciting a JOL in a delayed window of time instead of asking for a JOL immediately after studying. Specifically, this delayed-JOL effect (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991) is bounded in a better future memory performance match with prior metacognitive judgments, since relative accuracy was superior in the delayed condition than in the immediate condition.

An additional type of judgments – Confidence judgments – shall be described now. Specifically, confidence judgments are usually given after testing for the previously studied items (Dunlosky & Metcalfe, 2009). In metacognitive research that employs standard recognition paradigms related to memory performance, low-confidence responses are linked with random performance, whereas high-confidence responses are linked with near-perfect accuracy (Mickes, Hwe, Wais, & Wixted, 2011). Confidence judgments usually assume the same types of scale reported earlier (e.g., Likert).

Koriat (2016) proposes that like JOLs, confidence judgments are influenced by the fluency of how items are processed. For example, Finn and Tauber (2015) discuss a variety of diverse factors such as encoding and retrieval fluency that have an effect on how individuals attribute their confidence in terms of judgments. In addition, Busey, Tunniclif, Loftus and Loftus (2000) pointed that the accessibility of information can act as a cue for confidence judgments and recognition memory, thus sharing the same information eventually implies that they can predict each other accurately.

Nevertheless, accuracy is not equal across the different types of metacognitive judgments (Koriat, 2016), so a line should be drawn between prospective and retrospective judgments. At the same time, metacognitive measures have two independent aspects – *calibration* and *resolution* - to be addressed, since heterogeneous results have been found between the two (e.g., Koriat & Goldsmith, 1996)

The debate surrounding how one can measure metacognitive judgments has been intense throughout the past decades. However, not being necessarily inserted in this thesis' scope, one should denote some specificities of measuring metacognitive judgments.

*Absolute accuracy* or calibration can be defined as a match between overall metacognitive judgments and memory performance (Benjamin & Díaz, 2008; Koriat, 2016;

Rhodes, 2019). *Relative accuracy* or resolution refers to an item-by-item judgment and its relation to later memory performance for each item (Rhodes, 2019).

With relevance to this thesis, relative accuracy is usually assessed through a non-parametric correlational measure – Goodman-Kruskal (G) - proposed by Nelson (1984), ranging from -1 to +1. Particularly, this means that values near +1.0 are related to higher JOLs and subsequent better item recognition, whereas values near -1.0 are related to higher JOLs and subsequent poorer item recognition. We decided to use this measure since the bulk of research revolving around JOLs has used it, rendering it preferable in order to accommodate future results and discussion in current views in the literature.

**Predicting Other's performance.** Even though research has been conducted in metacognitive topics regarding individual's ability to infer theirs's and their peer's performance (e.g., Kruger & Dunning, 1999; Mata, Ferreira & Sherman, 2013), only a few of articles has assessed metamemory for others (Tullis & Fraundorf, 2017).

Before reporting some of the evidence found, some explanations are necessary to frame this topic. In particular, to forecast another individual's performance, one must have some insight into the other. This perspective-taking is usually self-based (Nickerson, 1999) and, therefore, based on beliefs and experience. For instance, Epley, Keysar, Van Boven, and Gilovich (2004) found that individuals, when trying to shift their perspective to another one's, utilize their own first and make small adjustments along the way. Consequently, this approach is normally insufficient to transform one's initial perspective completely, showing how egocentric and biased we initially are.

Intriguingly, as discussed earlier, some of the rationalizations behind JOLs that pinpoint that cues are important and affect this type of judgments (Koriat, 1997) can be adapted to explain how individuals judge the performance of others. Specifically, Tullis (2018) demonstrates that cues such as experience, fluency, and familiarity with a stimulus affect other-predictions in the same manner as self-predictions. Also, it has been claimed that underlying how people make these types of judgments are essentially the same processes that govern how one make self-assessments (Jost, Kruglanski, & Nelson, 1998)

In regard to the accuracy of predictions towards another one's performance, diverging findings have been presented. For instance, Vesonder and Voss (1985) showed that individuals are poorly accurate when predicting another's memory performance. Alternatively, Tullis and Fraundorf (2017) contested these results by demonstrating that individuals are able to predict another person's recall; however, in terms of JOL accuracy, the same pattern does not emerge. In other words, participants' estimations showed a (correct) belief for others' recall, nevertheless, being less accurate than estimations of own recall. At the same time, Hargis and Castel (2019) presented results that confirm this inferior accuracy pattern of predicting other's memory capabilities, insofar as meaningful information is given a priori.

**Effecst of group membership in Metamemory**

We now turn into the evidence that shows the spillover of the ORB to metamemory. Faces have been used as stimuli in these types of investigations (Busey et al., 2000; Sommer, Heinz, Leuthold, Matt, & Schweinberger, 1995; Watier & Collin, 2011) and, to the best of our knowledge, only a handful of investigations have tapped into ORB effects in metamemory (Arnold, 2013; Chen & Zhu, 2019; Hourihan et al., 2012; Ngueyn, Pezdek, & Wixted, 2017; Ngueyn, Abed, & Pezdek, 2018; Tullis, Benjamin, & Liu, 2014).

Specifically, Waitier and Collin (2011) used face-name targets associated with each other and demonstrated that individuals accurately predicted their memory for this type of relation. At the same time, Busey et al. (2000) showed a difference between types of metacognitive judgments induced by face characteristics (i.e., luminance), showing a diminishing monitoring efficacy of memory for these stimuli. Overall, participants were worse at recognizing faces tested in bright settings when they were firstly presented in a dim setting.

In relation to ORB, we are particularly interested in how metamemory works for own- and other-race faces and its connection to actual memory performance. Like in most studies, a replication of the ORB in recognition memory performance is found, meaning that superior recognition of own-race faces in detriment of other-race faces has been consecutively demonstrated. Nevertheless, the finding that is most striking corresponds to the observation of an ORB in metamemory where individuals are better at making predictions for their own-race faces, demonstrating that people can accurately forecast their memory performance for the latter (Chen & Zhu, 2019; Hourihan et al., 2012).

Specifically, Hourihan et al. (2012) work is particularly significant since it was one of the first to show this metamemorial ORB. Using a sample composed of White and Asian participants, they applied a standard memory paradigm that consisted of a study and test phase. In the study phase, they presented pictures of own- and other-race faces to each group of participants and, after each face presentation, a JOL was made. Afterwards, in the test phase, the studied items were once again presented together with distractor items. The main results of this experiment revealed, as commonly described in the literature, the superior recognition accuracy of White faces from White participants and vice-versa for the Asian participants/faces (i.e., ORB). At the same time, the participants' performance in terms of metamemory was shown to be more attuned to faces of their own-race, since participants' relative accuracy was better for the latter. Thus, the observed results can be triggered by a differential facial encoding that characterizes some of the theoretical accounts presented before.

Smith, Stinson, and Prosser (2004) devised an experiment as an attempt to mirror an actual eyewitness lineup identification in order to study the effects of Cross-race in erroneous eyewitness identification. By showing small videos of staged thefts and asking for pre and post-identification confidence judgments (i.e., prospective and retrospective judgments), the authors obtained a replication of the ORB again and observed higher confidence for own-race photos not only in pre-identification, as seen by Hourihan et al. (2012), but also in post-identification judgments.

Nevertheless, the above-mentioned results regarding metamemory have not been entirely replicated. For instance, Rhodes, Sitzman, and Rowland (2013) showed that individuals are insensitive to race, in the sense that they did not accurately discriminate between own- and other-race faces. In addition, Chen and Zhu (2019) engineered an experiment similar to Hourihan et al. (2012) with a different face manipulation where they utilized morphing to merge White and Asian faces, obtaining a set of three face conditions: ambiguous, White and Asian. The results showed a trend of higher JOLs for own-race faces when compared with other- and mixed-race faces. However, when they analysed relative accuracy for own-race faces, they found no significant differences between the three conditions. Moreover, Chen and Zhu (2019) demonstrated that the association between actual memory performance and JOLs was low (i.e., individuals were underconfident).

Additionally, Rhodes et al. (2013) manipulated the metacognitive control by allowing participants to control the time spent studying the stimulus. Consequently, participants' study time did not differ for own and other-race faces. Moreover, participants gave higher JOLs independently of having the opportunity to restudy the faces that they deemed necessary. In turn, this opportunity did not affect recognition performance of other-race faces, showing a disability to acknowledge that other-race faces may need additional study time.

Similarly, it has been found that the ORB remains unchanged and endures regardless of instructions or the ability to control the study of other-race faces (Tullis et al., 2014). Therefore, this metacognitive inability to supersede this effect is related to differential mechanisms. As proposed by Tullis et al. (2014), the lack of effective and deliberate control shown by individuals cannot be linked with the subjective insignificance of other-race faces, since they had instructions to individuate and had self-paced study time. In turn, individuals are consciously aware and invested in other-race faces, which is incompatible with Rodin's (1987) cognitive disregard theory.

Recapitulating, individuals do not always seem accurate in predicting their future memory, in terms of relative accuracy,

**Typicality**

Not all faces are good category exemplars (Rosch, 1973). Behind this intra-category goodness of fit of faces is the notion that categories assume a graded structure, making some exemplars more representative of that category than others (Ma, Correll, & Wittenbrink, 2018).

Specifically, *family resemblance* models assume that categories' best exemplars are considered to be the average one's in the category (Rosch & Mervis, 1975). In other words, face typicality is the way to be more categorically representative.

These differences in face category membership and the reliance on typicality or, in opposition, distinctiveness can be detrimental. Research in social cognition has shown that variations of Afrocentric features (i.e., Black prototypical characteristics) such as skin tone, lip thickness, among others (Blair & Judd, 2010; Blair, Judd, Sadler, & Jenkins, 2002) serve a socially nefarious purpose that results in a tendency of judging more negatively Black individuals and associating the latter with criminality and aggression (Kleider-Offut, Bond, & Hegerty, 2017). Indeed, it seems that these features are central to what is considered prototypical by facilitating the integration of these faces on race categories (i.e., Black individuals; Kleider-Offut et al., 2017).

Effects of typicality have been documented in research for the past few showing to be influential in recognition memory processes (Bruce, Burton, & Dench, 1994; Burton & Vokey, 1998; Sporer, 2001b; Valentine & Bruce, 1986a; Valentine & Bruce, 1986b). Normally, one can consider that not all faces seem equal to us. For example, Cohen and Carr (1975) demonstrated that highly distinctive (i.e., atypical) faces are rated as such, which in turn is in concordance with the hypothesis that some faces are more recognizable than others (Ford, 1958, cited in Sarno & Alley, 1997). For instance, atypical faces are far better remembered than typical ones and, that the interitem resemblance is a major factor of typicality (Light, Kayra-Stuart, & Holander, 1979; Winograd, 1981). Consequently, similarity is detrimental and, therefore, typicality of faces can have an effect in terms of recognition.

Vokey and Read (1992) identified two components of typicality, *familiarity* and *memorability* . Particularly, the authors propose that a face can have a degree of resemblance to another that an individual may have encountered (i.e., familiarity) or a degree of distinctiveness that makes it more memorable than others (i.e., memorability). In this case, an atypical face would be more memorable than familiar. Contrarily, Morris and Wickham (2001) argue that the results of Vokey and Read (1992) are an outcome of typicality measures and that metamemorial beliefs may play a role in typicality. Accordingly, individuals assume that distinctive faces are better recognized, which is consistent with evidence put forth by Cohen and Carr (1975).

Finally, Corneille, Huart, Becquart, and Brédart (2004) reported the effects that typicality can have on memory, showing that the mere presence of typical exemplars (i.e., faces that were moderately typical) leads to the overestimation of race by participants. Specifically, the authors presented faces that were slightly (i.e., 30%) typical exemplars of Caucasian faces which resulted in recollecting more Caucasian faces that those that were presented.

