THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse - Jean Jaurès*
Cotutelle internationale *University of Cape Town, Cape Town, South Africa*

Présentée et soutenue le *31 Janvier 2020* **par :**
TANIA WITTWER

The own-group bias in face processing: The effect of training on recognition performance

JURY

PR. JEAN-YVES BAUDOUIN	Université Lyon 2	Rapporteur
DR. HDR MAJA BECKER	Université Toulouse Jean Jaurès	Membre du Jury
PR. RAY BULL	University of Leicester	Rapporteur
PR. JACQUES PY	Université Toulouse Jean Jaurès	Co-directeur de thèse
PR. MELANIE SAUERLAND	University of Maastricht	Membre du Jury
PR. COLIN TREDOUX	University of Cape Town	Co-directeur de thèse

École doctorale et spécialité :
CLESCO : Psychologie
Unité de Recherche :
CLLE (UMR 5263) & ACSENT
Directeur(s) de Thèse :
Jacques Py et *Colin Tredoux*
Rapporteurs :
Jean-Yves Baudouin et *Ray Bull*

Table of Contents

List of Figures

List of Tables

Acknowledgments

To my supervisors, Professor Colin Tredoux and Professor Jacques Py, for their guidance, their valuable comments and input on my ideas, as well as for the opportunities I had for presenting my work at conferences, writing papers, and collaborating on many other projects.

To the jury of my defence: Professor Jean-Yves Baudouin, Doctor HDR Maja Becker, Professor Ray Bull, and Professor Melanie Sauerland for their interest in my work, the time taken to read this thesis, and their presence at my defence.

To the members of both CNRS CLLE (UT2J) and ACSENT (UCT) Laboratories for their support, meetings, and seminars. To the eyewitness group members over two years for all the interesting meetings and discussions. A special thanks to my research assistants and confederates in Cape Town: Aphiwe, Claudia, Faieeza, Isaac, Jason, Kaolin, Kelly, Micha, Mihkail, Rhiannon, Robyn, Sakina, Suzèl and Tsebo.

To my friends and colleagues of Cape Town, with some special mentions. Katie, Julian, Ruth, Lydia, Tamsyn, Jenny, and Miranda for being so helpful and supportive, for our more or less serious discussions over the year at work, on a hike, or on a getaway. Josh, for our morning coffee-working sessions, walks in the forest or in the mountains, and interesting conversations. You all helped me so much to practice and improve my English. An additional special thanks to Josh and Katie for the proofreading of the present thesis, and their enthusiasm in doing it.

A tou.te.s mes collègues et ami.e.s de Toulouse, pour leur soutien et encouragements, avec certaines mentions spéciales. Pierre-Vincent, pour avoir été d'une si grande aide, même lorsque le temps jouait contre nous. Michel, pour tous ces cafés et conversations matinales, toujours dans la bonne humeur. Anais, pour l'année durant laquelle j'ai été assez chanceuse de travailler avec toi, ainsi que pour tes précieux conseils. Malvina et Sarra, pour vos contributions dans le projet MisIdentification Contact, dans la construction de la base de données photographique, et pour tous nos nombreux échanges. Aurélie, pour la collaboration très constructive que l'on a eu l'opportunité de créer autour de notre projet d'étude concernant les élections présidentielles de 2017. Damien, pour avoir gentillement accepté de prêter son visage pour les illustrations de cette thèse. Jérémy et Laetitia, pour nos discussions constructives, divergeances, rires, pour votre soutien, et surtout pour avoir fait de notre bureau un environnement de travail idéal.

A mes ami.e.s hors du milieu académique, avec quelques mentions spéciales. Luce, pour avoir été et être encore ma plus grande source d'encouragement et ce, depuis notre mémoire de Master 1 en 2015. Bérengère, Gwenn, Mélissa, et Anne-Louise, pour votre soutien infaillible, votre confiance et vos mots rassurants et encourageants en toutes circonstances.

A ma famille. David et Kevin, pour être des frères si géniaux. Papa, Maman, pour m'avoir toujours soutenue et encouragée dans mes choix, et pour m'avoir toujours fait confiance. Vous méritez le plus grand des remerciements. À mes quatre grand-parents, ma tante - Béatrice, et mes cousines - Laurine et Valérie, pour leur soutien et leur fierté, pour avoir toujours cru en moi, et pour tous les moments passés ensemble.

Finally, thanks to you, whoever you are, for giving this thesis a read. I hope you will enjoy it as much as I enjoyed conducting this research, and writing about it.

Abstract

The own-group bias in face recognition (OGB) is the greater facility to distinguish and recognize people from one's own group at the expense of people from other-groups. The existence of the OGB has been studied for many years, however, very little research focuses on finding a way to decrease or eliminate it, through training. Reporting five studies involving memory or matching tasks, the aim of the present thesis was to develop and to explore to what extent training can decrease or remove the OGB. French White participants or South African White, Black and Coloured participants took part in different studies, using Black and White faces as stimuli. In each study, White participants from both countries presented the expected OGB prior to any intervention. However, the presence of the OGB in South African Black participants was detected only in one (matching task) study, instead recording a higher discrimination performance by Black participants for White faces in the other studies. As expected, South African Coloured participants did not display increased discrimination performance for any of the other stimuli groups, both being out-group stimuli. Results from the training studies revealed either (a) no effect of a distributed training in feature focus over 5 weeks; (b) an increase of the OGB after a focus on critical facial features; (c) a decrease of the OGB in a task-specific training using pictures whose quality had been manipulated, and; (d) an important implication of the presence/absence of the target in a field detection study. With some promising results, the present work contributes to our understanding of how training could be used to improve face-recognition, and especially other-group face recognition.

Keywords: own-group bias, face recognition, training, face matching, face memory

Résumé

Le biais intergroupe dans la reconnaissance des visages (own-group bias ; OGB) reflète l'idée selon laquelle il est plus facile de différencier et de reconnaître des personnes issues de son endogroupe plutôt que des personnes appartenant à un exogroupe. L'existence de ce biais a été étudié pendant des années, toutefois, peu de recherches se sont focalisées sur le développement d'un moyen pour le diminuer ou l'éliminer, en ayant recours à un entraînement. Présentant cinq études impliquant des tâches d'appariement ou de mémoire, l'objectif de cette thèse était de développer et d'explorer dans quelle mesure l'entraînement peut réduire ou éliminer l'OGB. Des participants français typés d'origine européenne, ou des participants sud-africains 'Black', 'White' et 'Coloured' ont participé à différentes études. Au cours de ces études, les participants ont tous été exposés à des stimuli des 'Black' et 'White'. Dans chacune des études, les participants typés d'origine européenne et les participants 'White' présentaient un OGB avant quelconque intervention. Cependant, le biais chez les participants 'Black' n'a été détecté que dans une seule des études (tâche d'appariement). Dans les autres études, ces derniers présentaient même une meilleure performance pour les stimuli 'White', au détriment des stimuli 'Black', pourtant de leur endogroupe. Comme attendu, les participants du groupe 'Coloured' n'ont présenté aucune différence de performance envers aucun des deux groupes de stimuli. Les résultats des différentes tâches d'entraînement ont montré (a) aucun effet d'un entraînement à la focalisation sur certains traits, entraînement distribué sur 5 semaines; (b) une augmentation de l'OGB après un entraînement visant à augmenter la focalisation sur des traits considérés comme critiques; (c) une diminution de l'OGB suite à un entraînement appliqué à une tâche spécifique utilisant des photographies dont la qualité à été altérée, et; (d) une meilleure performance lors d'une tâche écologique de détection lorsque la cible est présente, plutôt qu'absente. Avec des résultats prometteurs, ce travail de recherche contribue à notre compréhension des conditions de l'utilisation de l'entraînement lorsqu'il a pour but d'amélioration la reconnaissance des visages, et plus particulièrement la reconnaissance de visages dans des situations intergroupes.

Mots-clés : biais inter-groupe, reconnaissance des visages, entraînement, appariement de visages, mémoire des visages

General Introduction

"*Growing up I'd always have people telling me "Hey! I know an Asian girl that looks just like you!" or "Are you guys like friends? Or Sisters? 'Cause you look like you could be twins" I get it. Asians have similar features but that doesn't mean we all look the same. After years of struggling to find my own identity I realized I don't need to change my hair, or the color of my eyes or the way I dress to be different. We are all different. Unique in our own way and should be seen as such."*

Katie - Peter DeVito, "We all look the same" project.

Looking at a busy street, the terrace of a restaurant, or a university library anywhere in Toulouse or Cape Town, one would easily rely on the word "diversity" to give a global description of the crowd. Far away from being the case centuries ago, History - and colonialism in particular, with globalization and tourism as other examples - developed the demography of our countries. As a result, people mixed more widely across cultures and backgrounds, and the majority of modern societies became multicultural, multi-ethnic and multi-religious. Unlike more homogeneous countries, this diversity brought us to encounter people from many different groups[1], on a regular basis. While exposure to diversity is known to be positive, unexpected consequences appeared: We now have to be able to discriminate between and recognize

---

[1] Hereafter, 'group' refers to what is commonly described as 'race' or 'ethnicity'. This position statement and concepts definitions are discussed later on, p.29.

people from many groups. However, face recognition, which is already a difficult task, is even more so when it involves recognizing people from groups different from our own. Implications of varying degrees of seriousness can emerge from this difficulty. On one hand, one can confound two classmates, wrongly wave at someone, or even have a greater difficulty differentiating members of an unknown team during a basketball game. These social situations can be uncomfortable, and even result in the use of inappropriate comments. For instance, a great artwork collection was created by the photographer Peter DeVito (https://www.peterdevito.com) to raise awareness on the fact that Asian people do not all look the same and how they feel about being perceived as such. In this piece of work, he photographed Asian people with a sticker saying "we all look the same" on their faces and presented the pictures along with testimonies of the models on their experience related to this famous sentence. On the other hand, this difficulty results in a greater difficulty matching a photograph to an individual during passport control, and in an increase of misidentifications during forensic cases. According to The Innocence Project (https://www.innocenceproject.org), in the United States of America in 2019, 367 people were exonerated due to DNA tests after being wrongly detained, within which 69% involved eyewitness misidentification, among which 42% involved an eyewitness and a suspect from different groups. These situations show how important it is to understand all the processes implicated in face recognition, especially when it involves other-group individuals.

The greater difficulty to recognize faces from other groups is, in fact, an extension of the already difficult task of unfamiliar faces discrimination and recognition. Expertise for faces is primarily developed inside of our immediate social circle (e.g., family members and friends) from repeated exposure to, and contact with, specific people right from infancy. We are thus able to discriminate and recognize these familiar faces with an almost perfect accuracy, even without any differences for own-group or other-group faces (Laurence, Zhou, & Mondloch, 2016) as familiarity overcomes the effect of groups. While contact with familiar faces benefits from multiple exposures under numerous and variable conditions, thus developing expertise, it is more difficult to develop the ability to discriminate and recognize unfamiliar faces. Indeed, people who have only been met once do not benefit from an encoding as strong as that for familiar faces (Kramer, Young, & Burton, 2018), and therefore recognition accuracy is lower and errors more frequent, even for people from the same group. Discrimination and recognition of unfamiliar faces from other groups is even less accurate since repeated and more frequent contact with own-group faces results in a greater discrimination performance for own-group faces at the expense of other-group faces. This phenomenon is called the own-group bias (OGB) and although it is often described as a face recognition impairment, it actually is an evolutionary advantage resulting from the adaptation to our specific environment. Interestingly, adopted children who were from a different group than their adoptive parents and who grew-up in an environment in which their biological parents' group was not the majority group naturally developed a greater discrimination performance for faces of their adoptive parents' group, confirming the important role of contact (de Viviés, Kelly, Cordier, & Pascalis, 2010; Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005).

As a consequence of perceptual contact, an efficient representation of faces is built in memory (Valentine, 1991). Indeed, the more perceptual contact one has, the more the representation of faces is fine and strong, and so the more accurate and easy discrimination and recognition performances are. However, perceptual contact seems not to be sufficient on its own to explain the presence of the OGB, especially since own-group expertise has been developed over the years. One is not just exposed to many faces, but also interacts with other individuals. Social contact (i.e., interactions with other-group individuals) has been studied, and explains only 2% of the variance in the OGB (Meissner & Brigham, 2001). However, the relationship between social contact and the OGB seems to be moderated by prejudice (Brigham, Bennett, Meissner, & Mitchell, 2007), while prejudice on its own also does have a main effect on OGB (Pettigrew & Tropp, 2006). Considering this relation between contact and prejudice, training participants to individuate faces, some authors found a reduction of implicit racial bias in children (Xiao et al., 2015) as well as in adults (Lebrecht, Pierce, Tarr, & Tanaka, 2009), the latter even resulting in a reduction of the OGB. These results show that perceptual and social contact, prejudice, and face discrimination and recognition are inter-related, and therefore improving or reducing one of these might have a positive effect on the OGB.

Other-group faces are automatically categorized as such and this is deleterious for their later recognition, and individuation is needed in order to succeed at a face recognition task (Levin, 2000), but few studies have worked on the development of individuation training to increase face discrimination and recognition. In these studies, individuation is created by the association between a face and a name or number that has to be learned. When tested on new faces, in some studies trained participants presented a lower OGB after training relative to untrained participants (Elliott, Wills, & Goldstein, 1973; Goldstein & Chance, 1985), while other studies did not replicate these results as they revealed no effect of such training on the OGB (Hayward, Favelle, Oxner, Chu, & Lam, 2017; McGugin, Tanaka, Lebrecht, Tarr, &

Gauthier, 2011). While individuation seems promising, however, one does not always have semantic information to associate with a newly encountered face, and in an ecological application, relying on a visual individuation seems to be an interesting alternative.

The present thesis therefore shifts away from directly studying the influence of social contact and prejudice on individuation, and focuses on perceptual individuation and development of discrimination skills instead. The present thesis focuses on the development of training for adults to achieve several purposes: (1) to involve a visual individuation through the focus on what makes a face distinguishable from the others; (2) to develop a set of generalizable skills that could be applied to novel faces.

As a function of expertise, observers automatically direct their gaze on facial features considered as critical, namely features that are perceived as being the more relevant to discriminate faces within a group. Directing visual exploration of White observers through fixation crosses, Hills and Lewis (2011) observed an improved recognition of Black faces when the focus was made on the bottom part of faces (i.e., their critical features). The OGB was even eliminated after training participants to focus on critical features through the use of pictures of morphed faces (Hills & Lewis, 2006). The first study of the present thesis is based on these findings, and aimed to decrease the OGB by instructing and training White French participants to focus on the bottom part of Black and White faces. Although participants actually modified their visual exploration as a function of training, the OGB present prior to training unexpectedly increased, implying a greater number of false alarms (i.e., inaccurate recognition) for Black faces. The explanation might be, among others, that they had not created individuation, but instead created a focus on differences between our two groups rather than in-between each group. In this regard, and considering that few studies used multiple training sessions of individuation and succeeded in the reduction of the OGB (Goldstein & Chance, 1985; Malpass, Lavigueur, & Weldon, 1973), the second and third studies aimed to induce individuation differently. Over three weeks of training participants were asked to focus on the most relevant part of each face, namely the feature/part which would help them to better memorize this face. They were constrained on the parts of faces they were asked to consider while looking at them: top versus bottom, internal versus external, or features versus configuration. This study was conducted in South Africa and France, with some modifications. Discrimination performance, tested one week before and one week after the first and last session of training, revealed that White participants of both countries presented an OGB in pre-test while Black and Coloured South African did not. This OGB was still present in both studies during the post-test, revealing no effect of training. These results suggest that individuation induced in this way did not help participants to perform better, exposure to many faces did not decrease the OGB, and that directing participants to focus on specific parts of faces did not help either.

Since the OGB did not decrease as a function of visual individuation training, and that Matthews and Mondloch (2018) found an effect of training on a sequential matching task but not on a memory face recognition task, the last two studies I conducted used matching tasks. That is, the fifth study aimed to test task-specific training on South African Black, White and Coloured participants. Considering the difficulties also present during a matching task, and the importance of success in this task for passport and border officers for example, participants were trained to match pictures of identities depicted under different levels of degradation (i.e., manipulated through pixelation). Again, only White participants presented an OGB in the pre-test. In the post-test, untrained White participants still displayed an OGB, while trained participants presented a significantly higher discrimination performance for Black faces than untrained participants. Training was therefore efficient to reduce the OGB on a task on which they had been trained. The sixth study then sought to test if a similar type of training might be adapted to improve performance in a field detection task. South African White, Black and Coloured participants were trained to match different views of identities, with the expectation that they would develop a better representation of variability within an identity, and they were then asked to complete a field detection task. For the first time, Black participants also presented an OGB along with White participants prior to training, suggesting that the nature of the task and/or the larger sample size might have some implications for the finding of the OGB in this group. Results on the field task suggest that participants performed significantly better when the target was present at the expense of target-absent conditions. This suggests that field studies might involve different processing than laboratory studies which use pictures or video. These two studies highlight the differences that might be present between memory and matching tasks, as well as the relevance of training for a specific task, and expands on the utility of field studies and the consideration of present and absent conditions while studying face discrimination and recognition performance.

The present thesis presents some limited results on the usage of training to decrease the OGB, and is written in three main parts: first, a theoretical background is provided, then the experimental studies are detailed, and then a general discussion and a general conclusion are provided. In the former section, three chapters are presented about face processing (Chapter

1), explanatory concepts and theories of the OGB (Chapter 2), and presentation and criticism of learning and training studies on both own-group and, especially, other-group faces (Chapter 3). In the experimental section, the ethical considerations and implications of the thesis are presented, the database and common measures used are described, and a detailed presentation of rationale, method, results and discussion of each study are detailed. Finally, a general discussion presents the overall findings in regard to the existing theory, while a general conclusion closes the present work and proposes new perspectives for future research.

Context of the present research

The present thesis has been conducted as a joint doctoral contract between the universities of two different countries: L'Université de Toulouse, in France, and the University of Cape Town, in South Africa. The project was a part of the Chaire d'attractivité "MisIdentification Contact" carried out by Professor Jacques Py and Professor Colin G. Tredoux. The studies presented in this thesis have been conducted in the two countries, either at the Université de Toulouse (Study 1, Study 2), or at the University of Cape Town (Study 4, Study 5 and Study 6), while Study 3 was conducted online, across France. In respect to the policy of the University of Cape Town, the proposal of the present thesis (i.e., defence of the detailed project including the presentation of planed studies) was presented in front of a scientific and ethic committee prior to any data collection at UCT. As the research is presented in English, instructions given in French for studies conducted in Toulouse have been translated by the author when needed.

Due to the nature of the topic of OGB and the international dimension of my research, it is necessary to first question the concept of 'race' and its meaning and acceptance, and then to determine the descriptive and categorization terminologies used throughout the thesis, and finally, to situate the present work in its demographic contexts.

On the usage of racial terms

As Malpass and Kravitz (1969, p.330) specified in the eyewitness testimony field of research, race is "merely a shortland way of referring to differences in physiognomy that correlate with and are cues to the race of the stimulus person". The term ‘race’ has been used in a phenotypical (i.e., physical and morphological appearance) and a societal meaning in the field of eyewitness testimony, thereby distancing itself from any biological definition or association with negative attitudes and behaviors from colonialism used to classify groups and justify White supremacy (Fluehr-Lobban, 2018; Templeton, 2013). The concept of phenotypic race categorize rather than classify people regarding their shared features within a group, and looking at a crowd makes one easily understand how individual differences can be intuitively assembled to define different groups (Corcos, 2016). Indeed, some features tend to be more shared within a group than between groups. However, since talking about race is a sensitive issue, the term ‘race’ is not being used outside of the present section[2], and it is therefore replaced by ‘group’, for a very similar meaning.

Being the most distinctive feature, skin color was historically used to name racial groups since the 17’s, defining groups as: Black (or Negroes; now usually replaced by African American), White (now sometimes replaced by European or Caucasian), but also Yellow (later replaced by Oriental, or Asian), Red (later replaced by Native American) and Brown (later replaced by Malay or Indian; Corcos, 2016; Fluehr-Lobban, 2018). Even though terms such as ‘Negro’ are outdated and offensive today, there is no international consensus on the acceptability of racial group names, which are used differently in different countries. For instance, ‘race’, ‘Black’, ‘White’ and ‘Coloured’, are terms usually used and accepted in South Africa, even if its recording starts to be discussed from its relativity to the Historical context (i.e.,

---

[2] I acknowledge that other authors have also questioned the usage of the term ‘race’ (e.g., Mukudi & Hills, 2019; Valentine, Lewis, & Hills, 2016), specified its meaning (e.g., Sporer, 2001), or made a distinction between ‘race’ and ‘ethnicity’ (e.g., McKone et al., 2019).

Apartheid). In France, it is a complete different story. Indeed, the term ‘race’ is so problematic that in July 2018, the Assembly voted in favor of the removal of the term ‘race’ from the first article of the French Constitution (i.e., the article which defines the core values of the Republic), eventually replacing the sentence "assure l’égalité devant la loi de tous les citoyens sans distinction d’origine, de race ou de religion" (ensures the equality of all citizens without any distinctions of origins, race or religion) by "sans distinction de sexe, d’origine ou de religion" (without any distinction of sex, origins or religion), introducing at the same time the term ‘sex’ into the constitution (see the report of the National Assembly meeting, http://www.assemblee-nationale.fr). Even if the process to remove the word may take a long time to come into effect, and that it would actually be removed only from the first - and not the other - articles of the constitution, this change clearly shows the sensitivity of this term in France and more broadly in Europe (for a discussion, see Goldberg, 2006). Considering these elements, terms as ‘personne typée européenne’ (typical European looking person) and ‘personne typée africaine’ (typical African looking person) are preferred by my French research team.

Group categorization for research purposes

Considering the topic of the present thesis, categorizing participants within such groups is both necessary, and challenging. With the aim to study the OGB, we have to know which category both participants and stimuli belong to. Without this information it is not possible to compare discrimination performance between groups, and thus, to study the OGB. However, retrieving this information from participants is challenging. That is, whereas participants in South Africa are asked to self-report if they are Black, White, Coloured, or from another

Table 1 – *Fitzpatrick classification (Fitzpatrick, 1988).*

I	Pale white skin, blue/hazel eyes, blond/red hair, always burns, does not tan
II	Fair skin, blond to light brown hair, blue eyes, burns easily, tans poorly
III	Darker white skin, blue or light brown eyes, brown hair, tans after initial burn
IV	Light brown skin, brown hair, burns minimally, tans easily
V	Brown skin, eyes and hair, rarely burns, tans darkly easily
VI	Dark brown or black skin, eyes and hair, never burns, always tans darkly

group on a regular basis in administration forms, asking this question in France is impossible. In fact, recording such information is considered very sensitive by the CNIL (Commission Nationale de l'Informatique et des Libertés; The French Data Protection Authority) which, under the European Policy, specifies that it is forbidden to collect or use racial information (among other type of information, as religious belief for instance), except when justified with solid arguments. To get around this issue, my research team and I developed the idea of relying on a medical categorization from skin tolerance to the sun (i.e., 'phototype', the classification of Fitzpatrick, 1988), and to combine it with questions specifying the country or continent of birth of participants' relatives (i.e., parents, grand-parents). All together, this information led us to categorize people as White or non-White, since studies conducted in France only involved White participants because of the lack of representativeness of other groups. Although this method has some limitations, and is especially not reliable for use in South Africa[3], it was quite a useful and adaptive way to record this information, at least in France.

To summarize, in the present thesis 'group' is preferred to 'race' when referring to the bias as the own-group bias (OGB). Information on race, phototype and country of birth of relatives

---

[3] This point is discussed in Study 4, p.139, and in the general discussion p. 225.

are self-reported and optional[4], and questions were adapted to the place of data collection. Groups studied in the present thesis are: White (i.e., phototypes I, II and III), Coloured (i.e., phototype IV) and Black (i.e., phototypes V and VI). When reporting observations from published studies, terms the original authors refer to are preferred so as not to distort the point. When a distinction has to be made between different types of groups, 'race' or 'racial group' might be used for clarity.

Group representativeness in France and South Africa

In South Africa, data on group are usually recorded, and made available. Whereas the country is predominantly Black, Cape Town has quite a big community of Coloured people, and the University of Cape Town (UCT) has been predominantly White before 2017 (Figure 1). In 2017, UCT registered an equal number of Black and White students, and from 2018 and onward, Black will be the majority group.

This reversal is more representative of the actual demography of the entire country; the minority of Coloured students, however, is not representative of the population of Cape Town. These data are important and interesting to take into account for studies on the OGB when considering the effect of contact.

In France, such data are not recorded because it is considered sensitive. The only available data concerns immigration; that is, in 2018, of the 6.5 million people living in France, 9.7%

---

[4] Although left optional for ethical considerations, missing data resulted in the exclusion of the participants from the final sample.

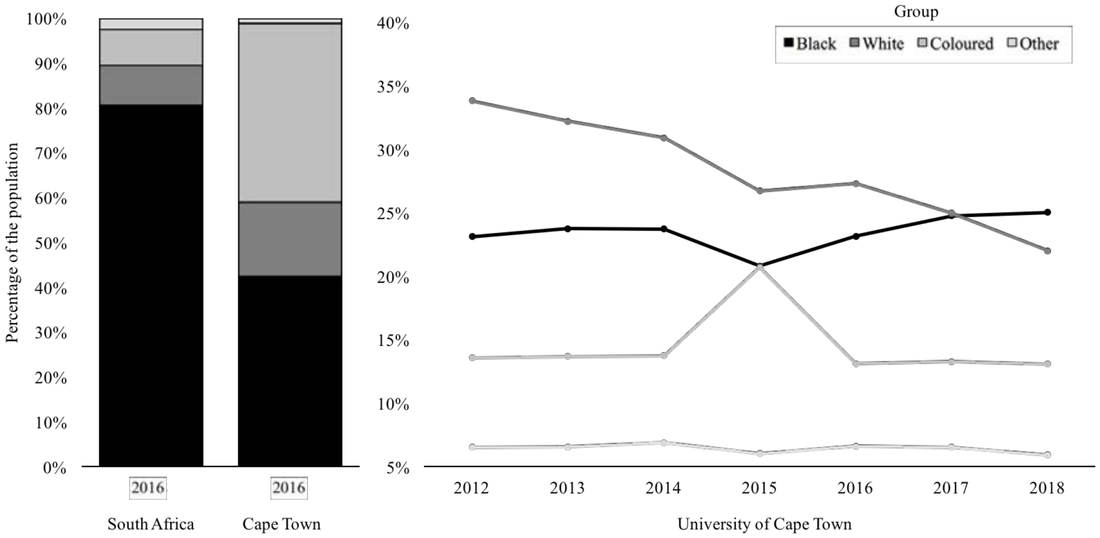

Figure 1 – On the left, group representativeness in South Africa (55’653’654 citizens) and Cape Town (4’005’016 residents) in 2016 (http://cs2016.statssa.gov.za). On the right, group representativeness at the University of Cape Town between 2012 and 2018, considering 18’969 South African students in 2018 (international students and undisclosed data are not included; https://www.uct.ac.za/main/about/finance/annual-statements).

were immigrants, among which 37% obtained French nationality. The majority of these immigrants were from Africa (46.1%), and especially from Maghrebian countries (https://www.insee.fr/fr/statistiques). In the absence of these data, one could assume with great confidence that the majority of the population of the French mainland is White, even if many French people from overseas departments and territories, who are mostly Black, are living in the mainland.

Section I

Theoretical background

Chapter 1

Theories of face processing

1.1 Faces as special stimuli

Faces, defined as specific and complex objects, are composed of first and second order relational properties (Diamond & Carey, 1986). 'First-order relational properties' correspond to the features that make a face a face, relative to other visual stimuli, containing eyes, noses, mouths, etc. 'Second-order relational properties' correspond to the distinctive relations between the features, especially spatial relations, on the basis of which individuals are unique and distinctive from each other. As a face is easily recognizable as being a face, and since we are used to recognize familiar faces with great - even perfect - accuracy, it is usually believed that people are experts in face recognition. In reality, people are only experts at recognizing familiar faces, and have a greater difficulty in recognizing unfamiliar faces (Young & Burton, 2017; 2018).

1.1.1 The effect of familiarity in face discrimination

When encoding a new face, two steps are involved: first, the image-base pictorial representation is encoded, as for any stimuli; and then the identity-specific representation of the face is encoded. The identity-specific representation, which is the result of a combination of multiple exposures to a specific face (Menon, White, & Kemp, 2015), is the cause of familiarity and, thus, of the better discrimination and recognition performance. Familiar faces (i.e., faces from familial, social and professional circles), benefit from multiple exposures under different lighting, distance, emotions or poses, resulting in strong representations in memory (Burton, 2013; Johnston & Edmonds, 2009). Variability, induced by multiple exposures, results in a greater representation of faces in memory due to a better encoding of the identity-diagnostic features of each face (Bruce, 1994), which are necessary to recognize new images of an identity (Andrews, Jenkins, Cursiter, & Burton, 2015). Celebrities and famous people, whose faces are encountered in different conditions as well, are also considered as familiar faces (Johnston & Edmonds, 2009). Conversely to familiar faces, unfamiliar faces are faces from individuals newly encountered for a brief and often unique time, a face to which the observer has never been exposed to before. As unfamiliar faces are encountered and encoded under a single variation defined by the angle, emotion and other parameters present during the encoding, their recognition is image dependent (Bruce, 1982; for a review see P. J. Hancock, Bruce, & Burton, 2000). Observers are therefore impaired in their capacity for identity generalization on the basis of so few cues, and are more likely to fail to match two pictures of the same person (Burton, 2013).

The superiority of familiar face processing has been demonstrated in sorting studies, during which participants are asked to sort photographs into piles where each pile corresponds to an identity. These studies revealed that sorting unfamiliar faces always results in a greater number of piles than the actual number of identities, while familiar faces are usually sorted into the correct number of piles (Jenkins, White, Montfort, & Burton, 2011). This greater accuracy for familiar faces is also present for other-group faces, suggesting that familiarity overcomes group differences in face discrimination (Laurence et al., 2016; Zhou & Mondloch, 2016). More broadly, familiar faces benefit from a greater discrimination and recognition performance than unfamiliar faces, regardless of the group of the faces and participants, and of the task used (i.e., face sorting, face matching, face detection, and/or memory face recognition).

1.1.2 Perceptual processing of faces

When encountering a novel face, an observer can rely on two types of processing: featural or configural (also known as holistic). Featural processing consists of the exploration of features independently, and configural processing considers the face as a whole - that is, the face features all together and, more importantly, their inter-relations. These two processes, although in opposition, are not conflicting, and it is possible to rely on either of the two independently from the other (Collishaw & Hole, 2000), or to use both simultaneously for an even stronger encoding (Hayward, Rhodes, & Schwaninger, 2008; Tanaka & Simonyi, 2016). Nevertheless, configural processing is overall known to result in a stronger encoding, and therefore in better recognition (Tanaka & Farah, 1993), while using featural processing alone appears to be inadequate for an efficient encoding, resulting in greater difficulties during recognition.

Configural processing, which results from a visual experience, is automatic and mainly used for familiar faces, and own-group faces (Le Grand, Mondloch, Maurer, & Brent, 2001). However, regardless of the processes used, own-group faces are still better recognized than other-group faces (Sadozai, Kempen, Tredoux, & Robbins, 2018), and own-group and other-group faces appear to be processed differently (K. J. Hancock & Rhodes, 2008). While own-group faces are more likely to be processed configurally, other-group faces are more likely to be processed featurally (Hugenberg & Corneille, 2009; Michel, Caldara, & Rossion, 2006; Michel, Rossion, Han, Chung, & Caldara, 2006; Sadozai et al., 2018; Tanaka & Simonyi, 2016). Conversely, featural processing is more likely to be used for unfamiliar other-group faces because of a deficit in the perceptual experience reducing the use of configural processing (Byatt & Rhodes, 2004; G. Rhodes, Locke, Ewing, & Evangelista, 2009; Tanaka, Kiefer, & Bukach, 2004; Walker & Tanaka, 2003).

Even though other-group face discrimination is impaired by the use of a featural processing, it remains likely that focusing on the more relevant features of other-group faces might be a way to improve one's performance by counteracting the deleterious effect of the featural processing in directing observers to the discriminating features within other-group faces. Establishing what the critical features for each group are is therefore necessary. Some studies on face description revealed that White observers more often describe hair color, texture and shape, and eye color, while Black observers provide a more global description, including description of hair position, size and whites of the eyes, eyebrows, ears and chin (Ellis, Deregowski, & Shepherd, 1975). By contrast, another study (Shepherd & Deregowski, 1981) revealed that while describing an European face, one would use more descriptors on hair color, length and texture than for African faces for whom skin tone and nose breadth are more likely to be described, no matter their group membership (Scottish and Zimbabwean in their study). Studies using eye-tracking gave more direct evidence on face exploration behavior, and revealed different results whether observers' and/or stimuli groups were more affected by a specific strategy. Despite differences, it remains that eyes are commonly the most frequently and automatically encoded feature, followed by the nose, and then the mouth (Althoff & Cohen, 1999; Heisz & Shore, 2008; Henderson, Williams, & Falk, 2005; Hsiao & Cottrell, 2008); this reveals the eyes as being the key feature of the learning and recognition of faces

(Sekiguchi, 2011; Stephan & Caine, 2007). There are, however, two inconsistent findings in the literature: the first being that features focused on are different in regard to the observer's group; the second is that the features focused on are dependent on the stimulus group. That is, White observers in several studies have focused mostly on the eyes, while Black and Asian observers have focused mostly on the nose, then the mouth, regardless of the stimulus group (Blais, Jack, Scheepers, Fiset, & Caldara, 2008; Goldinger, He, & Papesh, 2009; Tan, Stephen, Whitehead, & Sheppard, 2012). In other studies, eyes and hair are the features all groups of observers focused on when exploring a White face, while participants focused on noses and mouths when exploring Black and Asian faces (Arizpe, Kravitz, Walsh, Yovel, & Baker, 2016; Hills, Cooper, & Pake, 2013; Hills & Pake, 2013; Goldinger et al., 2009). These differences seems to be mostly due to differences in methodologies and/or analysis (Arizpe et al., 2016). However, since expertise is developed mostly from exposure to, and experience with, own-group faces, it is unlikely that observers spontaneously focus on diagnostic other-group face features (Hills & Pake, 2013). To develop this expertise, or at least to improve face recognition and to develop strong representations of other-group faces, contact with other-group individuals is fundamental.

1.2 Face representation in memory: The multi-dimensional face space model

The most famous model of face representation in memory is the Multidimensional Face Space (MDFS) model of Valentine (1991; Valentine & Endo, 1992; Valentine et al., 2016). According to this model, every new encountered face is represented in a multidimensional space, and the ease (and accuracy) of its retrieval is determined by its location within the space and its relation to the other stored faces. To visualize the model, one can imagine two straight lines (i.e., dimensions) intersecting in their origin point: one representing the range of variability existing on feature A, the second representing the range of variability existing on feature B. Assuming that the features' variabilities are normally distributed within a defined population, the direct space close to the intersection point between those two dimensions represents the central tendency of representation of the defined population. Since it exists in a great number of dimensions, each dimension representing a possible feature characteristic (e.g., eyes colour, eyes shape, eyes white), the model is in reality very complex and difficult to represent. Taking into consideration every single dimension, each new encountered face is represented within the face space as a unique point relative to its own values on each of these dimensions. Sharing close-to-identical feature specificities, and thus being similar on average to the majority of faces of a population, average faces are represented close to the central tendency, and close to each other. On the contrary, distinctive faces are separated by a larger distance from the others, and are represented as more isolated from the central tendency. Distinctive faces are therefore more easily recognized because they are dissimilar from the majority, while for the majority, their closeness is more likely to cause confusion during information retrieval. Applying the MDFS model, the presence of the OGB is explained by the way other-group faces are represented in memory. Because the representation model is refined as a function of exposure, it has been developed for, and is more adapted to, own-group faces, resulting in an impairment of other-group faces recognition. One consequence of the lack of exposure to other-group faces is that one does not focus on the more relevant or critical dimensions to discriminate other-group faces, and instead uses identical dimensions that are neither relevant, nor efficient, as for own-group faces. Indeed, some dimensions might be more or less usual, namely other-group faces differ on different dimensions than own-group faces do, even if they share other dimensions. As a consequence, other-group faces are represented close together, considering only their representation on the developed dimensions, in the absence of other more appropriate dimensions. In other terms, other-group faces are clustered in a different area than own-group faces as their central tendency involves different features or feature variations. It is thus more difficult to individuate them in regard to the whole group, than it is for own-group faces.