**Effects of typicality in ORB.** Even tough typicality effects have been widely investigated in social cognition and applied settings, they have been largely ignored in research on the ORB. Still, a few studies have been conducted that relate to this issue. Valentine and Endo (1992) found that British participants were more sensitive to own-race than other-race faces (i.e., participants exhibited the ORB), but that difference was not observed for Japanese participants (i.e., they showed superior recognition of other-race faces), which they explained as resulting from globalization and contact with Western culture. At the same time, participants were better in terms of recognition of distinctive faces (i.e., atypical faces) than typical ones, demonstrating sensitivity to deviations in distinctiveness as well. This result can be organized within an explanation advanced by Valentine and Endo (1992):

> The recognition advantage for own-race faces arises because an own-race face is likely, on average, to be more distant from its closest neighbour, therefore making own-race faces easier to identify. The effect of distinctiveness reflects the difference in exemplar-density around typical and distinctive faces. Although it is necessary to assume that the absolute exemplar density of other-race faces is greater, the difference in exemplar density of distinctive and typical other-race faces could be equivalent to the difference for own-race faces. (pp. 693-694)

Another important example corresponds to Chiroro and Valentine (1995) investigation which shows that meaningful contact with the other-race can act as a moderator of the ORB since participants with reported high contact with other-race faces had a reduced bias toward the latter.

**Effects of typicality in metamemory.** The usage of distinctiveness or typicality as concepts in the metamemory literature has been scarce, but still some studies have been done. Particularly, Watier and Colin (2012) propose that distinctiveness can be considered as a source of memory monitoring information. Moreover, Sommer et al. (1995) devised an electrophysiological and behavioural experiment to assess the relation between JOLs, facial distinctiveness, and brain processes. Specifically, the authors demonstrated that these prospective judgments were accurate in predicting face recognition performance. Also, facial distinctiveness was found to be influential on JOLs ratings, which is in line with the assumption that distinctiveness can be utilized as a source for judgments.

## Thesis' Overview

Considering the abovementioned review, this thesis strives to further investigate how accurate a person can be about their future recognition performance of own and other-race faces and if the typicality of the latter – being atypical or typicality – could have an effect on the ORB and participants' metamemorial predictions. Across three experiments, we systematically explored the participants' awareness of the ORB and of typicality effects and its subsequent accuracy in relation to recognition memory.

The first experiment initiates the exploration of this objective by employing a standard recognition paradigm and asking for participants' predictions before and after the recognition phase. The second experiment aimed at extending the previous experiment, using the same experimental design, while also assessing participants' predictions about others' memory. The third experiment manipulated participants' beliefs abouts who they are thinking about when they are making these predictions for others. In particular, we manipulated the other person by controlling for their race.

Overall, we expect that people will be better at remembering own-race than other-race faces. We expect the same pattern for typicality (i.e., superior recognition of atypical than typical faces). Additionally, we pose some questions that these experiments might untangle:

a) Are people aware of the ORB and typicality effects for themselves and others?

b) Are people metacognitively accurate as demonstrated by Hourihan et al. (2012)?

c) Can people capture the differences in typically and translate that to their memory and metacognitive judgments?

d) Are they as good for typicality as they are for race?

## Experiment 1

Adapted from Hourihan et al. (2012), this first experiment was designed to understand how accurate a person is about their future memory performance. Moreover, we explored the effects of typicality within own- and other-race faces.

So as to accomplish the above and following the procedure and techniques by Hourihan et al. (2012), we used a within-participant design with four experimental conditions [2 (Race: Own vs Other) x 2 (Typicality: High vs Low)]. For the purpose of accomplishing the aforementioned objectives, we employed a standard recognition paradigm with two measures of metamemory, JOLs and confidence judgments, for each face. We expect that recognition performance will be better for Own-race faces than Other-race faces (i.e., ORB).

**Method**

**Participants.** In this experiment, fifty-six undergraduates (46 females and 9 males; $M_{age} = 20.9$, $SD_{age} = 6.44$; one participant did not report their age) from the University of Lisbon were recruited in exchange for course credit. All participants were Caucasian.

**Stimuli.** To fulfil the objective mentioned above, a set of stimuli was gathered from a myriad of databases[1], as suggested by Sergent (1986), in order to control for different methodological aspects. We selected 60 Caucasian (30 Low-typicality and 30 High-typicality) and 60 Black (30 Low-typicality and 30 High-typicality) pictures of male faces from a pool of 320 faces previously pretested on their level of racial prototypicality (see Figure 3). The pictures consisted of only the targets' faces and hair presented against a white background. All faces displayed a neutral expression and a direct gaze. All photographs were presented in grayscale in a size of 382 x 330 pixels. We presented the pictures with a slight difference in luminosity and contrast between study and test phases to reduce pictorial recognition (see Bruce, 1982).

The 120 faces were randomly divided into two lists of 60 faces each. In each of these lists, there were 30 Caucasian faces (15 Low and 15 High-typicality) and 30 Black faces (15 Low-typicality and 15 High-typicality). These lists were counterbalanced across participants such that each participant was equally likely to see a given face as either a target face (i.e., presented in the study phase) or a distractor face (i.e., not presented in the study phase).

---

[1] The stimuli was composed of 82 images of faces from the Chicago Face database (CFD – Ma, Correll, & Wittenbrink, 2015), 11 faces from the NimStim face database (Tottenham, Borscheid, Ellertsen, Marcus, & Nelson, 2002) , 13 faces from the FACES database (Ebner, Riediger, & Lindenberger , 2010), 12 faces from the Meissner face database (Meissner et al., 2005) and lastly, 2 faces from the mr2 face database (Strohminger, Gray, Chituc, Heffner, Schein, & Heagins, 2016). These faces were selected for a previous study and used for this thesis.

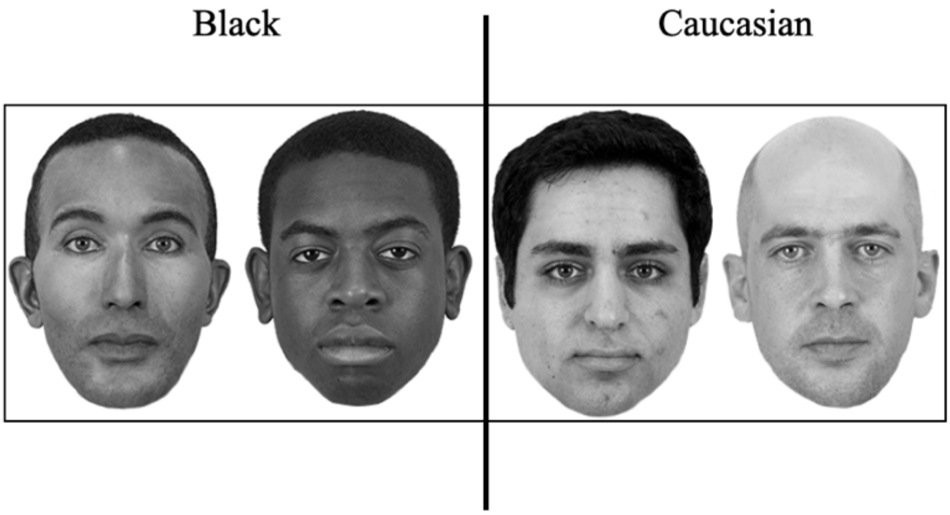


*Figure 3*. These 4 faces are an example of the stimuli utilized and were retrieved form the CFD (Ma et al., 2015). In the left panel, two Black faces are depicted, the first being highly atypical (i.e., Low-typicality) and the other highly typical (i.e., High-typicality). In the right panel, two Caucasian faces are presented in the same manner.

**Procedure.** Participants were tested in lab sessions of up to eight participants at a time. They sat in front of computer screens in individual workstations. All instructions and stimuli were presented on the computer. The experiment was programmed and ran in E-Prime, Version 2.0 (Psychology Software Tools, Pittsburgh, PA). Participants were first informed that their participation was entirely voluntary and that their responses were anonymous and confidential. They were then told that they would see several faces, one at the time, and pay close attention to them for a later recognition memory test. They were also informed that they should make a recognition prediction for each face. Each trial began with a fixation cross ("+") presented in the screen center for 500 ms. Then, a target face was displayed for 3 seconds, followed by a 750-ms black screen. The order of presentation of the 60 faces was randomized anew for each participant. Then, the prediction (JOL) screen appeared. At the top of this screen, a sentence instructed participants to "Please indicate how likely you think it is that you will recognize the face you have just seen in a later phase of this experiment." A rating scale ranging from 1 ("I am sure that I will NOT remember this face") to 9 ("I am sure that I WILL remember this face") was displayed in the center of the screen. Participants responded by pressing a key from 1 to 9. This screen remained visible until participants responded.

After completing the study phase, participants received the instructions for the recognition phase. They were told that they would see a set of faces, some of which they had seen in the study phase (target faces) and some not (distractor faces), and that their task would be to make two judgments for each face. Specifically, each test trial started with a 500-ms fixation cross that was immediately followed by a face. First, participants had to indicate whether they had seen the face in the previous study phase or not. If they recognized the face, they pressed the "M" key, and if they did not, they pressed the "C" key. After pressing one of these two keys, a screen with a confidence scale appeared on which participants had to rate their confidence in the decision they had just made. They were asked to indicate, "How confident are you in the response you just gave?" using a 9-point scale ranging from 1 ("Absolutely Uncertain") to 9 ("Absolutely Confident"). Half of the faces were target faces (60) and half were distractor faces (60). The order of presentation was randomized anew for each participant. Following the recognition test, participants were debriefed and thanked.

## Results

**Analytic strategy.** Following the standard practice in the ORE literature (see Meissner & Brigham, 2001), we computed the signal detection measure *d'* to assess recognition accuracy

(Green & Swets, 1966; Macmillan & Creelman, 2005)[2]. Specifically, *d'* is a discriminability measure that refers to how individuals differentiate “old” (i.e., previously seen items) and “new” (i.e., distractor items) faces. Higher *d'* values correspond to greater sensitivity. To account for perfect performance (i.e., hit rate = 1 and false alarm rate = 0) we applied a correction of 0.5/n and (n – 0.5)/n was to hit (HR) and false alarm (FAR) rates of 0 and 1, respectively (Macmillan & Kaplan, 1985; Stanislaw & Todorov, 1999).

In order to avoid confusion regarding metacognitive judgments in the analysis, we adopted Nguyen et al.’s (2018) nomenclature to address prospective (JOLs) and retrospective confidence judgments. Specifically, since we are measuring confidence in two ways, the terms *predictive* confidence for JOL ratings and *postdictive* confidence for retrospective confidence seem fit to avoid any semantic issues.

Regarding metamemory, we calculated gamma correlations to assess the relationship between predictive confidence and subsequent recognition memory performance (Nelson, 1984). Briefly, positive gamma correlations show that participants gave higher ratings to “old” items whereas negative gamma correlations show that higher ratings were given to "new” items.

To examine the relationship between recognition accuracy and postdictive confidence, confidence-accuracy characteristic (CAC) curves were computed (Mickes, 2015; Nguyen et al., 2018). Overall, CAC analysis reveals the reliability of eyewitness testimony by assessing the likelihood of a given identification being correct in relation to a particular confidence

---

[2] We also calculated the response criterion (*C*), which refers to how participants establish a threshold in order to respond to the given task. Specifically, this threshold can be stringent (i.e., more “new” responses to face items), assuming positive values or liberal (i.e., more “old” responses to face items, assuming negative values. For this thesis, the evaluation of response criterion results will not be reported in the Results section, in order to enhance fluidity and comprehension of the given section.

judgment (Ngueyn et al., 2017). CAC analysis can tap into the reliability of own- and other-race recognition performance and its relationship with postdictive confidence.

Briefly, CAC curves were obtained by calculating the average proportion of correct identifications (number of hits/number of hits + number of false alarms) separately for each confidence level. Since participants did not use the entire scale range (1 to 9), some levels had insufficient observations. Thus, we aggregated ratings between 1 and 3 to create a “low” confidence level, between 4 and 6 to create a “medium” confidence level, and between ratings of 7 and 9 to create a “high” confidence level. This procedure is commonly done to create more stable estimates when there are too few responses made at each confidence level (see e.g., Nguyen et al. 2018, Wixted, Mickes, Clark, Gronlund, & Roediger, 2015). This procedure does not affect CAC curves calculations, since it is insensitive to scale size (Mickes, 2015).

All statistical analyses reported in this thesis were performed with the open source software Jamovi (Version 1.6.3.0; https://www.jamovi.org).

**Predictive confidence.** A 2 (Race: Own vs. Other) x 2 (Typicality: High vs. Low) repeated measures ANOVA revealed a large main effect of Typicality, $F(1, 55) = 37.60$, $p < .001$, $\eta_p^2 = .41$, indicating that participants gave higher ratings to Low-typicality faces ($M = 4.93$, $SE = .15$) than High-typicality ones ($M = 4.51$, $SE = .15$). Although there was a trend showing higher ratings for Own- ($M = 4.83$, $SE = .16$) than Other-race faces ($M = 4.61$, $SE = .16$), the main effect of Race did not reach the conventional level of significance $p < .05$, $F(1, 55) = 3.68$, $p = .06$, $\eta_p^2 = .06$. Interestingly, we obtained a significant interaction between Race and Typicality, $F(1,55) = 4.73$, $p = .034$, $\eta_p^2 = .08$ (see Figure 4). Simple effects tests (Bonferroni-corrected) showed that the difference between Low-typicality and High-typicality faces is larger for Own- ($t(108.7) = 6.02$, $p < .001$) than for Other-race faces ($t(108.7) = 3.11$, $p = .014$).

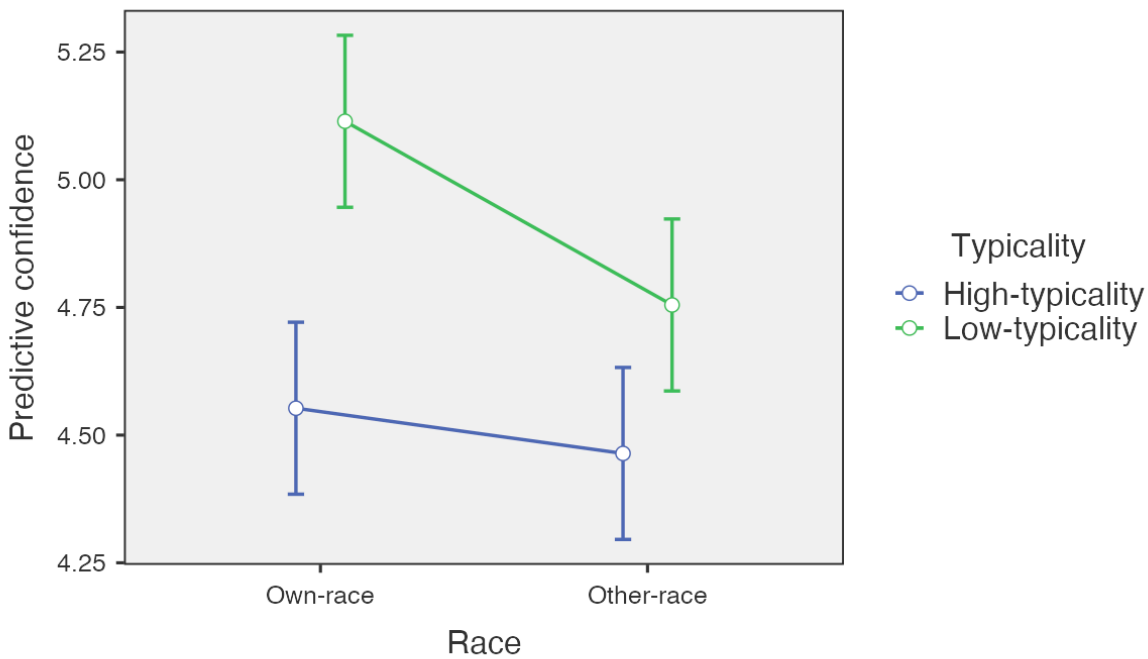


*Figure 4.* Mean predictive confidence plotted as a function of Race and Typicality. Error bars represent +/-1 standard error around the mean.

**Recognition accuracy (*d*').** Mean *d'* scores, HR and FAR for each condition can be found in Table 1. Below we will only present the analysis of *d'* score as this is our measure of interest. A 2 (Race: Own vs Other) x 2 (Typicality: High vs Low) repeated measures ANOVA on *d'* scores yielded a main effect of Race, showing that participants recognized Own-race faces ($M = 1.46$, $SE = .09$) more accurately than Other-race faces ($M = 1.02$, $SE = .09$), $F(1,55) = 29.59$, $p < .001$, $\eta_p^2 = .35$. Thus, these results replicate the ORB. The Typicality main effect was also significant, $F(1,55) = 79.25$, $p < .001$, $\eta_p^2 = .59$. Mirroring the pattern observed for predictive confidence, recognition accuracy was higher for Low-typicality faces ($M = 1.50$, $SE$ =. 08) than High-typicality faces ($M = .98$, $SE = .08$). There was no significant interaction between Race and Typicality ($F(1,55) = 3.30$, $p = .075$, $\eta_p^2 = .06$).

Table 1

*Mean d' scores, HR and FAR as a function of Race and Typicality.*

| | Own-Race | | Other-Race | |
|---|---|---|---|---|
| | High-typicality | Low-typicality | High-typicality | Low-typicality |
| *d'* scores | 0.82 (0.10) | 1.23 (0.10) | 1.14 (0.10) | 1.78 (0.10) |
| Hit rate (HR) | 0.63 (.02) | 0.59 (0.02) | 0.62 (0.02) | 0.68 (0.02) |
| False alarm rate (FAR) | 0.35(.02) | 0.20 (0.02) | 0.25 (0.02) | 0.14 (0.02) |

*Note: Standard Error (SE) around the mean are reported within parentheses*

**Predictive relative accuracy.** For all conditions, gamma correlations were positive and significantly greater than zero (all t-tests > 3.70 and all *p*'s < .001). Gamma correlation were then compared in a 2 (Race: Own vs Other) x 2 (Typicality: High vs Low) repeated measures ANOVA. This analysis revealed no significant main effect of Race, F (1,50) = .02, p = .893, ηp2 = .00, nor of Typicality, F(1,50) = 1.92, p = .172, ηp2 = .04. The interaction between these two factors was also not significant, F (1,50) = .05, p = .819, ηp2 = .00.

**Postdictive confidence.** Ratings were analysed using a 2 (Race: Own vs Other) x 2 (Typicality: High vs Low) repeated measures ANOVA. Results revealed a significant main effect of Race, $F(1, 55) = 18.63$, $p < .001$, $\eta_p^2 = .25$, showing that participants gave higher ratings for Own-race faces (*M* = 6.27, *SE* = .16) than for Other-race faces (*M* = 5.89, *SE* = .16), and a significant main effect of Typicality, $F(1, 55) = 13.53$, $p < .001$, $\eta_p^2 = .20$, indicating that participants attributed higher ratings to Low-typicality faces ($M$ = 6.25, $SE$ = .16) than High-typicality ones ($M$ = 5.91, $SE$ = .16). Unlike predictive confidence ratings, we did not find a significant interaction between Race and Typicality for confidence predictions ($F(1, 55) = 2.51$, $p = .119$, $\eta_p^2 = .04$).

**Postdictive CAC analysis.** A 2 (Race: Own vs Other) x 2 (Typicality: High vs Low) x 3 (Confidence level: Low, Medium, High) was performed on the proportion of correct identifications. This analysis revealed a significant main effect of Confidence level, $F(2, 26)$ = 12,27, $p < .001$, $\eta_p^2 = .49$). Post-hoc tests (Bonferroni-corrected) showed that proportion correct was higher at High confidence levels than at Medium confidence ($M_{Diff}$ = .14, $SE$ = .05, $p$ = .013) and Low confidence ($M_{Diff}$ = .22, $SE$ = .05, $p < .001$). There was no difference between Medium and Low levels of confidence ($M_{Diff}$ = .08, $SE$ = .05, $p = .274$).

**Predictive and Postdictive confidence relationship.** We also analysed the relationship between JOLs and confidence. Specifically, we found that across the four experimental within-participant conditions, both predictive and postdictive confidence were significantly and positively correlated (all $r$'s > .409 and all $p$'s < .001).

## Discussion

Taken together, the results of Experiment 1 show a number of interesting findings. First predictive judgments suggest that participants are aware of the ORB and can anticipate it, even though the ratings of Own-race faces only marginally differed from the ratings of Other-race faces. In terms of Typicality, participants are more sensitive to Low-typicality than High-typicality faces, in the sense that their estimations for future memory are higher for the former than the latter. Therefore, participants seem to have a (correct) belief that atypical faces are more memorable than typical (i.e., being aware of the effect of typicality). Intriguingly, we also found an interaction between Race and Typicality, showing a greater awareness of the effect of typicality for Own-race rather than Other-race faces. Consequently, this is an indication that, when participants make estimations for their future memory performance, they have an expectation that is congruent with it, which can be translated into suitable metacognition in this situation.

In relation to recognition memory, the same pattern of results emerged as participants are more accurate in recognizing Own- and Low-typicality than Other- and High-typicality faces. No interaction was found which means that there are no significant differences of discrimination accuracy (i.e., sensitivity) of High- and Low- typicality faces of Own- and Other-race faces.

Although relative accuracy regarding predictive confidence is above chance level, participants did not display a metamemorial ORB, because there were no differences between Own- and Other-race metamnemonic accuracy, which is in contrast with Hourihan et al.'s (2012) results. Together with the results for predictive confidence, it is implicated that while participants are aware of the ORB, that does not translate into superior relative accuracy for own-race faces.

Given that these past results are consistent across the present experiments, we will further discuss them within the General Discussion (except for predictive relative accuracy in Experiment 3).

As for postdictive confidence, the results suggest that participants' ORB awareness is also present after performance. Nevertheless, they seem incapable of reducing the ORB. Specifically, ratings were higher for Own- than Other-race faces and for Low-typicality than High-typicality faces. Thus, participants also exhibit a belief regarding the effects of typicality.

The postdictive confidence-accuracy relationship tells us that participants' confidence is predictive of accuracy for own- and other-race faces. Therefore, participants' judgments can be considered reliable even though their recognition accuracy was impaired for other-race faces. In other words, *a posteriori* performance estimation translated in postdictive confidence reveal that when participants' confidence is high, the likelihood of that high confidence matching a correct identification (i.e., judging a face presented in study phase as "old" in the test phase) is greater. This was observed independently of race and typicality, meaning that high postdictive confidence for Other-race and High-typicality faces is as reliable as Own-race and Low-typicality faces.

At the same time, predictive and postdictive confidence were significantly related, meaning that when participants gave higher predictive confidence estimations, they also gave higher postdictive confidence estimations, implying perhaps a similar mechanism or cue that may be behind them.

## Experiment 2

In the second experiment, we introduced some novelties to the identical experimental design that was used in Experiment 1, adding two new dependent variables, Study Time[3] and predictive confidence for Others.

Concerning Study Time, we adapted the design of Tullis et al., (2014) and gave the possibility of control of their own study to the participants. That is, each participant studied the stimuli presented for as long as they deemed necessary. Results regarding Study time, normally, show that participants do not differentiate the time that they spend studying own- and other-race faces (e.g., Rhodes et al., 2013).

[3] Regarding Study Time, since all of the statistical analyses did not reach the conventional level of significance $p < .05$, we will not report them in the Results Section.

In terms predictive confidence for Others, we tried to extend the predictive confidence judgments that participants made for themselves in Experiment 1 to Others. Our objective consists of exploring if participants can differentiate the effects of race and typicality in this type of judgment.

However, this prediction was made without knowledge of the other's environment or study time allocation, contrasting other studies (e.g., Koriat & Ackerman, 2010). At the same time, Koriat et al. (2004) have shown that by drawing out from their own personal estimations, the predictive confidence for Others mimic the latter. Therefore, we expect that participants predictive confidence for Self will be related with predictive confidence for Others, mimicking their self-forecasting.

**Method**

**Participants.** We recruited ninety-eight Caucasian participants (76 females and 22 males; $M_{\text{age}}$ = 23.4; $SD_{\text{age}}$ =7.96) from a large pool of paid participants consisting mostly of current and previous students from different University of Lisbon's faculties. They received a 10€ gift voucher in return for their participation in a 1-hour experimental session composed of 3 unrelated experiments. The present experiment was the first one in all sessions.

**Stimuli.** We used the same stimuli of the previous experiment.

**Procedure.** The procedure was similar to that of the previous experiment, with two modifications in the study phase. Instead of presenting each face for 3 seconds, as in Experiment 1, the faces remained on the screen until participants pressed the ENTER key. Additionally, participants were asked to make not one but two JOL predictions for each face. That is, after pressing the ENTER key, they first had to estimate the likelihood that they would recognize that face in a subsequent phase of the experiment.

Following this judgment, they estimated the likelihood that other person would recognize the same face – predictive confidence for Other – ("Please indicate how likely you think it is that another person will recognize the face you have just seen in a later phase of this experiment"). For this second judgment, we replaced the pronoun "I" in the labels of the 9-point scale with the pronoun "They" (1 -"I am sure THEY will NOT remember this face"; 9 - "I am sure THEY WILL remember this face").