Two versions of the MDFS are theorized: a norm-based model and an exemplar-based model. The norm-based model implies that each newly encountered face is encoded relative to a norm (i.e., prototypical face) which is located at the intersection point of dimensions. By contrast, the exemplar-based model implies that every newly encountered face is encoded relative to its distance from the other faces rather than from a norm. That is, every new face is represented according to its similarities and dissimilarity to every single other face already represented, and not in relation to a specific norm. These two versions are similar, since the area around the intersection point represents the central tendency and the density of faces is higher close to it. The difference between these two versions is that the first provides a prototype relative to which every new face is encoded, while the second does not, and faces are all relative to each other, creating several smaller groups instead of one (Figure 1.2).

Even though the two models have received empirical support (see Valentine et al., 2016), the exemplar-based model appears to be more appropriate for explaining the OGB (Valentine & Endo, 1992). In the norm-based model, other-group faces are encoded in relation to an

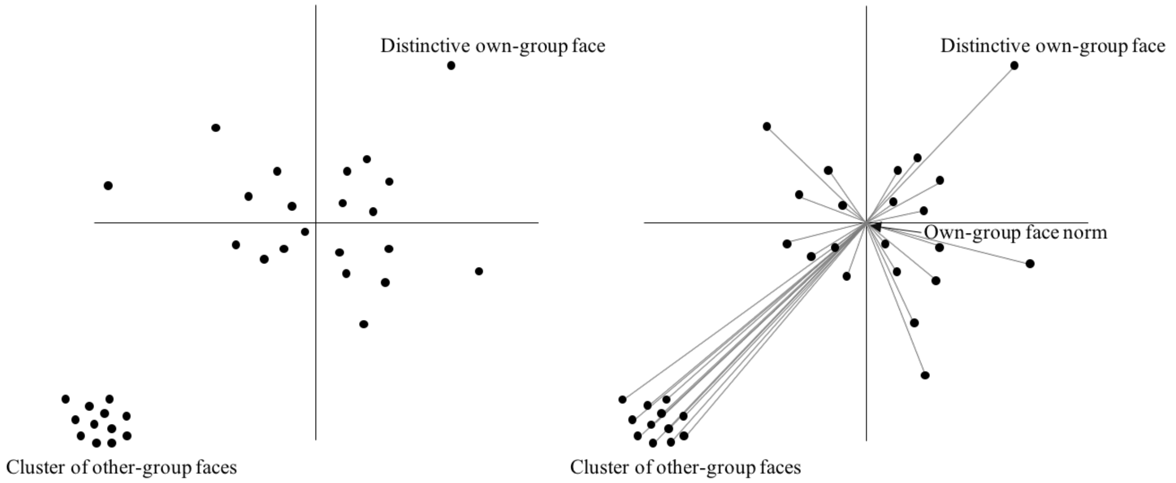

Figure 1.2 – Schematic representation of Valentine's multidimensional face space including own-group and other-group faces representation. On the left, the exemplar-based model where all faces are represented in relation to the others, and other-group faces are represented as a distinctive cluster since represented relatively to inappropriate own-group dimensions. On the right, the norm-based model where every face is represented according to an own-group face prototype (i.e., norm), other-group faces being in the same direction because lack of their own norm.

inappropriate norm (i.e., prototype). A prototype being missing for other-group faces, there is only a distinctiveness effect in the representation, but other-group faces are still represented in relation to the own-group norm. Other-group faces are thus represented as a dense cluster and from a greater distance to the norm, and the central tendency, without being able to distinguish one another within the group. In absence of a norm, it is difficult to discriminate any face of the group in relation to it. A question remains here on how this norm would be developed, and what the cutting point would be from which one would consider having enough exposure and expertise to have created this norm. For the example-based model, other-group faces are represented in relation to the other faces, and because these 'other faces' are mostly own-group faces represented along specific dimensions that are either not shared by other-group faces or on which other-group faces are different values, other-group faces are represented further from the other groups, creating another unique and indistinguishable 'other-group face cluster'. This cluster is actually the creation of a new central tendency, adapting all other representations around this area of a new other-group. According to the exemplar-based model, the greater exposure one has to other-group faces, the more expertise will grow, which results in more refined, and therefore better, representation of diversity and similarities of a group. As a result, new faces will be represented in relation to a higher number of relevant faces (i.e., faces from the other-group). The exemplar-based model allows the increase of other-group face discrimination by multiplying the relation between the exemplars whereas the norm-based model induces a creation of a new prototype in relation to which other faces will be represented.

Considering the MDFS and the OGB, Little, DeBruine, Jones, and Waitt (2008) suggested that there might be multiple face spaces, each of them being dedicated to a specific category, such as for group or even gender. These multiple face spaces could either be included into a more global face space as sub face spaces, or being independent face spaces where motivation would play a directive role on the relevant space to use (Valentine et al., 2016). Instead of being represented in relation to every other face, a newly encountered face would be first ‘sorted’ regarding its group characteristics, and then represented accordingly in the relevant subspace. In this configuration, expertise would lead to specification and the development of the relevant face space, including adapted dimensions and thereby central tendencies, for each group. Indeed, the central tendency is developed from the exposure and expertise with own-group faces; or the group one is exposed to the most (see a case of adopted children, Sangrigoli et al., 2005). As expressed above, to better represent faces in memory, one should be able to focus on more appropriate and critical features of other-group faces, as is efficiently done for own-group faces.

It is noteworthy that the way faces are encoded impact their representation and recognition accuracy. In fact, a face representation has to be similar enough to the encountered face, and dissimilar enough from representations of other faces to avoid confusion with the most similar representations of those. In this regard, the more a face is encountered, the better its representation would be in terms of its own specificities and dissimilarities to the others, and thus, the more accurately this face would be recognized. To be efficiently represented in the MDFS, faces have to be processed deeply, as expertise contributes to the development of the MDFS. Indeed, the MDFS becomes more refined with exposure and experience, and therefore one becomes more efficient with age to process own-group, familiar, and upright faces accordingly (Valentine et al., 2016). From this observation, it is of great interest to focus on how one could develop this expertise and an adapted representation of other-group faces.

Chapter 2

The Own-Group Bias in face discrimination and recognition

As early as 1914, Feingold acknowledged the existence of a greater similarity within groups than between groups, and conceptualized the first idea of the existence of an own-group bias, acknowledging that:

> Now, it is well known that, all thing being equal, individuals of a given race are distinguishable from each other in proportion to our familiarity, to our contact with the race as a whole. Thus, to the uninitiated American, all Asiatics look alike, while to the Asiatic all white men look alike (p. 50).

Later on, Howells (1938) mentioned the potential implication of 'race' in face recognition and the first experimental demonstration of the OGB was made by Malpass and Kravitz (1969) who observed a higher recognition performance for own-group faces than for other-group faces. Following these results, the concept of a 'racial bias in eyewitness identification' was first introduced by Malpass (1974) to name this recognition performance difference between own-group and other-group faces. Since then, scientific literature has reported this effect using terms such as *cross-race bias*, *cross-race effect*, *other-race bias* or *other-race effect* (non-exhaustive list).

As face recognition could be affected by the difference between an observer and a target face such as age (own-age bias, see M. G. Rhodes & Anastasi, 2012) or gender (own-gender bias, see Herlitz & Lovén, 2013) the own-group (i.e., 'race') bias is considered as one of the most robust effects in psychology. In (2001), Meissner and Brigham conducted a meta-analysis of 39 studies covering 30 years of research, and concluded that own-group faces are 1.4 times more likely to be correctly identified than other-group faces. Furthermore, other-group faces are 1.56 times more likely to be incorrectly identified than own-group faces. The OGB is usually characterized by a greater proportion of false alarms (i.e., inaccurate identifications) for other-group than own-group faces, often with no or little differences for hits (i.e., accurate identifications) such that own-group faces are more accurately recognized than other-group faces (Bertone, Mélen, Py, & Somat, 1995; Meissner & Brigham, 2001; Meissner, Brigham, & Butz, 2005; M. G. Rhodes & Anastasi, 2012; Slone, Brigham, & Meissner, 2000). Beside accuracy, participants are more likely to present a liberal bias (i.e., tendency to answer more positively than negatively at a recognition task) toward other-group faces than own-group faces, suggesting a greater confusion of faces a retrieval (Slone et al., 2000; Wilson, Bernstein, & Hugenberg, 2016). Overall, decision times are similar for decisions involving other-group faces and own-group faces, while confidence is usually higher for own-group faces than for other-group faces (Meissner & Brigham, 2001; Wright, Boyd, & Tredoux, 2001; 2003).

2.3 The universality of the own-group bias

Although a very large number of studies have been conducted using Black and White participants from the United States of America (see Meissner & Brigham, 2001), the literature has been expanded to include far more diverse populations, and the OGB has been demonstrated to be universal as it is shared by many groups in many countries (Sporer, 2001, Table 2.1). For instance, Black, East/Southeast Asian, Hispanic, and White children, all living in California, were tested in a yes/no recognition paradigme and presented a higher recognition performance for own-group faces than for the three other-group faces when gathered together as one 'other-group' (Gross, 2009, 2014), however, while separated (i.e., one group at the time), Black adults show a similar performance for Black and White faces (Gross, 2009). Anglo-American (i.e., White American), Black-American, and Hispano-American participants also presented an OGB in a field study; again, each in favor of their own-group at the expense of the two other groups (Platz & Hosch, 1988). Black, White, Latino, and Asian participants from California were also tested on Black and White faces with an old/new recognition paradigm (Teitelbaum & Geiselman, 1997). Although the OGB was found for both Black and White participants, no differences in the discrimination performance were found for Latino and Asian participants, for whom the presented stimuli were both from other-groups. Hispanic participants from Texas tested on Black and Hispanic faces also presented an OGB (MacLin, MacLin, & Malpass, 2001). Although these studies have been conducted in the same country using different participant groups, other studies were conducted using participants and/or stimuli from two different countries, and found similar results. Indeed, samples of Japanese and British participants (Valentine & Endo, 1992), Singaporean and Canadian Asian and White participants (W.-J. Ng & Lindsay, 1994, study 2), Black and White English and South African participants (Wright et al., 2001, 2003), Black and White English and Zimbabwean participants (Chiroro & Valentine, 1995), Black and White South African participants (Chiroro, Tredoux, Radaelli, & Meissner, 2008), White German and Turkish participants (Sporer &

Horry, 2011), or White German participants (Singmann, Kellen, & Klauer, 2013) were also tested with stimuli from the own-group and at least one of the other-groups. While all studies revealed an OGB, A. H. Ng, Steele, and Sasaki (2016) tested European Canadians and first-generation East Asian Canadians using White and East Asian stimuli, and revealed an OGB in study 1, but in study 2 where social categorization was manipulated, they observed an OGB in the European participants but not in the East Asian participants. These studies suggest that the OGB is found worldwide, and not only from the majority group to the minority, but also in the reverse direction. Nevertheless, studies which did not find the OGB (A. H. Ng et al., 2016; Wright et al., 2003) explained this failure by the presence of cross-group contact and/or categorization. These explanations are developed later on in the present chapter. Black participants from Wright et al.'s study (2003) even presented a 'marginal' higher recognition for White rather than Black faces, in contradiction to the expected direction of the own-group bias. This study was conducted using Black and White students from South Africa and White students from England who completed an old/new standard recognition task using Black and White faces as stimuli. However, in their earlier study (Wright et al., 2001), the authors found an OGB in the expected directions for both Black and White English participants and Black and White South African participants during a field study. The authors suggested that the targeted population were of different natures in the two studies, being students in one study (2003) and a convenience sample recruited in shopping malls in the other (2001). The fact that the majority of the population in South Africa is Black while the majority of students at the University of Cape Town was White at the time contributed to this explanation. Additionally, Black South African participants displayed a great amount of inter-group contact, which is another potential explanation of the absence of any OGB in this study (Wright et al., 2003). Later studies conducted in South Africa have also shown different results. A study conducted at the University of Pretoria (Chiroro et al., 2008) demonstrated an OGB in the expected direction for both Black and White students, while a study conducted at the University of Cape Town revealed an OGB only for Black participants but not for White participants (Derbyshire, 2018). Another study (Goodman et al., 2007) revealed an OGB for White South African participants (children and adults), right from school age (9-10 years old). Seutloali (2014) studied discrimination performance of Black, White and Coloured students from the University of Cape Town, while using the same groups as stimuli. White participants presented a better performance for White rather than Black faces, but without any significant difference between White and Coloured faces. Coloured participants demonstrated a greater discrimination performance for their own-group faces than for Black faces, without differences between own-group and White faces. Black participants performed similarly for Black and White faces, suggesting no OGB, and even displayed a greater performance for Coloured faces. These differences could be due to the evolution of the diversity within South African universities, and it raises interesting questions for this population for OGB studies compared to more homogeneous populations.

Studies on the OGB in France has generally shown consistent results. Indeed, when White French participants were tested, they presented an OGB toward African faces (Brunet, 2017, 2018) and Maghrebian faces (Hajji, 2018). Another study conducted by Bataille (2018) revealed that White and Maghrebian participants presented an OGB, with an overall greater discrimination performance for own-group rather than other-groups, and with a greater difficulty for African faces rather than own-group faces or White/Maghrebian faces. It is noteworthy that the Maghrebian participants had been living in France for years.

2.3.1 The development of the own-group bias during infancy

Face discrimination and recognition are dependent on the group of people we are usually exposed to, generally from the first months of life. As for face processing in general, the OGB also roots in infancy, and differences have been observed between the processing of own-group faces and other-group faces at a very young age (Quinn, Lee, & Pascalis, 2018; Sugden & Marquis, 2017). Infants show a preference for own-group rather than other-group faces from 3-months old (Kelly et al., 2005), but their difficulty to discriminate other-group faces starts from 6-months old (Kelly, Quinn, et al., 2007), and it is only from 9-months old onward that the OGB becomes robust (Anzures et al., 2013; Kelly, Quinn, et al., 2007). Infants' early visual preference is dependent on their growing environment in terms of exposure to different faces from their first month of life (Heron-Delaney et al., 2011). However, other-group adopted children also present an OGB against faces of their biological parents' group (i.e., in the same direction as that of their adopted parents) and are, presumably because of the visual environment they grew up in, either less able to recognise people from their biological parents' group than their adoptive parents' group (de Viviés et al., 2010; Sangrigoli et al., 2005), or equally able to recognize people from both own-group and other groups (De Heering, De Liedekerke, Deboni, & Rossion, 2010). Interestingly, discrimination performance for other-group faces decreased as age increased, which is not found for own-group faces (Sugden & Marquis, 2017), even if this finding is only found when age is treated as a categorical rather than continuous variable. However, in older childhood, before adulthood, children present

Table 2.1 – *Non-exhaustive list of studies on the OGB with different populations.*

	Participant group	Stimuli group
Bataille, 2018	White French and Maghrebi living in France	Black, White, Maghrebi
Brunet, 2017	White French	Black, White
Brunet, 2018	White French	Black, White
Chiroro et al., 2008	Black, White South African	Black, White South African and American
Derbyshire, 2018	Black, White South African	Black, White
Gross, 2009	Black, East Asian, Hispanic and White, living in the US	Black, East Asian, Hispanic and White, living in the US
Gross, 2014	Black, East/Southeast Asian, Hispanic and White, living in the US	Black, East/Southeast Asian, Hispanic and White, living in the US
Hajji, 2018	White French	White, Maghrebi
MacLin et al., 2001	Hispanic living in the US	Black , Hispanic
W.-J. Ng & Lindsay, 1994, study 2	Asian and White from Singapore and Canada	Asian, White
A. H. Ng et al., 2016	White and Asian Canadian	White, East Asian
Platz & Hosch, 1988	Anglo-American, Black-American and Hispano-American	Anglo-American, Black-American and Hispano-American
Seutloali, 2014	Black, White, Coloured South African	Black, White, Coloured
Singmann et al., 2013	White German	White, Turkish/Arabic
Sporer & Horry, 2011	White German and Turkish	White German,Turkish, Black American, White American
Teitelbaum & Geiselman, 1997	Black, White, Latino and Asian living in the US	Black, White
Valentine & Endo, 1992	Japanese, British	Japanese, British
Wright et al., 2001	Black and White English and South African	Black, White English and South African (confederates)
Wright et al., 2003	Black and White English and South African	Black, White

a lower OGB than adults (Goodman et al., 2007), most likely suggesting that the OGB is a consequence of the homogeneity of our specific environment and is developed over time (see also Hills & Lewis, 2018). Indeed, our visual environment from early childhood is implicated in the development of recognition abilities. The OGB is thus presumably due to a lack of exposure to other-group faces. However, in relation to what has been shown by studies with adopted children, a recent study revealed that social contact present during childhood is related to an absence of the OGB, making exposure crucial during childhood to prevent the appearance of the OGB, and especially before 12 years old (McKone et al., 2019).

From these studies, it seems clear that exposure to many faces, and contact, have great importance in the development of own-group and other-group face discrimination and recognition performance.

2.4 Inter-group contact in the own-group bias

2.4.1 Perceptual contact: Importance of contact quantity

An explanation of the OGB is that perceptual expertise present for own-group faces is also used to process other-group faces. However, as the critical features of each group are different, to improve other-group face discrimination and recognition an observer should develop more useful and adaptive face processing in this regard (Brigham & Malpass, 1985; Brigham et al., 2007; Slone et al., 2000). The OGB is thus reversible, and increasing contact with other-group faces should decrease the OGB as experience and exposition increase (Goldstein & Chance, 1985; Michel, Caldara, & Rossion, 2006). In addition to adopted children who developed a life-long expertise (De Heering et al., 2010; Sangrigoli et al., 2005), local Black children from a mixed-group school in Zimbabwe (Chiroro & Valentine, 1995) and expatriate Asians who have been living in Germany for an average of 22 months (Wiese, Kaufmann, & Schweinberger, 2014) tested on White (i.e., other-group faces) and respective own-group faces showed a better recognition for other-group faces than did control group members with lower contact. A study on White children growing-up in a segregated suburbs and schools in the US revealed a stronger OGB than those growing-up in a mixed-suburb and school (Cross, Cross, & Daly, 1971). Crucially, bi-racial participants from mixed White American and African American parents also performed equally for Black faces and Whites faces (Goodman et al., 2007). These studies suggest that increasing exposure to members from another group, an easy feat from a perceptual point of view, is necessary to counteract the OGB. In addition to developing the ability to focusing on critical features of other-group faces (Hills & Lewis, 2011; Hills et al., 2013), expertise also leads to the development of a more configural processing (Chance, Goldstein, & McBride, 1975; K. J. Hancock & Rhodes, 2008; Sadozai et al., 2018).

2.4.2 Social contact: Importance of contact quality

Perceptual contact might, however, not be enough to improve other-group face recognition performance, and it is difficult to separate perceptual from social contact since the latter implies the former in a non-mutual way. Social contact implies interaction with other-group members, and this interaction leads to a deeper processing of faces, as implied by individuation. However, studies on the relationship between contact and the OGB are not consistent (see K. J. Hancock & Rhodes, 2008; W.-J. Ng & Lindsay, 1994) and contact actually explains only 2% of the variance in the OGB (Brigham et al., 2007; Meissner & Brigham, 2001). Social contact might have a beneficial effect on the OGB due to the individuation processing it implies (Levin, 2000), and some authors distinguish different levels of contact, differentiating contact with friends from contact with classmates or with neighbours (Brigham, 1993; McKone et al., 2019; Pettigrew, 1997). In this conception, contact with friends, being the deeper contact, is positive and requires individuation. Contact with classmates only requires individuation, and contact with neighbours can be limited to perceptual contact without interaction or individuation. In their study, McKone et al. (2019) observed that the level of contact did not have any effect on the presence or not of the OGB, while the age of the contact is highly related to it, revealing a high correlation between the amount of contact before 12 years old and the magnitude of the OGB. Their study provided no support for the high-quality contact theory, but a strong support toward an early contact, concluding that mere exposure to other-group faces and high-quality contact have a greater effectiveness during childhood but not during adulthood. This idea was not supported by Lavrakas, Buri, and Mayzner (1976) in which a relationship was found between discrimination performance for other-group faces when participants reported having friends from the other-group, rather than being exposed to other-group faces in the neighbourhood or through attendance at a mixed school. Results regarding the relationship between recognition performance and contact are not always in the same direction, however, and this might be a result of the measurement used or of the difficulty in measuring contact. Indeed, contact can be measured through questionnaires, but can also be implied and retrieved from demographic data (i.e., diversity or segregation), the latter being more useful to measure perceptual contact while the former is more useful in measuring social contact.

The implication of attitudes (i.e., prejudice) is also important when studying social contact. Although studies show correlations between higher prejudice and poorer other-group face recognition (Ma, Yang, & Han, 2011; Walker & Hewstone, 2008, for a review on contact, Pettigrew and Tropp, 2006) found that 94% of the sample analysed showed a negative correlation between contact and prejudice, suggesting that the more contact one has with other-group individuals, the less prejudice they have against these individuals. Prejudice therefore appears to be a moderating variable of contact with other-group individuals (Brigham et al., 2007), however, and it is not clear which of the two influences the other, although it is more likely that this relationship is the result of a mutual effect and influence.

2.5 Face individuation versus categorization

While encountering a new face, an observer tends to categorize or individuate it. Categorization is considering a face as being a member of a social category rather than as a unique identity (Wilson et al., 2016). In contrast, individuation refers to noticing the unique features of individual faces, and is a consequence of a perceptual expertise, suggesting that the observer is able to discriminate within a group and to distinguish one member from another (McGugin et al., 2011). Individuating a face is likely to produce a greater recognition accuracy, whereas categorization disrupts the attention that would be paid to individual features and undermines the processes which lead to individuation, therefore resulting in a greater recognition difficulty (MacLin & Malpass, 2003). Other-group faces are more likely to be categorized as a member of a group rather than being individuated, and their recognition is impaired. Few theories have tried to explain what would lead an observer to individuate or categorize a face, especially an other-group face.

2.5.1 A perceptual categorization: The feature-selection model

Levin (1996, 2000) proposed the feature-selection model as an alternative to the MDFS (see section 1.2). This model postulates that 'race' (i.e., group) is considered as a feature in itself because of its salience, and therefore results in an automatic categorization of other-group faces. That is, because group membership is automatically detected from an identifiable and physical feature (i.e., skin color), it acts as an anchor to categorization. This process leads to an absence, or a significant reduction, of individuation and impairs face recognition. Levin also suggested that it is more useful to individuate a racial majority group, and less useful to rely on individual differences for members from a racial minority group, since group is coded at the expense of individual information. This is more important when the majority group is one's own-group. Other-group faces have to be in the minority to observe the presence of the group-feature encoding, resulting in the encoding of the group for other-group faces but not for own-group faces. On that point, Levin suggested that for any observer, group is coded as either other-group or not other-group, because group is not relevant for the coding of own-group faces: therefore, coded as the presence or absence of the other-group specificity, other-group faces display this feature and own-group faces do not display it, which leaves the possibility of finding individual features to encode own-group faces.

2.5.2 A social categorization: The ingroup/outgroup model

Similarly to the feature-selection model, and consistent with the MDFS, the in-group/out-group model (IOM) from Sporer (2001) is another categorization model. When encountering an own-group face, the observer focuses on relevant and useful dimensions for individuation (i.e., features relevant to distinguish one face from another). Conversely, while encountering an other-group face, the face is automatically categorized as a member of an out-group on the basis of either physical or non-physical cues. In terms of the feature selection model (Levin, 1996, 2000), out-group members are categorized and in-group members are individualized. Own-group faces still are better remembered because, first, they already benefit from an individuation due to expertise so group membership might not be as obvious as it is for other-group faces, and then because configural processing is done. However, noticing an out-group membership cue while encountering an own-group face leads to categorization. With the present model, Sporer bears a new social implication of group belonging which can be different from a physical feature. The out-group cue is salient, and an observer might focus on it at the expense of more relevant dimensions common to out-group but different from in-group faces. According to the present theory, an other-group face is not looked at carefully and an observer does not pay attention to the correct or critical dimensions since the out-group membership cue has been encoded instead. Other-group faces are therefore perceived as more homogeneous than own-group faces, and recognition is impaired.

2.5.3 An integrative model: The Categorization-Individuation Model

On the basis of the distinction between perceptual and social categorization, Hugenberg, Young, Bernstein, and Sacco (2010) developed the categorization-individuation Model (CIM). The model involves the inclusion of, and inter-relation between, three phenomena: categorization, experience, and motivation. *Categorization* refers to the activation of a social category rather than individuation of members of a category, which results in a confusion between the group members. Critically, the authors argue that this activation can be detrimental for both own-group and other-group faces. Categorization in the CIM echoes the above presented models. *Experience* refers to prior exposure/experience in discriminating among other-group faces, related to the development of perceptual expertise, namely the use of identity-diagnostic features in relation to the perceived group. Finally, *motivation* refers to the appearance of a selective attention which defines the relevance of the individuation processing of a face. To induce individuation on this basis, Marsh, Pezdek, and Ozery (2016) conducted a study on bi-cultural Latino-American participants. Priming one of their two cultures, the authors found that participants presented a greater recognition performance for the primed group, regardless of their own or the other ‘racial’ group. The authors suggested that priming did create an in-group and an out-group, which led to the individuation of faces from the primed group, therefore resulting in a better recognition performance. When uni-cultural Americans were primed with Latino culture to create an in-group membership, the authors did not find any effect. This study showed that priming is of interest in removing the OGB only if participants already self-identified with the primed group. A fair suggestion here is that bi-cultural participants have prior experience discriminating among their two cultural groups, which makes individuation possible for both groups.

Motivation is related to the perceived importance and relevance of individuating a face. Therefore, in-group and own-group faces are usually perceived as such, and in-group faces are individuated while out-group and other-group faces are perceived as less important to be remembered and are categorized instead. One can also be motivated to individuate a specific identity, in a specific context. This is usually related to familiarity development and is effective for own-group and other-group faces. In the CIM, observers would be capable of individuation in terms of perceptual processing, but will only do so to individuate faces that are perceived as important or relevant to be individuated or remembered. A study conducted by Hugenberg, Miller, and Claypool (2007) found that motivation to individuate can be induced by a simple instruction. Indeed, while raising awareness of the OGB and asking participants to pay careful attention to faces, especially to other-group faces, they found no OGB (study 1). However, while instructing participants to individuate faces, without mentioning the existence of the OGB, the authors observed an OGB (study 2). They explained that in the second study participants may have tried to individuate regardless of a conscious usage of different and relevant dimensions for both groups of faces. In the first study, they found a lower own-group recognition performance in the group who received instruction than in the control group, so it must be acknowledged that participants might have put in more effort when attempting to memorize other-group faces, and this was at the expense of memorizing own-group faces.

In summary, an observer has to be motivated to individuate other-group faces to protect from automatic categorization. They also have to have prior experience with other-group faces discrimination as being exposed to faces on a regular basis results in a certain experience and ability to discriminate within the exposed faces. This results in the ability to induce individuation, whereas categorization inhibits this (G. Rhodes et al., 2009) and it might not be possible to do so with completely new faces which the observers has never been exposed to. Based on this model, it seems that motivation and individuation can easily be induced, while it seems more difficult to do so for experience. Once perceptual expertise is developed, the OGB might only be a matter of instruction or easy manipulation. It is thus important to focus on the development of this expertise.

Chapter 3

Discrimination and recognition performance improvement as a function of training

Considering the importance of face recognition on a regular basis, and especially the consequences of an other-group face misidentification, a field of research has focused on the understanding and development of face discrimination and recognition learning. The OGB being present from visual face processing and encoding, many studies have developed tasks under the scope of perceptual learning. As with other visual objects, faces can be learned, and expertise could be achieved from the exposure to many faces during practice sessions. As a consequence of expertise acquisition, an observer would give automatic (i.e., quick) and accurate answers during a recognition task with unfamiliar faces, as is done for familiar faces (Young & Burton, 2017). Indeed, perceptual learning through practice would create a generalizable ability to discriminate and recognize faces (Wallraven, Whittingstall, & Bulthoff, 2013). For clarity, in the upcoming section, a distinction is made between research on learning, and research on training. Learning, hereafter, mostly implies learning specific identities that have to be recognised at a later time through practice[5], whereas training refers to the development of specific face processing skills that could afterward be engaged in a task with completely novel faces (i.e., generalization). The former mainly resulted in better comprehension of the development of familiarity and face representation in memory, while the second is anchored in a more applied frame.

---

[5] Although training is a part of learning, the term ‘practice’ would be preferred when reporting studies on learning.

3.6 Face recognition improvement as a function of training

3.6.1 Learning identities from practice

Usually, familiarity is built up from the presentation of an identity under different conditions (Burton, 2013). Considering that familiarity improves face discrimination and recognition performance, many studies have focused on learning. That is, when participants were asked to study pictures of a specific identity from multiple images, these identities profited from an induction of familiarity, resulting in a greater recognition performance (Bindemann & Sandford, 2011; White, Burton, Jenkins, & Kemp, 2014). A study conducted by Menon et al. (2015) suggested that the presentation of two pictures of an identity already resulted in a stronger representation in memory, and therefore in a positive effect on participants' matching performance. Although other studies found a beneficial effect of exposure even to little identity variability (Hussain, Sekuler, & Bennett, 2009; Matthews, Davis, & Mondloch, 2018), an even greater performance is obtained from the presentation of two pictures instead of one picture, and of using different views at encoding and when testing in a memory task (Longmore, Liu, & Young, 2008). Performance has also been found to be higher when identities have been learned from exposure to multiple images of one identity, rather than the repetition of one image (Murphy, Ipser, Gaigg, & Cook, 2015). These results suggest that learning is incremental and that the more variability of a face one is exposed to, the more learning is effective, due to the familiarity increasing (Clutterbuck & Johnston, 2002, 2004, 2005; Dowsett, Sandford, & Burton, 2016; Laurence & Mondloch, 2016). Indeed, the more variable the images of an identity are, the greater discrimination performance is achieved (Matthews et al., 2018). Ritchie and Burton (2017) compared performance after a face-name association learning task involving ten different pictures of identities, either highly variable (i.e., pictures retrieved from the Internet) or less variable (i.e., still images from one interview video sequence). Performance was found to be greater after the presentation of the highly variable rather than less variable faces (experiment 1A). While reproducing this task with the sole modification being the usage of a less variable face in the test phases of both learning conditions, results revealed no differences regardless of the variability of the images they had been trained on (experiment 1B). These results suggest that the highly variable condition built up a better representation of that face so that the presented unseen images are as well recognized as when these pictures were more physically similar to the learned ones.

Interestingly, even when participants were not asked to actively learn faces, participants presented a greater discrimination performance from multiple exposures to an identity (Andrews et al., 2015), therefore suggesting that faces were incidentally learned, and that the identities benefited from a good representation in memory so that the novel image was better recognized than a novel identity. This result is not surprising considering that when one encounters a new person, this person becomes familiar and better recognized without the explicit intention of learning. In fact, regardless to the intention of learning, participants appear to build an average of the exposed faces by extrapolating the diverse information they have encountered, resulting in a good representation in memory (Kramer, Ritchie, & Burton, 2015; Menon et al., 2015).

When it came to the recognition of faces learned within a certain variation but asked to be recognized under another variation, the learning effect seemed to be weaker. An interesting body of research concluded that matching and recognition under different lighting or facial emotions can benefit from learning or training with different poses and face views (Chen & Liu, 2009; Liu, Bhuiyan, Ward, & Sui, 2009). Learning faces under different emotion or lighting conditions, however, did not improve face matching or recognition, suggesting that variations in pose has a greater impact on face processing than emotion or lighting. Another study aimed to test whether showing different emotions or different intensity of an emotion during a face-name association learning task would improve subsequent recognition performance (Liu, Chen, & Ward, 2015). Their participants presented with a slightly lower performance after the presentation of several emotions, rather than one emotion, during learning, but an improved performance when compared to a neutral expression. Hussain et al. (2009) also concluded that learning is direction specific, as it is not transferable from upright to inverted faces, or the reverse. It has to be acknowledged that upright and inverted faces do not imply the same processing, which could explain this latter result. While tested on standardized pictures of identities from different views, participants were not capable of face invariant extraction (Longmore et al., 2008), suggesting that the inclusion of ecological variations, for instance using pictures taken on different days and in different contexts (Kramer et al., 2015; Matthews et al., 2018; Murphy et al., 2015), results in a better representation of faces, leading to a better recognition performance for those identities.

Whereas these studies were conducted on own-group faces, additional studies tested learning on other-group faces as well. Tuttenberg and Wiese (2019) tested learning performance for own-group and other-group faces with Caucasian and East Asian participants from a British university. A sorting task was used for identity learning as it exposed the participants to the presentation of variable images of each identity. Participants were then tested with either a matching task (experiment 1), or a memory task (experiment 2), and these tasks included the presentation of learned and novel faces. Caucasian participants presented an OGB during sorting and matching, even if overall, learnt identities of both groups were better discriminated than new faces. East Asian participants did not perform any better for either group, and did not present a better learning effect for other-group compared to own-group faces. In the second experiment, the results presented the same pattern, with a better recognition for own-group rather than other-group faces for Caucasian participants, but no differences for East Asian participants. This study suggests that both own-group and other-group faces can benefit from learning; however it does not achieve a sufficiently strong representation of other-group faces, to remove the OGB in Caucasian participants. The authors acknowledged that the absence of OGB in East Asian participants could be due to their presence in the UK for quite some time. Additional studies demonstrated that identity learning worked on own-group faces, but less so on other-group faces. Proietti, Laurence, Matthews, Zhou, and Mondloch (2019) asked Caucasian participants to either actively learn (i.e., active matching task) or passively learn (i.e., no action needed from participants) own-group and other-group faces. Tested with an old/new recognition task afterward, participants demonstrated an overall higher performance after the active learning but not after the passive learning, but without any modification of the OGB present in either group. Hayward et al. (2017) tested whether the OGB would decrease after an individuating face-name association task using multiple - and variable - pictures of each identity. Caucasian and Chinese participants, from Australia and Hong-Kong respectively, took part in the study. In line with previous findings, participants presented with a better performance for own-group rather than other-group faces during the test phase, especially Caucasian participants, while it does not seem to be present for Chinese participants.

Cavazos, Noyes, and OToole (2019) asked Caucasian and East Asian participants in America to complete either contiguous or distributed face learning from multiple and variable images (experiment 1) or a single but repeated image (experiment 2) of each identity. The contiguous learning consisted of the presentation of the four face images of each identity, one after the other, while distributed learning consisted of the presentation of the face images of all the identities intermixed with one another. Learned identities were presented along with new identities in an old/new recognition task after learning. Results revealed an interaction effect between participant group and stimulus group in both experiments and in both manipulated conditions, suggesting an OGB. A main effect of learning type with a greater performance after the distributed rather than the contiguous learning was found in experiment 2, but not in experiment 1. While comparing the two experiments, the authors observed a greater performance for participants who learned identities from multiple images than from the repetition of a single image. The authors suggested that the lack of effect from the distributed learning, where several images of each identity were presented, might be explained by the fact that these identities have not been perceived as belonging to one identity (i.e., following the idea of 'telling people together'). In addition, a significantly better recognition of learned faces, rather than novel faces, was observed whether participants were presented with an identical image of one identity multiple times, or with multiple images per identity; this has also been found after training involving an individuation task where the presentation of multiple images resulted in a higher accuracy rate than when exposed to a single image (Matthews & Mondloch, 2018). Finally, to induce familiarity, McKone, Brewer, MacPherson, Rhodes, and Hayward (2007, experiment 3) asked Caucasian participants to learn four Asian faces, presented as being 'friends', through multiple exposure during a label-face association learning task. They observed an OGB at the beginning of the task, which then decreased as familiarity increased with learning, resulting in an almost perfect recognition performance of the learned faces at the end of the learning trial and with no difference between own-group and other-group faces. In addition, the authors found that when manipulating upright and inverted faces, participants relied on configural processing for other-group faces after the training. The elimination of the OGB is as a function of learning, and without familiarity induction, other studies found an OGB unrelated to the mobilized visual processing (Sadozai et al., 2018).

These studies, which show that one is capable of learning face identities, are not surprising. Indeed, any new encountered face has the potential to become familiar through exposure, as seen in how friends and colleagues faces become familiar despite no conscious intention to be learned. Moreover, learning identities is efficient for increasing face discrimination and recognition for those faces; however, it usually does not reduce the OGB, except when the creation of familiarity (i.e., calling identities 'friends') has been clearly stated (McKone et al., 2007). One could therefore imagine that training participants solely on other-group faces would lead to an increase of the related performance, which would become similar to the performance found for untrained own-group participants, resulting in an absence of an OGB. This is a very important consideration, and training should include both own-group and other-group faces to ensure that all changes are correctly studied and understood, and to attribute the effect of learning to an overall face recognition performance increase or a reduction of the OGB. These studies, however, omitted an important aspect of face processing: generalization. That is, face learning as it has been presented so far, does not imply a generalization perceptual learning acquisition, such that one would be able to better recognize novel identities, and not only learned identities. Learning specific faces does not help to develop a better representation of a whole category of faces, and especially for faces from other-groups.