**Results**

**Predictive confidence for Self.** A 2 (Race: Own vs Other) x 2 (Typicality: High vs Low) repeated measures ANOVA revealed a significant main effect of Race, $F(1,97) = 9.77$, $p = .002$, $\eta_p^2 = .09$, showing that participants' ratings were higher for Own- ($M = 4.81$, $SE = .13$) than Other-race faces ($M = 4.54$, $SE = .13$). Consistent with Experiment 1, there was a main effect of Typicality, $F(1,97) = 77,14$, $p < .001$, $\eta_p^2 = .44$, indicating that participants' ratings were higher for Low-typicality ($M = 4.92$, $SE = .13$) than High-typicality faces ($M = 4.43$, $SE = .13$). However, there was no significant Race x Typicality interaction, $F(1,97) = 2.70$, $p = .104$, $\eta_p^2 = .03$).

**Predictive confidence for Others**. A 2 (Race: Own vs Other) x 2 (Typicality: High vs Low) repeated measures ANOVA revealed a significant main effect of Race ($F(1,97) = 19.49$, $p < .001$, $\eta_p^2 = .17$) and Typicality ($F(1,97) = 98.74$, $p < .001$, $\eta_p^2 = .50$), indicating that participants' ratings for others mimic the same pattern as for self-ratings. Specifically, participants gave higher ratings for Own- ($M = 5.09$, $SE = .13$) than Other-race faces ($M = 4.76$, $SE = .13$) and Low-typicality ($M = 5.17$, $SE = .12$) than High-typicality faces ($M = 4.69$, $SE = .12$), respectively. No significant interaction between Race and Typicality was found ($F(1,97) = 1.53$, $p = .220$, $\eta_p^2 = .02$).

**Predictive confidence: Self vs Others.** We compared participants' predictive confidence ratings for Self and Other in a 2 (Race: Own vs Other) x 2 (Typicality: High vs Low) x 2 (Target: Self vs Other) repeated measures ANOVA. A main effect of Target emerged, $F(1,97) = 7.39$, $p = .008$, $\eta_p^2 = .07$, meaning that participants gave higher ratings for Other ($M = 4.67$, $SE = .12$) than Self ($M = 4.93$, $SE = .12$).

Moreover, a Pearson correlation analysis between the two forms of predictive confidence judgments revealed that Self and Other ratings were all positive and significantly correlated across for all of the experimental conditions (all *r'*s > = .718 and all *p*'s < .001).

**Recognition accuracy (*d*').** Mean *d'* scores, HR and FAR can be found in Table 2. As in Experiment 1, we will only report the analysis for *d'* scores. Results in Experiment 2 of participants' sensitivity were analyzed through a 2 (Race: Own vs Other) x 2 (Typicality: High vs Low) repeated measures ANOVA. A main effect of Race ($F(1,97) = 33.31$, $p < .001$, $\eta_p^2 = .25$) showed that participants' recognition accuracy was superior for own-race ($M = 1.58$, $SE = .06$) than other-race faces ($M = 1.28$, $SE = .06$), replicating the ORB. The main effect of Typicality was again significant, $F(1,97) = 66.24$, $p < .001$, $\eta_p^2 = .41$, in that Low-typicality faces ($M = 1.66$, $SE = .06$) were more accurately recognized than High-typicality faces ($M = 1.19$, $SE = .06$). No significant interaction between Race and Typicality was found ($F(1,97) = .234$, $p = .630$, $\eta_p^2 = .00$).

Table 2

*Mean d' scores, HR and FAR as a function of Race and Typicality.*

| | Own-race | | Other-race | |
|---|---|---|---|---|
| | High-typicality | Low-typicality | High-typicality | Low-typicality |
| *d'* scores | 1.33 (0.07) | 1.83 (0.07) | 1.05 (0.07) | 1.50 (0.07) |
| Hit rate (HR) | 0.69 (0.02) | 0.68 (0.02) | 0.66 (.02) | 0.64 (0.02) |
| False alarm rate (FAR) | 0.25 (0.02) | 0.13 (0.02) | 0.30(.02) | 0.17 (0.02) |

*Note: Standard Error (SE) around the mean are reported within parentheses*

**Predictive relative accuracy.** Gamma correlations were once again computed in order to assess participants predictive accuracy. One-sample *t*-tests showed that gammas were significantly different from zero (all *t*-tests > 4.78 and all *p*'s < .001). A 2 (Race: Own vs Other) x 2 (Typicality: High vs Low) repeated measures ANOVA showed no significant main effect of Race ($F(1,87) = .21$, $p = .650$, $\eta_p^2 = .00$) or main effect of Typicality ($F(1,87) = 2.60$, $p = .111$, $\eta_p^2 = .03$), nor the Race x Typicality interaction ($F(1,87) = 1.99$, $p = .162$, $\eta_p^2 = .02$).

**Postdictive confidence.** As in Experiment 1, ratings were analysed through a 2 (Race: Own vs Other) x 2 (Typicality: High vs Low) repeated measures ANOVA. The same pattern of results emerged with a significant main effect of Race, $F(1, 97) = 23.38$, $p < .001$, $\eta_p^2 = .19$, meaning that participants gave higher ratings to Own ($M = 6.62$, $SE = .12$) than Other-race faces ($M = 6.30$, $SE = .12$) and a significant main effect of Typicality, $F(1, 97) = 27.55$, $p < .001$, $\eta_p^2 = .22$, indicating that participants gave higher ratings to Low-typicality ($M = 6.64$, *SE*

= .12) than High-typicality faces (*M* = 6.28, *SE* = .12). No significant Race x Typicality interaction was observed ($F(1, 97) = 2.30$, $p = .133$, $\eta_p^2 = .02$).

**Postdictive CAC analysis.** A 2 (Race: Own vs Other) x 2 (Typicality: High vs Low) x 3 (Confidence level: Low, Medium, High) was performed on the proportion of correct identifications. Results revealed a significant main effect of Confidence level ($F(2,36) = 50.82$, $p < .001$, $\eta_p^2 = .74$). Post-hoc tests (Bonferroni-corrected) showed that the proportion correct was higher at High confidence levels than at Medium confidence ($M_{Diff} = .20$, *SE* = .03, $p < .001$) and Low confidence ($M_{Diff} = .38$, *SE* = .03, $p < .001$). There was also a difference between Medium and Low levels of confidence ($M_{Diff} = .18$, *SE* = .03, $p < .001$).

Moreover, results showed a significant main effect of Typicality ($F(1,18) = 6.59$, $p = .019$, $\eta_p^2 = .27$), meaning that the proportion of correct identifications was higher for Low-typicality (*M* = 0.75, *SE* = .03) than High-typicality faces (*M*= .65, *SE* = .03).

**Predictive and Postdictive confidence relationship.** We found that across the four experimental within-participant conditions, both predictive and postdictive confidence were significantly and positively correlated (all *r*s > .418 and all *p*'s < .001).

**Discussion**

The results of Experiment 2 displayed a series of novel findings. With regard to predictive confidence for Others, participants gave higher ratings for Own- than Other-race faces and for Low-typicality than High-typicality faces, mimicking the pattern of predictive confidence for Self. Evidently, individuals project their own predictive judgments when making predictive other judgments as predictive confidence for Self and Others was significantly correlated.

Moreover, participants' ratings were higher for Other than Self. Therefore, the addition of a different type of metacognitive judgment showed that these predictions are inflated towards the other, meaning that individuals have an expectation that the other person will be better than them in terms of future memory performance, regardless of race or typicality.

In terms of postdictive confidence ratings, the results mirrored those obtained in Experiment 1. For the relationship concerning postdictive confidence and accuracy, the differences between Experiment 1 and 2 are rooted in an increasing reliability of own- and other-race faces recognition rendered as confidence increases as well. Specifically, the proportion of correct identifications was significantly different across confidence levels, meaning that while confidence went up so did the proportion of correct identifications. The same pattern for the relationship between predictive and postdictive confidence was found.

## Experiment 3

In the second experiment, we introduce an additional predictive confidence judgment aimed at Others. Although participants seemed to project their estimations for others, we actually do not know who the Other person in participants mind was. Therefore, we employed a modification on who the other person was when making predictive confidence judgments for Others. For some participants, in the Other condition, the other person was Caucasian and for others was Black.

Specifically, this manipulation serves the objective of trying to direct the participants to think about the other person as an expert when their race matches the face that's being presented. Therefore, Caucasian's (or Blacks) are more likely experienced with their own-race faces than other-race, and since greater contact with race plays a role in terms of expertise (Hancock & Rhodes, 2008), it may trigger an assumption of expertise of the other person.

Additionally, we tried to understand if participants' awareness of race and typicality effects transfers to the Other when the latter is Black or Caucasian.

## Method

**Participants.** In this experiment, 179 participants (129 females, 48 males and 1 Non-binary; $M_{age}$ = 27.1, $SD_{age}$ = 8.72;) took part in this experiment. Participants were recruited from the same large pool of participants that we used for Experiment 2, however we limited the participation to individuals who had not participated in Experiment 2. They received a 10€ gift voucher in exchange for the completion of 3 unrelated experiments.

**Stimuli.** The face stimuli were the same 60 Caucasian faces (30 low and 30 highly prototypical) and 60 Black faces (30 low and 30 highly prototypical) of the previous experiments.

**Procedure.** The procedure mirrored that of Experiment 2, with one important modification. In the predictive confidence for others, we manipulated the other person's race by telling participants in one condition that the other person was Caucasian ("Please indicate how likely you think it is that another Caucasian person will recognize the face you have just seen in a later phase of this experiment") and participants in the other condition that the other person was Black ("Please indicate how likely you think it is that another Black person will recognize the face you have just seen in a later phase of this experiment"). Another modification was that faces were presented for 3 seconds in the study phase, as in our first experiment.

**Results**

The data from 6 non-Caucasian participants were not include in the analyses reported below.

**Predictive confidence for Self.** Replicating Experiment 2, a 2 (Race: Black vs Caucasian) x 2 (Typicality: High vs Low) repeated measures ANOVA showed a significant main effect of Race, $F(1,172) = 13.80$, $p < .001$, $\eta_p^2 = .074$; participants' ratings were higher for Caucasian ($M = 5.28$, $SE = 0.11$) than Black faces ($M = 5.05$, $SE = 0.11$). Results also revealed of results the Typicality main effect found in previous the experiments, $F(1,172) = 106.50$, $p < .001$, $\eta_p^2 = .38$, indicating that participants' ratings were higher for Low-typicality ($M = 5.38$, $SE = .11$) than High-typicality faces ($M = 4.94$, $SE = .11$). The interaction between Race and Typicality was significant, $F(1,172) = 21.80$, $p < .001$, $\eta_p^2 = .11$). Post-hoc (Bonferroni-corrected) pairwise comparisons showed that the difference between Low-typicality and High-typicality faces is larger for Caucasian ($M_{Diff} = .63$, $SE = .06$); ($t(334) = 10.91$, $p < .001$) than for Black faces ($M_{Diff} = .28$, $SE = .06$); ($t(334) = 4.90$, $p <.001$).

**Predictive confidence for Others**. A 2 (Race: Black vs Caucasian) x 2 (Typicality: High vs Low) X 2 (Other: Caucasian vs. Black) mixed model ANOVA with Other manipulated as a between-participants factor revealed the following significant effects. We found a significant main effect of Race, $F(1,171) = 9.08$, $p = .003$, $\eta_p^2 = .05$, and main effect of Typicality, $F(1,171) = 74.63$, $p < .001$, $\eta_p^2 = .30$; Higher ratings were attributed to Black faces ($M = 5.49$, $SE = .10$) than Caucasian faces ($M = 5.27$, $SE = 5.27$), and Low-typicality faces ($M = 5.54$, $SE = .09$) had higher ratings than High-typicality ($M = 5.23$, $SE = .09$). More importantly, there was a significant interaction between Race and Other, $F(1,171) = 51.62$, $p < .001$, $\eta_p^2 = .23$. Post-hoc (Bonferroni-corrected) comparisons showed that when the Other was

Caucasian, participants gave higher ratings for Caucasian than Black faces ($t(171) = 2.98$, $p = .020$); however, when the Other was Black, participants rated Black faces higher than Caucasian faces ($t(171) = 7.15$, $p < .001$) (See Figure 6).

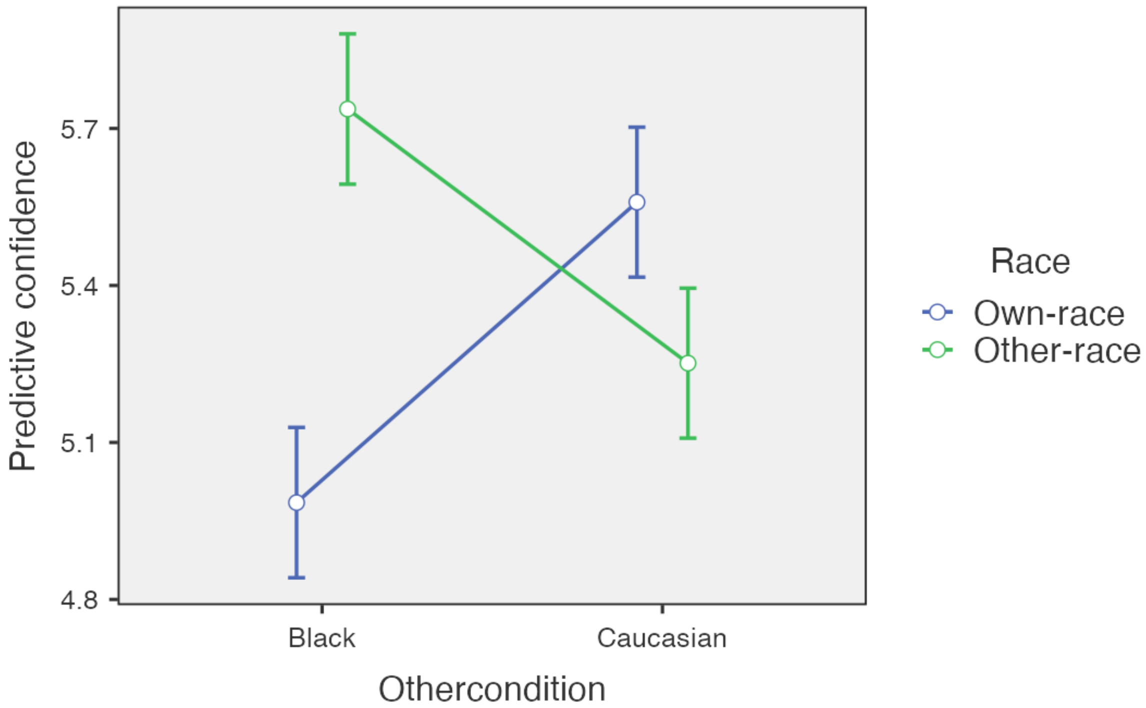


*Figure 6.* Mean predictive confidence as a function of Race and Other condition. Error bars represent +/-1 standard error around the mean.

We also obtained a three-way interaction between the Race, Typicality, and Other, $F(1,171) = 12.88$, $p < .001$, $\eta_p^2 = .07$ (see Figure 7). We ran separate repeated measures ANOVAS for each Other condition. For the Caucasian condition, the interaction between Race and Typicality was significant, $F(1,84) = 6.46$, $p = .013$, $\eta_p^2 = .07$. Post-hoc (Bonferroni-corrected) comparisons showed that the difference between Low-typicality and High-typicality is larger for Caucasian faces ($M_{Diff} = .50$, $SE = .08$) ($t(173) = 6.47$, $p < .001$) than Black faces ($M_{Diff} = .23$, $SE = .08$) ($t(173) = 2.98$, $p = .020$). For the Black condition, the interaction between Race and Typicality was also significant $F(1,84) = 75.00$, $p < .001$, $\eta_p^2 = .47$. Post-hoc

(Bonferroni-corrected) comparisons showed that, for Black faces, participants' gave similar ratings for High- and Low-typicality faces ($M_{Diff}$ = .12, *SE* = .06) (*t*(168) = 1.92, *p* < .343); conversely, for Caucasian faces the participants gave higher ratings for Low-typicality than High-typicality faces ($M_{Diff}$ = .63, *SE* = .06) (*t*(168) = 10.18 *p* < .001);

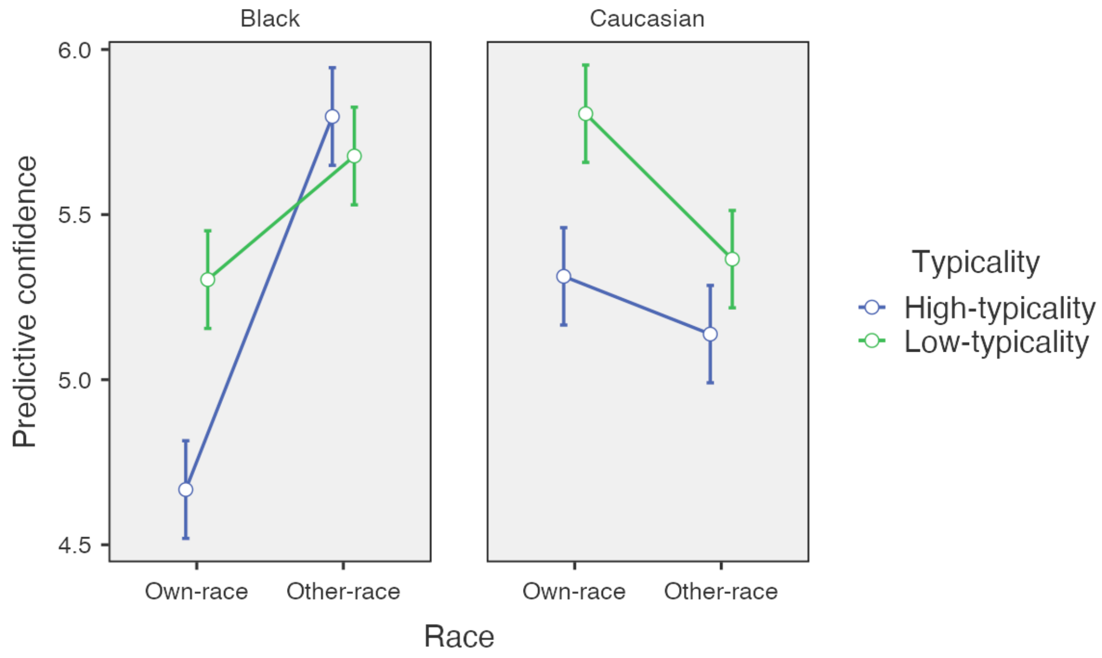


*Figure 7.* Mean predictive confidence as a function of Race, Typicality and Other condition. Error bars represent +/-1 standard error around the mean.

**Recognition accuracy (*d*').** Mean *d'* scores, HR and FAR for each condition can be found in Table 3. Below only the analysis of *d'* score will be reported, since this is our measure of interest. A 2 (Race: Black vs Caucasian) x 2 (Typicality: High vs Low) repeated measures ANOVA yielded the expected main effect of Race, *F*(1,171) = 72.62, *p* < .001, $\eta_p^2$ = .30), with more accurate recognition for Caucasian (own-race) faces (*M* = 1.85, *SE* = .04) than Black (other-race) faces (*M* = 1.52, *SE* = .04). Results also revealed a strong significant main effect of Typicality, *F*(1,171) = 128.41, *p* < .001, $\eta_p^2$ = .43, indicating that Low-typicality faces (*M* = 1.93, *SE* = .04) were more accurately recognized than High-typicality faces (*M* = 1.45, *SE* =

.04). Again, there was no significant interaction between Race and Typicality, $F(1,171) = 1.41$, $p = .236$, $\eta_p^2 = .01$.

Table 3

*Mean d' scores, HR and FAR as a function of Race and Typicality.*

| | Own-race | | Other-race | |
|---|---|---|---|---|
| | High-typicality | Low-typicality | High-typicality | Low-typicality |
| *d'* scores | 1.59 (0.05) | 2.12 (0.05) | 1.31 (0.05) | 1.74 (0.05) |
| Hit rate (HR) | 0.69 (0.01) | 0.72 (0.01) | 0.68 (.01) | 0.69 (0.01) |
| False alarm rate (FAR) | 0.19 (0.01) | 0.10 (0.01) | 0.27(.01) | 0.16 (0.01) |

*Note: Standard Error (SE) around the mean are reported within parentheses*

**Predictive relative accuracy.** Gamma correlations revealed a completely different pattern when compared with Experiment 1 and 2. A 2 (Race: Black vs Caucasian) x 2 (Typicality: High vs Low) repeated measures ANOVA was performed, showing a main effect of Race ($F(1,141) = 5.78$, $p = .021$, $\eta_p^2 = .04$) and a main effect of Typicality ($F(1,141) = 16.05$, $p < .001$, $\eta_p^2 = .10$). Therefore, participants showed superior relative accuracy for Caucasian (own-race) ($M_{gamma}$= .32, $SE$ = .03) than Black (other-race) ($M_{gamma}$= .23, $SE$ = .03) and Low-typicality ($M_{gamma}$= .35, $SE$ = .03) than High-typicality faces ($M_{gamma}$= .20, $SE$ = .03), respectively. No significant Face x Typicality interaction was found, $F(1,141) = 1.99$, $p = .162$, $\eta_p^2 = .02$. Thus, the metamemorial ORB was observed.

**Postdictive confidence.** Similarly, to Experiment 1 and 2, ratings were analysed through a 2 (Race: Black vs Caucasian) x 2 (Typicality: High vs Low) repeated measures ANOVA. The same pattern of results of the previous experiments emerged, with a significant main effect of Race, $F(1, 172) = 64.16$, $p < .001$, $\eta_p^2 = .27$, meaning that participants made higher ratings for Caucasian (own-race) ($M = 6.83$, $SE = .09$) than Black (other-race) faces ($M = 6.45$, $SE = .09$). Results also revealed a significant main effect of Typicality, $F(1, 172) = 110.28$, $p < .001$, $\eta_p^2 = .39$, indicating that participants gave higher ratings to Low-typicality ($M = 6.90$, $SE = .09$) than High-typicality faces ($M = 6.39$, $SE = .09$). No Race x Typicality interaction ($F(1, 172) = 2.24$, $p = .136$, $\eta_p^2 = .01$) was observed.

**Postdictive CAC analysis.** A 2 (Race: Black vs Caucasian) x 2 (Typicality: High vs Low) x 3 (Confidence level: Low, Medium, High) was conducted on the proportion of correct identifications, showing a significant main effect of Confidence level ($F(2,58) = 34,49$, $p <.001$, $\eta_p^2 = 0.54$). Post-hoc tests (Bonferroni-corrected) depicted that the proportion correct was higher at High confidence levels than at Medium confidence ($M_{Diff} = .18$, $SE = .04$, $p < .001$) and Low confidence ($M_{Diff} = .34$, $SE = .04$, $p < .001$). There was also a difference between Medium and Low levels of confidence ($M_{Diff} = .16$, $SE = .04$, $p < .001$).

Moreover, results showed a significant main effect of Typicality ($F(1,29) = 7.77$, $p = .009$, $\eta_p^2 = .21$), meaning that the proportion of correct identifications was higher for Low-typicality ($M = 0.78$, $SE = .02$) than High-typicality faces ($M = .71$, $SE = .02$).

**Predictive and Postdictive confidence relationship.** As in Experiment 1 and 2 in the four experimental within-participant conditions, both JOLs and confidence were significantly and positively correlated (all $r$'s > .323 and all $p$'s < .001).

**Discussion**

The results from Experiment 3 were consistent with the other experiments, although, conveying us a key finding regarding predictive confidence for others. The manipulation implemented towards the Other (i.e., Black or Caucasian) yielded an inverted pattern for predictive confidence when compared to Experiment 2. Overall, participants gave higher ratings to Other rather than Own-race faces, meaning that they did not project their own estimations. This finding is qualified in terms of the relation between Race and the Other condition, in the sense that participants do not consider the Other person as more biased than them. Participants estimated that when the Other is Caucasian, he will be better for Caucasian faces (Own-race) whereas when the Other is Black, he will be better for Black faces (in this case, Own-race). Possibly, participants' estimations reflect a well-adjusted naïve theory of memory.