3.6.2 Learning critical features of other-group faces

One explanation for the OGB is that one does not focus on the most critical parts of other-group faces. Some authors developed perceptual redirection as a way to improve discrimination performance by shifting the focus to critical features on other-group faces. Howells (1938) observed higher recognition performance when only the top half of a face remained visible (as opposed to when only the bottom half remained visible), and when participants were told about the importance of the eyes. Although it should be noted that this observed difference was not statistically tested, and that the article does not provide any information about the stimuli and participant groups. Hills and Lewis (2011, study 1) explored the effect of guided attention of White participants on critical features of Black and White faces using a fixation cross directed to either the bottom or the top half of Black and White faces. They concluded that when the fixation cross was located at the bottom of the face, Black faces were better encoded and recognized than White faces. However, a fixation cross located in the top face region resulted in a better performance toward White faces than Black faces. In the second study (Hills & Lewis, 2011, study 2), the added effect of a delay between encoding and recognition was explored with White participants. The addition of a delay resulted in a moderated or negated effect, highlighting the short-term nature of the fixation cross effect. Finally, they observed that when directing observers' attention to the bottom half of Black faces, the OGB was reduced. Hills et al. (2013) extended these findings to Black observers. Using the same methodology with top and bottom fixation crosses, they confirmed the previous findings;

namely a better performance at recognizing White faces than Black faces after a fixation cross was located in the top half of a face, while performance was better with Black faces than for White faces after a fixation cross was located in the bottom half of a face. Crucially, they found the same pattern for Black participants, therefore resulting in the elimination of the OGB. Another study asked whether caricature would help participants to focus on critical features (Rodriguez, Bortfeld, & Gutierrez-Osuna, 2008). In this study, Caucasian participants were asked to recognize East Asian and Indian caricatured faces (i.e., face for which distinctive features are exaggerated) during an old/new recognition task. Results revealed that caricatures resulted in a higher recognition performance for Indian faces than the control group, while no differences were observed for own-group faces or East Asian faces. Also, an OGB was present toward Indian faces but not toward East Asian faces in the control group, while the OGB was no longer present for Indian faces but did appear for Asians in the caricature group. The authors explained the differences in discrimination performance for the two other-groups in terms of prior exposure to them.

These studies reveal that focusing on critical features does have an effect in improving discrimination performance for other-group faces; however, it does not have any effect for own-group faces. This result is not surprising considering that one already knows what the critical features of one's own-group are, and already directs attention to these without any intervention. The above literature has shown that it is possible to increase familiarity in order to improve discrimination performance, and that instructions/tasks can develop a specific way of encoding other-group faces that is beneficial for face discrimination. However, none of these tasks clarify whether there will be a generalization of this specific ability, and if it can be obtained without directing attention using a fixation cross for example. It is thus interesting to explore the contribution of training on the development of skills to improve face recognition.

3.7 Training as a function of face recognition improvement

As specified at the beginning of the present chapter, I consider training as it implies the completion of an identity-independent task, which aims to develop discrimination and recognition abilities that are generalizable. This implies that a post-test with novel faces would capture the effect of training, testing the generalizability of the skills acquired from training.

3.7.1 Own-group face discrimination and recognition generalization

Some authors have tried to improve own-group face recognition using training. For instance, a study conducted by Wallraven et al. (2013) asked participants to complete a learning trial of different faces and then complete an old/new recognition task with new faces over six sessions. The first session included a pre-training task to measure the recognition performance baseline. The second and third sessions included training during which participants learned faces for a subsequent identification in an old/new recognition task that combined learnt faces with new faces to test generalization. During the learning phase, participants were restricted in their visual movement. The final three sessions were for the post-training measurements, which involved a test of learned faces, a generalization test, and a persistence test several days later. Results showed that performance improved with training and skills were generalized to novel faces, suggesting that participants learned how to learn faces in this specific gaze restricted tasks. It remains, however, although these findings are important, the training only improved recognition while face exploration was restricted; one therefore cannot draw conclusions regarding the effect of such training for unconstrained visual exploration.

Paterson et al. (2017) trained participants to either focus on different internal, external features of a face, or both (i.e., or freely; study 2). Participants were trained to focus on these features while encountering new faces, as instructed. Their visual exploration was recorded using an eye-tracking device. Participants completed lineup identifications before and after the training, during which eye movements were also recorded. Eye-tracking data confirmed that after training participants did focus on the facial features they had been trained to focus on. When external features remained unchanged between encoding and recognition of post-task (study 2), participants in the internal feature training group performed worse than the other two groups. However, when external features were modified, results showed a better performance after training on internal features (study 1), but these results were not replicated under identical conditions in study 2. It remains interesting that internal features seem to be more important than external features, and are helpful when recognizing faces whose external features have been modified. Also, in the present study, participants and stimuli were either Caucasian or Asian, but there is no mention of the OGB, nor a presentation of differences in the performance of both participant groups or stimuli groups. Even though the OGB is usually the result of an interaction and one can assume that results were confounded, the authors did not acknowledge this as a potential limitation in their study. An interesting point of this study is that the visual exploration during pre-training and post-training remained free (as opposition to the usage of fixation cross, see Hills & Lewis, 2011; Hills et al., 2013). Unrestricted gaze during face exploration has been shown to be better for facial recognition, as restriction is detrimental to face learning and recognition, while free eye-movements facilitates face learning (Henderson et al., 2005). Free exploration of faces is important when one considers that even if faces have similar features (i.e., ‘first-order relational properties’), every face is original and might have relative distinctiveness located in different parts of faces. In addition, with an unrestricted exploration, it remains possible to use configural processing, while this is more difficult, even impossible, with a restricted exploration.

On another point, Woodhead, Baddeley, and Simmonds (1979) evaluated the effect of training provided to professionals to enhance face recognition. The training was organized independently of the authors, and few details are given about the content of this training, other than that "*a typical instruction was to exaggerate a feature or shape in the mind, like a cartoonist*" (p. 335) and the authors were 'convinced' by its design. They tested the performance of the participants before and after training, and found no evidence of a better performance after training. In the same manner, Dolzycka, Herzmann, Sommer, and Wilhelm (2014) trained participants to complete different face cognition tasks (e.g., face perception, face memory, face speed processing), which are considered as global face processing learning tasks. Although participants presented a successful recognition of faces they had been trained on, they presented no effect of the training on their face recognition accuracy in the generalization task.

Whereas training own-group face discrimination and recognition seems to be difficult, there is a larger scope for improvement for other-group faces. In fact, it might be possible to train people so that they perform as good for other-group faces as for own-group faces when performance for the latter remains unchanged.

3.7.2 Other-group face discrimination and recognition generalization

Few studies have sought to use training to develop specific abilities to discriminate faces that could then be generalizable to new faces, with an overall intention to improve other-group faces discrimination and recognition. Since this is the aim of the present thesis, inclusion criteria have been drawn up to select past studies which attempted to achieve this purpose as well.

The first criterion is that only *studies on adults* are included. Therefore, two studies which met the other criteria were excluded. The first study, conducted by Heron-Delaney et al. (2011), aimed to train 6-month old Caucasian infants on Caucasian and Asian adult faces for three months, using a book that had to be presented to them according to a specific schedule. After training, the now 9-months infants did not present any OGB from the presentation of novel faces, while untrained 9-months infants did. Infants can thus be protected from the appearance of the OGB, if frequently exposed to other-group faces. Another study conducted by Brigham, Bennett and Butz (2005, cited in Brigham, 2008) tested college students performance after a training task based on the famous ‘memory game’ (i.e., a task implying individuation). Whereas they did not test the OGB prior to training, it was present after training on other-group faces, but not when trained on own-group faces. These results are surprising, and the authors suggested that the task required participants to dwell on featural processing more than configural processing, thereby: (1) replacing their automatic configural processing for own-group with featural processing, which resulted in a decrease of own-group face recognition; (2) not creating any modification in other-group face processing since it was already featural. The second inclusion criterion is that studies had to *report training effect*. Malpass et al. (1973, experiment 2) did not meet this criteria and their study was removed because the results of training effect in face recognition performance were not presented and the authors only concluded that "*no effect of training on visual recognition was evident*" (p. 289). Indeed, even if face recognition was tested, they sought to explore the effect of training on face description. The second experiment they conducted in this paper was not included either since they tested the effect of different types of feedback on face recognition, therefore not implying any face processing training. The two additional criteria were that the training task had to be a *visual task* (e.g., discrimination, matching, recognition), and the effect of training had to be tested with novel faces, implying a *generalization* from training.

A total of nine articles are included, collectively reporting 10 studies (Table 3.1) which are hereafter presented and discussed together. The majority of these studies tested individuation training through face-label association (Elliott et al., 1973; Goldstein & Chance, 1985; Matthews & Mondloch, 2018; McGugin et al., 2011; Stahl, 2010; Tanaka & Pierce, 2009), the others tested training through feature modification (Hills & Lewis, 2006; Lavrakas et al., 1976). All included studies used White participants, sometimes along with Black American participants as a control group (Lavrakas et al., 1976) or alongside non-White/Black/Asian participants confounded with White participants (Tanaka & Pierce, 2009). For other-group stimuli, participants were either tested on Black faces (Hills & Lewis, 2006; Lavrakas et al., 1976; Lebrecht et al., 2009; Matthews & Mondloch, 2018; McGugin et al., 2011; Tanaka & Pierce, 2009); Hispanic faces (McGugin et al., 2011; Tanaka & Pierce, 2009), or Asian faces (Elliott et al., 1973; Goldstein & Chance, 1985; Lebrecht et al., 2009; Stahl, 2010). Participants were either tested with one or two other-group(s).

Summary of the characteristics of the studies

Studies used different methods to measure discrimination and recognition performance, and to measure the effect of training. While Elliott et al. (1973) only measured post-test performance, others tested and reported data from before and after training. Some studies used a control group who had no training (Elliott et al., 1973; Goldstein & Chance, 1985; Hills & Lewis, 2006; Lavrakas et al., 1976), while other studies only assessed other-group performance measures and the absence of own-group faces meant they could not measure an OGB directly (Lebrecht et al., 2009; Lavrakas et al., 1976; Matthews & Mondloch, 2018; McGugin et al., 2011). Then, five studies used distributed or longitudinal training involving several sessions (Goldstein & Chance, 1985; Lebrecht et al., 2009; Matthews & Mondloch, 2018; McGugin et al., 2011), and two tested long-term effects (Goldstein & Chance, 1985; Lavrakas et al., 1976). Two studies aimed to test event-related potentials (ERPs) associated with face learning, and also reported OGB measures as a training effect even if this was not their main purpose (Stahl, 2010; Tanaka & Pierce, 2009). Then, the majority of the studies used an old/new recognition task before and after training, sometimes with differences in the encoding phase. Three studies used a standard old/new recognition task with an automatic presentation of some faces during the encoding phase, and the presentation of those faces again along with new faces during the recognition phase (Elliott et al., 1973; Goldstein & Chance, 1985; Lebrecht et al., 2009), while the others added a task during the encoding phase such as rating attractiveness (Hills & Lewis, 2006), or a discrimination task consisting of a sequential matching task (McGugin et al., 2011).

Table 3.1 – *Main characteristics of the 10 included studies in the training comparison, with specification of participants and stimuli group, training tasks content and number/organization of session, type of measurement, and main results.*

	Participant group	**Stimuli group**	**Training task**	**Session(s)**	**Measurement task**	**Main results**
Elliott et al., 1973	66 White American	Caucasian and Asian[6]	Face-Number association Either trained (on Asian or Caucasian faces) or untrained	Unique	Old/new on either Caucasian or Asian, in post-test	Participants trained on other-group faces performed better in post-test than participants trained on own-group faces or left untrained
Lavrakas et al., 1976	42 White American, 6 Black American	Black	Concept detection Either trained (simple; conjunctive) or untrained	Unique	Face detection, in pre-test and post-test + 1 week delay	The two trained groups performed better than the untrained group in the immediate post-test, but not in the delayed post-test
Goldstein & Chance, 1985	8 White American	Japanese, Chinese and White	Face-Number association, either trained or untrained	6 sessions over 2-3 weeks	Old/new recognition task, in pre-test and post-test + 2-4 days, 1 month and 5 month delay	Trained participants performed better with other-group faces than untrained participants in the three post-test measures
Hills & Lewis, 2006	124 White Welsh	White and Black, 2 images per identity	Feature focus task, either trained (critical; not critical; blob) or untrained	Unique	Old/new recognition task, in pre-test and post-test	Trained participants in the critical feature conditions performed better than the two other trained groups and the untrained participants
Lebrecht et al., 2009	20 Caucasian Canadian	African American, Chinese and Caucasian	Face -Label association, either individuation or categorization	5 sessions, every second day	Old/new recognition task, in pre-test and post-test	Participants trained to individuate performed better than participants trained on categorization. All participants presented a better performance in post-test relative to pre-test
Tanaka & Pierce, 2009	19 Caucasian Canadian, 4 Other	African American and Hispanic	Face -Label association, either individuation or categorization	5 sessions within 5 consecutive days	Old/new recognition task, in pre-test and post-test	Participants did not perform better in the individuation than the categorization training, however they all presented a better performance after training
Stahl, 2010	20 Caucasian German	Asian and Caucasian, 3 images per identity	Face-name association using the three views, all trained	5 sessions over 14 days	Old/new recognition task with race categorization during encoding, in pre-test and post-test	Trained participants presented a lower OGB after training; however this was due to a worse performance for own-group faces rather than a better performance for other-group faces
McGugin et al., 2011	39 Caucasian American	African American and Hispanic	Face-Number association, either individuation or visual judgment	3 sessions within a week	Old/new recognition task and discrimination task, in pre-test and post-test	Participants in the individuation training, but not those in the visual judgment training, performed better in the discrimination task but not in the recognition task
Matthews & Mondloch, 2018, exp. 1	20 Caucasian Canadian	Black, 6 images per identity	Face-Number association, all trained	7 sessions, over 2 weeks	Sequential matching task and old/new recognition task, in pre-test and post-test	Training participants performed better after training and trained identities were better recognized than untrained identities
Matthews & Mondloch, 2018, exp. 2	20 Caucasian Canadian	Black	Face-Number association, all trained	6 sessions, over 2 weeks	Sequential matching task and old/new recognition task, in pre-test and post-test	Trained participants did not perform any better after training and learned identities were still better recognized than new identities

In their study, Matthews and Mondloch (2018) used an old/new recognition task to test whether learnt faces during training would be better remembered than novel faces, but assessed a sequential matching task to test for generalization performance. For pre-test and post-test measures, Lavrakas et al. (1976) presented three targets sequentially in the encoding phase, followed by the presentation of six pictures simultaneously, amongst which the targets were presented one by one. Participants always completed identical tasks before and after training, with different stimuli (except for Lebrecht et al., 2009).

Training tasks and results

Elliott et al. (1973) asked their participants to learn pairs of face-number associations, and to then recall the associated number when the face was presented alone. Corrective feedback was given, and participants were trained with either own-group or other-group faces. Untrained participants did not complete any such tasks. A better performance was observed for other-group faces in post-test for participants trained on other-group faces than for participants trained on own-group faces or untrained. Participants trained on own-group faces showed no improvement. These results suggest that training encouraged participants to individuate other-group faces, resulting in a greater representation of this group of faces, whereas this ability already existed for own-group faces, which explained the absence of effect when trained on own-group faces. Using a similar training task, but conducting the six sessions over two to three weeks, Goldstein and Chance's (1985) participants presented with an improved performance over time when tested at 2-4 days, one month, and five months after training. No differences were observed for own-group faces; however, trained participants presented a greater discrimination performance for other-group faces than untrained participants in every post-test measures. Unexpectedly, untrained participants also presented an improvement, suggesting that a repeated exposure had an effect on the reduction of the OGB through the improvement of other-group face recognition over time.

In addition to an individuation condition, and in replacement to a passive control group, two studies used an alternative categorization condition. Caucasian participants were asked to either complete an individuation task as presented in the two studies above, or to complete a categorization task where an identical letter was associated with all members of one group (instead of a different letter for each face as in the individuation condition). Participants completed the individuated task with one group, and the categorization task with the other, both being other-group stimuli (Hispanic or African-American). Participants completed five sessions every second day, and corrective feedback was provided. In one study, despite all participants showing a significantly improved performance after training, participants on the individuation condition presented with a higher recognition performance than participants in the categorization condition, improved over the training sessions, and showed a reduced OGB relative to their OGB before training (Lebrecht et al., 2009). In another study, all participants performed better after training; however, there were only marginal (i.e., not significant) differences between the individuation and categorization training conditions (Tanaka & Pierce, 2009). Despite a similar procedure, results might have been different because of the additional manipulation of implicit bias measures (Lebrecht et al., 2009) which would have heightened the effect of individuation.

As an alternative to categorization, or a control group, McGugin et al. (2011) used an eye-luminance judgment task. This task aimed to induce as much difficulty and learning as individuation, but without any intention to give attention to the presented identities. With this task, participants had to learn which of the two eyes was brighter than the other. Individuation thus was assessed with face-name association learning, whereas participants in the eye-luminance judgment condition were presented pictures with a brighter eye, and the localisation of the brighter eye as a label (i.e., left or right). Participants completed the individuation task on one stimuli group (i.e., Hispanic or African American), and the eye-luminance judgment on the other stimuli group, and completed three sessions within a week. Results revealed that perceptual discrimination, as measured by a delayed matching task, was improved by individuation, but recognition, measured by an old/new recognition task, was not. Eye-luminance judgment had no effect on post-test measures. It was not possible to compare performance pre- and post-training because the study method required the addition of pictures at several stages. Three additional studies trained all participants with an individuation task. In her study, Stahl (2010) presented three pictures (i.e., frontal, three-quarters, profile) of one identity associated with a name for five sessions over 14 days. Participants were either trained to individuate own-group or other-group faces. Results are not detailed in her manuscript, but she concluded that a lower OGB was seen after training; and that is was a consequence of a decrease in performance for own-group faces rather than an increase for other-group faces. Matthews and Mondloch (2018) trained their White participants to individuate Black faces with a face-number association task, either using multiple images per identity (experiment 1), or one image per identity (experiment 2). In their first experiment, participants presented with a greater discrimination performance after training, while there was no difference in experiment 2 where they presented one picture per identity. The authors concluded that recognition performance benefited from multiple exposures to various faces, whereas presenting a unique image many times did not result in any improvement. However, they confirmed an effect of learning such that learned identities were better recognized than novel faces, regardless of their exposure condition (i.e., multiple or single image).

Non-individuation training tasks also revealed an improvement in other-group face recognition. Lavrakas et al. (1976) used a sequential presentation of stimuli during which participants were asked to state if it represented a concept (i.e., followed a visual pattern) and, if yes, to name the concept. Untrained participants were asked to rate the attractiveness of the same stimuli in order to ensure attention to them. Trained participants performed better than untrained participants in the immediate post-training test, but not one week later, suggesting a short-term rather than a long-term effect of training. Hills and Lewis (2006) asked their participants to either focus on critical features (i.e., critical features of Black faces: bottom halves of faces), non critical features (i.e., top halves of faces), or on a color blob placed on the face. A group of participants was also untrained. They observed a strong elimination of the OGB as a function of the critical feature focus training, while no differences were observed for the three other groups between before and after training. From these two studies, it appears that visual exploration is another way to individuate faces, and it is also efficient for improving other-group faces recognition and to reduce the OGB.

Limitations and perspectives

Although these studies offer promising results, they have limitations, and some questions remain unanswered. First, the stimuli used differed across studies, from pictures cut out from newspapers (Elliott et al., 1973), to school yearbooks (Lavrakas et al., 1976), databases (Hills & Lewis, 2006; Lebrecht et al., 2009), to those artificially constructed using specific software (Lavrakas et al., 1976; Hills & Lewis, 2006), to those retrieved from the Internet (Matthews & Mondloch, 2018). However, limitations occur when faces do not seem realistic (Hills & Lewis, 2006), when an identical set of faces is used before and after training (Lebrecht et al., 2009), or when Japanese faces are replaced by Chinese faces in a post-training measure (Goldstein & Chance, 1985). Indeed, measuring the effect of training using identical pictures in the pre- and post-training testing to measure discrimination performance (Lebrecht et al., 2009) is a limitation since results obtained after the training could have been biased by familiarity from the first exposure. Using different Asian groups (i.e., Chinese and Japanese) as one unique other-group for Caucasian participants, Goldstein and Chance (1985) might have induced bias in their 5-month measure, even if the authors justified that these faces can be confounded. Nevertheless, regardless of the origins and the types of pictures, it is important to highlight that when directly measured, every study found an OGB prior to any intervention.

Another limitation comes from the sample sizes. In fact, only four of the studies had more than 20 participants (Table 3.1), and this meant even smaller samples sizes for each condition. Although small numbers of participants might be sufficient to find an effect, it has to be considered that this effect could be exaggerated by the low number of participants. For instance, Goldstein and Chance (1985) had only eight participants, four trained and four untrained. The use of within-subject measures still had value for the overall result, although it might not be sufficient to detect a generalizable effect.

Some studies only used other-group performance measures, and the absence of own-group faces in the method meant they could not directly measure an OGB (Lebrecht et al., 2009; Lavrakas et al., 1976; Matthews & Mondloch, 2018; McGugin et al., 2011). Only testing recognition performance of other-group faces did not undermine the aims of those studies, especially since between group differences and/or pre-test and post-test differences were presented; nevertheless, inclusion of own-group faces would have been useful. When tested with two other-groups, participants did not transfer the individuation process to the untrained group, as they only improved their recognition of the group they had been trained on. These studies suggest that when asked to individuate members of one group during training, participants would still be motivated to individuate this specific group, therefore resulting in a greater recognition performance. In the two studies using individuation versus categorization (Lebrecht et al., 2009; Tanaka & Pierce, 2009), it would have been interesting to add a control group, since previous studies found a difference in individuation relative to a control group (Elliott et al., 1973; Goldstein & Chance, 1985), but not between individuation and categorization. The absence of such a difference is likely due to the fact that both studies (Lebrecht et al., 2009; Tanaka & Pierce, 2009) demonstrated a better performance after training than before, regardless of the type of training (i.e., individuation; categorization). Therefore, a categorization task requires a processing of other-group faces that might be useful - but not sufficient - for impacting other-group faces recognition performance.

According to these studies, training people to individuate members of an other-group results in a generalization effect, namely a better recognition of members of this group after training. Whereas most of these studies used training with labels (e.g., name or number), it is of some interest that individuation and improved performance can be achieved after more visual-based training. From the few studies conducted on training, one would expect a quite quick effect of training, sometimes as a result of a single or very few sessions.

Rationale of the present research

Using both French and South African participants, the present research had the potential to test training effects on different populations, which is even more interesting when one considers the demographic differences in these populations. Indeed, the diversity present in South Africa, in Cape Town, and more especially at UCT (see Figure 1), offers great potential for studying the development of the experience of individuation (Hugenberg et al., 2010), such as through the exposure to lecturers and staff from different groups, to workers from different groups, or more broadly the exposure to other groups members off campus. This diversity is less present in France though, resulting in the majority of the data collection occurring in South Africa. Mere exposure is thus of some interest in the consideration of the overall increase of discrimination and recognition performance. Even if studies already concluded on an effect of contact on the OGB, past work on training revealed that participants presented with a better performance after being exposed to measurement or categorization tasks (Goldstein & Chance, 1985; Lebrecht et al., 2009; Tanaka & Pierce, 2009). Even if contact itself is not sufficient to decrease the OGB, it did still have a visible effect. In the current thesis, I considered that previous studies found that training would improve discrimination and recognition performance, and I therefore designed different training tasks to decrease the OGB, and especially to increase other-group face recognition. It is important to bear in mind that a decrease in the OGB is expected to be a function of the increase of other-group faces discrimination and recognition performance, and not of a decrease in discrimination and recognition performance of own-group faces. Then, if own-group face discrimination and recognition performance does increase, a proportional increase for other-group faces would be expected as well.

As most of the studies whose training focused on individuation gave promising results, and considering that associating semantic information with a face may not be the most efficient training method for a visual task involving unfamiliar faces, I developed training tasks which included different kinds of processing. Indeed, from the findings of Hills and Lewis (2006) on the reduction of the OGB as a function of training, and the observations of an increase of own-group (Paterson et al., 2017) and other-group (Hills et al., 2013; Hills & Lewis, 2011) on the importance of directing the gaze to critical features, I developed training methods that focus on how to identify the critical information and on where to focus visually in order to increase participants' performance. This training was developed for memory tasks, while matching tasks were designed from different observations on the positive effects of multiple exposures of a single identity under different variations (Longmore et al., 2008; Menon et al., 2015). The training tasks designed in the present thesis were therefore based on the development of visual procedural skills that could then be used spontaneously when encountering a face from the other-group that participants have been trained on. Only short-term effects have been tested in the current work, namely that no delay or more than a week was left between training and the post-test OGB measure, usually presented right at the end of training.

Section II

Experimental part

Experimental specifications: Ethics, material and measures

Ethics considerations

All studies presented in the manuscripts received approval from the Ethics committee of University of Cape Town during the proposal presentation on the 2nd of May 2018 (Appendix A). Study 1 also received additional ethics approval from the CERNI (Comité d'éthique pour la Recherche Non Interventionnelle) of the University of Toulouse Jean Jaurès, in April 2019 (Appendix B).

Call for volunteers and compensation

When conducted at the University of Toulouse Jean Jaurès (France), calls for volunteers were spread through social media or by direct approach. In Study 1, participants who were registered in a specific Psychology course received 0.5 points of compensation for their participation. When conducted at the University of Cape Town (South Africa), announcements were spread through the Student Research Participation Program (SRPP) of the Psychology department. The program gives points to students for participating in studies, and certain number of points is required to complete the course. Points allocated through the SRPP program is 1 point per 30 minutes of study, and therefore is different for each study.

Inclusion and exclusion criteria

Common inclusion criteria were: older than 18 years old, and not have any prior participation in another study conducted by the same researcher. Specific exclusion criteria for Study 1 were that participants had to not be wearing glasses or heavy makeup, given the use of an eye-tracking device. Because of the topic of the present thesis, the most important inclusion criterion was based on group belonging information. Indeed, in France White participants only were included while in South Africa, White, Black and Coloured participants were included. To avoid any sensitive situations, this was not specified in any of the announcements, and all participants were included and completed the studies, regardless to their group. The inclusion criterion was only implemented immediately prior to data analysis. An additional inclusion criterion used in South Africa is that participants had to be living in the country for at least two years. This excluded students from international exchanges who could have biased the samples.

Consent and debrief forms

Once participants had arrived at the study collection place or on the online page, suitable explanations about the study were given prior to data collection and a consent form had to be read and completed. Consent forms were very similar across studies, with only a modification of the short description and objective of the study in which they were participating. Consent forms therefore specified: a short explanation of the study, their freedom to leave the study at any time without any consequences, a guarantee of anonymity of the collected information, the aimed usage of the data (publication, presentation) and the data storage plan (Appendix C).

At the end of every study, except study 2, participants were directed to a debriefing document (Appendix D). They were told that it was allowed to ask any question at any time during the study, but that some questions may have to be answered after the study. In study 2, since it was completed during a course, I debriefed participants directly at the end of their participation, and presented the results to them one week after the last session. Participants were told that they were free to request results of the study they participated in, by email, under simple request.

Risks in participating

To our knowledge, no risks were involved in participating in any of the present studies. None of the participants reported any discomfort or negative physical or psychological feelings during or after their participation. If any of the participants would have experienced any discomfort, they were informed who to contact and that discussion was made possible with the researcher, a supervisor or a specific person in the department.

Database used across studies

Since every study conducted in the presented manuscripts often used the same face databases, the present section gives all the needed details about the database content prior to description of stimuli manipulation (i.e., raw pictures prior to any modification for study purposes), detailed in as needed for each of the studies. The database used is specified in the method sections of every study. Only pictures of males were used in the studies. Initial sorting of images were done at the outset, selecting only good quality pictures, to make available as material over the studies. In regard to ethics and anonymity policy, examples of the images in the database will not be given here, except for an example of a young male who gave his explicit consent for this purpose.

Although an own-gender bias exists, it is noteworthy that only pictures of males were used in the present work since the own-group. It has to be acknowledged that even if it could have been of higher validity to use pictures of both genders, for practical reasons (namely variability of the used database), only pictures of males were used.

UCT2007 database

The most consistently used database in this thesis was the one collected in Cape Town shopping malls in 2007, and maintained by Colin G. Tredoux. This database contains photographs of more than 500 pictures of males and females of different groups (Black, White, Coloured, Indian, Asian) and of a wide range of ages. Each person was photographed in various poses (i.e., frontal neutral, frontal casual, three-quarter neutral and profile neutral views, where neutral means no emotional expression). Pictures were taken against a grey uniform background. The present database thus offers very good quality pictures of different people. However, the number of young people, especially the White men, was too small to cover all the needs over the studies, and other databases were added when needed.

UCT2005 database

This database contains pictures from about 2000 pictures of male and female of different groups (Black, White, Coloured, Indian, Asian) and of a wide range of age. They were collected at the University of Cape Town between 1999 and 2005 and maintained by Colin G. Tredoux at the University of Cape Town. Pictures are of lower quality, namely of poorer resolution, no lighting control, and no uniform background. For each person, pictures from neutral frontal and thee-quarter frontal were taken. When used in the studies, lighting of these pictures were modified to better match the main database quality.

UT2J2017 database

Additional stimuli were retrieved from a database of more than about 100 male and female from different group collected and maintained by myself[7] in Toulouse Jean Jaurès under the supervision of Jacques Py and Colin Tredoux. This database contains pictures of males and females from different groups and of a wide range of ages. Every person was taken with frontal (neutral and casual), three-quarter neutral and profile neutral poses. Additional pictures of full body from different angles, videos of gait and voice recording are also a part of the database but were not used in the present work. Pictures were all taken against a uniform green background.

Radboud Face Database (RaFD)

A Dutch database (Langner et al., 2010) of about 67 pictures of White males and females was also used when necessary. This database contains good quality pictures of White males and females from frontal, three-quarter and profile poses, displaying different expressions (neutral, casual, happiness, anger, sadness, surprise, etc.).

---

[7] Acknowledgments go to Malvina Brunet, Sarra Hajji, Agatha Bataille for their help in the project.

Measures and analyses specifications

All data analyses were performed using R statistical programming language (R Core Team, 2013), and with a number of add-on packages, as required, which are declared as needed.

The signal detection theory and other measures

1.1 Signal Detection Theory

In the present work, the OGB is measured using the differences between discrimination performance (or sensitivity) of each stimulus group, according to participants group. That is, the discrimination performance measure from the signal detection theory (SDT; Stanislaw & Todorov, 1999) was used. Since it was problematic to assume that my data follow a normal distribution, the non-parametric measure A' was used to assess discrimination performance. Signal detection theory takes as a key measure the discrimination, for example, of an old face (i.e., previously presented; *signal*), from a new face (i.e., not presented previously; *noise*) in a participant's memory. Responses can either be a hit – accurately saying ‘old’ to an old stimulus, a false alarm – inaccurately saying ‘old’ to a new stimulus, a miss – inaccurately saying ‘new’ to an old stimulus, or correct rejection – accurately saying ‘new’ to a new stimulus. A high value of discrimination performance reveals a good accuracy in discriminating signal from noise and therefore, a good ability to make accurate decisions. However, a low value of A' expresses a confusion between signal and noise. In the present thesis, discrimination performances of two group of stimuli are compared, to explore the presence of a bias toward either of the groups.

Beside discrimination performance, response bias (hereafter B'' for the non-parametric measure of it), reflects two possible behaviors while answering: a liberal or a conservative response bias. The former is expressed by a greater tendency to answer ‘yes’ rather than ‘no’, with a value closer to -1. The later is expressed by a greater tendency to answer ‘no’ rather than ‘yes’, with a value closer to 1. A value close to 0 indicates the absence of any response bias.

1.2 Effect size and thresholds

Cohen’s *d* effect size for within-subject measures was calculated, creating a function using the formula reported by Lakens (2013, formula 10), and sometimes directly with the R *psych* package (Revelle, 2017). Between-subject Cohens’ *d* estimates were computed using the R *esc* package (Lüdecke, 2019a). Threshold alpha for statistical significance was fixed at 5%, and confidence intervals are 95%. All graphs report 95% confidence intervals except otherwise specified.

Specification of analytic models

Across the studies, similar types of statistical model were computed. When analyzing continuous outcome variables, such as discrimination performance, mixed linear models were used. For binary outcome variables, such as accuracy, mixed logistic models were used. Fixed and random main or interaction effects are specified, as needed. Only the best model is presented for each analysis, to keep it easier to read and understand. However, when including more than one predictor variable, several models were tested, reducing from greater to simpler complexity. Analyses of variance (ANOVA) were sometimes used to determine which of two models should be used, and when resulting on a non significant difference, the more theoretically meaningful model was kept. Packages used to compute regression measures were *car* (Fox, Friendly, & Weisberg, 2013), *lme4* (Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017), *lsmeans* (Lenth, 2016) for post-hoc measures and *sjPlot* (Lüdecke, 2019b) for graphics. Discrimination performance and response bias measures were computed using the *psycho* (Makowski, 2018) package. The packages *esc* (Lüdecke, 2019a) and *metafor* (Viechtbauer, 2010) were used to compute meta-analysis coefficients.

Overview of the studies conducted

In the present research work, seven studies were conducted. One of them is presented in the overview, but is not developed further, for obvious reasons, presented later on.

Study 1 was an attempt to erase the OGB in 30 White French participants (22 women, M_{age} = 22.73, SD_{age} = 4.91), by instructing participants to focus on what should be the most critical features of Black faces: the bottom halves of faces. An eye-tracking device was used as a manipulation check to record visual patterns of exploration. Firstly, participants completed an old/new recognition task with both Black and White faces with eye-movement recording. Secondly, they completed a training task during which they had to pair a target picture with its duplication among other faces in which only the bottom halves of faces were modified. Finally, they were asked to complete another old/new recognition task, using different pictures of Black and White faces. Comparing visual patterns of exploration before and after training, participants significantly looked more at the bottom halves of faces after the training, even if they overall still focused more on the top halves than the bottom halves. However, unexpectedly and as a result of training, participants' OGB significantly increased after the training, while being already present before the training. False alarms were impacted in particular, with a significantly higher rate for after-training than before-training, meaning it was higher when participants focused more on the lower halves of faces. The present study concluded that focusing on the bottom halves of faces does not improve overall recognition nor does it reduce the OGB of participants.

Study 2, which was conducted alongside Study 1, was an attempt to use multiple training tasks split over three weeks to reduce the OGB in 11 White French female participants (M_{age} = 20.45, SD_{age} = 0.52). The sample contains only the experimental group, since practical difficulties made control group data unusable. The study was conducted over five weeks, including a baseline measure of the OGB, three weeks of training and another post-training measure of the OGB. In every session, participants were asked to complete an old/new recognition task. Participants were split into two groups, where each group received different instructions during the encoding part of the old/new task according to what they were asked to focus on: internal versus external parts of the faces during the first training session; top versus bottom part of the faces during the second training session; and featural or configural exploration of faces during the third training session. Results revealed the presence of an OGB prior to the training, but no improvement after the training. However, the study faced several issues: absence of a control group, incomplete randomization and incomplete stimuli standardization, and use of only a White group in the experimental sample.

Study 3 is different from the others. Since sets of pictures used to asses the OGB were not counterbalanced in the two previous studies, the presence of the OGB using these sets was tested in an online study. The final samples contained 88 participants in Set 1 (72 women, M_{age} = 22.19, SD_{age} = 5.77) and 93 in Set 2 (M_{age} = 22.58, SD_{age} = 7.26). Results revealed a presence of the OGB in both sets, suggesting that the results observed in the two previous studies were not due to a difference in the difficulty of the set used.

Study 4 was a replication and an improvement of Study 2, and was conducted with 51 South African participants (41 women; 20 Black, 17 White, 14 Coloured[8]; M_{age} = 20.63, SD_{age} = 3.24). The overall method of the study was similar: Five sessions split into two measuring sessions (pre and post-test) and three training sessions. However, the face images were better standardized, a control group was added, participation was individualized (instead of group participation as in Study 2), while randomization and stimuli presentation time were modified. A control group was asked to complete the standard old/new recognition task without any specific instruction, in contrast to the training group. Results revealed that in the pre-test, only White participants presented an OGB. No effect of training was found, resulting in an OGB being present for both trained and untrained White participants. Black participants, surprisingly, presented a better discrimination performance for White rather than Black faces, significant for the trained participants in the post-test. The present results suggest that the training had no positive effect of erasing the OGB in White participants, and even created a better discrimination performance of White faces for Black participants.