At the same time, participants seem that they do not differentiate the effects of typicality for the Other. Particularly, participants believe that when the Other is Black his/her awareness of effect of typicality within race is impaired (i.e., subsequent memory would be equal for High-typicality and Low-typicality Black faces), meaning that regardless of being atypical or typical, Other's future memory would be equally accurate for the latter. However, the results for when the Other was Caucasian mimic those of predictive confidence for Self across experiments.

Additionally, in contrast to the previous experiments, results regarding predictive relative accuracy rendered a different pattern. Specifically, participants' predictive confidence was found to be more accurate for both Own-race and Low-typicality faces rather than Other-race and High-typicality faces. This is consistent with Hourihan et al. (2012) result, meaning that participants showed a metamemorial ORB.

The results regarding postdictive confidence, postdictive confidence-accuracy relationship and predictive and postdictive confidence relationship were similar to those obtained before, hence being discussed in the following section.

## General discussion

The present thesis's main goal was to investigate the relationship between metamemory judgments and actual memory performance, particularly exploring participants' awareness of their face recognition memory performance before and after the latter while manipulating race and typicality. Concomitantly, we tried to understand if they are accurate when making predictive and postdictive estimations for themselves and another person, and whether the estimations for other person vary as a function of their race (i.e., Caucasian or Black).

Thus, across three experiments, we systematically asked participants' predictive confidence regarding their future recognition memory. Results showed that participants' JOLs were higher for own- than other-race faces and for atypical than typical faces. Globally, JOLs' difference between atypical and typical faces was higher for own-race than other-race faces.

Concerning relative predictive accuracy, we did not find in Experiments 1 and 2 an indication that participants' predictive judgments were accurate regarding race and typicality; still, participants predicted their future memory performance above chance level. In contrast, in Experiment 3, the results showed that participants accurately predicted their memory superiority for own-race (vs. other-race) and for atypical (vs. typical) faces. Nevertheless, our global findings suggest that participants are not accurate when predicting their recognition memory.

Overall, recognition accuracy results showed, as expected, the replication of the ORB. Specifically, in all three experiments, participants were better at recognizing own-race than other-race faces. Likewise, the pattern for typicality observed was superior recognition of low-typicality than high-typicality faces.

In Experiment 2, we requested participants to predict future recognition memory for others as they had done for themselves. Results mimicked those of predictive confidence for Self, as higher ratings were given to own- than other-race faces and to low than high-typicality faces. Interestingly, we found a significant correlation between Self and Other

In Experiment 3, we found that participants' ratings were higher for Caucasian faces when the Other was Caucasian. The opposite pattern emerged when the Other was Black (i.e., higher JOLs for Black than Caucasian faces). Importantly, we found that participants did not differentiate low-typicality and high-typicality faces in the Black condition, whereas they did so in the Caucasian condition.

Finally, the results regarding postdictive confidence showed that, in general, participants' postdictive confidence was reliable in the sense that higher accuracy was related to higher confidence judgments. Additionally, we found that both predictive confidence and postdictive confidence were significantly correlated in all experiments.

The findings regarding predictive confidence are in line with past research that displays this metacognitive awareness of the ORB (e.g., De Lozier, 2015), showing that participants are aware *a priori* of the ORB and can anticipate it as well as the effects of typicality. Conversely, other research (Rhodes et al., 2013; Experiment 2) has found that people are unaware of the ORB, nevertheless showing a trend for numerically higher JOLs for own- than other-race faces.

The metacognitive awareness of the ORB seems to rest on people's prior beliefs of this bias. Particularly sustained in the cue-utilization approach (Koriat, 1997), participants may have a prior belief that own-race faces are easier to remember than other-race faces and, therefore, translating this belief in greater confidence for own-race faces. The same could be argued in relation to typicality. Rendered in a belief that atypical faces are more memorable than typical ones, participants seemed to be more aware of the differences of typicality for own-race faces.

At the same time, encoding fluency has been pointed out as an important cue that affects JOLs (Mueller et al., 2014, Yang et al., 2018). These findings can be accommodated in terms of encoding fluency as well, in the sense that other-race faces tend to be superficially encoded, resulting in poor representations of those faces in memory. In contrast, the facility of encoding of own-race faces suggests a basis for the predictive confidence pattern in relation to those faces. In the same direction, the typicality findings seem to rest on a superior encoding of atypical (i.e., distinctive) than typical, informing participants that this ease (i.e., fluency) can be used to predict their future performance.

Intriguingly, this "fined-tuned" awareness is present overall, nevertheless participants' were more aware in capturing differences in typicality for own-race faces is a contrasting finding with Valentine and Endo's (1992) predictions where it is assumed that the distribution of atypical and typical faces in terms of exemplars is similar in own and other-race faces. However, our findings show that people believe that they will remember more atypical than typical faces for faces of their own race. Therefore, this awareness may correspond to a broader distribution of exemplars concerning atypical and typical own-race faces and a clustering for atypical and typical for other-race faces in memory.

Taken together, these findings seem consistent with a dual-basis view of JOLs that considers the interplay of beliefs and fluency (Undorf, Zimdhal, & Bernstein, 2017). Importantly, we assert that a more general "mechanism" can be behind these findings operating in conjunction with beliefs and fluency. Schwarz (2004) points out that people have naive theories about how memory functions operate with prior beliefs and fluency. Supported in Tversky and Kahneman's (1973) availability heuristic, people tend to infer greater exemplar' frequency from the easiness of trying to remember them when making predictions.

Applying this logic to our findings, by having more access to low-typicality exemplars' since they are easier to remember (i.e., more distinctive and more salient) and coupled with the superior encoding of own-race faces, people become aware of the differences regarding typicality. In terms of recognition accuracy, our findings present a connection to this good naïve theory about the functioning of memory.

Firstly, the replication of the ORB is consistent with a myriad of experiments (e.g., Hourihan et al., 2012; Rhodes et al., 2013; Tullis et al., 2014). Additionally, our results fall in line with our expectations, and with Valentine and Endo's (1992) work, where people were more accurate for own-race and low-typicality faces than other-race and high-typicality faces.

Even though this thesis's scope does not comprise a full discussion and integration of the ORB theoretical accounts, our findings appear to be in contrast with the MDS framework (Valentine, 1991, 2001; Valentine & Endo, 1992). Specifically, the difference between low-typicality and high-typicality recognition accuracy was numerically larger for own-race than for other-race faces, although not finding an interaction between race and typicality across experiments. Therefore, our findings showed a trend that mimics the pattern obtained for predictive confidence, suggesting greater sensitivity of participants for typicality effects in own-race faces.

An explanation that can accommodate the findings regarding recognition memory and predictive confidence corresponds to a social-cognitive view of in-group heterogeneity and out-group homogeneity (Judd & Park, 1988; see also categorization-individuation model, Hugenberg, Young, Bernstein, & Sacco, 2010). Particularly, it appears that this awareness and sensitivity of people towards differences in typicality within own-race faces is deep-rooted on the in-group's individuation. Consequently, insensitivity to similarities of typicality within other-race faces seems to stem from the out-group's categorization.

While people's predictive confidence shows an awareness of the ORB and typicality that rests on an adequate naïve theory of memory, they are still inaccurate. In general, our findings are at odds with Hourihan’s (2012) work (except for Experiment 3) that previously showed a metamemorial ORB, nevertheless being in line other research (Chen & Zhu, 2019; Hourihan, Fraundorf, & Benjamin, 2013; Rhodes et al., 2013). One hypothesis that can explain these incongruent findings may lay on the differential measures adopted by the different authors. For example, Hourihan et al. (2012) used the $d_a$ measure (Benjamin & Díaz, 2008) to analyze the relative accuracy of JOLs and found that relative accuracy was superior for JOLs regarding own-race faces than other-race faces. However, using the same measure, Hourihan et al. (2013) did not find the same results.

In our view, the key finding of this thesis concern participants' predictions towards others. Firstly, as previously demonstrated in the literature (Epley et al., 2004; Koriat et al., 2004), our results showed that people judge others grounded on their own. Since results of predictive confidence for self and others are positively and significantly correlated, it seems plausible that people project their naïve theories for others.

More importantly, the direct comparison between these judgments exhibited an interesting pattern because participants gave higher JOL ratings for others than self. This is an interesting but contrasting finding since it has been documented that people tend to enhance themselves and their performance in relation to others (Kruger & Dunning, 1999; Snyder, Stephan, & Rosenfeld, 1976). In contrast, Kruger (1999) demonstrated that people tend to rate themselves below-average when they estimate their ability as being reduced (i.e., below-average effect). At the same time, Palermo et al. (2017) showed that people are relatively aware of their face recognition abilities. Our findings collectively indicate that people might acknowledge that their face recognition abilities are poor and, thus, estimate superior recognition memory performance for others. This is, in part, consistent with work in other contexts (*e.g., general knowledge questions*, Johansson & Allwood, 2007), but it is the first demonstration of these differences between self and other predictions for the ORB and typicality effects.

The manipulation introduced in Experiment 3 brought another layer to the discussion mentioned above, in the sense that according to each Other condition (Caucasian vs. Black), participants adjusted their predictions. Specifically, our findings pinpoint that people do not think of others as being more biased than them. This is clear from the interaction between Race and Other condition, where the inverted pattern obtained tells us that participants gave higher ratings for others when the race in the Other condition was congruent with the facts presented.

Overall, our findings of predictive confidence for self and others show a consistent pattern. People appear to apply their beliefs and their experience of encoding fluency with Own- and Other-race faces and atypical and typical faces to these judgments. A mechanism that could behind these findings do not seem to stem from a self-enhancement necessity, since participants rated them as “below-average” in a Krugerian way (Kruger, 1999), in comparison to others. Nor can it be attributed to a projection of self-ratings to other-ratings, since we observed an inverted pattern between Experiments 2 and 3. Therefore, a fit candidate might be that people have a well-adjusted naïve theory of how the memory functions in relation to the ORB and typicality.

Nevertheless, people predicted that when the other was Caucasian, the other person would capture the differences in terms of typicality in own-race faces (i.e., Caucasian faces), whereas the opposite was predicted when the other was Black. This might indicate that this metacognitive awareness translated in a well-adjusted naïve theory cannot untangle these fine-tuned distinctions when the other is Black but only when the other is Caucasian and congruent with the race of the participant.

One possibility might be related to the homogenization of the out-group (Judd & Park, 1988). Not being able to individuate the facial differences of other-race faces may lead to the colloquial notion of "they all look alike," and if the other-race examples are perceived as homogeneous, people may not be aware of those differences. Thus, it is plausible to think that participants apply a belief that can be informally translated into the following: "They all look alike to me, and I have difficulties in distinguishing them, but to them, they all are different and, in this way, they will be better no matter what.". In other words, people may assume that when the Other is Black, the other person will be equally good for low-typicality and high-typicality faces of their own-race (i.e., Black faces), since participants themselves do not capture those differences on other-race faces.

Finally, our findings regarding postdictive confidence showed that these judgments reliably predicted recognition accuracy. Moreover, it showed that the proportion of correct identifications was higher for high confidence judgments. These findings are in line with other research (Ngueyn et al., 2017; Ngueyn et al., 2018) by exposing the non-effect of race in this case. Specifically, postdictive confidence predicted both own-race and other-race faces reliably.

Furthermore, behind the observed relationship across experiments between predictive (i.e., JOLs) and postdictive confidence, typicality may be acting as a cue to make these judgments.

**Limitations and Future directions**

Some caveats must be taken into consideration regarding this thesis, while at the same time, future studies are considered. For example, in experiment 3, we manipulated who the other person was in terms of race (Caucasian vs. Black), although our sample only comprised

White participants. Therefore, it should be of the utmost interest to compose a sample of both White and Black participants (fully crossed-race design) to thoroughly compare the results and understand the naïve theories both groups possess. While it has been noted that the ORB effect is greater for White participants (Meissner & Brigham, 2001), Experiment 4 should be derived from this cross-race design to fully assess the cross-over interaction that we obtained in Experiment 3.