Study 'X'[9] was different from the previous ones, as it tried to remove the OGB in matching tasks rather than memory tasks. In this regard, 74 participants[10] completed two pairing tasks and a multiple view matching training task. The pairing task, completed in pre-test and post-test, presented pictures of identities that were either display grey scaled *and* in color

---

[8] These categories were constructed on the basis of the usage of a skin gradient phototype scale (Fitzpatrick, 1988), as for Study 2. However, in Study 2 where only White participants (being the majority) were kept, Black, White and Coloured populations were expected to be the studied groups in the present study.

[9] Since the present study is not developed in the thesis and in order to keep a logical track, it does not have any number.

[10] Descriptive data available for only 45 South African participants: 13 Black, 15 White, 17 Coloured, M_{age} = 21.13, SD_{age} = 2.03. In the present study as for the following ones, categories were self-declared by the participant. Participants self-declared as Asian, Indian or Other were excluded from the sample.

(i.e., pair) or grey scaled *or* in color (i.e., unique exemplar). The training task consisted of a multiple-view matching task first from a frontal view to a three-quarter view followed by a three-quarter view to a profile view, which increased in difficulty (i.e., similarity). The control group was asked to play a game and answer to questions about it. In the post-test, participants were split into three groups: those completing an immediate measure, a delayed measure, or a transferred measure. The first two groups were asked to complete the same task as in the pre-test (i.e. OGB measures, but with different pictures), while the third one was asked to complete a field task, which involved looking for two targets in the campus library. A number of issues were found in this study, which lead to the conclusion that it was impossible to analyze the data: a very small sample size, especially when broken down by condition and group, the design which might have been better as separate studies and in the first task there was a clear ceiling recognition effect in pre-test (minimum accuracy = 96%). This study is therefore not reported in the present thesis. However, the next two studies were designed to resolve the problems that have been mentioned, and so it is useful to bear this brief description in mind.

Study 5 aimed to test a matching training task, being the equivalent of the baseline measuring condition from the previous study. In this regard, 140 South African participants (108 women, 40 Black, 69 White, 31 Coloured, M_{age} = 19.30, SD_{age} = 1.49) were asked to complete a standard matching task, followed by the training and a post-test matching task. The two matching tasks were measures of the OGB and were designed as follows: For each pair of pictures (one original, one pixelated), participants had to state if the pictures displayed the same person (i.e., match) or a different person (i.e., mismatch). Training consisted of a matching task using images modified under different levels of pixelation: original resolution, followed by a light degree of pixelation, to a strong degree of pixelation (i.e., the latter being of the same pixelation level than pictures in the pre-test and post-test tasks). The control group completed the same task as in Study ‘X’ (i.e., the game and questions mentioned earlier). Despite a high accuracy rate prior to training, an OGB was found for White participants. In the post-test, trained White participants did not display any OGB while untrained White participants continued to present the bias. Interestingly, the discrimination performance for Black stimuli was significantly higher for the trained White participants than the untrained ones, while no differences were observed for White stimuli. To summarize, training was successful in removing the bias in a task-specific way: Participants were better at pairing pictures with or without pixelation after the training. The next step would be to explore the effect of such training on a different task, in order to explore the transfer of skills developed as a result of the training.

Study 6 was created to address the question of the interest of a training for a field task. The present study was completed by 166 South African participants (143 women, 63 Black, 77 White, 43 Coloured; M_{age} = 19.72, SD_{age} = 2.36) and involved the help of two confederates (one Black, one White) as targets, placed in the library for identification[11]. Participants first completed a matching task similar to the one in the previous study, however matching pictures from two different views: frontal and profile. The training task was the same as the one used in Study 5: Matching a frontal view to a three-quarter view followed by a three-quarter view to a profile view, with increasing difficulty. The control group completed the same game and questions as specified above. In the post-test, participants completed a field detection task: Looking for two targets (i.e., the confederates) in a designated area of the campus library.

---

[11] Two confederates were also needed in the field part of Study ‘X’.

During the field task, participants were asked to look for the people in the pictures, and could look at the pictures as many times as they wanted until they made a decision. In the pre-test, White participants presented once again an OGB. For the first time, an OGB was also present in Black participants. During the field task, an other-group bias was present, however, only for White trained participants. A significant effect of the presence of the target was also found: Participants detected significantly more accurately the targets when they were present, than they detected their absence. Confidence was also related to accuracy of the choice, and Black and White participants were significantly more confident in making their decisions for their own-group target than the other-group target.

Study 1

Training participants to focus on critical facial features does not remove own-group bias

A version of the material reported in the present chapter has been published in Open Access in September 2019 (Wittwer, Tredoux, Py, & Paubel, 2019).

From previous studies, I argue that to better encode a White face for later recognition, an observer should focus more on the top half of a face, and to better recognize a Black face, an observer should focus more on the bottom half of a face, regardless of the group to which the observer belongs. The present study addresses this idea directly. I aimed to decrease the OGB of White participants, presenting perceptual training during which participants were asked to focus on critical discriminating features of Black faces: The bottom halves of the faces. Such a training should encourage participants to develop a deliberative way of looking at faces, and involve paying greater attention to critical and diagnostic features. Participants' awareness of critical features should be positively modified to achieve better discrimination. The increase of discriminating processing resulting from the training should transfer to encoding and thus, recognition processes. Instead of directing attention using a fixation cross (Hills et al., 2013; Hills & Lewis, 2011, study 1; Hills & Pake, 2013, study 2) the training task I constructed was an attempt to induce a spontaneous visual pattern of exploration akin to that done by Hills and Lewis (2006). I used a feature replacement technique within an eigenface software program to create differences only in the bottom halves of faces (i.e., nose, mouth or both) in both Black and White faces. I used eye-tracking recording to establish the effect of training on the modification of visual patterns of exploration, as a manipulation check (Paterson et al., 2017). White faces were included in the present training task as stimuli, in addition to Black faces. The aim was to test if the visual patterns or exploration, while not explicitly directed by a fixation cross, would be modified in favor of an increase of the time spent on the bottom halves of faces. I expected that such an increased focus would be independent to what would spontaneously occur, which I anticipated to be preferential focus on the eyes, as found in previous studies.

1.1 Hypotheses

The aims of this study were thus to (1) explore the effect of attention-focused training on spontaneous visual patterns of exploration, (2) decrease the OGB through training, (3) explore the relationship of modified visual patterns of exploration to potential decreases in the OGB. Since I used White participants, I expected to see an initial visual pattern of exploration mainly focused on the top rather than on the bottom halves of the face for both own and other-group faces since that is what they would ordinarily do, when faced with face stimuli, usually White. Secondly, I expected participants to focus more on the bottom halves than on the top halves of faces as a function of the training task. Finally, I expected an elimination of the OGB after training, as a function of changed visual patterns of exploration.

1.2 Method

1.2.1 Population

Required sample size was computed with GPower 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009). Hills and Lewis (2006) reported a very large effect size for removal of the OGB (d = 4.2), which implied a sample size of 4 for my study, but an alternate calculation based on a mean difference and MSE reported on p.1000 of their study resulted in an effect size of d = .80, and with α = .05, power = .80, and the correlation among the repeated measures conservatively at 0. With these parameters, I computed that I required 19 participants but I over-sampled because I was not sure about the correlation between the measures over time, and was not entirely convinced by the effect size estimated by Hills and Lewis (2006). I thus targeted 30 participants. I recruited 39 participants (30 Women, M_{age} = 22.77, SD_{age} = 4.99) for the study, but keeping only White participants in the analysis itself, the final sample included 30 White participants (22 Women, M_{age} = 22.73, SD_{age} = 4.91). Participants were recruited on the campus of the University of Toulouse Jean Jaurès by direct interaction or through social media. I was interested in White participants only, however, this information was not specified during the recruitment according to ethics considerations. All participants had normal or corrected-to-normal vision. The exclusion criteria of not wearing glasses or heavy make-up were made explicit during recruitment, since their presence makes eye-tracker calibration very onerous. One-third of the participants received course credit for their participation while the rest participated voluntarily, without explicit reward.

1.2.2 Design

This study had a two variables factorial within-subject design: the OGB measure (before; after the training) and stimulus group (own; other-group).

1.2.3 Material

Stimuli

A first sorting from UCT2007 database was made to select photographs of suitable quality (i.e., that had clearly neutral facial expressions, whose eyes were not closed, and whose frontal or three-quarter views were well standardized). From these, photographs of 140 different young males with neutral expressions (70 White and 70 Black – hereafter, respectively, referred as own and other-group) were randomly chosen: 40 for the pre-training task, 60 for the training task, and 40 for the post-training task. In both pre- and post-training tasks, 20 faces served as targets while 20 served as foils. Targets were presented from a frontal view during the encoding phase and from three-quarter view during the recognition phase, along with three-quarter view foils. The use of alternate views at encoding and recognition was intended to minimize picture recognition, and thus constitute a test of face recognition rather than picture recognition (Bruce, 1982).

For the training task, 60 trials (30 photographs of each group) were constructed. In each trial, six derivations of an original picture were generated using a face synthesis program, among which one was randomly designated as the target (ID; Tredoux, Nunez, Oxtoby, & Prag, 2006). Synthetic faces are typically created from statistical models of real face images, and the software I used allowed feature replacement/modification holistically, through statistical sub-models of features. A trial thus consisted of the presentation of six photographs in an array: the target picture alongside the five other derived images. An image of the target identical to that in the array was presented next to the array, indicating the picture participants were asked to search for in the array. For each trial, the nose (*n* = 20), the mouth (*n* = 20), or both features (*n* = 20) were modified, to constitute the derivations.

Apparatus

Feature derivation photographs of the training task were generated with ID software (Tredoux et al., 2006), and controlled to be realistic and uniform so that the target did not stand out from the other array members. Stimuli for the pre-training and post-training tasks were modified using GIMP 2.8.14 software (GNU Image Manipulation) as 1270x720 pixels. The training task was displayed on a 21" Screen, with E-prime 3.0 software (Psychology Software Tools, Pittsburgh, PA). The tasks measuring the OGB were displayed on the same screen through Experiment Center 3.6 (SMI, Teltow, Germany) and eye-movements were recorded using a SMI RED250 mobile eye-tracker (SMI, Teltow, Germany) installed under the computer screen at 60 centimeters from the participant. The lighting in the room was identical over the sessions. The sampling rate was set at 250 Hz frequency. The calibration was effected prior to each of the OGB measurement tasks, using a 5-point calibration procedure.

1.2.4 Procedure

The study was presented in three phases: two old/new recognition tasks were used as OGB measures, during which eye movements were also recorded, and a training task without eye-movement recording (Figure 1.1). In each of the old/new recognition tasks, 20 stimuli (10 own and 10 other-group) were presented for 3 seconds, one second apart, in the encoding phase. A fixation cross, located in the middle of the screen was presented in-between two stimuli: approximately on the bridge of the nose for frontal stimuli and on the cheekbone for three-quarter stimuli. After a 5-minute filler task (word puzzle completion), the 20 previously seen stimuli were presented again, interleaved with 20 new stimuli, in the recognition phase. Participants had to decide for each stimulus, if it had been presented or not during the first phase. No time pressure was applied. After the first old/new task (i.e., pre-training), the participants completed the training task. First, to induce motivation, they were told:

> Research shows that focusing on the bottom part of a face (nose, mouth, cheeks, shape and volume) improves face recognition. Indeed, the bottom part of a face is important for its global configuration and memory representation. You will now be asked to complete a training task to help you focus on that part of a face.

Participants then completed 60 training trials: for each trial, they had to decide which of the six faces in the array matched the target. They were told that they would see several blocks of faces (i.e., trials), and that in each trial, a target face and six faces would be displayed. Their goal was to find, among the six faces, the one identical to the target. They were also told that their answer would be collected from the numeric keypad on the keyboard, and corrective feedback would be displayed. After feedback, participants were required to press the space bar to continue with the next trial. Decision time was not restricted and pictures location within the array were randomized over the study. They were then told how to use the numeric keypad to record their answers, and started the task after an example. After each choice, a feedback popped-up as ‘correct’ or ‘incorrect’ with the incorrect face enclosed in a red rectangle or the

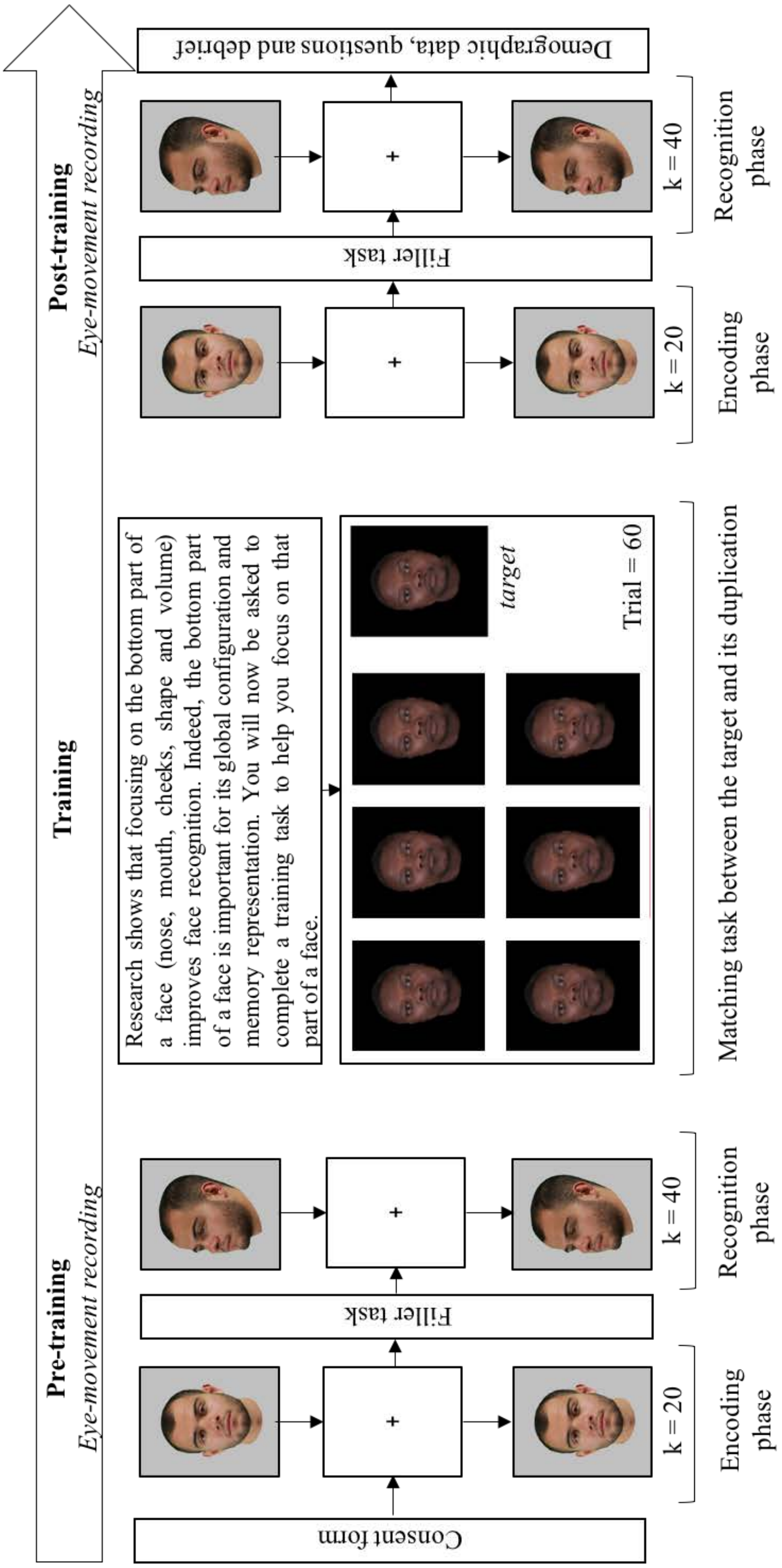

Figure 1.1 – Flow chart of the procedure of Study 1. Every participant completed the three tasks, all task used equally faces of Black and White people. Pictures in pre and post training were not used during the study, but presented here for illustration purpose with the consent of the model. Picture in training are used in the study.

correct face enclosed in a blue rectangle, where appropriate. In the case of an incorrect answer, the selected face was enclosed in a red rectangle and the right answer was displayed enclosed in a blue rectangle, and displayed simultaneously. Once all 60 trials were completed, participants completed the second old/new task (i.e., post-training) with eye-movement recording, and finally answered some demographic questions. The entire experiment lasted between 45 and 90 minutes, and eye-tracker calibration was effected twice in the session: once before each old/new task. At the end of the tasks, participants were asked if they have already seen one of the stimuli faces prior to the present study, and were given access to the debriefing form that they could read. They were free to ask me any questions they may have had.

1.2.5 Measures

In order to analyse the data, two areas of interest (AOIs) were defined for each face: top half versus bottom half (Figure 1.2). Dwell time (i.e., cumulated time of fixations) was used to express the amount of time spent on each half of the face.

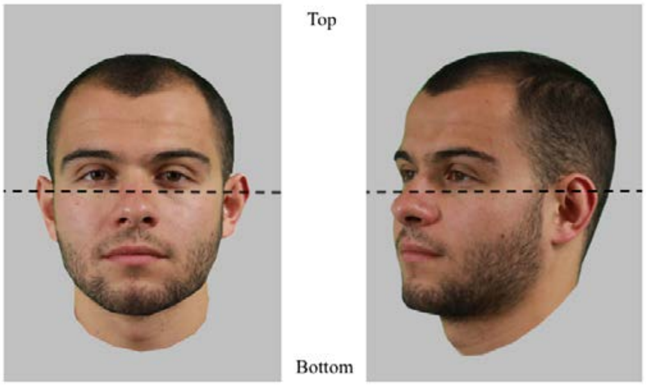

Figure 1.2 – The two AOIs (Areas of Interest) used to assess eye movement patterns, demonstrated on frontal and three-quarter views. The coordinate of the separation is situated as close to the eyes as possible and was defined for each face, individually. Pictures were not used during the study, but presented here for illustration purpose with the consent of the model.

Table 1.1 – *Mixed linear regression coefficient table, participant as random effect, and dwell time as dependent variable.*

	b	df	Std error	*t*	*p*
Intercept	2.20	200	.04	60.02	**<.001**
Top/Bottom	1.09	182	.05	21.81	**<.001**
Own/Other	.17	182	.05	3.41	**<.001**
Pre/Post	.70	182	.05	14.04	**<.001**
Top/Bottom X Own/Other	-.18	182	.07	-2.56	**.010**
Top/Bottom X Pre/Post	-.93	182	.07	-13.16	**<.001**
Own/Other X Pre/Post	-.09	182	.07	-1.28	.202
Top/Bottom X Own/Other X Pre/Post	.08	182	.10	0.76	.446

Note. Significant *p*-values are in **bold**

1.3 Results

1.3.1 Visual pattern of exploration - manipulation check

Eye-movement recording failed for three participants, thus the final sample for eye-tracking analyses was 27 participants (6 men, M_{age} = 22.78; SD_{age} = 5.18). A normality distribution check showed that the dwell time data was normally distributed after the training, but not before. I thus used a logarithmic transformation for both pre- and post-training data to normalize, and to allow comparisons. A mixed linear regression was run with participant as a random effect and results showed participants spontaneously looking significantly more at the top (M = 3.29, SD= .10) than the bottom (M = 2.29, SD = .27) halves of faces before the training regardless of stimulus group, as expected. Then, also as expected, time spent on the bottom half of faces significantly increased from before training (M = 2.29, SD = .27) to after training(M = 2.95, SD = .21) for both own and other-group faces (β = -.93, $t(182)$ = -13.16, p < .001, d = 2.75; Table 1.1).

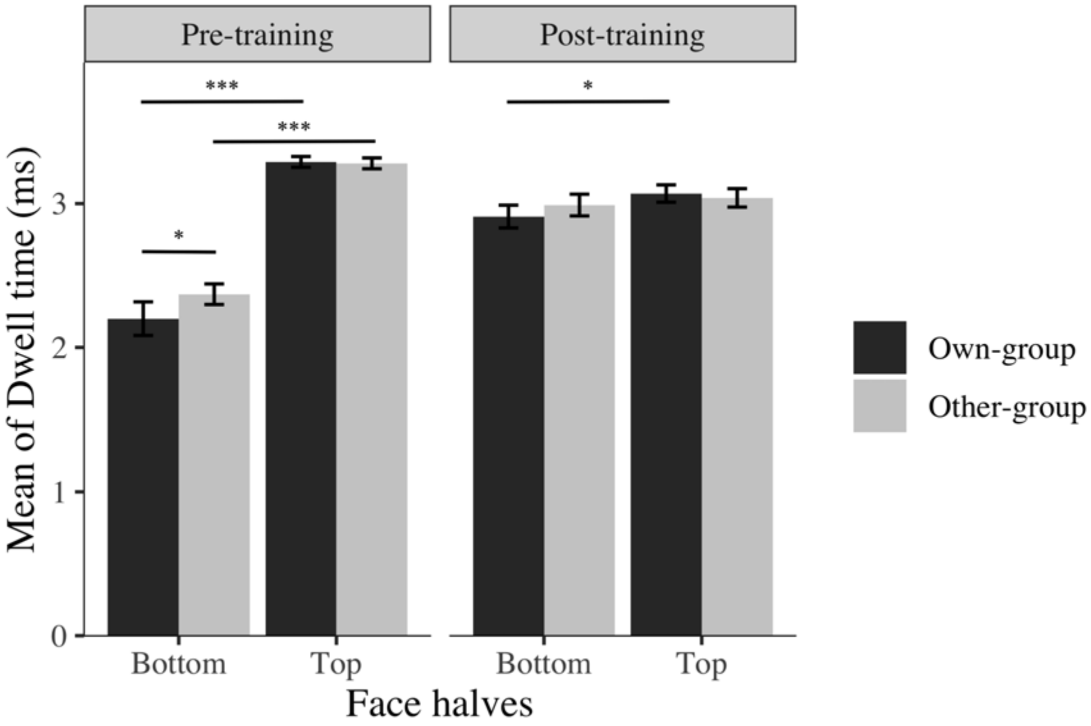

Figure 1.3 – Average dwell time on top and bottom halves of faces across stimulus group (own; other group) and time (pre; post-training). I bars are 95% confidence intervals.
***$p < .001$; *$p < .05$

The training worked as expected, participants focused more on the bottom halves of face after training compared to before the training, even though they still focus on the top halves of faces more or equally as much as they focused on the bottom halves of the face (Figure 1.3).

As exploratory analysis, I explored the location of the first fixation and the time to first fixation (TTFF) on the bottom halves of faces. A mixed linear model taking proportion of first fixations located on the bottom halves of faces as dependent variable, and stimulus group and time (pre-training, post-training) as fixed effects, with participants as a random effect, was run. Results showed a significant effect of stimulus group (β = -.13, $t(82)$ = -3.95, p < .001), and of time (β =-.37, $t(82)$ = -11.30, p <.001) and of their interaction (β =.13, $t(82)$ = 2.72, p =

Table 1.2 – *Mean ratio of first fixation (FF) located to the bottom halves of faces and log transformed mean time to first fixation (TTFF) on the bottom halves of faces by stimulus group (own; other) and time (pre; post).*

	Pre-training		Post-training	
	Own-group	Other-group	Own-group	Other-group
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
FF on bottom halves	.02 (.03)	.03 (.04)	.26 (.19)	.39 (.24)
TTFF on bottom halves	3.11 (.16)	3.13 (.15)	2.69 (.15)	2.52 (.35)

.008). *Post-hoc* analyses revealed a significantly higher proportion of first fixations located on the bottom halves of faces after than before the training for both own-group faces (β = .24, $t(88) = 7.25$, $p < .001$, $d = 2.18$) and other-group faces (β = .37, $t(82) = 11.30$, $p < .001$, d = 2.57; see Table 1.2 for descriptive data). There were no differences between stimulus groups before the training (β = .01, $t(82) = .14$, $p = .887$, $d = .29$) while a higher proportion of first fixations was directed to the bottom halves of other-group rather than own-group faces after the training (β = .13, $t(83) = 3.95$, $p < .001$, d =.60).

An additional mixed linear model taking time to first fixation (i.e., time to notice a specific feature from the display of the stimulus) as dependent variable was conducted with stimulus group and time as fixed effect along with participants as random effect. Results show a very similar pattern at the previous analyses: a main effect of stimulus group (β = .16, $t(78) = 3.37$, $p = .001$), of time (β =.59, $t(78) = 12.14$, $p < .001$) and of their interaction (β = -.15, $t(78)$ = -2.15, $p = .035$). *Post-hoc* analyses revealed a significantly quicker time to notice the bottom halves of faces after than before the training, for both own-group faces (β = -.44, $t(78) = -.102$, $p < .001$, $d = 1.90$) and other-group faces (β = -.59, $t(78) = -12.14$, $p < .001$, $d = 2.44$; see Table 1.2 for descriptive data). In addition, time to notice the bottom halves of other-group faces regarding to own-group faces is also significantly quicker after the training (β = -.16, $t(78)$ = -3.37, p = .001, d = .54), but no differences between the two groups prior to training was observed (β = -.02, $t(78)$ = -.33, p = .739, d = .13).

These results, on dwell time, location of first fixations and time to first fixation, suggest that participants direct their first fixation more to the bottom halves of faces, fixate longer on the bottom halves of faces and take less time to notice the bottom halves of faces after than before the training, indicating an effect of training on attentional eye movement patterns. The difference between own and other-group faces, revealing a higher proportion of first fixations, for longer periods of time with a quicker attention directed to the bottom halves of other-group than own-group faces suggests an effect of learning from the training: participants appear to have recognized that it is more relevant to focus on the bottom halves of other-group faces than own-group faces, and consciously adapted their strategy accordingly.

1.3.2 Own-group bias - recognition data

Mixed linear models were tested, with participants as a random effect, to explore recognition performance (A') across time and stimulus group. Results showed a significant main effect of time (β = - .13; $t(90)$ = -3.49, p < .001), stimulus group (β = .09; $t(90)$ = 2.32, p < .001), and an interaction effect (β = .16; $t(90)$ = 2.95, p = .003) on discrimination performance. An interaction effect was also present for B'', namely for the interaction between time and stimulus group (β = .19; $t(90)$ = 2.15, p = .031). *Post-hoc* analyses were performed, and these revealed the presence of an OGB before training, as expected – there was significantly higher discrimination performance for own than other-group stimuli (M_{own} = .75, SD_{own} = .14;

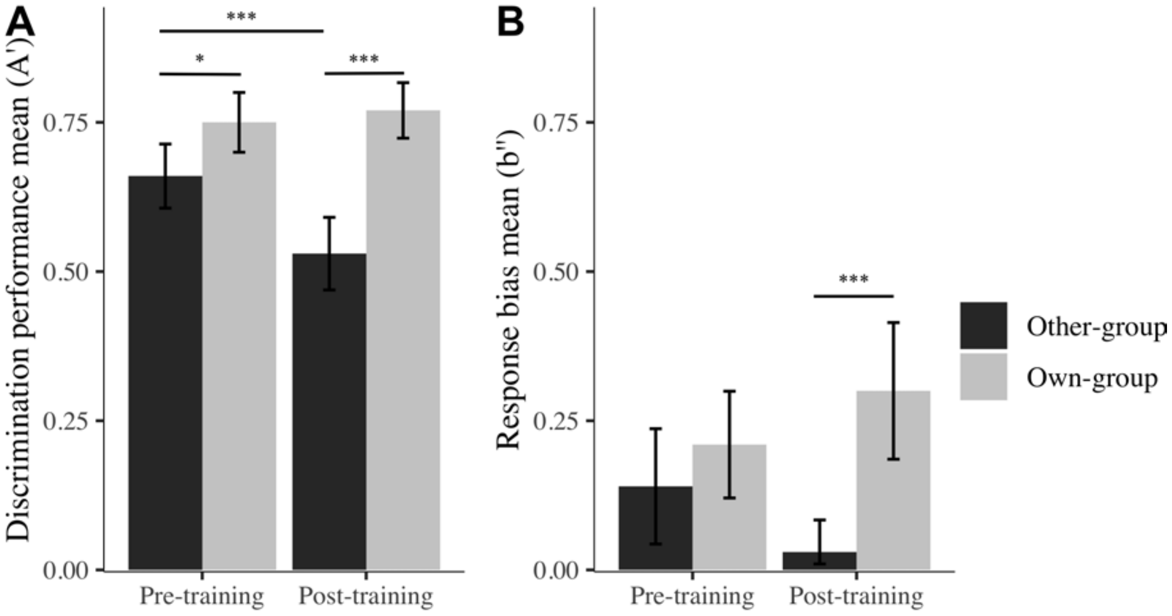

Figure 1.4 – Mean A' (A) and B'' (B) across stimulus group (own; other) and time (pre; post). I bars are 95% confidence intervals.
***$p < .001$; *$p < .05$

M_{other} = .66, SD_{other} = .15; β = -.09; $t(93)$ = -2.29, p = .025, d = .62; Figure 1.4). However, and contrary to my expectations, the OGB became stronger after training, showing a better performance on own rather than other-group stimuli (M_{own} = .77, SD_{own} = .13; M_{other} = .53, SD_{other} = .17; β = -.24; $t(93)$ = -6.39, $p < .001$, d = 1.60). Whereas discrimination performance for own-group stimuli was similar across time (β = -.03; $t(93)$ = .67, p = .505, d = .74), discrimination performance for other-group stimuli decreased significantly from before to after training (β = .13; $t(93)$ = 3.43, $p < .001$, d = .81).

In terms of decision criterion (i.e., B''), a one-sample t-test revealed that participants presented a conservative response bias (i.e., a tendency to answer 'no' more often than 'yes' during the recognition task) for both stimulus groups prior to training (M_{other} = .14, SD_{other} = .27; $t(29)$ = 2.80, p = .010, d = .51; M_{own} = .21, SD_{own} = .25; $t(29)$ = 4.60, $p < .001$, d = .85).

After training, the response bias was still significantly different from zero for own-group faces (M_{own} = .30, SD_{own} = .32; $t(29) = 5.10$, $p < .001$, $d = .93$), while no bias was revealed for other-group faces (M_{other} = .03, SD_{other} = .15; $t(29) = 1.16$, $p = .260$, $d = .21$). The difference in response bias for own and other-group stimuli only became significant after training (Figure 1.4), congruent with the fact that there was no bias toward other-group faces, while there was a strong conservative bias toward own-group faces. These differences were nonetheless not significant on measures taken before and after the training. To summarize, participants' discrimination performance decreased after the training for other-group faces, but at the same time their response bias changed from conservative to unbiased toward other-group faces.

To further explore the potential reasons behind this OGB increase, separate mixed linear models were run for hits, false alarms, correct rejections and misses (percentages). I observed an effect of time ($\beta = -1.43$, $t(90) = -3.21$, $p < .001$), stimulus group ($\beta = .97$; $t(90) = 2.57$, $p = .010$), and their interaction ($\beta = 1.87$; $t(90) = 3.51$, $p < .001$) on correct rejections. However, no effects were found for these factors on hits. I found effects of time ($\beta = 1.43$; $t(90) = 3.88$, $p < .001$), stimulus group ($\beta = -\ .93$; $t(90) = -2.52$, $p = .012$), and their interaction ($\beta = -1.93$, $t(90) = -3.70$, $p < .001$) on false alarms, while observing no effect of any of the factors on misses.

It seems that changes in discrimination performance were not driven by decisions taken when the target was present (i.e., hit or miss) but by decisions taken when the picture was a foil (i.e, correct rejection or false alarm). Indeed, from *post-hoc* exploration (Figure 1.5), I noticed that the difference in A' is due to a significantly higher rate of false alarms after the

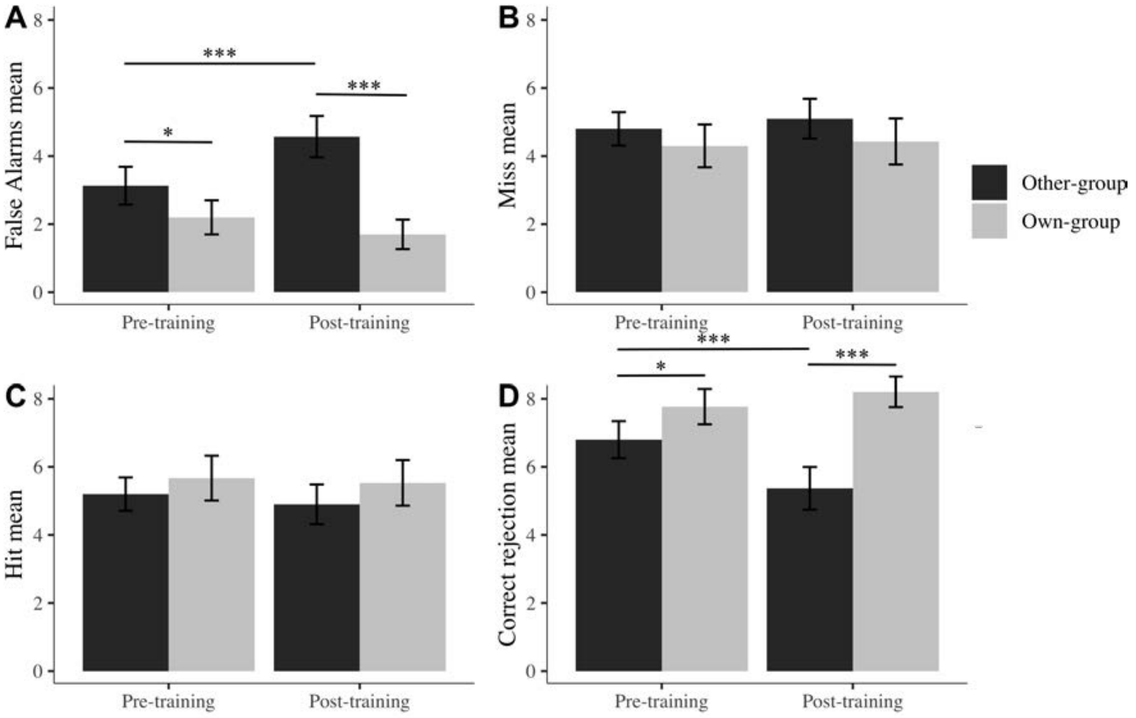

Figure 1.5 – Mean percentage of false alarms (A), misses (B), hits (C) and correct rejections (D) across stimulus group (own; other) and time (pre; post). I bars are 95% confidence intervals.
***$p < .001$; *$p < .05$

training for other-group stimuli and a lower rate of correct rejection after the training for same group stimuli. There were also significantly more false alarms toward other than own-group stimuli and more correct rejections of own than other-group stimuli. Both differences were larger after the training, in comparison to before training.

1.3.3 The modification of the OGB as a function of the modification of visual exploration

An additional aim of this research was to explore the relationship between the OGB and the focus on the bottom halves of faces. I made the hypothesis that the OGB should decrease as a function of a modification intended by the intervention to the visual pattern of exploration.

I computed the correlation between A' and the time spent on the bottom halves of faces, for both own and other-group stimuli after the training. Contrary to my hypothesis, neither correlation displayed a significant relation between the time spent on the bottom half of the face and discrimination performance for both other-group (r = .36, t(25) = 1.92, p = .066, 95%CI [-.02, .65]) and own-group (r = -.32, t(25) = -1.70, p = .102, 95%CI [-.63, .07]) after the training. However, this could be due to a lack of power, and it is noteworthy that no significant correlations were evident prior to training for own-group (r = .03, t(25) = .14, p = .887, 95%CI [-.36, .40]) or for other-group faces (r = - .06, t(25) = -.32, p = .750, 95%CI [-.43, .32]).

1.3.4 Accuracy in the training task

Even though I had no clear expectations about performance during the training task itself, it seemed useful to explore the training data as it might have shed light on the results reported above. Interestingly, participants performed significantly better on other-group than on own-group trials (M_{other} = .75, SD_{other} = .16; M_{own} = .36, SD_{own} = .11; t(28) = -14.10, p < .001, d = 2.85). Since feedback was displayed after each answer, I assume that participants were aware of their own performance.

1.4 Discussion and conclusion

In the present study, I first confirmed through eye-movement data that training participants to pay attention to the bottom halves of faces modified their visual pattern of exploration in favor of spending significantly more time focusing on the bottom halves of faces. The training program I implemented did not reverse the pattern of visual exploration, but created a better balance between time spent on the top and bottom halves of faces, with more attention being paid to the bottom halves of faces in particular, from the first fixation onward, after the training. Thus, training made participants pay more attention to the bottom halves of faces as a function of training. However, a difference appeared between processing of own-group and other-group faces. Before the training, participants focused more on the top than the bottom halves of faces, for both groups of faces. After the training, they focused more on the bottom than the top halves of faces than before the training, however the difference between dwell time on top and bottom remained significant only for own-group faces but not for other-group faces. The same pattern was found for the proportion of first fixations directed to the bottom halves of faces. These results suggest that participants learned the processing change intended by the training task, namely that focusing on the bottom halves of faces would help to increase recognition performance. These results also suggest that training raised awareness that focusing on the bottom halves of other-group faces increased their performance while it decreased it for own-group faces. Therefore, after the training and under an unconstrained visual exploration, participants directed their first fixation more to the bottom halves of other-group than own-group faces, and focused more on the bottom than top halves of faces for other-group faces while they focused equally on the two halves for own-group faces. My results also corroborate previous findings that observers spontaneously direct first fixations to the eyes (in my work, to the top halves of faces), even when they are asked and trained to focus on the bottom halves of faces.

However, unlike Hills and Lewis (2006), I did not observe a decrease of the OGB as a function of training. In fact, I found a significant increase of the OGB, which on closer inspection seemed due to an increase in false alarms toward other-group faces, while discrimination performance for own-group faces did not change significantly as a function of training. This increase could be explained by a training effect, albeit not one I expected. Indeed, regarding the results from the training task (i.e., better performance on other than own-group faces), I assume that participants were aware of their higher performance for other-group faces due to the feedback they were given after each trial. In addition to their awareness that bottom halves of faces are more discriminating for other-group than own-group faces, highlighted by the changes in visual exploration, participants may have become 'over-confident' in their capacity to discriminate other-group faces, but less confident in their discrimination ability for own-group faces. This hypothesis is supported by the absence of response bias toward other-group faces after training, while a conservative bias was observed towards other group faces before training. This overconfidence in other-group face discrimination performance may have resulted in more errors toward other-group faces while making more careful decisions towards own-group faces, as shown in the SDT measure of criterion, regardless of the visual strategy used for each group of faces.

The recognition of own-group faces in the present study was not affected by the time spent on the bottom halves of faces, although participants did spend more time on the bottom halves for own-group faces after training. However, one could increase the sample size and explore the promising observations made in the correlation analysis of time spent on the bottom half of faces and discrimination performance. Modifying the pattern of visual exploration does not seem to be enough to remove the OGB, and the individuation hypothesis is not only perceptual since individuation tasks can for example be to give a label (Lebrecht et al., 2009; Tanaka &

Pierce, 2009), describe physically (Malpass et al., 1973), rate attractiveness (Hills & Lewis, 2006) or rate distinctiveness (Hills et al., 2013; Hills & Lewis, 2011). Moreover, the difference between the present study and previous studies may have to do with the fact that participants were trained to develop a spontaneous way of looking at faces, and their first fixation was thus not constrained by a fixation cross (Hills et al., 2013; Hills & Pake, 2013). Some studies have, however, observed that the first two fixations are more important for recognition than complete gaze patterns (Hsiao & Cottrell, 2008). Finally, the present study used a different training regimen to that used by Hills and Lewis (2006), modifying only noses and/or mouths, while they modified noses, mouths, chins and checks. This difference might explain the failure here to replicate their results, since their more extensive modifications could have led to a different type of processing than that which I induced.

I would like to acknowledge some limitations of the present study, and encourage further studies to address the limitations, and to pursue my findings. First, I tested White participants only, not offering a complete cross-over design. Black participants were not included in the present study, not allowing me to explore if Black participants would have a spontaneously higher pre-training dwell time directed to the bottom halves of faces than White participants. Including Black participants could be done with the aim of exploring the effect of focusing on the top halves of faces (being critical for White faces), in order to eliminate an OGB from Black participants. Second, I did not use a control group, which could have added value to my results. Finally, in order to support my suggestion on the distinction I made between the effect of the instruction and the training task, one could design a study teasing out the effects of each, allowing conclusions to be drawn perhaps on the relative effect of each on the results observed in the present study. It has already been demonstrated that without clearly pointing out the difficulty raised from a cross-group recognition task, a simple instruction asking to raise general awareness to be careful at completing a recognition task does not result in a decrease of the OGB (Hugenberg et al., 2007).

In conclusion, the present study showed that training encouraged participants to focus on an apparently diagnostic facial features of other group faces did not assist recognition of other-group faces. In fact, such a strategy increased the OGB. One of the explanations could be that instead of improving face encoding by getting participants to pay attention to more discriminating features, the present training may have restricted the processing, thus reducing attention to other important parts of the face. Since own-group faces already profit from configural processing, increased time spent on the bottom halves of the face is not likely to impair discrimination performance.

In further studies, one should consider developing more thoroughgoing configural processing through training. It could also be of some interest to explore the same task but conversely: train Black participants to focus more on the top halves of faces to see whether that reduces the OGB increasing their discrimination performance for other-group faces. None of the other studies conducted in the present thesis is a direct follow-up of this one.

Study 2

Training White French participants for three weeks does not remove the own-group bias

Based on the observations of previous studies, I sought to explore to what extent a training regimen conducted over three weeks would decrease the OGB when participants were asked to focus on different parts of faces, and especially on parts found to be useful in increasing discrimination performance. I therefore conducted an experiment over five weeks which included a pre-training measure of the OGB (i.e., pre-test), three weeks of training, and a post-training measure of the OGB (i.e., post-test). I expected to observe a positive effect of distributed/longitudinal training as seen in previous studies which designed training that included multiple sessions (Goldstein & Chance, 1985; Lebrecht et al., 2009; Matthews & Mondloch, 2018; McGugin et al., 2011; Stahl, 2010; Tanaka & Pierce, 2009). I therefore developed a task in which participants had to encode faces in respect to different instructions. Although these tasks utilize featural processing, one is likely have strong encoding from the focus on the most critical characteristics of faces (Rodriguez et al., 2008). Previous studies (Hills & Lewis, 2006; Hills et al., 2013; Hills & Lewis, 2011) revealed that observers who focused on the bottom halves of Black faces showed a better discrimination performance than those focusing on the top halves of Black faces[12], so the first training sessions aimed to test whether an instruction to focus on bottom halves (i.e., rather than top halves) of faces would increase other-group faces discrimination performance. Additional studies showed promising

---

[12] Because data collection of this study was achieved in parallel to Study 1, Study 1 results were not available when designing this study, therefore were not considered in the design of this study.

results in favor of a better discrimination performance when focusing on the internal facial features (Fletcher, Butavicius, & Lee, 2008; Kemp, Caon, Howard, & Brooks, 2016; Paterson et al., 2017) which led to the design of the second training session: direct participants to focus on internal parts of faces to increase overall face discrimination performance. Finally, even though participants rely more on featural processing for other-group faces (Michel, Rossion, et al., 2006; Sadozai et al., 2018; Tanaka & Simonyi, 2016), the reliance on configural processing results in stronger encoding and thus in greater discrimination and recognition performance (Tanaka & Farah, 1993). In this regard, the third, and final, training session assessed whether participants would be better at recognizing faces after a configural instruction rather than a featural instruction. For this study, I collected data in a classroom while lecturing, which gave me the opportunity to test a multiple sessions training regimen without a great cost. However, this led to some difficulties that are discussed later on.

2.1 Hypotheses

In the present study, I expected to (1) find an OGB in pre-training task, and (2) an absence of OGB in post-training. In addition, differences within the training tasks were expected.

2.2 Method

2.2.1 Population

Twenty seven third year psychology students of the University of Toulouse Jean Jaurès took part in the study (M_{age} = 21.00, SD_{age} = 1.50[13]). They were recruited and completed the study during five weeks of an applied social psychology course which took place over six

---

[13] Data on characteristics are missing for two participants.

weeks. The sample size was therefore constrained by the size of the class. The final sample contained 11 White females (M_{age} = 20.45, SD_{age} = .52), since participants with incomplete participation (i.e., who missed at least one session), as well as participants who did not self-report being phototype I, II or III (i.e., White) were excluded due to the inclusion criteria of the present study. Participants did not receive any compensation or course credit for their participation[14].

2.2.2 Design

The present study was a two variables factorial within-subject design: Stimulus group (own; other-group) and the OGB measure (before; after the training).

2.2.3 Material

Stimuli

The experimental material consisted of frontal pictures of 200 males (100 Black and 100 White) from database UCT2007 and database UT2J2017. The 80 photographs used for pre-training and post-training tasks were the same faces at in Study 1, however clothes remained visible, background was not standardized, and stimuli were all from frontal view only. The additional 120 photographs used over the training tasks were randomly picked from the same UCT2007 and UT2J2017 databases among the unused ones. In every session, faces were presented full, as a whole. Instructions were written on paper sheets and given to the participants.

---

[14] Please note that data from an additional control group (N = 26) started to be collected from another classroom; however the research assistant misunderstood the instructions and prematurely ended the study. Data are partial and therefore excluded from the analyses. The collection process is described in the procedure section below.

Apparatus

A Power-Point (Microsoft Office) presentation was used to display the stimuli to the entire group of participants at the same time. Individual answer grids were used to collect participants' responses from the recognition phases. Pages with 20 face shapes (Figure 2.1) were used for the training sessions as a manipulation check. Training instructions were presented on paper, asking participants to either focus on the faces in terms of: (1) internal *vs* external features, (2) top *vs* bottom features, or (3) featural *vs* configural aspects. A questionnaire was also administered after the first and the last sessions, to investigate the strategies that participants used to encode and recognize faces. These questions were not intended to be explored in the present thesis as we took the opportunity to collect data on visual strategies for another project.

2.2.4 Procedure

The study was administered over five weeks, at the beginning of the weekly class for about 10 to 25 minutes for each of the sessions, the first and the last ones being the longest.

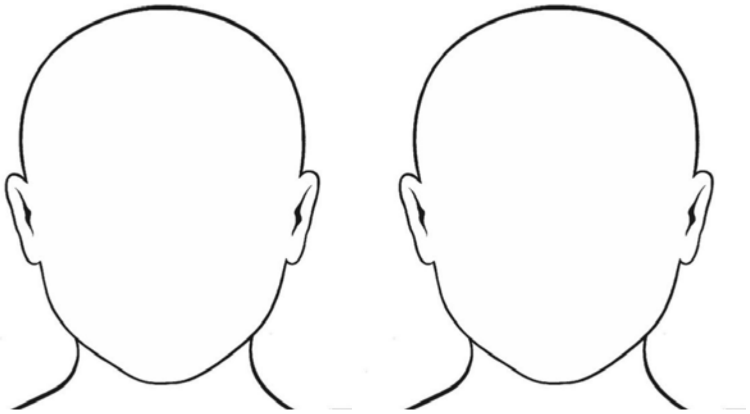

Figure 2.1 – Face shapes for the training task. Participants were asked to put a cross on or circle the part of the face they were focusing on during the coding phases.

At the beginning of every session, all the necessary documents (filler task, questionnaire, face shapes and answer forms) were given to the participants. Instructions for the training tasks were randomly distributed around the classroom and the completion of face shapes was used as a manipulation check to ensure the understanding and following of the right training instruction (i.e., "*to put a cross on, or circle the part of the face shape that you focused on*"). Participants had to write down their anonymous number on each of the documents. Over the session, maximum effort was made to keep the room quiet and to avoid people disturbing or copying one another.

Session 1 The first session counted as the pre-training session. Once the study was explained to the participants and the consent forms were signed, they were asked to look at a series of faces (k = 20) displayed on a PowerPoint presentation and to try to remember them. Pictures were presented for 10 seconds each, one second apart from each other[15]. After the presentation, participants were asked to complete a word puzzle for 5 minutes. After the 5 minutes, another slide show presented the 40 pictures for the recognition phase and participants had to circle ‘yes’ or ‘no’ for each face, answering the question "*Have you already seen this person before?*". After the presentation of the pictures, participants had to complete a questionnaire on the strategies they believed they used to encode/recognize faces. They were finally told not to talk about the study to each other or to anyone else before the end of the five weeks. The control group completed the exact same procedure at the same time in another classroom with a research assistant.

---

[15] The presentation time was chosen subjectively expecting to give enough time to the participants to draw onto the face shapes during training sessions and was kept identical over the five sessions for standardization purposes.

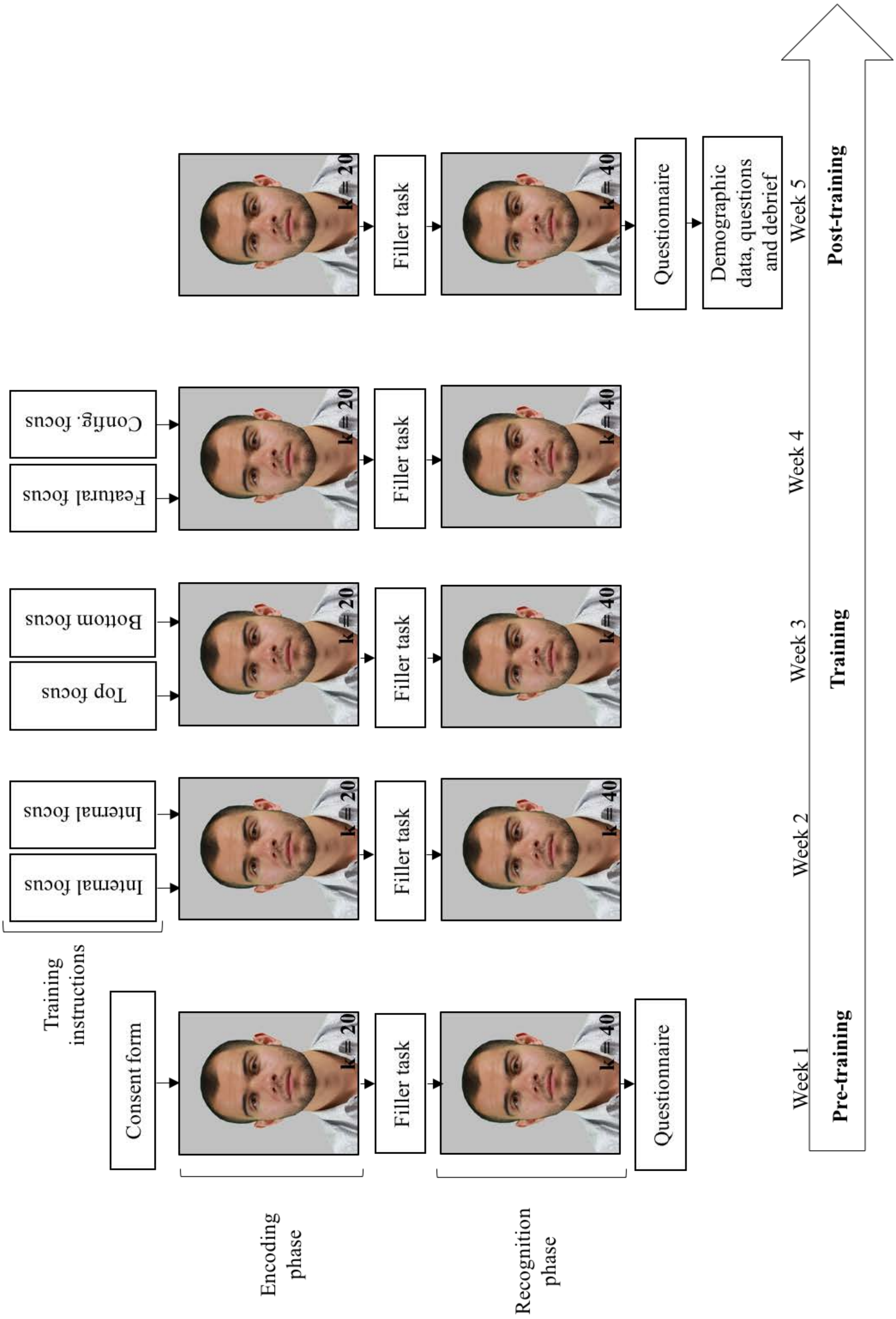

Figure 2.2 – Flow chart of the procedure: five old/new recognition tasks over five weeks using color pictures from frontal neutral of young males with clothing present. Pictures were not used during the study, but presented here for illustration purpose with the consent of the model.

Session 2 The second session consisted of the first session of the training. During the encoding phase, participants were asked to circle or put a cross on, using the sheet with the face shapes and according to their training group, the part of the face they found the most important in order to recognize the person. The first group was asked to focus on the bottom part of the faces while the second was asked to focus on the top part of the faces. Following the encoding phase with the face shape completion and a 5-minute anagram completion task, participants completed the recognition phase. The instructions during the training sessions specified on which part of the face participants had to focus on.

Session 3 The third session was the second part of the training. Participants were asked to either focus on the external or the internal part of faces during the encoding phase while completing the face shapes. After a 5-minutes crossword puzzle, they completed the recognition phase.The instructions during the training sessions specified on which part of the face participants had to focus on.

Session 4 The fourth session was the third and last training session. Participants were asked to focus on faces either configurally or featurally while completing the face shapes. The instructions were the following, respectively:

> Your aim is to focus on the face as a whole, on how features are arranged and organized to form a whole. For example, focus on the distance between both eyes, between eyes and mouth, etc.; what is the global shape and size of the face? What is its organization? Is the chin and jaw squared, or rounded, or more like a Y?; and many other possibilities. Don't limit yourself to the examples.

or:

> Your aim is to focus on all the features of the faces one by one, each independently from the others. For example, focus on the color and shape of the eyes, the mouth, the length and width of the nose, the chin, etc. Are the ears big? Or prominent? And many other possibilities. Do not limit yourself to the examples.

After a 5-minute spot the differences task, the recognition phase was completed.

Session 5 The fifth session was the last session of the study, and thus counted as a post-training session. As in the pre-training session, participants were not asked to focus on any specific aspect of the faces or to complete any of the face shapes. After the encoding phase, they completed a 5-minute task mixed between anagrams and calculations and completed the answer grid during the recognition phase. After the recognition task, they were once again asked to complete the questionnaire about their strategies, and to provide some demographic information (gender, age, phototype, country of birth of grand-parents). Participants in the control group were supposed to complete the same task; however data are missing since the research assistant faced issues during the completion of this session.

At the end of the fifth session, participants were directly debriefed, and results were presented one week after, during the last lecture. They were free to ask me any questions they may have had, during the debrief or the presentation of the results.

2.3 Results

2.3.1 Discrimination performance over the five sessions

In order to explore the presence of the OGB across sessions, a mixed linear model using main effects and interaction effect between the sessions and stimulus group on the discrimination performance was conducted with participant as a random effect. The model revealed a significant effect of stimulus group ($F(1, 99) = 38.278$, $p < .001$), but not session ($F(4, 99) = 1.265$, $p = .289$) nor the interaction effect between the two ($F(4, 99) = .254$, $p = .907$) on discrimination performance. *Post-hoc* analyses showed the presence of an OGB, namely a better performance for own-group faces than other-group faces, in session 1 ($\beta = -.131$, $t(109) = -3.089$, $p = .003$, $d = 1.18$), session 2 ($\beta = -.127$, $t(109) = -2.978$, $p = .004$, $d = 1.41$), session 4 ($\beta = -.114$, $t(109) = -2.672$, $p = .009$, $d = 1.33$) and session 5 ($\beta = -.110$, $t(109) = -2.588$, $p = .011$, $d = .81$), but not in session 3 ($\beta = -.079$, $t(109) = -1.863$, $p = .065$, $d = .88$; Figure 2.3, see Table 2.1 for descriptive data). In addition, there were no differences in discrimination performance between session 1 (i.e., pre-training) and 5 (i.e., post-training) for both the own-group faces ($\beta = .053$, $t(109) = 1.250$, $p = .722$) and the other-group faces ($\beta = .032$, $t(109) = .749$, $p = .944$).

Therefore, even if I validate the hypothesis of the presence of an OGB prior to any training, I cannot conclude an effect of training on the elimination of the OGB. Indeed, the OGB is still present in the last session.

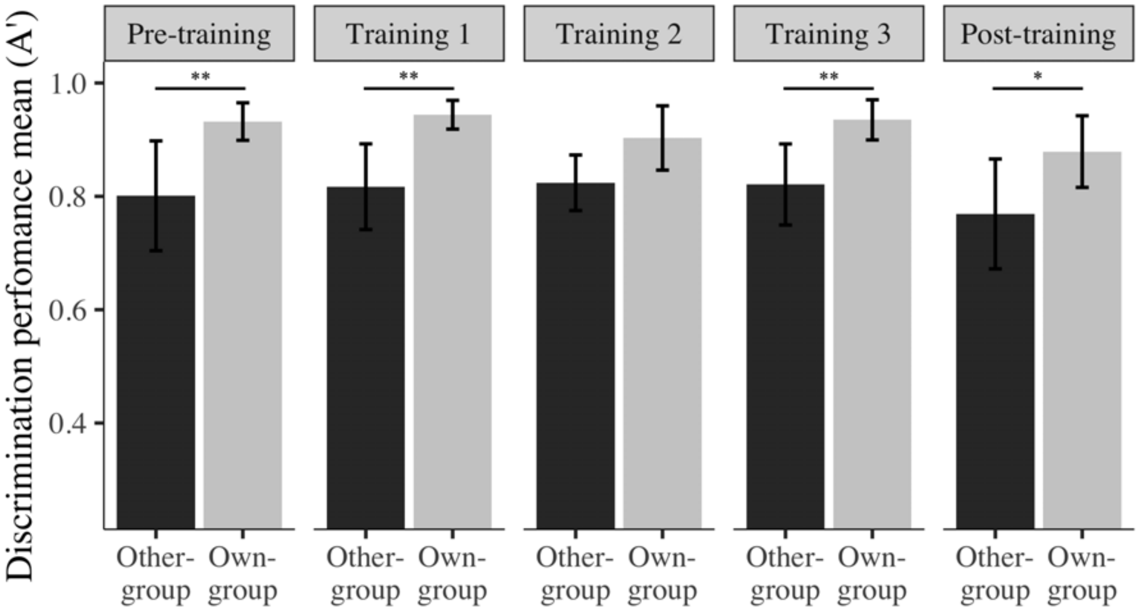

Figure 2.3 – Discrimination performance (A′) across stimuli group (own; other) in each session (session 1-5). I bars are 95% confidence intervals.
$^{**}p < .01$; $^{*}p < .05$

2.3.2 Hit and false alarms over the sessions

I also conducted additional mixed linear model analyses to predict first hits and secondly false alarms, using the same main and interaction effects between session and group stimulus as well as including participant as a random effect. The model to predict hits showed a significant main effect of stimulus group ($F(1, 99) = 21.41$, $p < .001$) but no significant main effect of session ($F(4,99) = 1.76$, $p = .144$) or the interaction between session and stimulus group

Table 2.1 – *Mean and Standard deviation of discrimination performance (A′) across stimulus group (own; other) in each session (sessions 1-5).*

	Session 1	Session 2	Session 3	Session 4	Session 5
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Black	.80 (.16)	.82 (.13)	.82 (.08)	.82 (.12)	.77 (.16)
White	.93 (.06)	.94 (.04)	.90 (.10)	.94 (.06)	.88 (.11)

Table 2.2 – *Mean, standard deviation and within-subject t-test between stimuli group (own; other) in each session (session 1-5) for hit and false alarms.*

		Own-group	Other-group					
		M (SD)	*M (SD)*	β	*t*	*df*	*p*	*d*
Hits	Session 1	8.18 (1.47)	6.73 (1.49)	-1.45	-2.08	109	**.040**	.98
	Session 2	8.27 (1.27)	6.91 (2.30)	-1.36	-1.95	109	.054	.76
	Session 3	7.46 (2.51)	7.46 (1.51)	0	0	109	1.00	0
	Session 4	8.64 (1.50)	6.55 (1.70)	-2.09	-2.99	109	**.004**	1.31
	Session 5	7.55 (1.51)	5.55 (2.38)	-2.00	-2.86	109	**.005**	1.03
FA	Session 1	.64 (1.03)	2.00 (1.73)	1.36	3.01	109	**.003**	.99
	Session 2	.36 (.65)	2.00 (1.18)	1.64	3.61	109	**<.001**	1.79
	Session 3	.64 (.92)	2.64 (1.12)	2.00	4.41	109	**<.001**	1.96
	Session 4	.91 (.94)	1.64 (1.29)	.73	1.60	109	.112	.65
	Session 5	1.18 (.87)	1.82 (1.78)	.64	1.40	109	.164	.48

Note. Significant *p*-values are in **bold**

(F(4,99) = 1.57, p = .188). *Post-hoc* analyses revealed that mean of hits is overall higher for own-group than other-group faces, except for session 3, which may explain the absence of OGB (Table 2.2). The model to predict false alarms revealed the same pattern: A significant main effect of stimulus group (F(1,99) = 43.26, $p < .001$) but no significant main effect of session (F(4,99) = .72, p = .581) or of the interaction (F(4,99) = 1.83, p = .129). *Post-hoc* analyses revealed that the mean of false alarms is, in opposition, overall lower for own-group than other-group faces (Table 2.2), staying quite similar for other-group faces (except for session 3) but increasing over time for own-group faces. There are therefore more hits and less false alarms for own-group than other-group faces.

2.3.3 Discrimination performance within and between training types

From observations made in previous studies which were considered to develop the present training, it was relevant to conduct exploratory analysis on type of training. Indeed, instructions leading to a focus on the internal part of faces, or the configural processing might have resulted in a higher discrimination performance than other instructions. In order to explore the effect of the type of training on the performance (i.e., six types of training, two by session), a mixed linear model was constructed to predict the discrimination performance from the main effects and interaction effect between training type and session, with participant as a random effect. The model resulted in no significant effect of training type ($F(1, 46) = .868$, $p = .357$), session ($F(2, 51) = .118$, $p = .889$) nor the interaction between session and training type ($F(2, 48) = .845$, $p = .436$). The performance is thus not better after any of the type of training than after another, which does not support my hypothesis. However, in regards to the previous results, particular attention was paid to session 3. Session 3 was the second training, where participants had to focus either on the internal or the external part of faces. A *t*-test was conducted, and revealed no differences between the two training conditions ($M_{ext} = .83$, $SD_{ext} = .10$; $M_{int} = .90$, $SD_{int} = .07$; $t(20) = -1.77$, $p = .09$, $d = .82$).

2.4 Discussion and conclusion

The present study aimed to explore the effect of a training regimen conducted over three weeks and using different focus instructions. The presence of the OGB was measured one week before the first training session, and one week after the thirda, training session. As expected, I found an OGB during the first measure: Participants presented a better discrimination performance for own-group faces than for other-group faces. The OGB was still present after training, as participants did not improve their performance as a function of training.

When comparing the different instructions across the sessions, I observed no differences in outcomes, suggesting that none of the training instruction resulted in a better performance over the others. The absence of differences as a function of instruction could have several explanations. First, it was confirmed that participants respected the instructions, as was verified as they were asked to put a cross on the image, or circle the part of the face that would help them to better recall each face, compliance with the instructions was verified. All participants followed the instructions they were given, suggesting that they did focus on the part of the faces they were asked to focus on. A possible limitation to explain the absence of a training effect in this study is the small sample size (N = 11), as well as the absence of a control group, which both limit the interpretation of the results of this study. Further, although the presence of clothing on the pictures could have been used as a retrieval cue by participants, it does not appear to be the case considering the OGB was found, and none of the face resulted in a 100% accuracy rate.

Since this study has been replicated as a part of the present thesis (see Study 4) with improvements to rectify the observed issues, further discussion of the content of training and conclusions of this study are presented together with those of Study 4.

Study 3

Independence of the sets of pictures to measure the OGB

In Study 1 and Study 2, the sets of stimuli used to assess pre-training and post-training measures of OGB were fixed (i.e., defined and constructed prior to data collection) and were not counterbalanced between the two measures. Set 1 was used for pre-training, and Set 2 was used for post-training, hence the conclusions might have been affected by a greater difficulty across sets. It was therefore of interest to explore whether an OGB can be found with both sets, and more importantly, if there were any differences in the discrimination performance observed between the two sets. An online study has been conducted in France to explore this concern.

3.1 Hypotheses

In the present study, I expected to find an OGB in both sets along with no significant difference in participants' discrimination performance between the two sets.

3.2 Method

3.2.1 Population

Three hundred and fifty nine participants were recruited online through several French social media, among which 230 completed the study entirely (188 women; M_{age} = 22.32, SD_{age} = 6.53). Forty participants were removed because they were not of phototype I, II or

III, and nine because they declared having seen at least of one the picture or person before the completion of the study. The final sample contained 88 participants for evaluation of Set 1 (72 women, M_{age} = 22.19, SD_{age} = 5.77) and 93 for Set 2 (82 women, M_{age} = 22.58, SD_{age} = 7.26).

3.2.2 Design

This study had a two variables factorial mixed design with set (Set 1; Set 2) as between-subject and stimulus group (own; other-group) as within-subject.

3.2.3 Material

Stimuli

In total, pictures of faces of 80 males were used: 40 in each set. In each set, pictures were of 20 different targets (10 Black and 10 White) and 20 different foils (10 Black and 10 White). There were the same as the pictures used in Study 2: All poses were frontal, with neutral expressions, and clothing was unedited on purpose. Pictures of targets were used twice: during the encoding, and during the recognition phases. No modifications were applied to pictures between encoding and recognition, as it was the case for Study 2.

Apparatus

Stimuli were modified using GIMP 2.8.14 software (GNU Image Manipulation) as 1270x720 pixels, and the survey was built on Qualtrics online survey tool (Qualtrics, Provo, UT).

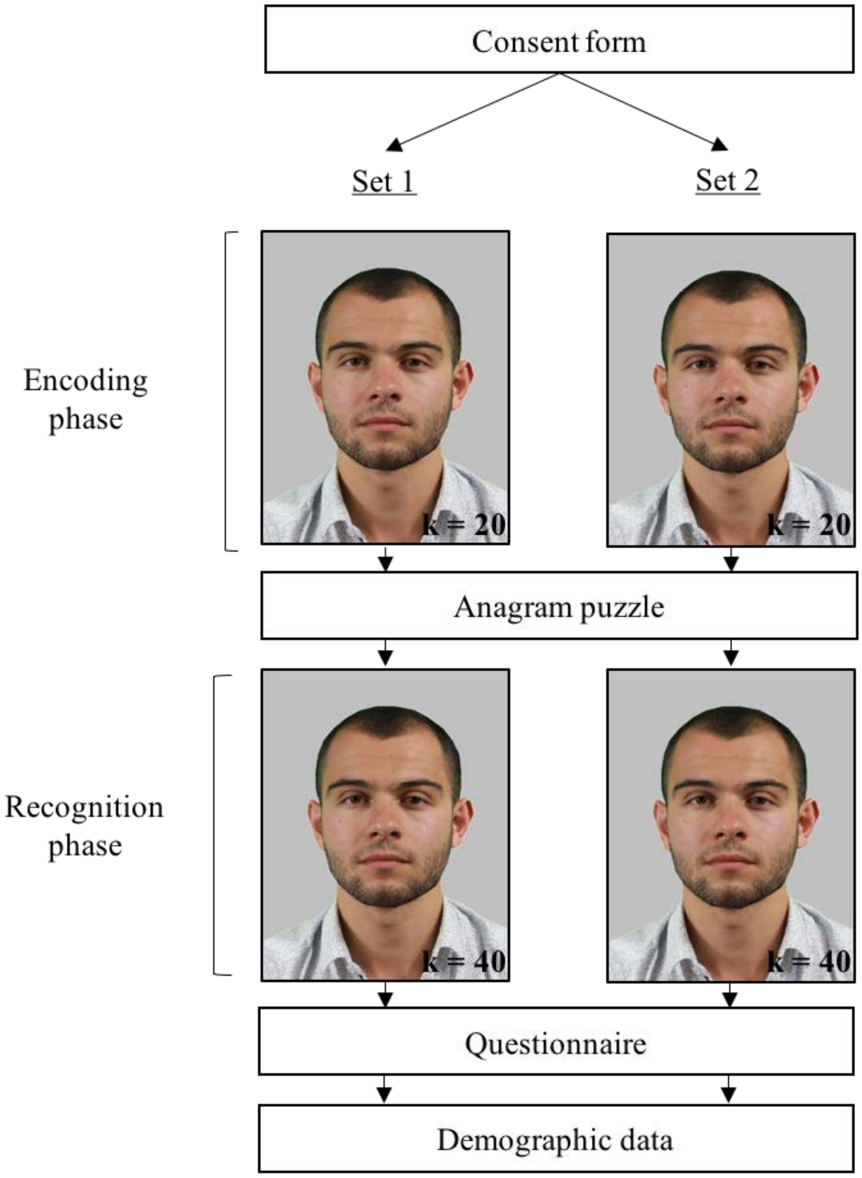

Figure 3.1 – Flow chart detailing procedure for Study 3. All participants completed identical consent forms, anagram puzzles, questionnaires and demographic data. Participants were randomly assigned to one of the two sets. Pictures were not used during the study, but presented here for illustration purpose with the consent of the model.

3.2.4 Procedure

The procedure of the present study was an old/new standard recognition task (Figure 3.1). Once opening the survey link online, participants were asked to read a consent form and give their agreement to participate. A first control question was displayed to ensure that they were using a computer to complete the study as it was specified in the consent form. Otherwise, they were directed to the end of the survey, specifying the reason why they would not be permitted to complete the study. This was done to ensure the presentation of pictures from a comparable size with previous studies and across participants. After this question, the two sets were randomly assigned to participants. Instructions were then given:

> You will complete two phases in the study. The first one is a memorization phase during which you will learn faces which you will be asked to recognize afterward during the second phase. Please pay attention carefully to the presented faces. Faces will be displayed automatically one after the other after a defined time, you have no action to do until the end of the first phase. Once you are ready, go the next page and pictures will start to be presented to you.

Pictures of the 20 target faces were presented automatically for three seconds each, one second apart from each other. After the encoding phase, participants completed a 5-minute anagram puzzle, and the instruction for the recognition phase was displayed automatically at the end of the five minutes:

> Now, you will be presented a number of faces. Some of them have already been presented to you during the first phase. For each face, you will be asked to state if it has been presented to you during the first phase or if it has not. If is has been

> presented before, you will select ‘yes’. If it has not been presented before, you will select ‘no’. There is no time limit. Once you are ready, you can start.

The 20 target faces along with the 20 new faces were displayed. Once the task was completed, participants were asked to complete the same questions about their strategies as in study 2. These data regarding strategies will, however, not be analyzed in the present thesis[16]. Finally, participants were asked to provide demographic information: gender, age, education, profession, phototype, countries of birth of themselves, their parents and grand parents and were given a debriefing form after the completion of those questions. The task lasted between 10 and 20 minutes.

3.3 Results

3.3.1 Discrimination performance over the two sets

To explore the discrimination performance of own-group and other-group faces, two mixed linear models predicting discrimination performance were run. The first one aimed to explore main effect of stimulus group, adding participant as random effect, on the discrimination performance for Set 1. The second model was identical, but applied on data from Set 2. Set 1 resulted in a significant effect of stimulus group ($F(1, 88) = 28.04$, $p < .001$), with a significantly lower discrimination performance for other-group faces ($M = .86$, $SD = .09$) rather than own-group faces ($M = .91$, $SD = .07$; $\beta = -.05$, $t(89) = -5.27$, $p < .001$, $d = .58$). Set 2 revealed an identical pattern, namely a significant main effect of stimulus group ($F(1, 93) = 46.16$, $p < .001$) with a significant lower discrimination performance for other-group faces ($M = .85$, $SD = .10$) rather than own-group faces ($M = .92$, $SD = .07$; $\beta = -.07$, $t(94) = -6.76$,

---

[16] The opportunity was taken to collect data on visual strategies, but no hypothesis in this regard were planned to be answered in the present thesis.