## References

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

Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review*, *12*(6), 1043–1047. https://doi.org/10.3758/BF03206441

Arnold, M. M. (2013). Monitoring and meta-metacognition in the own-race bias. *Acta Psychologica*, *144*(2), 380–389. https://doi.org/10.1016/j.actpsy.2013.07.007

Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. *Handbook of memory and metamemory*, 73-94.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, *18*(8), 706-712. https://doi.org/10.1111/j.1467-9280.2007.01964.x

Blair, I. V., & Judd, C. M. (2010). Afrocentric Facial Features and Stereotyping. In *The Science of Social Vision* (Vol. 15, pp. 306–320). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195333176.003.0019

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 and Social Psychology*, *83*(1), 5–25. https://doi.org/10.1037/0022-3514.83.1.5

Blake, M. (1973). Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *12*(3), 311–319. https://doi.org/10.1016/S0022-5371(73)80075-1

Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-Racial Identification. *Personality and Social Psychology Bulletin*, *15*(1), 19–25. https://doi.org/10.1177/0146167289151002

Brigham, J. C., Meissner, C. A., & Wasserman, A. W. (1999). Applied issues in the construction and expert assessment of photo lineups. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *13*(S1), S73-S92. https://doi.org/10.1002/(SICI)1099-0720(199911)13:1+<S73::AID-ACP631>3.0.CO;2-4

Brigham, J. C., & Ready, D. J. (1985). Own-race bias in lineup construction. *Law and Human Behavior*, *9*(4), 415–424. https://doi.org/10.1007/BF01044480

Bruce, V., & Young, A. W. (2012). *Face perception*. Psychology Press.

Bruce, V., Burton, A. M., & Dench, N. (1994). What's Distinctive About a Distinctive Face? *The Quarterly Journal of Experimental Psychology Section A*, *47*(1), 119–141. https://doi.org/10.1080/14640749408401146

Burton, A. M., & Vokey, J. R. (1998). The Face-Space Typicality Paradox: Understanding the Face - Space Metaphor. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *51*(3), 475–483. https://doi.org/10.1080/713755768

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin and Review*, *7*(1), 26–48. https://doi.org/10.3758/BF03210724

Byatt, G., & Rhodes, G. (2004). Identification of own-race and other-race faces: Implications for the representation of race in face space. *Psychonomic Bulletin and Review*, *11*(4), 735–741. https://doi.org/10.3758/BF03196628

Cohen, M. E., & Carr, W. J. (1975). Facial recognition and the von Restorff effect. *Bulletin of the Psychonomic Society*, *6*(4), 383–384. https://doi.org/10.3758/BF03333209

Chen, J., & Zhu, X. (2019). The Cross-race Effect on Face Recognition and Judgments of Learning. In *Proceedings of the 3rd International Conference on Culture, Education and Economic Development of Modern Society (ICCESE 2019)* (Vol. 310, pp. 672–675). Paris, France: Atlantis Press. https://doi.org/10.2991/iccese-19.2019.147

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 Section A*, *48*(4), 879–894. https://doi.org/10.1080/14640749508401421

Corneille, O., Huart, J., Becquart, E., & Brédart, S. (2004). When Memory Shifts Toward More Typical Category Exemplars: Accentuation Effects in the Recollection of Ethnically Ambiguous Faces. *Journal of Personality and Social Psychology, 86*(2), 236–250. https://doi.org/10.1037/0022-3514.86.2.236

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*(4), 1–17. https://doi.org/10.1167/10.4.16

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*(2), 107–117. https://doi.org/10.1037/0096-3445.115.2.107

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition.* Sage Publications, Inc.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*(4), 374–380. https://doi.org/10.3758/BF03210921

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES - A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods, 42*(1), 351-362. doi:10.3758/BRM.42.1.351.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327–339. https://doi.org/10.1037/0022-3514.87.3.327

Finn, B., & Tauber, S. K. (2015). When Confidence Is Not a Signal of Knowing: How Students' Experiences and Beliefs About Processing Fluency Can Lead to Miscalibrated Confidence. *Educational Psychology Review*, *27*(4), 567–586. https://doi.org/10.1007/s10648-015-9313-7

Fiske, S. T., & Neuberg, S. L. (1990). A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation. *Advances in Experimental Social Psychology*, *23*(C), 1–74. https://doi.org/10.1016/S0065-2601(08)60317-2

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–197. https://doi.org/10.1038/72140

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes. *Vision Research*, *38*(15–16), 2401–2428. https://doi.org/10.1016/S0042-6989(97)00442-2

Goldsmith, M., & Koriat, A. (2007). The Strategic Regulation Of Memory Accuracy And Informativeness. *Psychology of Learning and Motivation - Advances in Research and Theory*, *48*, 1–60. https://doi.org/10.1016/S0079-7421(07)48001-X

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* John Wiley.

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology*, *99*(1), 45–56. https://doi.org/10.1348/000712607X199981

Hargis, M. B., & Castel, A. D. (2019). Knowing What Others Know: Younger and Older Adults' Perspective-Taking and Memory for Medication Information. *Journal of Applied Research in Memory and Cognition*, *8*(4), 481–493. https://doi.org/10.1016/j.jarmac.2019.09.004

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*(4), 208–216. https://doi.org/10.1037/h0022263

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*, 575–604. https://doi.org/10.1146/annurev.psych.53.100901.135109

Horry, R., Cheong, W., & Brewer, N. (2015). The other-race effect in perception and recognition: Insights from the complete composite task. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 508–524. https://doi.org/10.1037/xhp0000042

Hourihan, K. L., Benjamin, A. S., & Liu, X. (2012). A cross-race effect in metamemory: Predictions of face recognition are more accurate for members of our own race. *Journal of Applied Research in Memory and Cognition*, *1*(3), 158–162. https://doi.org/10.1016/j.jarmac.2012.06.004

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2013). Same faces, different labels: Generating the cross-race effect in face memory with social category information. *Memory & Cognition, 41*(7), 1021–1031. https://doi.org/10.3758/s13421-013-0316-7

Hugenberg, K., & Corneille, O. (2009). Holistic processing is tuned for in-group faces. *Cognitive Science*, *33*(6), 1173–1181. https://doi.org/10.1111/j.1551-6709.2009.01048.x

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review, 117*(4), 1168–1187. https://doi.org/10.1037/a0020463

Johansson, M., & Allwood, C. M. (2007). Own-other differences in the realism of some metacognitive judgments. *Scandinavian Journal of Psychology*, *48*(1), 13–21. https://doi.org/10.1111/j.1467-9450.2007.00565.x

Jost, J. T., Kruglanski, A. W., & Nelson, T. O. (1998). Social Metacognition: An Expansionist Review. *Personality and Social Psychology Review*, *2*(2), 137–154. https://doi.org/10.1207/s15327957pspr0202_6

Judd, C. M., & Park, B. (1988). Out-Group Homogeneity: Judgments of Variability at the Individual and Group Levels. *Journal of Personality and Social Psychology*, *54*(5), 778–788. https://doi.org/10.1037/0022-3514.54.5.778

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8), 759–763. https://doi.org/10.1038/77664

Kleider-Offutt, H. M., Bond, A. D., & Hegerty, S. E. A. (2017). Black Stereotypical Features: When a Face Type Can Get You in Trouble. *Current Directions in Psychological Science*, *26*(1), 28–33. https://doi.org/10.1177/0963721416667916

Koriat, A. (1997). Monitoring One's Own Knowledge during Study: A Cue-Utilization Approach to Judgments of Learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (2016). Metacognition: Decision-making processes in self-monitoring and self-regulation. *The Wiley Blackwell handbook of judgment and decision making*, *1*, 356-379.

Koriat, A., & Ackerman, R. (2010). Metacognition and mindreading: Judgments of learning for self and other during self-paced study. *Consciousness and cognition*, *19*(1), 251-264.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*(4), 643–656. https://doi.org/10.1037/0096-3445.133.4.643

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*(3), 490–517. https://doi.org/10.1037//0033-295x.103.3.490

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, *131*(2), 147–162. https://doi.org/10.1037/0096-3445.131.2.147

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, *77*(2), 221–232. https://doi.org/10.1037/0022-3514.77.2.221

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lavrakas, P. J., Buri, J. R., & Mayzner, M. S. (1976). A perspective on the recognition of other-race faces. *Perception & Psychophysics*, *20*(6), 475–481. https://doi.org/10.3758/BF03208285

Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(6), 1364–1382. https://doi.org/10.1037/0278-7393.22.6.1364

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*(4), 559–574. https://doi.org/10.1037/0096-3445.129.4.559

Lewis, M. B., & Hills, P. J. (2018). Perceived race affects configural processing but not holistic processing in the composite-face task. *Frontiers in Psychology*, *9* (AUG), 1–15. https://doi.org/10.3389/fpsyg.2018.01456

Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(3), 212–228. https://doi.org/10.1037/0278-7393.5.3.212

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. doi: 10.3758/s13428-014-0532-5

Ma, D. S., Correll, J., & Wittenbrink, B. (2018). The effects of category and physical features on stereotyping and evaluation. *Journal of Experimental Social Psychology*, *79*(June), 42–50. https://doi.org/10.1016/j.jesp.2018.06.008

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates Publishers.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98*(1), 185–199. https://doi.org/10.1037/0033-2909.98.1.185

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120. https://doi.org/10.1146/annurev.psych.51.1.93

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13*(4), 330–334. https://doi.org/10.1037/h0028434

Malpass, R. S., & Lindsay, R. C. (1999). Measuring lineup fairness. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *13*(S1), S1-S7. https://doi.org/10.1002/(SICI)1099-0720(199911)13:1+<S1::AID-ACP678>3.0.CO;2-9

Mata, A., Ferreira, M. B., & Sherman, S. J. (2013). The Metacognitive advantage of deliberative thinkers: A Dual-process perspective on overconfidence. *Journal of Personality and Social Psychology*, *105*(3), 353–373. https://doi.org/10.1037/a0033640

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, and Law*, *7*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3

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*(5), 545–567. https://doi.org/10.1002/acp.1097

Metcalfe, J., & Dunlosky, J. (2008). Metamemory. In *Learning and Memory: A Comprehensive Reference* (pp. 349–362). Elsevier. https://doi.org/10.1016/B978-012370509-9.00159-5

Michel, C., Corneille, O., & Rossion, B. (2007). Race categorization modulates holistic face encoding. *Cognitive Science*, *31*(5), 911-924. https://doi.org/10.1080/03640210701530805

Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition*, *14*(1), 55–73. https://doi.org/10.1080/13506280500158761

Mickes, L. (2015). Receiver operating characteristic analysis and confidence–accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition*, *4*(2), 93–102. https://doi.org/10.1016/j.jarmac.2015.01.003

Mickes, L., Hwe, V., Wais, P. E., & Wixted, J. T. (2011). Strong Memories Are Hard to Scale. *Journal of Experimental Psychology: General*, *140*(2), 239–257. https://doi.org/10.1037/a0023007

Minear, M. and Park, D. C. (2004). A lifespan database of adult facial stimuli. Behavior Research Methods, Instruments, & Computers, 36(4):630–633.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural Face Processing Develops more Slowly than Featural Face Processing. *Perception*, *31*(5), 553–566. https://doi.org/10.1068/p3339

Mondloch, C. J., Elms, N., Maurer, D., Rhodes, G., Hayward, W. G., Tanakao, J. W., & Zhou, G. (2010). Processes underlying the cross-race effect: An investigation of holistic, featural, and relational processing of own-race versus other-race faces. *Perception*, *39*(8), 1065–1085. https://doi.org/10.1068/p6608

Morris, P. E., & Wickham, L. H. V. (2001). Typicality and face recognition: A critical re-evaluation of the two factor theory. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *54*(3), 863–877. https://doi.org/10.1080/713755992

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, *70*(1), 1–12. https://doi.org/10.1016/j.jml.2013.09.007

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin and Review*, *20*(2), 378–384. https://doi.org/10.3758/s13423-012-0343-6

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Nelson, T. O., & Dunlosky, J. (1991). When People's Judgments of Learning (JOLs) are Extremely Accurate at Predicting Subsequent Recall: The "Delayed-JOL Effect." *Psychological Science*, *2*(4), 267–271. https://doi.org/10.1111/j.1467-9280.1991.tb00147.x

Nelson, T. O., Gerler, D., & Narens, L. (1984). Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General*, *113*(2), 282–300. https://doi.org/10.1037/0096-3445.113.2.282

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of Self-Paced Study Time and the "Labor-in-Vain Effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 676–686. https://doi.org/10.1037/0278-7393.14.4.676

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 26, pp. 125–173). New York, NY: Academic Press.