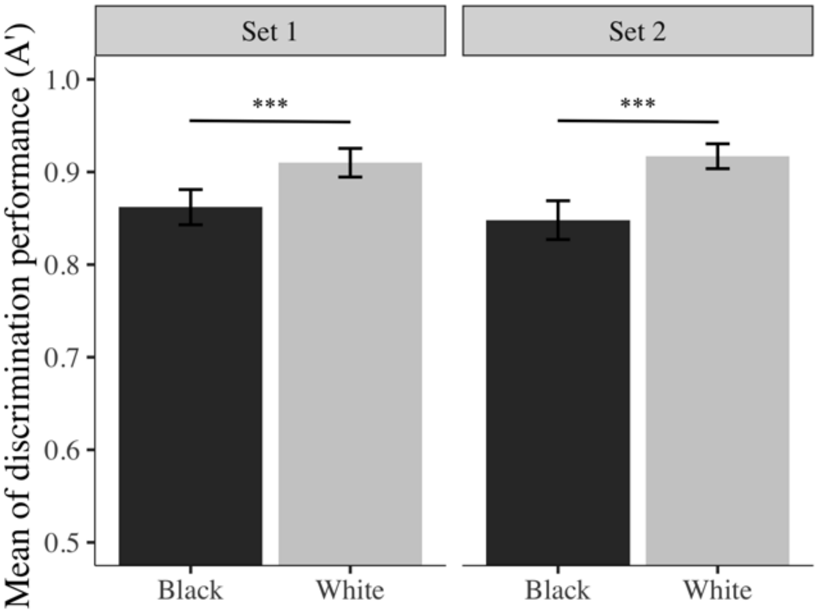

Figure 3.2 – Mean of discrimination performance (A') across stimulus group (own; other) in each set (Set 1; Set 2). I bars are 95% confidence intervals.
***$p < .001$.

$p < .001$, $d = .82$; Figure 3.2). A third mixed linear model was run to explore the difference between the two sets in terms of discrimination performance. The model included main effects and the interaction effect of set and stimulus group on discrimination performance, along with participant as random effect. The model resulted in a significant effect of stimulus group (F(1, 180) = 72.70, $p < .001$), but not of set (F(1, 179) = .10, p = 755), therefore confirming the absence of differences between the two sets in terms of presence of the OGB.

3.4 Discussion and conclusion

Study 3 aimed to test whether two different sets of pictures had different discrimination difficulties, and if they would therefore affect the presence or absence of the OGB. Participants completed a standard old/new recognition task online of either Set 1 or Set 2 stimuli. As expected, the OGB was found in both sets, and there was neither a significant difference in the magnitude of the OGB, nor in the discrimination performance for each group (i.e., own or other) between the two sets. This study suggests that the absence of a training effect in Study 1 and Study 2 was not due to the use of a biased set of pictures. However, in the following studies, stimuli sets were counterbalanced across participants to improve the study design.

This study raised additional questions and concerns about the materials used in the field of face recognition studies - indeed, it is important to consider that the OGB might be moderated by the materials used (i.e., pictures). Considering that as well as group, distinctiveness (Valentine, 1991), gender (Herlitz & Lovén, 2013), age (M. G. Rhodes & Anastasi, 2012), or angle (Bindemann, Attard, Leach, & Johnston, 2013) have an impact on face discrimination and recognition performance, and it is noteworthy that even though the OGB is strong enough to be found across these variables, it remains a question of the extent of standardization. Two points of view can be conflicting here. First, one could argue that the more standardized pictures and sets are, the better it is to study an effect. Thus, a standardized set of pictures aiming to capture the OGB could be created, and validated; for instance, as is done to measure face matching (e.g., Glasgow Face Matching Test, Burton, White, & McNeill, 2010) or face recognition performance (e.g., Cambridge Face Memory Test, Duchaine & Nakayama, 2006) with White faces. In opposition, one could argue that the more variability there is, the better it is, since a standardized set of pictures would not be representative of a general population. Whereas there is not a straightforward, and ultimately correct, answer, when studying the effects of training, it is of some interest to have a standardized set of pictures. In fact, controlling for variability from the faces, improved performance after training A or training B could directly be attributed to either the content of training, or to individual differences. It could also be of some interest to explore different sets consisting of different types of pictures and to see under which conditions the OGB is, or is not found.

Study 4

Training White South African participants for three weeks does not remove the own-group bias

Based on the limitations observed in Study 2 (i.e., absence of a control group, unstandardized stimuli, group assessment, and absence of randomization, see p.136), I aimed to replicate the study in South Africa in a manner that addressed these limitations. This study was conducted in Cape Town, giving access to different racial groups: Black, White, and Coloured. This study was similar in design to Study 2, but a control group was added and the material was improved (e.g., standardizing pictures: faces were superimposed on a grey background, and were cropped from the neck upwards to exclude clothing). Photographs were also transformed between the encoding and recognition phases to counteract possible reliance on artifactual picture cues rather than face cues (Bruce, 1982) as done in Study 1. The procedure included individual assessment, random presentation, set counterbalancing, and a shorter presentation time.

When discussing Study 2, I wondered whether the absence of an effect for training in Study 2 could have been due to the issues mentioned above. That is, no theoretical discussions on the context of the training were considered before conducting this study. However, limitations from the training design are acknowledged in the discussion section.

4.1 Hypotheses

In this study, I first expected to find an OGB for White participants as found in Study 2. Secondly, I expected to find an OGB for Black participants (i.e., a better performance for Black than White faces) while no difference was expected for the Coloured participants since the stimuli were both other-group faces for these participants. Thirdly, the OGB was expected to be lower after training than before, and lower in for trained than untrained participants. This was expected from both the modification in the study design, viz., the inclusion of a new type of participant (i.e., from a different country and group), and from the increase of the sample size, even if no training effect was found in Study 2.

4.2 Method

4.2.1 Population

One hundred and two Psychology undergraduate students from the University of Cape Town took part into the study[17]. They were recruited over two recruitment drives through the Student Research Participation Program (SRPP). Desired sample size was a minimum of 33 to find an OGB (i.e., 11 per group), according to results reported in Study 1. However, because of the high dropout rate and time constraints and despite the two collection sessions (i.e., twice five weeks), the expected sample size was not attained. Indeed, 44 participants were removed because did not complete every session, and seven met the exclusion criteria (had ancestors from Asia/India or had seen one of the pictures before) or chose not to answer the demographic questionnaire. The final sample contained 51 participants (41 women; 20

---

[17] Since descriptive data were collected during the last session, they are missing for the global sample.

Black, 17 White, 14 Coloured[18]; M_{age} = 20.63, SD_{age} = 3.24). Participants were awarded with five course credit points for the completion of all the five sessions.

4.2.2 Design

The present study has a three variables mixed factorial design with training (training; no training) as between-subject variable and stimulus (Black; White) and the OGB measure (pre-test; post-test) as within-subject variables.

4.2.3 Material

Stimuli

Photographs of a total of 200 males were used in this study. The 80 stimuli for pre- and post-tests sets were identical to the ones used in Study 2; however both frontal and three-quarter views were used and the clothes and background were removed (i.e., the same type of manipulation as for Study 1). As opposed to Study 2, due to concerns about the homogeneity of facial appearance in those images, a new random selection of young males from the frontal view was made for the 120 training stimuli within the UCT2007 database. Frontal views were used for all encoding phases in every task, three-quarter views were used for recognition phases in pre- and post-test tasks, and vertically flipped grey-scaled faces were generated for training recognition phases.

---

[18] Categories constructed on the basis of the phototypes: I (n = 1), II (n = 5) and III (n = 11) for White; IV (n = 14) for Coloured; V (n = 14) and VI (n = 20) for Black. Countries of birth of the ancestors were checked according to the exclusion criteria. This point will be discussed further.

Apparatus

Stimuli were modified using GIMP 2.8.14 software (GNU Image Manipulation) as 1270x720 pixels. Each session was completed on 17" to 21" Computers[19]. The entire study was built using Qualtrics online tool (Qualtrics, Provo, UT), except for the face shape task which was completed on paper sheets (see Figure 2.1, Study 2). The questionnaire on encoding/recognition strategies was also presented after the first and last session, and will not be analyzed here.

4.2.4 Procedure

Participants had to register for a time slot which would be on the same day and at the same time over the five weeks. However, in cases of particular practical impediments from a participant, arrangements were made to pick another time within more or less 24h of the initially allocated time. When participants arrived at the lab for the first session, they were given consent forms and told what their participation would entail. The procedure was identical to Study 2, with the following exceptions: instructions were provided directly on the screen of a computer for each participant; the presentation of faces was randomized between the participants within each session; the sets of faces used during the first and last session were randomized between the participants; answers from the recognition task were recorded directly through Qualtrics and a control group was added (Figure 4.1). During the training sessions, the control group completed an identical recognition task as for the pre-test and post-test sessions. Instructions were identical to the ones in Study 2, but translated into English.

---

[19] Screen size was different, but that did not affect the size of the presented stimuli.

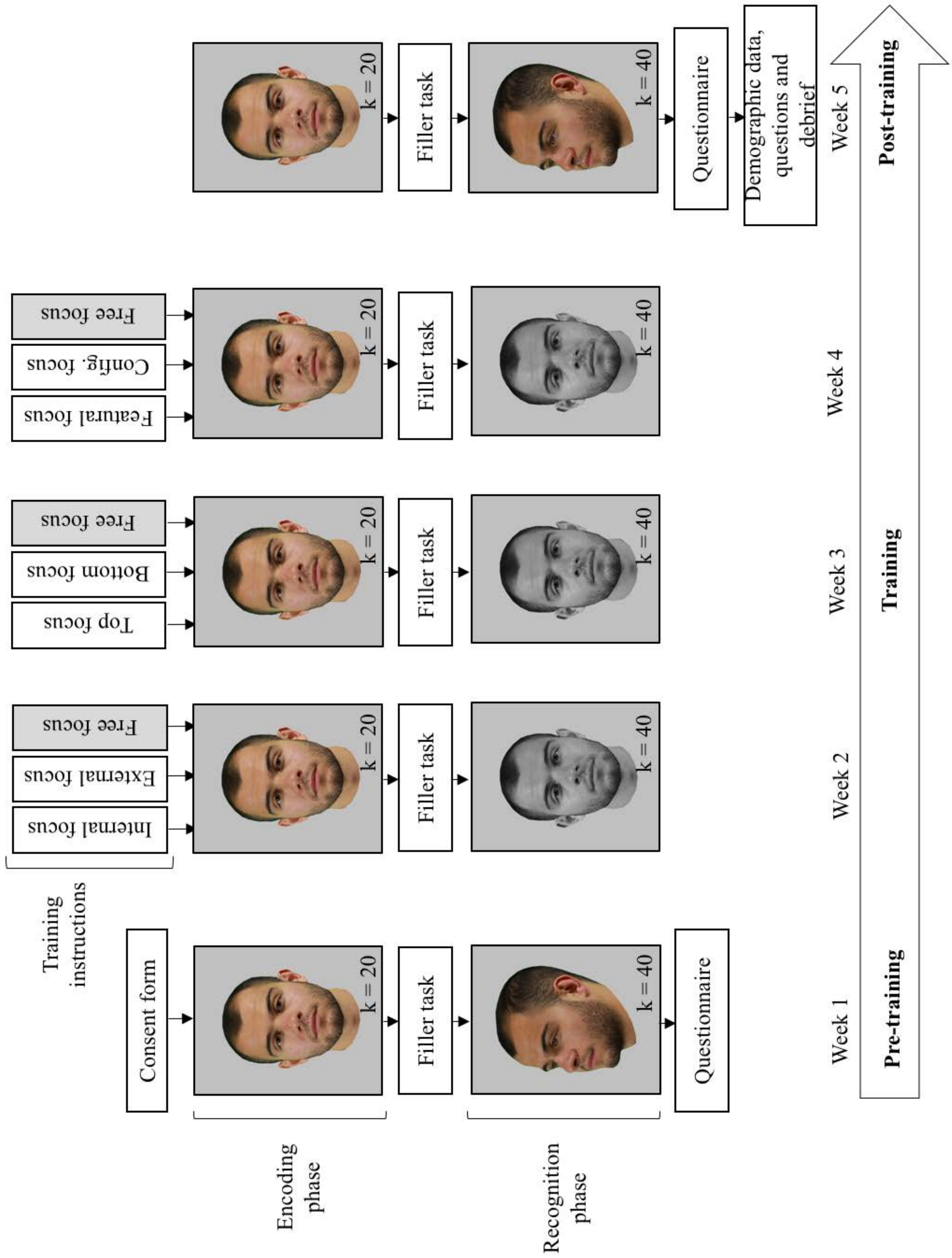

Figure 4.1 – Flow chart of the procedure for Study 4: five old/new recognition tasks over five weeks using frontal neutral color pictures of young males. Different instructions were used: two different training, and no training (i.e., free focus). Pictures were not used during the study, but presented here for illustration purpose with the consent of the model.

Participants were split into different training or no-training groups on the basis of the ‘racial’ group they were perceived to belong to in order to ensure a balance[20]. At the beginning of each session, participants received the same instruction:

> There will be two phases in the present study. The first one is a memorization task during which you will have to learn a certain number of faces, which you will be asked to recognize in the second task. Please study the following faces. All of them will be presented one by one automatically after a certain delay. You do not have to do anything during this task.

Then, the 20 targets (10 for each group) were randomly presented for three seconds each, one second apart. After a 5-minute filler task (anagram puzzle, simple calculations, Ravens' matrices, spot the differences or word classification), the second instruction was presented:

> Now you will see a certain number of faces, some of which have been shown to you in the previous task. If the face has already been shown to you, select ‘yes’. If the face has not already been shown to you, select ‘no’. There is no time limit.

Target face were presented again, one by one and randomly with 20 new faces included. Decision time was recorded, but not limited. For each picture, participants answered the question "*Did you see this person during the first phase [of the experiment]?*".

During the training sessions, the task was identical for the training and the no training groups. The no training group was asked to complete identical tasks as in pre-test and post-

---

[20] Since the self-reported phototype and origin questions were asked during the last session, perceived ‘racial’ group was used to distribute participants equally among the different groups. However, only self-reported information were used for data analyses. I acknowledge the ethical limitations of such a method.

test. However, the training groups received the following instructions before the encoding phase (e.g., here for the internal focus instruction):

> There will be two phases in this task. The first one is a memorization task during which you will have to learn a certain number of faces, which you will be asked to recognize in the second phase. Please study the following faces. All of them will be presented one by one. You have to click the red arrow to continue the task and see the next face. The training requires that you complete another task in between seeing individual faces. Look at the sheet of paper in front of you in which you see blank face shapes, and also make sure you have a pen. Each face shape represents the face you will have just seen in the training session that is about to commence. *For this training session you must focus only on the internal part of the face as it is shown in the following example*[21]. When the first photograph has been presented, you need to put a cross on, or circle on the face shape the part of the face stimuli that you focused on. That means you should not put a cross on the ear or the hair for example even if the whole face is presented. I am asking this to make sure you have followed the instructions correctly. You need to complete the same task for each photograph. If you have any questions, please call the researcher.

Then, the face were randomly presented for three seconds each, and then blanked. Participants had to manually click on a button to start the 3-second presentation of the next stimulus. This manipulation by the participant aimed to let sufficient time to write down on the face shapes

---

[21] This part was replaced by the external part for the second group of Session 2; upper or lower for Session 3; configuraly (that means that your aim is to focus on the face as a whole, on how features are arranged and organized to form a whole, for example, the distance between the eyes, or how are the features are organized to form a face) or featuraly (that means that your aim is to focus on all of the features of the face, one by one, each independently from the others) for Session 4. Examples with pictures of faces were also provided to make clear the part participants were asked to focus on.

without being disturbed by the presentation of the next stimulus.

At the end of the tasks, participants were asked if they have already seen one of the faces prior to the present study, and were given access to the debriefing form that they could read. They were free to ask me questions they may have had. The study was administered over five weeks, for about 10 to 25 minutes for each of the sessions.

4.3 Results

4.3.1 Discrimination performance in the pre-test

A mixed linear model including the main effects and interaction effect of participant group and stimulus group with participant as a random effect was created, in order to analyze discrimination performance in the first session. The model revealed no main effect of participant group ($F(2, 51) = .07$, $p = .935$) nor of the interaction between participant group and stimulus group ($F(2, 51) = 2.47$, $p = .094$) but a significant main effect of stimulus group ($F(1, 51) = 9.61$, $p = .003$). *Post-hoc* analyses revealed a significantly better recognition in White participants for White faces ($M = .84$; $SD = .10$) than for Black faces ($M = .68$; $SD = .15$; $\beta = -.16$, $t(54) = -3.51$, $p < .001$, $d = 1.28$). No significant differences were present for Black participants ($M_{bk} = .74$; $SD_{bk} = .15$; $M_{wh} = .77$; $SD_{wh} = .13$; $\beta = -.04$, $t(54) = -.92$, $p = .360$, $d = .21$) or Coloured participants ($M_{bk} = .73$; $SD_{bk} = .16$; $M_{wh} = .77$; $SD_{wh} = .13$; $\beta = -.04$, $t(54) = -.82$, $p = .42$, $d = .28$; Figure 4.2). White participants thus displayed an OGB, but unexpectedly, not Black participants.

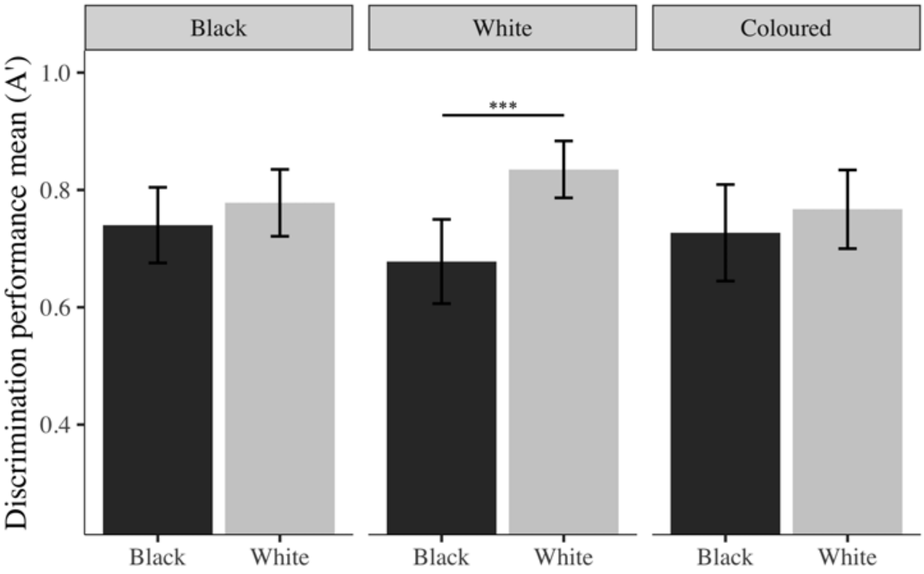

Figure 4.2 – Discrimination performance (A') in pre-test across stimuli group (Black; White) for each participant group (Black; White; Coloured). I bars are 95% confidence intervals.
***p < .001

4.3.2 Discrimination performance in the post-test

In order to explore the effect of training on the recognition performance during the last session (i.e., post-test), a mixed linear model was created. The best model included all main effects and a three-way interaction between training, participant group and stimulus group, in addition to participant as a random effect. The model revealed only one significant effect: The main effect of stimulus group (F(1, 51) = 10.80, p = .002). I observed no significant main effects of training (F(1, 51) = .76, p = .389), participant group (F(2, 51) = .06, p = .945) nor of the interaction between training and participant group (F(2, 51) = .74, p = .483), the interaction between training and stimulus group (F(1, 51) = .44, p = .511) nor the interaction between stimulus group and participant group (F(2, 51) = 1.42, p = .251). *Post-hoc* analyses revealed that an overall better recognition of White faces than Black faces (Figure 4.3). Unexpectedly, Black participants recognized White faces better than Black faces, significantly for the training group (M_{wh} = .78, SD_{wh} = .09; M_{bk} = .65, SD_{bk} = .18; β = -.13, t(55) = -2.21, p = .032, d = .96) but not significantly for the untrained group (M_{wh} = .75, SD_{wh} = .19; M_{bk} = .66, SD_{bk} = .21; β = -.089, t(55) = -1.878, p = .066, d = .45). This matches the observation already made during the pre-test. The same pattern is present for White participants, resulting in an OGB: A significantly better recognition for White faces than Black faces for untrained group (M_{wh} = .80, SD_{wh} = .09; M_{bk} = .66, SD_{bk} = .14; β = -.114, t(55) = -2.176, p = .034, d = 1.13), but also for the trained group (M_{wh} = .77, SD_{wh} = .30; M_{bk} = .67, SD_{bk} = .14; β = -.150, t(55) = -2.607, p = .012, d = .45). No differences were present for Coloured participants, neither in the trained group (M_{wh} = .81, SD_{wh} = .10; M_{bk} = .74, SD_{bk} = .14; β = .037, t(55) = -.691, p = .492, d = .58), or in the untrained group (M_{wh} = .63, SD_{wh} = .21; M_{bk} = .71, SD_{bk} = .16; β = -.001, t(55) = -.019, p = .985, d = .43; Figure 4.3).

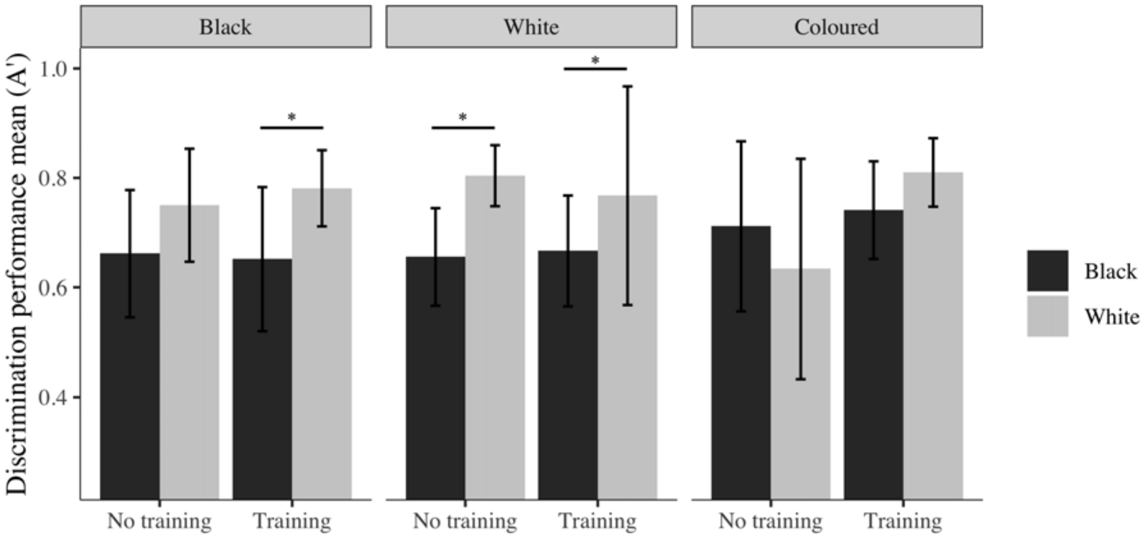

Figure 4.3 – Discrimination performance (A') in the post-test across stimuli group (Black; White) for each participant group (Black; White; Coloured) and training condition (training; no training). I bars are 95% confidence intervals.
$*p < .05$

Table 4.1 – *Mean, standard deviation and between-subject t-tests between training condition (training; no training) of discrimination performance (A') in post-test, across stimulus group (Black; White) and participant group (Black; White; Coloured).*

Part	Stim	Training	No training					
Group	Group	*M (SD)*	*M (SD)*	*b*	*t*	*df*	*p*	*d*
Black	Black	.65 (.18)	.66 (.21)	-.01	-.11	81	.912	.05
	White	.78 (.09)	.75 (.19)	-.03	-.41	81	.685	.20
White	Black	.67 (.14)	.66 (.14)	.03	.43	79	.671	.07
	White	.77 (.27)	.80 (.09)	-.01	-.07	79	.944	.15
Coloured	Black	.74 (.14)	.71 (.16)	-.08	-.99	73	.326	.20
	White	.81 (.10)	.63 (.21)	-.12	-1.41	73	.162	1.09

Post-hoc analyses also revealed no differences of discrimination performance between trained and untrained groups (Table 4.1). Results therefore suggest no effect of training, except for the difference between Black and White face discrimination performance of Black participants, which was significantly different only in the trained group. However, the effect size is smaller and the *p*-value marginal, suggesting a lack of statistical power.

4.3.3 Response bias in the pre-test and the post-test

One sample *t*-tests conducted on the pre-test response bias measure revealed that Black and White participants display a conservative bias (i.e., a tendency to answer ‘new’ more often than ‘old’ in the recognition phase) toward both Black and White faces, while Coloured participants only displayed a response bias towards Black faces (Table 4.2). That means, Black and White participants were more likely to falsely recognize both Black and White faces, whereas Coloured participants were only more likely to have this behavior toward Black faces, but not White faces.

Table 4.2 – *Mean, standard deviation and one sample t-test (against 0) of response bias (B″) in pre-test across stimulus group (Black; White) for each participant group (Black; White; Coloured).*

Part group	Stim group	*M (SD)*	*t*	*df*	*p*	*d*
Black	Black	.26 (.36)	3.22	19	**.005**	.72
	White	.24 (.49)	2.14	19	**.045**	.48
White	Black	.09 (.13)	2.80	16	**.013**	.68
	White	.35 (.44)	3.27	16	**.005**	.79
Coloured	Black	.23 (.29)	2.98	13	**.011**	.80
	White	.17 (.36)	1.77	13	.100	.47

Note. Significant p-values are in **bold**

During the post-test, the trained White group still displayed a conservative bias towards White faces, but not Black faces, while it was the opposite for the untrained group (Table 4.3), suggesting that their discrimination performance changed for the cross-group faces in regard to training. For the Black group, the conservative bias was only significant in the untrained group, not in the trained group, suggesting that training resulted in no response bias. In addition, the Coloured trained group only displayed a conservative bias toward Black faces. A mixed linear regression model testing the main effects and interaction effect of training and session along with the interaction between participant group and stimulus group (using participants as random

Table 4.3 – *Mean, Standard deviation and one sample t-test (against 0) of response bias (B″) in post-test across stimulus group (Black; White), participant group (Black; White; Coloured) and training condition (training; no training).*

Part	Stim	Training					No training				
Group	Group	*M (SD)*	*t*	*df*	*p*	*d*	*M (SD)*	*t*	*df*	*p*	*d*
Black	Black	.36 (.46)	2.07	6	.084	.78	.31 (.45)	2.47	12	**.029**	.69
	White	.16 (.38)	1.12	6	.306	.42	.35 (.46)	2.77	12	**.017**	.77
White	Black	.21 (.41)	1.37	6	.221	.52	.18 (.23)	2.53	9	**.032**	.80
	White	.62 (.39)	4.18	6	**.006**	1.58	.17 (.43)	1.28	9	.232	.41
Coloured	Black	.33 (.30)	3.46	9	**.007**	.28	.26 (.19)	2.68	3	.075	1.34
	White	.17 (.61)	.89	9	.396	1.09	.15 (.12)	2.60	3	.081	1.30

Note. Significant p-values are in **bold**

Table 4.4 – *Assignment of participants across training types, training session (1; 2; 3) for each participant group (Black; White; Coloured).*

		Black	White	Coloured	Total
Training 1	Control	13	10	4	27
	External	1	3	7	11
	Internal	6	4	3	13
Training 2	Control	12	10	5	27
	Bottom	4	4	5	13
	Top	4	3	4	11
Training 3	Control	12	10	5	27
	Configural	5	5	3	13
	Featural	3	2	6	11

effect) revealed no significant effect of training ($F(1, 47) = 0.04$, $p = .835$), session ($F(1, 148) = 1.11$, $p = .294$), the interaction between training and session ($F(1,148) = .83$, $p = .361$) or the interaction between stimulus group and participant group ($F(5, 102) = 1.52$, $p = .189$) on response bias.

4.3.4 Discrimination performance during training

The distribution of participants between the training group and training condition was quite balanced across participant groups (Table 4.4), with one or two exceptions. I expected to find a better performance for training on internal face features than training on external face features or no training in session 2, for training on top halves of faces than on bottom halves of faces or no training in session 3 and for training on configural processing than training on a featural processing or no training on session 4.

Therefore, regardless of the small sample size, three simple mixed linear models were run: One for each session, to analyze discrimination performance by training type, with participant as a random effect. None of the training types significantly predicted discrimination perfor-

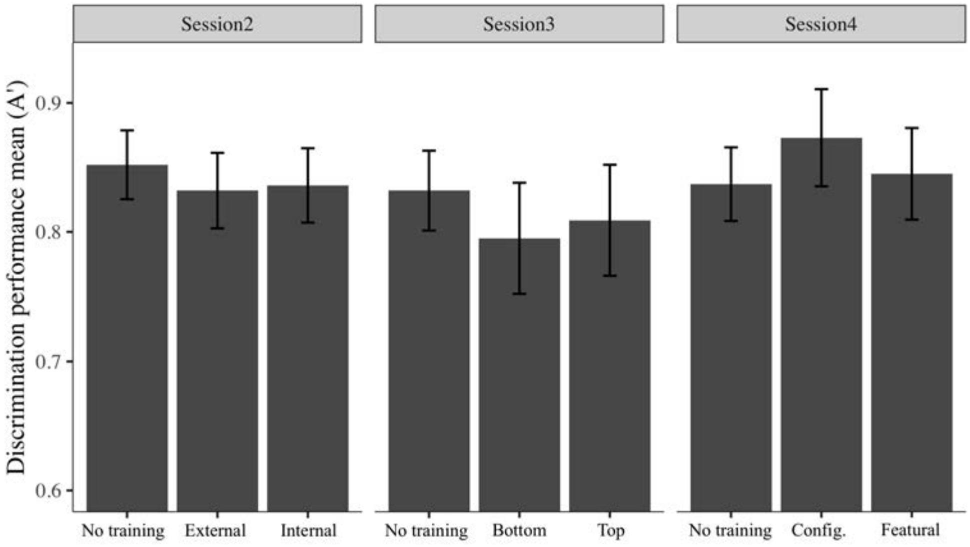

Figure 4.4 – Discrimination performance (A') across training type and training session (1 (Session 2); 2 (Session 3); 3 (Session 4)). I bars are 95% confidence intervals.
***$p < .001$; **$p < .01$; *$p < .05$

mance in session 2 ($F(1, 48) = 1.444$, $p = .246$), session 3 ($F(2, 48) = .178$, $p = .837$) nor session 4 ($F(2,48) = .293$, $p = .747$; Figure 4.4).

4.4 Discussion and conclusion

This study, created as a replication of Study 2, also tested to what extent a three week training regimen would reduce the OGB after the improvement in the material/protocol, which included a bigger sample size and the application to a different population. As expected, and as in Study 2, an OGB was present at pre-test for White participants. This suggests that White participants in France and South Africa present a bias in favor of own-group at the expense of other-group (in this case when the other-group is Black) and that this is regardless of their position as a racial majority or minority group in their country. Unexpectedly, Black participants did not present any bias, although a slightly higher discrimination performance was found for White faces than for Black faces. Coloured participants did not present any bias either, as expected; however, even if Coloured faces were not included as stimuli in any of the studies of the present thesis, it has to be acknowledged that a previous study found a greater performance for their own-group rather than for Black faces, but not for White faces (Seutloali, 2014). Considering that the expected minimum sample size of 11 participants per group was reached in the pre-test prior to group allocation for training (20 Black, 17 White, and 14 Coloured participants), I expected to find an OGB for Black participants as well. The absence of a bias in Black participants concurs with previous findings on the general population in South Africa (Wright et al., 2001) or in another South African University (Chiroro et al., 2008). However, previous research conducted at the same University as the current study sometimes found the OGB for White participants but not Black (or Coloured) participants (Seutloali, 2014; Wright et al., 2003), or for Black participants but not for White participants (Derbyshire, 2018). Considering the heterogeneity of these findings, they are discussed in the general discussion (p.225) with the results of the studies mentioned below.

In post-test, both trained and untrained White participants still present an OGB, as found in Study 2, and trained Black participants presented a significantly higher discrimination performance for other-group rather than own-group faces. Although not statistically significant, trained Coloured participants presented a greater discrimination performance for White than Black participants (as in pre-test), whereas the reverse result was present for untrained Coloured participants. These results suggest that there was no effect for training, at least not in the expected direction, and an overall decrease of the discrimination performance for own-group decisions as a function of training. In addition, there were no differences for the different training instructions, in line with the results of Study 2. However, this comparison may be under-powered, as once broken down by participants group and training type, sub samples were from three to 13 participants each. The under-powering of the study is a result of the dropout rate of my participants. Indeed, considering the total sample (N = 102), I expected to have a sufficient number of participants to draw stronger conclusions on any training effect if I had a lower dropout rate. During data collection, half of the sample withdrew, raising the dropout rate to approximately 11% between every two sessions. Although I conducted the data collection twice to address this issue, it remains that considering time constraints and the results from Study 2 that showed an absence of any training effect, I decided against repeating the study a third time and to analyze the data as it was. In Study 2, I also lost half of my sample as 13 of the 27 students did not attend at least one of their sessions, and their presence was compulsory for their data to be included. The issue of participants volunteering for, and completing, multiple-session studies is a limitation to point out.

When comparing this study and Study 2, as well as previous studies exploring training effects, differences in training design have to be acknowledged. First, previous studies on multiple-sessions training (Goldstein & Chance, 1985; Lebrecht et al., 2009; Matthews & Mondloch, 2018; McGugin et al., 2011; Stahl, 2010; Tanaka & Pierce, 2009) asked participants to complete between three and seven sessions across consecutive days or over a maximum of two weeks, whereas my participants came weekly (i.e., implying a 6-7 days delay) over five weeks. This delay might have been to long to induce any skills extraction from the tasks. This idea is also supported by the study conducted by Lavrakas et al. (1976) which revealed an elimination of the OGB when tested immediately after training, whereas when tested after one week of delay the training effect disappeared.

In previous studies, participants were also asked to complete the exact same task over the sessions, whereas I asked my participants to complete different tasks at each of the sessions, which also needs to be considered as a limitation. These two differences might have had a deleterious effect on learning, and particularly on consolidation. In addition, post-test was administered one week after the last training session, whereas an immediate measure might have given different results. Then, the absence of feedback in my training regimen might have reduced the full potential of training. Indeed, experience itself seems to be insufficient without feedback to reinforce learning (Estudillo & Bindemann, 2014), and past studies revealed an increase of performance as a result of feedback (Hussain et al., 2009; White, Kemp, Jenkins, & Burton, 2013). In my studies, the lack of feedback on participants' success/failure following each instruction meant participants did not know which of the instruction would help them most to discriminate and recognize faces. They may therefore have reverted to their automatic strategy after training, causing the absence of a generalization effect. I believe feedback, and as was the case for Study 1, would have led participants to use the more successful strategy taught in training, and hopefully would have resulted in an improvement of discrimination performance.

The design of training also has several limitations. First, it involved many different steps and instructions. Then, although it made theoretical sense to ask participants to focus on the internal features, or to use a configural exploration for both stimuli groups, this was different for the third training regimen. As the bottom halves of faces are more relevant for processing Black faces whereas the top halves of faces are more relevant for processing White faces (Ellis et al., 1975; Hills et al., 2013; Shepherd & Deregowski, 1981), asking participants to focus on those halves regardless of stimuli group could have resulted in an absence of discrimination performance difference. However, regarding to the results from Study 1, this is unlikely. In addition, the way of introducing configural or featural processing (i.e., simple instructions to follow) might not have ensured such processing. Indeed, studies usually use blurred versus scrambled faces (Collishaw & Hole, 2000; Sadozai et al., 2018) or upright versus inverted faces (Bartlett & Searcy, 1993; Bruce, 2008) to respectively create configural versus featural processing. To conclude on training, it would have been interesting to either repeat the same task in different parallel studies (e.g., train participants to only focus on the internal part of faces during each session of a longitudinal training), or to develop an optimum regimen only, which included the internal and configural focus instructions.

To control for generalization, I also used different picture manipulations in between the two OGB measure tasks (i.e., pre-test and post-test) and the training tasks. Indeed, in both pre-test and post-test, faces were presented from a frontal view during encoding and a three-quarter view during recognition. In the three training tasks, faces were presented from frontal view during encoding, and from frontal view but vertically flipped and grey-scaled during recognition. In this manner, the observed effect of generalization in post-test, if observed, could not have been a consequence of training to recognize faces from their three-quarter view when encoded from their frontal view, as a result of practice.

The group categorization used in this study was also not optimum, and this strategy needs to be improved if used in future studies. When I adapted the design for South Africa, I did not consider the differences in asking for racial groups in the language that is commonly used there, given practices in France (where it is not permitted in law to ask questions about race membership directly). South Africans on the other hand, are used to giving this information and are indeed required to do so by law in many situations. I thus used the same phototype scale as for studies 1, 2 and 3, and asked about the country of birth of parents and grandparents. In these studies, conducted in France, it was quite an efficient way of controlling my inclusion criterion (participants had to be White on the phototype scale, Fitzpatrick, 1988), without asking for racial information directly. However, in the case of other racial groups, the Fitzpatrick's classification (1988) lacks relevance and exhaustiveness. Assuming that White people could be phototypes I, II and III, Coloured people phototype IV, and Black people phototypes V and VI, this is inadequate. Specifically, it was a limitation for Coloured participants, whom may have categorized themselves as phototype V and defined themselves as Coloured. The country of birth was completely irrelevant in South Africa to resolve such a hypothetical case – Coloured people are descendants of indigenous Khoi-San (who have lived in South Africa for thousands of years), slaves from Indonesia who arrived over 300 years ago, and the European settlers who arrived from the 17th century onwards, but almost every grandparent of Coloured participants was born in South Africa. As no perfect solution was available at the time, and data already collected, I decided to categorize my participants as planned, with the phototype scale. I thus kept phototypes I, II and III as White and removed participants with Asian ancestors, classified phototype IV as Coloured, and classified phototypes V and VI as Black.To solve the limitation of categorization, one could ask for racial group and phototype, using the former for categorization and the latter for comparison purposes (see general discussion, p.225), while keeping in mind the possibility of constructing a new categorization.

In the studies 5 and 6, I addressed issues observed in the initial studies. However, because of the limitations of such a prolonged training task, and the results from Study 1 showing the increase of the OGB after training participants to focus on the bottom halves of faces, I considered returning to a single session training model. In fact, it is easier to explore the presence or absence of an effect for training when training contains one unique task. That is, the following studies are not direct follow-ups to the two multiple-sessions studies. Also, in this research participants were randomly assigned to a group prior to the collection of their own group belonging which limited the distribution of a comparable sample size across conditions. This issue was addressed by asking group belonging prior to training condition attribution. Finally, as in Study 1, feedback was added to the training tasks for studies 5 and 6.

Study 5

Training South African participants through a task-specific visual practice does remove the own-group bias

Study 5 was pre-registered on the Open Science Framework prior to data collection (publicly accessible at `osf.io/kg8hd`).

Matching tasks are often performed by the police or control officers, and since these tasks can include degraded images (e.g., from CCTV footage), it is of some interest to consider a training regimen that would address this difficulty directly. Similarly to memory tasks, in matching tasks it is difficult to perform well with unfamiliar faces (for a review, see Fysh & Bindemann, 2017) and this results in an even higher inaccuracy with other-group faces (Kokje, Bindemann, & Megreya, 2018; Megreya, White, & Burton, 2011). While some studies have focused on the development of specific tasks to improve matching performance on unfamiliar own-group faces, none studied matching performance with other-group faces, and none developed training for generalization. Training people seemed promising since Forensic examiners have been demonstrated to be experts and perform better and with less error than the general population in matching different images of people, suggesting that training could offer promising improvements in these skills (White, Phillips, Hahn, Hill, & O'Toole, 2015). From what is known from studies conducted in the general population, the presentation of multiple exemplars would help to increase matching performance for unfamiliar faces (Bindemann

& Sandford, 2011; White, Burton, et al., 2014). However, this method is not viable in the field, as often only one image per identity is available (e.g., passport control). Alternatively, a feature-to-features task has been tested in professionals, and resulted in a small effect after training alongside other tasks (Towler et al., 2019), or no effect on face matching when compared to the effects of configural processing (Megreya, 2018). To the same end, Megreya and Bindemann (2018) asked Arab participants to focus on specific parts of faces during matching own-group and other-group (i.e., Caucasian) face pairs. For own-group face pairs, participants who were asked to focus on the eyebrows performed better than participants who were asked to focus on the eyes (i.e., no differences), or participants who focused on the ears and who performed the worst. No improvement was visible for other-group face pairs. Featural processing may therefore have some limitations, since it requires good quality images which is also rarely possible in the field.

Considering the unexpected ceiling effect observed in Study ‘X’ (96% accuracy; see p.101) when matching original color pictures to vertically flipped grey-scales pictures, I sought another alteration that could be applied but would be more difficult and would therefore prevent a ceiling effect. The present study was therefore inspired by Bindemann et al. (2013) who conducted four experiments on matching by manipulating image size and pixelation (i.e., resolution) to degrade quality. They found that unfamiliar face matching is strongly affected by picture resolution, with a large accuracy drop between original-to-original and original-to-pixelated pairs (experiment 1). They also observed that reducing the image size could partially reverse the deleterious effect of pixelation (experiment 2), especially when the pixelated face, relative the original face, was reduced in size (experiment 3). However, pixelation and pose (i.e., frontal versus profile) both have deleterious main effects but no interaction effect (experiment 4). I decided to manipulate picture resolution with a pixelation filter while keeping other pictures parameters (e.g., pose) stable and standardized. Since accuracy decreases as a function of resolution deterioration (Bindemann et al., 2013), training participant on a gradual pixelation would result in a better understanding of the extraction of facial features from a pixelated picture and thus a greater performance and generalization after training for matching original-to-pixelated pictures.

5.1 Hypotheses

First, I expected to find an OGB in the pre-test and training tasks for White and for Black participants: In the case of White participants, I expected replication of previous results, and for Black participants, the use of a larger sample size led to the expectation, although not very strong, that I would detect an OGB for them. Secondly, I expected a smaller OGB in the post-test as a function of training. Finally, I expected the decision time to be quicker for own-group rather than other-group trials, and quicker for the trained group than for the untrained group[22].

5.2 Method

5.2.1 Population

A total of 156 participants were recruited through the Student Research Participation Program (SRPP) of the University of Cape Town (117 women, M_{age} = 19.36, SD_{age} = 1.47). The sample size was constrained by the time that could be allowed for data collection. Participants were awarded with two course credit points after the completion of the study. The final sample

---

[22] Another hypothesis, which has been pre-registered, was that own-group trials in the post-test were expected to be completed more quickly than in the pre-test. However, it was judged to be irrelevant to test afterward since I cannot directly compare pre-test and post-test measure without including training.

contained data from 140 participants (108 women, 40 Black, 69 White, 31 Coloured, M_{age} = 19.30, SD_{age} = 1.49), while 16 participants were removed because they did not belong to one of the studied groups (Black, White, Coloured) or because they had claimed to have seen one of the pictures before the study.

5.2.2 Design

The present study has a three variables mixed factorial design with training (training; no-training) as a between-subject variable and stimulus group (White; Black) and the OGB measure (pre-test; post-test) as within-subject variables.

5.2.3 Material

Stimuli

A total of 724 pictures of 308 different males were used in the present study. They were retrieved from tree database: UCT2005, UCT2007 and RaFD. For the pre and post-tests, a total of 200 different male faces were used resulting in pictures of 50 different faces of each group in each OGB measure. All the pictures, which were grey-scaled, were cropped from the neck upwards, against a grey background, and were displayed in two versions: an original version or a highly pixelated version, resulting in a total of 400 different pictures. Mismatching pairs were constructed on a similarity basis by the author, selecting among the pictures not used for the training task, to create a total of 50 pairs for each group. Each of the two faces of each pair were presented as an original, at the same time as its pixelated match (same face) or pixelated mismatch (the picture selected to be similar).

Figure 5.1 – Example of the three levels of pixelation, from left to right: original, medium, high. Pictures were not used during the study, but presented here for illustration purpose with the consent of the model.

For the training task, a total of 108 different faces were used, all grey-scaled, cropped from the neck upward, on a grey background and divided into three versions: the original, a medium pixelation (i.e., one pixel is the size of 5x5 pixels of the original picture) and a high pixelation (i.e., one pixel is the size of 10x10 pixels of the original picture, Figure 5.1), resulting in 324 different pictures.

The pictures were chosen from the same database as for the pre-test and post-test tasks: UCT2005, UCT2007 and RaFD, on the basis of their similarity to each other. Eighteen trials (9 of each group) were constructed, broken down into three difficulty levels: easy, intermediate and difficult. The difficulty of each level was manipulated by increasing or decreasing picture similarity, as rated by two independent judges (one from each group: Black; White). In the easy level, faces were all men, but from different perceived age, build, hair color. In the intermediate level, they were more similar to each other, sharing the same perceived age, build, and hair color. In the difficult level, faces were even more similar to each other, sharing the same perceived age, build and hair color, hair shape, eye colour and skin tone. In order to construct these groups of faces (k = 6 faces per group), the two judges picked from the database a first pair of similar faces. Then, they had to find a third face that would match the two from the pair, and so on until a total of six faces was found. The difficulty level was constructed first in order to have the highest possibility of good matching.

Apparatus

All stimuli were modified using GIMP 2.8.14 software (GNU Image Manipulation) as 449x569 pixels. OGB measure and training tasks were constructed and displayed through E-prime 3.0 software (Psychology Software Tools, Pittsburgh, PA). The tasks were completed on computers with 17" to 21" screens[23]. An additional survey for demographic data and group self-attribution was created and displayed using Qualtrics online survey tool (Qualtrics, Provo, UT). The online game "QWOP" (`www.foddy.net/Athletics.html`) was also used as a filler task: The aim of this game is to make an ‘avatar’ athlete run, using a computer keyboard.

5.2.4 Procedure

In the present study, participants were asked to complete three tasks: the pre-test, the training (or no training) task and the post-test (Figure 5.2). After giving their consent, participants completed the demographic questionnaire to record information about their gender, age, study year, group and phototype. Group was asked with the question "Do you define yourself as", offering the following options: Black, White, Coloured, Mixed, Asian or Other (with an instruction asking them to specify). Only participants who selected Black, White or Coloured were kept in the sample for analyses, but other participants were allowed to complete the study, and their data discarded. At the end of the questionnaire, a participant number was

---

[23] Screen size was different, but that did not affect the size of the stimuli during the presentation.

awarded, and random assignment to experimental group (training; no training) was done by the software.

Pre-test and post-test OGB measures

The two measures of OGB, in the pre-test and the post-test, were identical in procedure, but two different sets of pictures were displayed. The sets of pictures were randomly composed from the constructed pair and were counterbalanced among the participants. In this regard, participants were told the following:

> During the current task, pairs of faces will be presented as follows [*illustration*]. Each pair could depict either the same person or two different people. For each pair, you then have to answer the question "*Do these pictures depict the same person*?"

They were then told to use the keys ‘F’ for ‘Yes, they do’ or ‘J’ for ‘No, they do not’ from the keyboard to record their answers. After an example, they were instructed:

> Please notice that to make your decision you must rely on the physical features of the person instead of pictures characteristics (lighting, size). Please complete the task as fast and as accurately as possible. If you have understood the instructions, please press the space bar to start the task. If you have any questions before starting the task, please call the researcher.

The task began, and 100 pairs were randomly displayed (50 matching and 50 not matching), for each stimulus group. All the pictures were presented once, together with its matching face, or its mismatching face. The two pictures for each pair were presented: either both with their match, or both with their mismatch. At the end of this task, the second (training; no training) task started according to the participants assignment.

Training task

The training task was a matching task involving two different levels of pixelation (medium; high) and the original version of each identity. In this regard, participants had to first match the original picture to the medium pixelation version and secondly, match the medium pixelation version to the high level of pixelation, according to the following instructions:

> In this task, a series of faces will be displayed on the screen. In each series, you will be shown a ‘target’ face, and six other photographs of varying quality. Your aim is to find, among the six pictures, which one is related to the target face. You will see the target face on the right-hand side of the screen and all the possible matches on the left-hand side of the screen. Use the numerical pad to record your answers. Feedback will then be displayed: If your answer is incorrect, you will be asked to restart the set. If your answer is correct, you will then proceed to the next step. Please note that you do not have any time limit to complete the task.

After an example and an explanation of the use of the numeric keypad for recording their answers, participants were told that the kind of feedback they would receive was a follows:

> If you choose the correct match, ‘correct, well done’ [displayed in blue, with the chosen face encased by the same color] will be displayed on the screen and you will proceed to the next set. If you choose an incorrect match, ‘incorrect, try again’ [identical as before but in red] will be displayed on the screen and you will

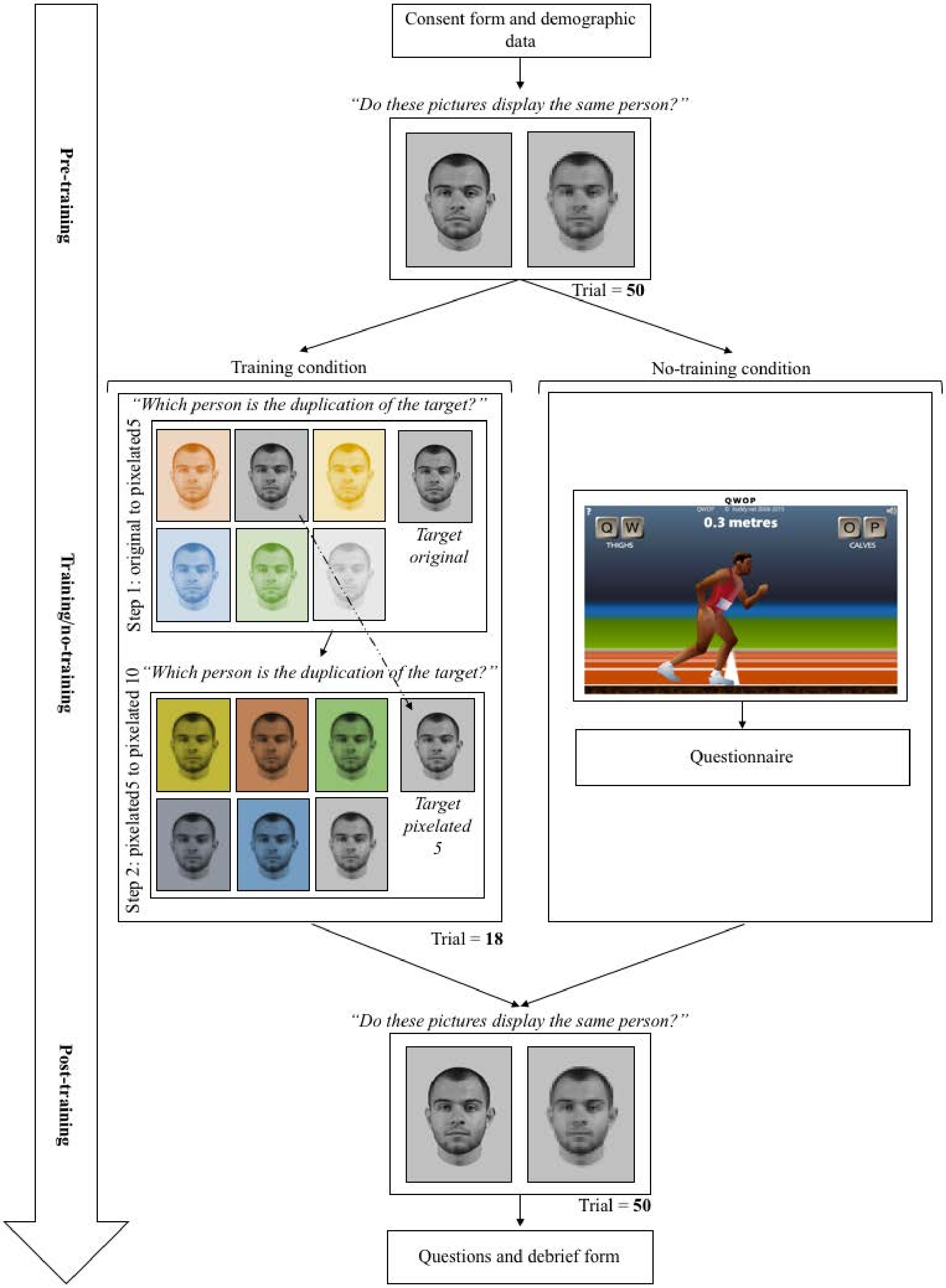

Figure 5.2 – Flow chart of the procedure for Study 5. Participants completed the pre-test task, followed by either the training or the no training task and, the post-test task. Pictures were not used during the study, but presented here for illustration purpose with the consent of the model.

> have to give a new answer, until you find the correct match. Note that no matter the nature of the feedback, you need to press the space bar either to go to the next set or to submit a new answer.

The task began right after the instructions. Participants were presented with a series of 18 trials in total, containing a target face on the right-hand side of the screen (randomly chosen from the six similar faces of the trial) and the six medium pixelated versions on the left-hand side of the screen. Once the correct decision was made, the target on its medium pixelated version (i.e., the correct answer from the six) was now displayed on the right-hand side of the screen with the six highly pixelated pictures on the left-hand side of the screen. After each decision, participants received feedback on their answer and had to try again in the case of an incorrect answer or could proceed to the next step in the case of a correct answer. The easy trials were completed first, followed by the intermediate ones and the difficult ones. The untrained group of participants played a game for 5 minutes, and answered to a questionnaire about the game. The task was of approximately the same duration of the training task.

At the end of the tasks, participants were asked if they have already seen one of the faces prior to the present study, and were given access to the debriefing form that they could read. They were free to ask any questions they may have had. The entire study lasted about 40 minutes.

5.3 Results

5.3.1 Discrimination performance in pre-test

Overall, the matching accuracy during the pre-test task was quite high (*M* = .94; *SD* = .04). A mixed linear model was constructed to predict discrimination performance with the main effects and interaction effect between participant group and stimulus group and participant as a random effect. The model resulted in no significant main effect of stimulus group ($F(1, 133) = 2.04$, $p = .155$), participant group ($F(2,133) = .47$, $p = .625$) or their interaction ($F(2, 133) = 1.58$, $p = .210$). However, I decided to conduct *post-hoc* analyses anyway considering from descriptive data that an OGB could be present for White participants but not visible because of the direction of Black participants (in favor of White faces) also observed in previous studies.

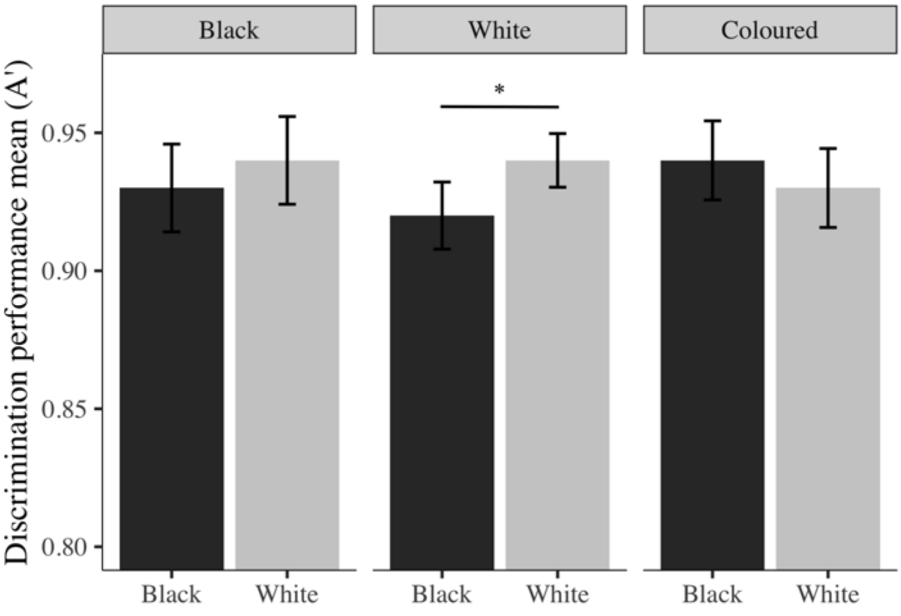

Figure 5.3 – Discrimination performance (A') in pre-test across stimulus group (Black; White) for each participant group (Black; White; Coloured).
$^{*}p < .05$

Post-hoc analyses therefore revealed the presence of an OGB for White participants (M_{wh} = .95, SD_{wh} = .04; M_{bk} = .92, SD_{bk} = .05; β = -.01, t(136) = -2.61, p = .010, d = .67) but not for Black participants (M_{bk} = .93, SD_{bk} = .05; M_{wh} = .94, SD_{wh} = .05; β = -.01, t(136) = -.69, p = .492, d = .20) or Coloured participants (M_{bk} = .94, SD_{bk} = .04; M_{wh} = .93, SD_{wh} = .04; β = .00, t(136) = .27, p = .791, d = .50; Figure 5.3). The OGB is therefore also found for White participants in a matching task but not for Black participants.

5.3.2 Discrimination performance in post-test

After the training, the overall performance in the trained group was higher (M = .96, SD = .04) than in the untrained group (M = .93, SD = .05), suggesting an effect of training. To test the effect of training, a mixed linear model was run using training as main effect, mains and interaction effect between stimulus group and participant group and participant as a random effect. The model resulted in a significant main effect of training (F(1,133) = 5.47, p = .021) but not of stimulus group (F(1, 133) = 2.24, p = .137), participant group (F(2, 133) = 1.09, p = .341) or their interaction (F(2,133) = .84, p = .432). *Post-hoc* analyses revealed an other-group bias for untrained White participants (M_{wh} = .93; SD_{wh} = .05; M_{bk} = .91, SD_{bk} = .07; β = -.19, t(139) = -2.68, p = .008, d = .33) but not for trained White participants (M_{wh} = .95, SD_{wh} = .04; M_{bk} = .94, SD_{bk} = .04; β = -.01, t(139) = -.376, p = .705, d = .05) suggesting an absence of the bias in the trained group (Figure 5.4). This effect was not present for the Black or Coloured participants (Table 5.1), who, as a reminder, did not display any OGB in pre-test.

In addition, while recognizing Black faces, White trained participants also showed better performance (M = .95, SD = .04) than untrained participants (M = .91, SD = .07; β = -.03,

Table 5.1 – *Discrimination performance (A′) mean, standard deviation and within-subject t-test across stimulus group (Black; White) for each participant group (Black; White; Coloured) and training condition (training; no training).*

Part	Stim	No training					
Group	Group	*M (SD)*	*b*	*t*	*df*	*p*	*d*
Black	Black	.92 (.05)	-.00	-.34	139	.732	.06
	White	.92 (.05)					
White	Black	.91 (.07)	-.19	-2.68	139	**.008**	.33
	White	.93 (.05)					
Coloured	Black	.94 (.04)	-.01	-.42	139	.675	.10
	White	.95 (.03)					
		Training					
		M (SD)	*b*	*t*	*df*	*p*	*d*
Black	Black	.94 (.05)	.00	.28	139	.780	.06
	White	.94 (.05)					
White	Black	.95 (.04)	-.01	-.376	139	.705	.05
	White	.95 (.04)					
Coloured	Black	.94 (.05)	-.01	.60	139	.553	.13
	White	.94 (.04)					

Note. Significant *p*-values are in **bold**

$t(198) = -2.82$, $p = .005$, $d = 1.14$), suggesting an increase in discrimination performance as a function of training for White participants. No differences were present for Black or Coloured participants (Figure 5.4).

5.3.3 Hit and False alarms in post-test

Further analyses were conducted on hits and false alarms in White and Black participants only, in order to explore where the differences were that extinguished the OGB in White participants in post-test, and made the Black participants display an equal performance in both training and no training groups. Complete descriptive data on hits, false alarms, correct rejections and misses are in Table 5.2. An initial mixed linear model was conducted to predict

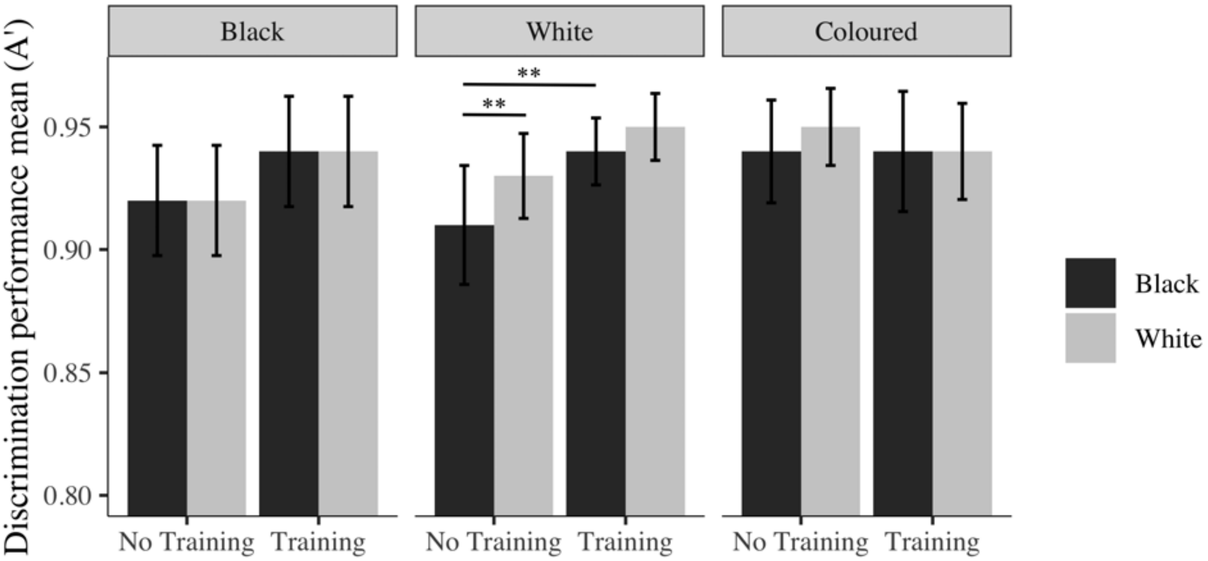

Figure 5.4 – Discrimination performance (A') in post-test across stimulus group (Black; White) and training group (training; no training) for each participant group (Black; White; Coloured). I bars are 95% confidence intervals.
$*p < .01$

hits in the post-test task, including the main effects and interaction of stimulus group and participant group, the main effect of training and participant as a random effect. The model resulted in a significant main effect of stimulus group ($F(2, 103) = 13.83$, $p < .001$) and of the interaction between stimulus group and participant group ($F(1, 103) = 5.76$, $p = .018$) but no significant main effect of participant group ($F(1, 103) = .48$, $p = .492$) or training ($F(1, 103) = 3.37$, $p = .69$). *Post-hoc* analyses were conducted on stimulus group and participant group without training, and revealed a significantly higher number of hits for Black participants on Black faces ($M = 23.39$, $SD = 1.82$) rather than White faces ($M = 21.89$, $SD = 2.41$; $\beta = 1.50$, $t(105) = 3.81$, $p < .001$, $d = .71$). There were no differences for White participants ($M_{bk} = 23.08$, $SD_{bk} = 2.31$; $M_{wh} = 22.75$, $SD_{wh} = 2.41$; $\beta = .32$, $t(105) = 1.07$, $p = .285$, $d = .43$). A second model, identical to the first but predicting false alarms rates, resulted in a significant

Table 5.2 – *Mean and standard deviations of hit, false alarms, misses and correct rejections in pre-test and post-test across stimulus group (Black; White) for each participant group (Black; White; Coloured) and training condition (training; no training).*

		Pre-test			
Part. group	Stim. group	Hit	False alarm	Miss	Correct rejection
		M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Black	Black	23.79 (1.32)	4.76 (3.68)	1.21 (1.32)	20.24 (3.68)
	White	22.84 (1.97)	3.26 (2.84)	2.16 (1.97)	21.74 (2.84)
White	Black	23.60 (1.57)	5.45 (3.32)	1.40 (1.57)	19.55 (3.32)
	White	22.68 (2.46)	3.48 (3.09)	2.32 (2.46)	21.52 (3.09)
Coloured	Black	23.63 (1.45)	4.43 (3.66)	1.37 (1.45)	20.57 (3.66)
	White	22.23 (2.90)	3.13 (2.65)	2.77 (2.90)	21.87 (2.65)

		Post-test			
Part group	Stim group	No training			
		M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Black	Black	23.05 (2.01)	5.74 (4.56)	12.26 (4.56)	1.95 (2.01)
	White	21.58 (2.29)	3.63 (3.37)	21.37 (3.37)	3.42 (2.29)
White	Black	22.66 (2.71)	5.03 (3.36)	19.97 (3.36)	2.34 (2.71)
	White	22.47 (2.72)	3.59 (2.98)	21.41 (2.98)	2.53 (2.72)
Coloured	Black	22.79 (2.15)	2.79 (2.72)	22.21 (2.72)	2.21 (2.15)
	White	22.64 (2.17)	2.43 (2.71)	22.57 (2.71)	2.36 (2.17)
		Training			
Black	Black	23.74 (1.59)	4.05 (3.05)	20.95 (3.05)	1.26 (1.59)
	White	22.21 (2.55)	2.74 (2.49)	22.26 (2.49)	2.79 (2.55)
White	Black	23.48 (1.79)	3.52 (2.87)	21.48 (2.87)	1.52 (1.79)
	White	23.03 (2.07)	2.79 (2.23)	22.21 (2.23)	1.97 (2.07)
Coloured	Black	22.94 (2.43)	3.56 (3.37)	21.44 (3.37)	2.06 (2.43)
	White	22.75 (2.41)	2.75 (2.86)	22.25 (2.86)	2.25 (2.41)

main effect of stimulus group ($F(1, 103) = 28.63$, $p < .001$) and training ($F(1, 103) = 4.91$, $p = .029$). However, there were no significant effects of participant group ($F(1, 103) = .30$, $p = .588$) nor the interaction between participant group and stimulus group ($F(1, 103) = 1.48$, $p = .227$). *Post-hoc* analyses were conducted on training only, without any distinction of participant or stimulus group, and resulted in an overall significantly higher false alarms rate for the untrained group ($M = 4.45$, $SD = 3.56$) than for the trained group ($M = 3.24$, $SD = 2.66$; $\beta = 1.21$, $t(106) = 2.18$, $p = .031$, $d = .39$). Additional *post-hoc* analyses were performed on stimulus group according to participant group, and showed a significantly higher false alarm rate for Black than White faces, for Black participants ($M_{bk} = 4.89$, $SD_{bk} = 3.92$; $M_{wh} = 3.18$, $SD_{wh} = 2.96$; $\beta = 1.71$, $t(105) = 4.09$, $p < .001$, $d = .50$) and White participants ($M_{bk} = 4.26$, $SD_{bk} = 3.19$; $M_{wh} = 3.18$, $SD_{wh} = 2.64$; $\beta = 1.08$, $t(105) = 3.37$, $p = .001$, $d = .37$). These results suggest that the absence of OGB in the trained group is not due to a hit rate, which is not significantly different according to the training group, but to a lower number of false alarms in the trained group as compared to the untrained group. These results are even more interesting for Black participants who, unexpectedly, also had a higher false alarm rate for Black than for White faces (i.e., in the same direction as White participants).

5.3.4 Response bias in the pre-test and the post-test

One sample *t*-tests of B$''$ were computed for each participant group toward each stimulus group, in the pre-test and the post-test, taking into account the training or no training group. Results revealed that the tendency of being liberal, namely a tendency to answer ‘yes’ more often than ‘no’ during the matching task, was significant for all the participant groups towards Black faces but not towards White face, in pre and post-tests, regardless of the training group, except for the Coloured group in the post-test condition (Table 5.3). An initial mixed linear model was run to predict the response bias in the pre-test, with main effects and an interaction between participant group and stimulus group along with participant as a random effect.

Table 5.3 – *Mean, standard deviation and one sample t-test (against 0) of B'' in the pre-test and the post-test across stimulus group (Black; White) for each participant group (Black; White; Coloured) and training condition (training; no training).*

Pre-test						
Part group	Stim group	*M (SD)*	*t*	*df*	*p*	*d*
Black	Black	-.47 (.64)	-4.44	35	**<.001**	.74
	White	-.13 (.67)	-1.18	36	.245	.19
White	Black	-.55 (.42)	-10.38	64	**<.001**	1.29
	White	-.13 (.63)	-1.68	63	.099	.21
Coloured	Black	-.43 (.60)	-3.89	28	**<.001**	.72
	White	-.00 (.67)	-.01	28	.988	.00

Post-test						
Part group	Stim group	No training				
		M (SD)	*t*	*df*	*p*	*d*
Black	Black	-.40 (.59)	-3.28	18	**.004**	.75
	White	.09 (.56)	.67	16	.509	.16
White	Black	-.41 (.55)	-4.17	30	**<.001**	.75
	White	-.23 (.68)	-1.95	31	.060	.34
Coloured	Black	-.09 (.77)	-.45	13	.659	.12
	White	.10 (.75)	.51	13	.617	.14
		Training				
Black	Black	-.44 (.64)	-2.98	18	**.008**	.68
	White	.01 (.72)	-.71	14	.487	.18
White	Black	-.40 (.61)	-3.64	30	**.001**	.65
	White	-.20 (.67)	-1.68	30	.104	.30
Coloured	Black	-.12 (.72)	-.65	15	.527	.16
	White	-.12 (.64)	-.71	14	.487	.18

Note. Significant *p*-values are in **bold**

The effect of stimulus group was significant ($F(1, 130) = 38.70$, $p < .001$), *post-hoc* analyses resulting in a higher bias (i.e., a more negative B″ value) toward Black than White faces, for Black participants ($\beta = -.33$, $t(135) = -2.94$, $p = .004$), White participants ($\beta = -.41$, $t(131) = -4.93$, $p < .001$), and Coloured participants ($\beta = -.41$, $t(135) = -3.29$, $p = .001$). However, there was no significant effect of participant group ($F(2, 132) = .69$, $p = .509$) or of the interaction between participant group and stimulus group ($F(2, 130) = .21$, $p = .814$).

A second model was run to predict the response bias in the post-test, with a main effect and an interaction between participant group and stimulus group, a main effect of training and participant as a random effect. The model resulted in no significant effect of participant group ($F(2, 128) = 2.48$, $p = .087$) but a significant effect of stimulus group ($F(1, 125) = 14.94$, $p < .001$). The interaction between stimulus group and participant group was not significant ($F(2, 125) = 2.81$, $p = .064$), the main effect for training was also not significant ($F(1, 128) = 0.102$, $p = .750$). *Post-hoc* analyses of the interaction of stimulus group and participant group, independent of the training, resulted in a significantly higher response bias towards Black faces than White faces for Black participants ($M_{bk} = -.42$, $SD_{bk} = .58$; $M_{wh} = .05$, $SD_{wh} = .64$; $\beta = -.48$, $t(133) = -4.01$, $p < .001$, $d = .77$), but no significant difference for White participants ($M_{bk} = -.40$, $SD_{bk} = .57$; $M_{wh} = -.22$, $SD_{wh} = .67$; $\beta = -.17$, $t(131) = -1.91$, $p = .059$, $d = .29$) or for Coloured participants ($M_{bk} = -.11$, $SD_{bk} = .74$; $M_{wh} = -.01$, $SD_{wh} = .70$; $\beta = -.11$, $t(130) = -.84$, $p = .401$, $d = .07$). In the pre-test and the post-test, participants still displayed the same bias: a liberal response bias which led them to have a higher tendency to answer ‘no’ in the target-absent trials. This bias was significantly present towards Black faces for both Black and White participants before and after the test, independent of the training, while Coloured participants displayed such a bias only in the pre-test condition.

5.3.5 Decision time in the post-test

Decision time was directly measured using software (E-prime), from the display of the trial until the decision was made. Time was measured in milliseconds, but were log transformed for the analysis since it was highly skewed. A mixed linear model was run to predict decision time, including a main effect and interaction of stimulus group and participant group with the main effect of training and participant as a random effect. The model resulted in no significant effect of stimulus group ($F(1,133) = .48$, $p = .491$), participant group ($F(2,133) = .75$, $p = .476$), training ($F(1,133) = 1.43$, $p = .235$) or the interaction between stimulus group and participant group ($F(2,133) = 2.84$, $p = .062$).

5.3.6 Discrimination performance and decision time during training

In order to understand the training effect, a mean number of trials to succeed, and response time were explored as dependent variables. Number of trial corresponds to the number of trials it took to get the right answer, since participants got feedback and had to start the trial again until selecting the right answer. A mixed linear model with main effects and interaction of participants group and stimulus group with participant as a random effect was run to predict the number of trials. None of the variables showed a significant prediction, neither participant group ($F(1,68) = 1.03$, $p = .314$), nor stimulus group ($F(1, 68) = 2.14$, $p = .148$) nor their interaction ($F(1, 68) = .54$, $p = .464$). However, descriptive data (Table 5.4) suggest an overall tendency to complete a higher number of trials for Black than White faces for successful matching. The expected own-group bias during the training was therefore not found.

The same type of mixed linear model was adapted to predict the log transformed decision time, which resulted in a significant effect of stimulus group ($F(1,68) = 83.50$, $p < .001$) but not of participant group ($F(1,68) = 2.98$, $p = .057$) or their interaction ($F(1,68) = .33$, $p = .723$). *Post-hoc* analyses revealed a significantly higher decision time for Black than White faces, for all participants: Black ($\beta = .08$, $t(71) = 5.03$, $p < .001$, $d = .59$), White ($\beta = .09$, $t(71) = 7.16$, $p < .001$, $d = .67$) and Coloured ($\beta = .07$, $t(71) = 4.03$, $p < .001$, $d = .64$; Table 5.4).