Ng, W. J., & Lindsay, R. C. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology*, *25*, 217-232. https://doi.org/10.1177/0022022194252004

Nguyen, T. B., Abed, E., & Pezdek, K. (2018). Postdictive confidence (but not predictive confidence) predicts eyewitness memory accuracy. *Cognitive Research: Principles and Implications*, *3*(1). https://doi.org/10.1186/s41235-018-0125-4

Nguyen, T. B., Pezdek, K., & Wixted, J. T. (2017). Evidence for a confidence–accuracy relationship in memory for same- and cross-race faces. *Quarterly Journal of Experimental Psychology*, *70*(12), 2518–2534. https://doi.org/10.1080/17470218.2016.1246578

Nickerson, R. S. (1999). How we know - And sometimes misjudge - What others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*(6), 737–759. https://doi.org/10.1037/0033-2909.125.6.737

Olson, I. R., & Marshuetz, C. (2005). Facial attractiveness is appraised in a glance. *Emotion*, *5*(4), 498–502. https://doi.org/10.1037/1528-3542.5.4.498

Omer, Y., Sapir, R., Hatuka, Y., & Yovel, G. (2019). What Is a Face? Critical Features for Face Detection. *Perception*, *48*(5), 437–446. https://doi.org/10.1177/0301006619838734

Palermo, R., Rossion, B., Rhodes, G., Laguesse, R., Tez, T., Hall, B., … McKone, E. (2017). Do people have insight into their face recognition abilities? *Quarterly Journal of Experimental Psychology*, *70*(2), 218–233. https://doi.org/10.1080/17470218.2016.116105

Palma, A. T. and Garcia-Marques, L. (under revision). Does Repetition Always Makes Perfect? Differential effects of repetition on learning of own-race and other-race faces. *Basic and Applied Social Psychology.*

Palma, T. A., Santos, A. S., & Garcia-Marques, L. (2018). The future is now: the impact of present fluency in judgments about the future. *Memory*, 26(2), 144–153. https://doi.org/10.1080/09658211.2017.1335328

Papesh, M. H., & Goldinger, S. D. (2010). A multidimensional scaling analysis of own- and cross-race face spaces. *Cognition*, *116*(2), 283–288. https://doi.org/10.1016/j.cognition.2010.05.001

Rezlescu, C., Susilo, T., Wilmer, J. B., & Caramazza, A. (2017). The inversion, part-whole, and composite effects reflect distinct perceptual mechanisms with varied relationships to face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(12), 1961–1973. https://doi.org/10.1037/xhp0000400

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology*, *80*(3), 313–331. https://doi.org/10.1111/j.2044-8295.1989.tb02323.x

Rhodes, G., Locke, V., Ewing, L., & Evangelista, E. (2009). Race coding and the other-race effect in face recognition. *Perception*, *38*(2), 232–241. https://doi.org/10.1068/p6110

Rhodes, M. G. (2019). Metacognition. *Teaching of Psychology*, *46*(2), 168–175. https://doi.org/10.1177/0098628319834381

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin*, *138*(1), 146–174. https://doi.org/10.1037/a0025750

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin and Review*, *16*(3), 550–554. https://doi.org/10.3758/PBR.16.3.550

Rhodes, M. G., Sitzman, D. M., & Rowland, C. A. (2013). Monitoring and control of learning own-race and other-race faces. *Applied Cognitive Psychology*, *27*(5), 553–563. https://doi.org/10.1002/acp.2948

Rodin, M. J. (1987). Who Is Memorable to Whom: A Study of Cognitive Disregard. *Social Cognition*, *5*(2), 144–165. https://doi.org/10.1521/soco.1987.5.2.144

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rossion, B. (2018). Humans Are Visual Experts at Unfamiliar Face Recognition. *Trends in Cognitive Sciences*, *22*(6), 471–472. https://doi.org/10.1016/j.tics.2018.03.002

Sarno, J. A., & Alley, T. R. (1997). Attractiveness and the memorability of faces: Only a matter of distinctiveness? *American Journal of Psychology*, *110*(1), 81–92. https://doi.org/10.2307/1423702

Schwartz, B. L., & Metcalfe, J. (2017). Metamemory: An Update of Critical Findings. In *Learning and Memory: A Comprehensive Reference* (Second Edi, Vol. 2, pp. 423–432). Elsevier. https://doi.org/10.1016/B978-0-12-809324-5.21056-0

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, *14*(4), 332–348. https://doi.org/10.1207/s15327663jcp1404_2

Sergent, J. (1986). Methodological constraint in neuropsychological studies of face perception in normals. In Bruyer, R., editor, The Neuropsychology of Face Perception and Facial Expression, pages 91–124. Lawrence Erlbaum Associates, Inc., Hillsdale: NJ.

Sergent, J., Ohta, S., & Macdonald, B. (1992). Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain*, *115*(1), 15-36. https://doi.org/10.1093/brain/115.1.15

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*(2), 146–154. https://doi.org/10.1037/h0087225

Snyder, M. L., Stephan, W. G., & Rosenfield, D. (1976). Egotism and attribution. *Journal of Personality and Social Psychology, 33*(4), 435–441. https://doi.org/10.1037/0022-3514.33.4.435

Sommer, W., Heinz, A., Leuthold, H., Matt, J., & Schweinberger, S. R. (1995). Metamemory, distinctiveness, and event-related potentials in recognition memory for faces. *Memory & Cognition*, *23*(1), 1–11. https://doi.org/10.3758/BF03210552

Sporer, S. L. (2001a). The cross-race effect: Beyond recognition of faces in the laboratory. *Psychology, Public Policy, and Law*, *7*(1), 170–200. https://doi.org/10.1037/1076-8971.7.1.170

Sporer, S. L. (2001b). Recognizing Faces of Other Ethnic Groups: An Integration of Theories. *Psychology, Public Policy, and Law*, *7*(1), 36–97. https://doi.org/10.1037/1076-8971.7.1.36

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers, 31*(1), 137–149. https://doi.org/10.3758/BF03207704

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., and Heagins, T. B. (2016). The mr2: A multiracial, mega-resolution database of facial stimuli. Behavior Research Methods, 48(3):1197–1204.

Sunday, M. A., & Gauthier, I. (2018). Face Expertise for Unfamiliar Faces: A Commentary on Young and Burton's "Are We Face Experts?" *Journal of Expertise*, *1*(1), 1–7. Retrieved from https://www.journalofexpertise.org

Susa, K. J., Meissner, C. A., & Heer, H. De. (2010). Modeling the role of social-cognitive processes in the recognition of own-and other-race faces. *Social Cognition*, *28*(4), 523–537. https://doi.org/10.1521/soco.2010.28.4.523

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, Calif: Brooks/Cole Pub. Co.

Tanaka, J. W., & Farah, M. J. (1993). Parts and Wholes in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*(2), 225–245. https://doi.org/10.1080/14640749308401045

Tanaka, J. W., & Gauthier, I. (1997). Expertise in object and face recognition. *Psychology of learning and motivation*, *36*, 83-125. https://doi.org/10.1016/S0079-7421(08)60282-0

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition*, *93*(1). https://doi.org/10.1016/j.cognition.2003.09.011

Tauber, S. K., Witherby, A. E., & Dunlosky, J. (2019). Beliefs about memory decline in aging do not impact judgments of learning (JOLs): A challenge for belief-based explanations of JOLs. *Memory and Cognition*, *47*(6), 1102–1119. https://doi.org/10.3758/s13421-019-00919-3

The jamovi project (2020). *jamovi*. (version 1.6.3.0)[Computer Software]. Retrieved from https://www.jamovi.org.

Tottenham, N., Borscheid, A., Ellertsen, K., Marcus, D. J., & Nelson, C. A. (2002). Categorization of facial expressions in children and adults: Establishing a larger stimulus set. Poster presented at the annual meeting of Cognitive Neuroscience Society, San Francisco, CA.

Tullis, J. G. (2018). Predicting others' knowledge: Knowledge estimation as cue utilization. *Memory and Cognition*, *46*(8), 1360–1375. https://doi.org/10.3758/s13421-018-0842-4

Tullis, J. G., Benjamin, A. S., & Liu, X. (2014). Self-pacing study of faces of different races: Metacognitive control over study does not eliminate the cross-race recognition effect. *Memory & cognition*, *42*(6), 863-875.

Tullis, J. G., & Fraundorf, S. H. (2017). Predicting others' memory performance: The accuracy and bases of social metacognition. *Journal of Memory and Language*, *95*, 124–137. https://doi.org/10.1016/j.jml.2017.03.003

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology,* *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language*, *92*, 293–304. https://doi.org/10.1016/j.jml.2016.07.003

Valentine, T. (1991). A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *43*(2), 161–204. https://doi.org/10.1080/14640749108400966

Valentine, T. (2001). Face-space models of face recognition. *Computational, Geometric, and Process Perspectives on Facial Cognition: Contexts and Challenges.*, (July), 83–113. https://doi.org/10.1006/jmps.1997.1193

Valentine, T., & Bruce, V. (1986a). Recognizing familiar faces: the role of distinctiveness and familiarity. *Canadian Journal of Psychology*, *40*(3), 300–305. https://doi.org/10.1037/h0080101

Valentine, T., & Bruce, V. (1986b). The effects of distinctiveness in recognising and classifying faces. *Perception*, *15*(5), 525–535. https://doi.org/10.1068/p150525

Valentine, T., & Endo, M. (1992). Towards an Exemplar Model of Face Processing: The Effects of Race and Distinctiveness. *The Quarterly Journal of Experimental Psychology Section A*, *44*(4), 671–703. https://doi.org/10.1080/14640749208401305

Vesonder, G. T., & Voss, J. F. (1985). On the ability to predict one's own responses while learning. *Journal of Memory and Language*, *24*(3), 363–376. https://doi.org/10.1016/0749-596X(85)90034-8

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition*, *20*(3), 291–302. https://doi.org/10.3758/BF03199666

Watier, N., & Collin, C. (2012). The effects of distinctiveness on memory and metamemory for face-name associations. *Memory*, *20*(1), 73–88. https://doi.org/10.1080/09658211.2011.637935

Watier, N. N., & Collin, C. A. (2011). Metamemory for faces, names, and common nouns. *Acta Psychologica*, *138*(1), 143–154. https://doi.org/10.1016/j.actpsy.2011.05.017

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

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., … Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(11), 5238–5241. https://doi.org/10.1073/pnas.0913053107

Wilson, J. P., Bernstein, M. J., & Hugenberg, K. (2016). *A synthetic perspective on the own-race bias in eyewitness identification.* In B. H. Bornstein & M. K. Miller (Eds.), *Advances in psychology and law: Vol. 2. Advances in psychology and law* (p.

241–270). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-43083-6_8

Wilson, J. P., Hugenberg, K., & Bernstein, M. J. (2013). The Cross-Race effect and eyewitness identification: How to improve recognition and reduce decision errors in eyewitness situations. *Social Issues and Policy Review*, *7*(1), 83–113. https://doi.org/10.1111/j.1751-2409.2012.01044.x

Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory, 7*(3), 181–190. https://doi.org/10.1037/0278-7393.7.3.181

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. III. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70*(6), 515–526. https://doi.org/10.1037/a0039510

Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica*, *114*(1), 101–114. https://doi.org/10.1016/S0001-6918(03)00052-0

Wright, D. B., & Stroud, J. N. (2002). Age differences in lineup identification accuracy: People are better with their own age. *Law and Human Behavior*, *26*(6), 641–654. https://doi.org/10.1023/A:1020981501383

Yang, C., Huang, T. S. T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language*, *99*, 99–110. https://doi.org/10.1016/j.jml.2017.11.005

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145. https://doi.org/10.1037/h0027474

Yonelinas, A. P. (1994). Receiver-Operating Characteristics in Recognition Memory: Evidence for a Dual-Process Model. *Journal of Experimental Psychology: Learning,* *Memory, and Cognition*, *20*(6), 1341–1354. https://doi.org/10.1037/0278-7393.20.6.1341

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Young, A. W., & Burton, A. M. (2018). What We See in Unfamiliar Faces: A Response to Rossion. *Trends in Cognitive Sciences*, *22*(6), 472–473. https://doi.org/10.1016/j.tics.2018.03.008

Young, A. W., Hellawell, D., & Hay, D. C. (2013). Configurational information in face perception. *Perception*, *42*(11), 1166–1178. https://doi.org/10.1068/p160747n

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and Motivation in Face Recognition: A Critical Review of Theories of the Cross-Race Effect. *Personality and Social Psychology Review*, *16*(2), 116–142. https://doi.org/10.1177/1088868311418987