5.4 Discussion and conclusion

Study 5 aimed to assess whether performance in an original-to-pixelated pictures matching task could be improved from gradient training. The overall observed performance in this study was quite high (i.e., 94% in average in the pre-test), which was unexpected. Indeed, in a study using the same paradigm of original-to-pixelated pictures matching, Bindemann et al. (2013) found an accuracy rate of approximately 62% for their easier pixelated conditions (i.e., original to low pixelation). This difference could be due because that in their study, as in many matching studies (e.g., Estudillo & Bindemann, 2014; Kokje et al., 2018; Kramer & Reynolds, 2018), they used pictures from the Glasgow Face Matching Test (Burton et al., 2010), in which

Table 5.4 – *Mean of number of trials and mean of log transformed decision time during the training task across stimulus group (Black; White) for each participant group (Black; White; Coloured).*

	Mean number of trials		Mean decision time	
	Black	White	Black	White
Part. group	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Black	19.11 (2.66)	18.63 (1.21)	3.70 (.14)	3.62 (.13)
White	18.64 (.90)	18.21 (.55)	3.62 (.13)	3.54 (.11)
Coloured	18.56 (.89)	18.44 (1.50)	3.63 (.11)	3.56 (.11)

matching pairs are created using two pictures of each identity which were taken a few minutes apart. The fact that I used the same images, but added pixelation alteration could account for some of the observed high accuracy since the pictorial structure of the image was unchanged. Due to the large sample size, differences emerged when analysing the data despite the high accuracy performance. Indeed, in the pre-test, I replicated both findings of an OGB in White participants, and the tendency for Black participants to perform better for White faces over Black faces. This suggests that the discrimination performance of participants were fairly similar in memory and matching tasks. Once again, Coloured participants did not present ant significantly better performance for either of the other-groups presented, although a slightly better performance for Black rather than White faces was observed, which was the opposite direction to that seen in Study 4. In the post-test, a better performance for both stimulus group was observed overall, with a significantly better discrimination performance for Black faces in the trained groups compared to the untrained groups. Crucially, trained White participants did not present any OGB anymore, while untrained White participants still displayed it. Training therefore succeeded in removing the OGB from White participants. Trained Black participants performed slightly better than untrained Black participants for both stimulus groups, and presented no differences between own-group and other-group faces. The elimination of the OGB in White participants is due to a significantly lower false alarm for Black faces as a result of training. In addition, all participants in the pre-test, and Black and White participants in the post-test, presented a liberal response bias; that is, they were more likely to answer that they had already seen Black faces during the encoding phase, suggesting a greater uncertainty in their decisions regarding Black faces rather than White faces.

Training therefore improved overall performance by reducing false alarms while hits remained stable, and erased the OGB from trained White participants. The training developed for this study aimed to develop a particular ability: Matching people from a good resolution to a degraded resolution. The observed effect of training is thus task-related as participants learned how to better deal with, and to resolve, an original-to-pixelated picture matching task as a function of practice. Even though a generalization was present in the post-test, this training would probably have little to no effect on matching tasks which do not use a pixelation degradation.

Study 6

Training South African participants to create three-dimensional representation of face does not improve accuracy in an ecological face detection task

Study 6 was pre-registered on the Open Science Framework prior to data collection (publicly accessible at `osf.io/59jfu`). Differences are present between the registration and the conducted analyses, they are acknowledged when present.

It is often noted that studying face discrimination during a matching task is of interest for passport border officers and cashiers. However, it has not been acknowledged that police officers sometimes have to look for suspects on the streets, and to do so from memory on the basis of a description or from pictures they can carry with them. The latter situation is the condition I was interested in for this study.

A limited number of studies have tested matching performance during a field task. While an observer is simultaneously presented with a live person and a video/picture, matching a live person from a video resulted in a high error rate (i.e., 22% in target-present condition; 18% in target-absent condition), and this increased by double with the presence of a disguise (Davis & Valentine, 2009). Another study revealed an absence of differences in memory and matching performances when the task required participants to match a live person to a picture, or a picture to another picture (Megreya & Burton, 2008), suggesting that a live situation does not present any advantage. Creating a situation where real cashiers were asked to match a picture on a credit card to its cardholder, Kemp, Towell, and Pike (1997) observed an overall error rate of 32.6%. However, their study involved uncontrolled cross-group situations that would have to be removed if replicated. Although participants from the general population are not familiar with this type of task, it has been assumed that passport control officers would present with a better performance, as forensic examiners do (see White et al., 2015). However, White, Kemp, Jenkins, Matheson, and Burton (2014) observed that passport officers, who are in charge of comparing passport photographs to passport holders, do not present any differences in their performance when compared to the general population, regardless of their experience. This result suggests that practice does not help, and the authors acknowledged that adding feedback would be helpful to train these professionals. This result is surprising considering that professionals are actually trained to perform matching tasks, yet when tested, they did not show improved performance as a function of training (Towler, White, & Kemp, 2014; Woodhead et al., 1979).

To my knowledge, no study has reported a matching field task including cross-group situations. It has been done, however, with a memory task with customers in a shopping mall or employees in a shop, as participants. These studies showed an OGB for Black and White South African and English participants (Wright et al., 2001) as well as for White and Mexican American participants (Platz & Hosch, 1988), even if the latter could not conclude on an OGB for Black participants because of the small sample size. In opposition, (Brigham, Maass, Snyder, & Spaulding, 1982) did not find any OGB in their Black and White American participants.

From these studies, there are three main conclusions: (1) matching tasks, whether they involve a live person or a picture, are difficult to achieve with great accuracy; (2) the OGB has also been demonstrated in field tasks; (3) to my knowledge, no one has explored the effect of training on performance during a field matching task as is has predominantly been studied with pictures. I therefore decided to create a training regimen that was expected to be helpful in the generalization from a field task where participants were asked to look for targets using pictures, which artificially created a spot-in-a-crowd task.

This training was therefore designed in order to develop face processing skills that would be useful to succeed in a field detection task. Considering that a picture is two-dimensional while a live person is three-dimensional, I was curious as to whether I could artificially create a three-dimensional representation of identities using pictures from different face views. In fact, it is more difficult to extract constant information and to match pictures across different views (Bindemann et al., 2013), and presenting both frontal and profile views at the same time does not improve performance (Kramer & Reynolds, 2018). The presentation of a three-quarter view, being the more relevant to create a three-dimensional representation (Hole & Bourne, 2010), has been shown to be at least as informative as the frontal view, whereas profile view is less informative (McKone, 2008; Stephan & Caine, 2007). Considering that one view provides limited information to extract from a face (Estudillo & Bindemann, 2014), and that being exposed to variable views facilitate recognition under different variations (Menon et al., 2015), it was interesting and relevant to explore whether exposing participants to the three types of views (i.e., frontal, three-quarter, profile) during a gradient matching task would help to extract a representation of the identity, and to extract the concept of three-dimensional representations from pictures, which were expected to be generalizable to new faces. The training task designed in this study takes its root in the concept of mental rotation, as though practicing the completion of a mental rotation with an object step by step (i.e., gradient variation) would help to fill in the gaps and improve performance in a task without intermediate steps. This idea was thus applied to faces in this study.

6.1 Hypotheses

First, I expected to find an OGB before training, resulting in better performance for own- than other-group faces for Black and White participants. If present, this would replicate findings established earlier for White participants, and explore once again the scope of the presence or absence of the OGB in a different task. For the Coloured participants, I did not expect to find any better performance on either of the stimulus groups, which would be consistent with previous results reported in this thesis. An OGB was also expected to be found during the field task for untrained Black and White participants. Then, I also expected quicker decision times towards own- rather than other-group faces in the pre-test, and in the training tasks, with less trials for own-group faces during training. During the field task, it was also expected that the presence/absence of targets will be detected more accurately and more quickly by trained than by untrained participants. Then, confidence was expected to be higher for own- than other-group decisions in the field task, and higher for the trained than the untrained participants, and in both cases I expected confidence to be related to accuracy[24].

---

[24] The hypothesis on confidence was not pre-registered; this was an involuntary omission, as it is one of the most important variables to explore in face memory tasks.

6.2 Method

6.2.1 Population

A total of 196 participants, recruited through the Student Research Participation Program (SRPP) of the University of Cape Town, completed the study (143 women, 63 Black, 77 White and 43 Coloured, 13 Indian/Asian/Mixed; M_{age} = 19.72, SD_{age} = 2.36). The sample size was not fixed since I was constrained by the time allocated to the data collection. The final sample contained data from 166 Participants (123 women, 59 Black, 69 White, 38 Coloured, M_{age} = 19.65, SD_{age} = 2.40), since 30 participants were remove according to the exclusion criteria (they were familiar with of the confederates prior to the study, were neither Black, White nor Coloured, because of missing data files (a problem occurred at the beginning of the data collection which resulted in some unrecorded data for the second task) or because of non-compliance with the field instructions (going out of the designed area, speaking to the confederate). Participants were awarded of two course credit points upon the completion of the study

Eight research assistants offered their help for data collection. They were third year students in Psychology and received three points from the same SRPP program as participants, for their assistance. They were recruited through an announcement made by a lecturer on my request. They were mainly recruited for the field task, but helped in the laboratory as needed. The field task was explained and shown to them before data collection to ensure that the execution of the task was closely comparable. One of the assistants acted as an observer to the other (or myself) for one or two trials before conducting the procedure on his/her own.

6.2.2 Design

The present study had a three variables factorial, and mixed design with training (training; no-training) as a between-subject variable, and stimulus group (Black; White) and target group (White; Black) as within-subject variables. Four possibilities were implemented in the field task in order to induce ecological variations according to the presence/absence of the targets: both present, both absent, one of each (counterbalanced).

6.2.3 Material

Stimuli

Overall, 524 pictures of 208 different males were used in the present study. The pictures used for the pre-test matching task were the same stimuli as for Study 5, with the exception that frontal and a profile views of the same face were used, across conditions, instead of pixelation levels. Only 50 (25 of each group) of the photographs were randomly selected from the 100 used in Study 5. Their matching profile pictures were then retrieved from the database they were initially from. The pictures from the training task were the same stimuli as those in Study 5: nine targets and 45 foils of each stimulus group (*N* = 108). Once again, they were not formatted to have different pixelation, but presented from three different angles: frontal, three-quarter and profile.

Confederates

The confederates were two young males in their twenties, short hair, average height, weight and build, without any distinctive feature and self-declared as White or Black. They were

Figure 6.1 – Pictures of the confederates who were targets during the field task.

recruited though shared acquaintance at the University of Cape Town. They were students in second year in the Architecture Department, and were paid R800 (about 47€) each for their help in the study. They signed a consent form specifying their agreement that their pictures would be used for research purposes, including scientific publication and conferences. One frontal[25] picture was taken from each confederate, one week prior the beginning of the Study. Background and clothes were removed and hair and neck were cut to have the same shape (Figure 6.1). Pictures were printed in color on 10x14cm glass paper in three sets. During the study, confederates were asked to wear simple clothes (monochrome pants and basic tee-shirt/hoodies) in order not to stand out from the crowd, and to keep as constant as possible their facial hair, and hair. They were asked to remove any distinctive personal accessories during the study (cap, watch, glasses, jewellery, etc.).

Apparatus

All stimuli were modified using GIMP 2.8.14 software (GNU Image Manipulation), at 449x569 pixels. The OGB measure and training task were constructed and displayed with E-prime 3.0 software (Psychology Software Tools, Pittsburgh, PA) and tasks completed on computers with

---

[25] Even though a three-quarter view is more informative and relevant to create three-dimension representation (Hole & Bourne, 2010), I decided to use frontal pictures there are mostly used in the field by police officers.

17" to 21" screens[26]. Two additional survey for demographic data and post-field questions, as well as a questionnaire for confederates, were created and displayed using Qualtrics online survey tool (Qualtrics, Provo, UT). The instant messaging mobile and desktop application WhatsApp (WhatsApp Inc., Menlo Park, CA) was used to communicate with confederates and research assistants throughout the study. The online game "QWOP" (`www.foddy.net/Athletics.html`) was used as a filler task: The aim of this game is to make an 'avatar' athlete run, using a computer keyboard.

Information collection sheet

To collect consent from participants, consent collection grids were printed, asking for the following: name, surname, student number, date, location, signature, "*I accept to participate in the study*" [circle 'yes' or 'no'] and "*I would like to receive an electronic copy of the consent form*" [circle 'yes' or 'no']. A participant's number code was constructed in which identification numbers (i.e., 'participant ID'; randomly created prior data collection) and group of the participant (training; no-training) were specified. For the field task, an observation grid was also constructed to record participant ID, date and time, the extent of occupation of the area they were located in (relative to seats), target presence/absence[27], participant's decision for each target (identification/rejection and confidence for each of the targets), overall decision time, and comments (Appendix E).

---

[26] Screen size differences did not affect the size of the stimuli during the presentation which were kept constant.

[27] Presence/absence condition was completed by me after the task, in order to keep the research assistants as often as possible blind to the condition.

Field environment

The field task was conducted in the main library of the upper campus of the University of Cape Town. This place was chosen because of the following advantages: being quite busy all day long, being a mixed and representative place of the group diversity at the University, and being protected against weather events. It is also a place where the confederates could sit alone, without appearing out of place, and where there was enough space for the participants to walk around without disturbing the other students, and in which the confederates could easily move from one spot to another. Only a specific area of the library, containing 142 seats, was used for the study (Figure 6.2).

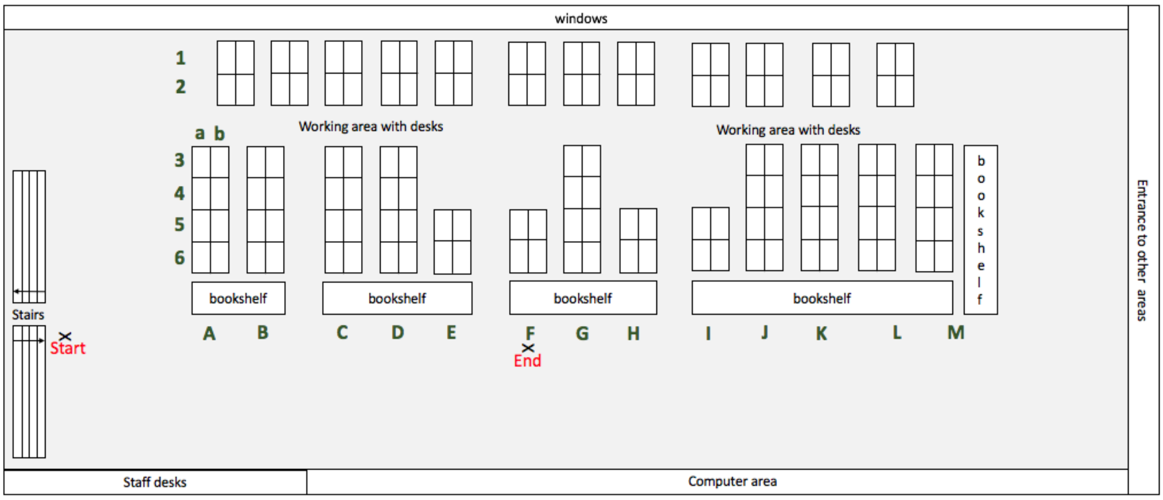

Figure 6.2 – Map of the section of the library used during the field task. Each of the smaller rectangles represents a desk (k = 142). Confederates were told to chose different seats throughout the study and to use the numbers and letters displayed on the schema to indicate their position for each participants (e.g., 1D,b). The study always started and ended at the same place (indicated on the schema by two crosses labelled 'start' and 'end').

6.2.4 Procedure

The study consisted of five different tasks: a demographic questionnaire, the pre-test OGB measure task, the training or no training task, the field detection task and a post-field questionnaire (Figure 6.3). On their sides, confederates completed another questionnaire, and the research assistants an observation grid. The entire experiment lasted between 45 and 90 minutes.

Demographic questionnaire

Once they arrived in the lab, some basic information was given to the participants:

> You will first have to read the consent form displayed on the screen. If you agree on the terms, please sign the paper which is on the desk. Then, you can scroll down on the computer page, and proceed to a questionnaire. After the questionnaire, you will have to complete three tasks: two here in the laboratory and one in the library. You will have all the information and instructions at the time. At the end of each task, please raise your hand since I need to come to start the next task. If you have any question or any problem during the study, please raise your hand and I will help you.

Participants completed the questionnaire and raised their hands when asked (i.e., when the group attribution number was display on the screen). I came to write down the number on the assignment tracking sheet. This number was randomly assigned by the software according to the self-reported group of the participant, to ensure an equal distribution across the two conditions. Then, the first task was started.

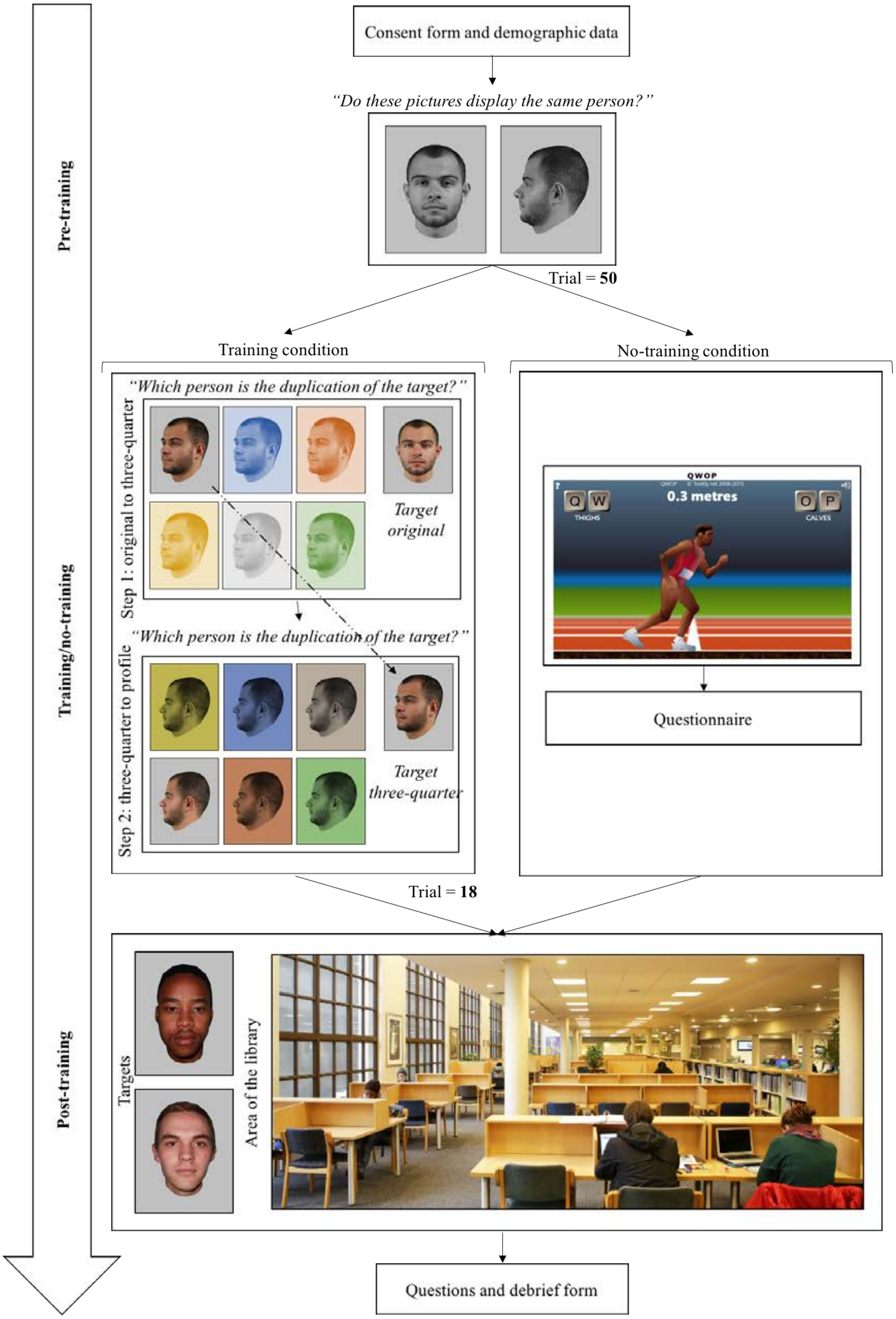

Figure 6.3 – Flow chart of the procedure for Study 6: Every participant completed a matching task as pre-test and a field detection task in post-test. In-between they were split into two groups: training versus no training. Pictures in pre-test and training were not used during the study, but presented here for illustration purpose with the consent of the model.

Pre-test OGB measure

In the pre-test face matching task, participants were asked to indicate whether the presented pairs of pictures displayed the same face (i.e., same identity) or did not. They had been told that a certain number of pairs of faces would be presented and that each pair would either depict the same person or depict two different people. For each pair, they had to answer the question "*Do these pictures depict the same person*?" using the keyboard: 'F' for 'Yes, they do' or 'J' for 'No, they do not'. An example was given to illustrate the task and they were told the following:

> Please notice that to make your decision you must rely on the physical features of the person instead of the pictures' characteristics (lighting, size). Please complete the task as fast and accurately as possible. If you have understood the instruction, please press the space bar to start the task. If you have any questions before starting the task, please call the researcher.

Then, the presentation of the 100 trials commenced in randomized order. Match and mismatch trials were randomly chosen in such a way that at the end of the task participants would have seen 25 match and 25 mismatch pairs for both Black and White faces. At the end of the task, participants called me, who opened the next task according to the assignment order: training or no-training.

Training task

The training task consisted in the matching of one frontal target to its three-quarter view and then, of the three-quarter view to its profile view. To this end, participants were told the following:

> In this task, a series of faces will be displayed on the screen. In each series, you will be shown a ‘target’ face, and 6 other photographs of different quality. Your aim is to find, among the six pictures, which one matches the target face. You will see the target face on the right-hand side of the screen and all the possible matches on the left-hand side of the screen. Use the numerical keypad to record your answers. Feedback will then be displayed: If your answer is incorrect, you will be asked to restart the set. If your answer is correct, you will proceed to the next step. Please note that you do not have a time limit to complete the task.

Then, an example was displayed, with further illustration on the usage of the numerical keypad. It was then specified:

> If you choose the correct match, ‘correct, well done’ [displayed in blue, with the chosen face encased by the same color] will be displayed on the screen and you will proceed to the next set. If you choose an incorrect match, ‘incorrect, try again’ [identical as before but in red] will be displayed on the screen and you will have to give a new answer, until you find the correct match. Note that no matter the nature of the feedback, you need to press the space bar either to go to the next set or to submit a new answer.

Then the task began. Participants first saw a target face from frontal view on the right-hand side of the screen, which was randomly chosen from the six faces, and then the same six faces but from three-quarter view were presented on the left-hand side of the screen. Then, the target was presented from its three-quarter view with the six possibilities of matching from the profile view. Trials were presented according to their difficulty levels (i.e., blocks): easiest trial first, then intermediate, thus difficult. In each block, there were six trials with three Black and three White faces. After each choice, participants received a feedback and proceeded accordingly (try again, or move on) to the next trial, then to the next block, until the completion of the 18 trials. Once the task was completed, participants raised their hand and I came to give them the field task instructions.

Field task

Participants side

Trained participants were then asked to read the following instructions:

> You are going to go to the library with the researcher or a research assistant (RA), and you will have to complete a field task. Once there, the researcher/RA will give you two pictures of two different people, that you will have to look for in a defined area of the library, as if you are looking for a friend to sit with. The researcher/RA will show you the defined area once in the library and will give you the pictures. Once you have the pictures, you can start the task. Bear in mind that you are allowed to look at the pictures as many times as you want and need. Be careful, because each of them may or may not be in the library! The task will be complete when you have made your decision: find one person, both or none.

> In any case, come back to the researcher/RA and give him/her your decision. If they are present, the researcher/RA will then ask you to show where is he or they are. Feel free to walk around in the library, and to look at the pictures as many time as you want to. The task will be ended after 10 minutes if you have not given your decision before. Because the task is double-blinded, the researcher/RA does not know if the people you are looking for are or are not in the library. If you recognize someone you know in the library, please do not talk to him/her. If someone talks to you, please tell him/her that you are completing a study and will go back to them afterwards. You will be free to do so after the study. If you have any question about the present instructions, you will be able to put them to the researcher/RA during the walk to the library. To ensure a good comprehension of the task, you will also have to explain to the researcher/RA what you will have to do. Once you are ready, please raise your hand.

Once the instruction had been read and the participants were ready, I reported the participant's ID number on the observation grid. The RA or me took the observation grid, the two pictures, and went to the library with the participant. Meanwhile, I sent the ID number to the confederates. During the walk to the library, the participant had to recall the instruction of the task. The RA made sure that everything was recalled, and corrected if necessary. Please note that the RAs were blind to the library condition (i.e., presence/absence of the confederates). Once at the library, the RA showed the defined area, stopped at the starting point, got the timer ready and gave the pictures - facing down, to the participant. Once the participant was ready, the task began. As soon as the participant looked at the pictures, the timer was started. The participant then walked freely in the area. The RA walked at the same pace, observing from away to verify that the participant did not interact with anyone. The RA stopped walking at the ending point to write the date, time and extent of occupation of the area, and waited still, observing, until the participant came to the ending point. The participant then went back to the RA once decisions have been made, and the timer was stopped. If the participant did not come back within 10 minutes after the beginning of the task, the RA stopped the task. Then, the time was recorded, and decisions were recorded using the following instructions:

> Now, do not look at the area. Is this person present or absent? [showing the picture of the White target (M), the RA circled 'identification' or 'rejection' accordingly]. From zero to 100 percent, how certain are you that this person is present/absent? [adapted from the previous answer]. Now, is this person present or absent? [showing the picture of the Black target (I)]. From zero to 100 percent, how certain are you that this person is present/absent?

In the case of an identification, the RA asked the participant to point out the target(s) ("*could you please show me now where the person(s) you identified is/are?*"). The RA had to verify the accuracy of the identification. In the case of an inaccurate identification, the RA wrote 'False ID' in the comment column of the observation grid[28]. As soon as the task was completed, the RA had to send a text in the RA WhatsApp group saying that the task was completed. First, this allowed me to forward it to the confederates (who were not looking at what was happening around them, from just after their questionnaire completion to receiving this text), so they could move or leave. Very often, the next participants was following within a short

---

[28] In order to make sure that the RA could verify the accuracy of the identifications, they were showed the pictures of the confederates before modification, and a short video clip (5seconds) of each of them showing different angles of their face, as many times as they wanted. If they had any doubt, they would have either asked to the person if he was the confederate, or recorded their position in the area. I made sure of the accuracy of the identification if needed, from the position, cross-checked with the position declared by the confederates themselves.

amount of time, which left no, or very few, time for the confederates to move and thus, it was very organized and directed by me.

Confederate side

The presence/absence of the confederates[29] was planned according to their availabilities and the data collection schedule, and their presence/absence was randomized as much as possible: A first scheduled was constructed with all the slots in which the confederates were both present, or only one was present. Then, sessions were organized accordingly. For every session they attended, they were asked to wear basic clothing and no distinctive accessories (cap, jewelry, etc.). At their scheduled time, they had to be sure that no one was completing the task before taking a seat. Once in, they took place in the defined area, trying to use as many different seats as possible over the study. When they received a text with the participant number, it meant that the participant just left the laboratory and thus, they had approximately 3 to 5 minutes to get ready for the task and to complete the questionnaire. The questionnaire was a record of: participant number, confederate's name, the extent of occupation of the area, location of their seat according to the schema, social behavior (e.g., sitting alone), and activity they planned to do for the next 10 minutes. Then, they had to act like they would do it if they were not helping for a study, and were asked not to look at the participants. Once the task was completed, I sent a text and confederates could either move from their actual seat to another for the next participant, or leave if it was the end of their session. Sometimes, two participants were directly following each other. In this case, they were told not to move seats. However, it was organized in such a way that they should move as much as possible after each participant.

---

[29] Confederates are also referred as 'targets' in the design and results sections.

I organized all of it, specifying once a participant was done and if they had time left to move before the next participant.

Questionnaire post-field

Once back at the lab, participants were asked to complete a final questionnaire. It contained a question on their decision for each target (presence/absence), one about if they had seen one of them before the task, and the final question checked that they had not seen any of the faces used in the pre-training and training tasks prior to the study. Finally, they had access to the debriefing form, could read it and ask any questions they may have had.

6.2.5 Measures

To explore data from the field task, an accuracy rate was calculated for each target: accurate decision or inaccurate decision. An accurate decision is either a hit (i.e., accurately find the target) or a correct rejection (i.e., accurately reject the presence of the target), while an inaccurate decision is either a false alarm (i.e., identity a foil instead of the target) or a miss (i.e., reject the presence while the target was actually present).

6.3 Results

6.3.1 Discrimination performance and response bias in the pre-test

Overall results showed a quite high discrimination performance, regardless of participant group, both Black faces (M = .90, SD = .05) and White faces (M = .91, SD = .04). A first mixed linear model was created to explore recognition performance across stimulus group and participant group, included as main effects and interaction effects. Results revealed no significant main effect of participant group (F(2, 165) = .26, p = .773) or stimulus group (F(1,167)

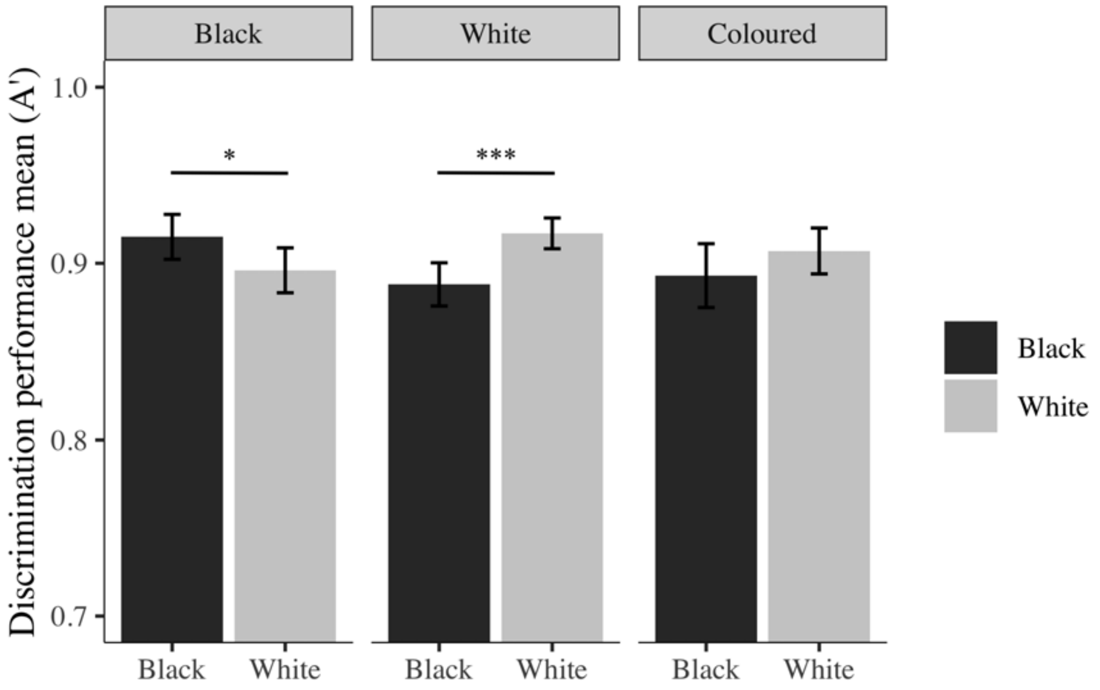

Figure 6.4 – Discrimination performance mean (A') across stimulus group (Black; White) for each participant group (Black; White; Coloured). I bars are 95% confidence intervals.
***$p < .001$; *$p < .05$

= 2.93, p = .089), but a significant interaction between participant and stimulus groups (F(2, 167) = 11.62, $p < .001$), suggesting the presence of an OGB (see Figure 6.4). *Post-hoc* analyses were performed and confirmed the presence of an OGB for both Black (β = .01, t(170) = 2.57, p = .011, d = .20) and White (β = -.03, t(170) = -4.20, $p < .001$, d = .67) participants. Black participants displayed a significantly higher recognition performance for Black faces (M = .91, SD = .05) than White faces (M = .90, SD = .05) while the converse was observed for White participants, with a higher recognition performance for White faces (M = .92, SD = .04) than Black faces (M = .89, SD = .05). Coloured participants, once again, did not present any performance differences for one group or the other (M_{wh} = .91, SD_{wh} = .04; M_{bk} = .89, SD_{bk} = .06; β = -.01, t(170) = -1.46, p = .147, d = .40).

Table 6.1 – *Mean, standard deviation and one sample t-test (against 0) of response bias (B″) in the pre-test across stimulus group (Black; White) for each participant group (Black; White; Coloured).*

Part. group	Stim. group	*M (SD)*	*t*	*df*	*p*	*d*
Black	Black	.14 (.53)	1. 98	58	.053	.26
	White	.09 (.46)	1.45	58	.152	.19
White	Black	-.02 (.47)	-.44	68	.663	.05
	White	-.02 (.53)	-.35	68	.728	.04
Coloured	Black	.23 (.50)	2.84	37	**.007**	.46
	White	.08 (.50)	.96	36	.345	.16

Note. Significant *p*-values are in **bold**

Response bias (B'') was then explored, and a one-sample *t*-test only revealed a significant conservative bias (i.e., a tendency to answer 'no' more often than 'yes' during the matching task) for Coloured participants towards Black faces ($t(37) = 2.84$, $p = .007$, $d = .46$, Table 6.1). A mixed linear model was performed to investigate the main and interaction effects of participant group and stimulus group on response bias. No significant effect was observed, neither of participant group ($F(2, 165) = 2.54$, $p = .082$), stimulus group ($F(1, 166) = 3.10$, $p = .080$) nor the interaction of both ($F(2, 166) = 1.30$, $p = .276$).

6.3.2 Decision time in the pre-test

Decision time was log transformed before performing the analyses, since it did not follow a normal distribution. A mixed linear model using participant group and stimulus group as main and interaction effects along with participant as a random effect was performed on log transformed decision time. The model resulted in no significant effect of participant group ($F(2, 165) = .46$, $p = .631$), but significant effects were found of stimulus group ($F(1, 167) = 33.17$, $p < .001$) and the interaction between participant group and stimulus group ($F(2, 167) = 3.37$, $p = .037$). *Post-hoc* analyses were performed and revealed a significantly higher decision time in White participants for Black faces (M = 4.01, SD = .24) rather than White faces (M = 3.96, SD = .23; β = .05, t(170) = 5.70, p < .001, d = .21) and a significantly higher decision time in Coloured participants for Black faces (M = 3.98, SD = .24) than White faces (M = 3.94, SD = .24; β = .04, t(170) = 2.80, p = .006, d = .17). There were no significant differences for Black participants (M_{bk} = 4.01, SD_{bk} = .20; M_{wh} = 3.99, SD_{wh} = .21; β = .02, t(170) = 1.77, p = .078, d = .10).

6.3.3 Discrimination performance and decision time in training

The number of trials to get the right answer was computed to be predicted according to the manipulated variables. A first mixed linear model using participant group and stimulus group as main and interaction effects with participant as a random effect was performed to predict the number of trials taken. A significant effect of stimulus group was observed (F(1, 83) = 8.77, p = .004) but no significant effect of participant group (F(1, 83) = .27, p = .467) or of the interaction between participant group and stimulus group (F(2, 83) = 2.41, p = .096) suggesting an overall higher number of trials for Black faces as compared to White faces. *Post-hoc* analyses revealed only a significant higher number of trials for White participants for Black faces (M = 19.47, SD = 1.81) rather than White faces (M = 18.25, SD = .55; β = 1.22, t(86) = 3.84, p < .001, d = 1.03). There were no differences for Black participants (M_{bk} = 19.11, SD_{bk} = 1.52; M_{wh} = 18.85, SD_{wh} = 2.24; β = .25, t(86) = .69, p = .491, d = .14) nor for Coloured participants (M_{bk} = 18.94, SD_{bk} = 1.18; M_{wh} = 18.52, SD_{wh} = .84; β = .42, t(86) = .96, p = .340, d = .42).

Table 6.2 – *Mean decision time (seconds) during the training across stimulus group (Black; White) for each participants group (Black; White; Coloured).*

	Part group		
	White	Black	Coloured
	M (SD)	*M (SD)*	*M (SD)*
Black	5.82 (1.68)	5.15 (1.75)	6.29 (1.50)
White	4.22 (1.25)	4.85 (2.48)	4.98 (1.59)

The decision time to complete the training followed an approximately normal distribution. A mixed linear model was run using the main effects and interaction effects of participant group and stimulus group with participant as a random effect. Decision time was significantly affected by stimulus group ($F(1, 83) = 59.50$, $p < .001$) and the interaction of stimulus group and participant group ($F(2, 83) = 9.21$, $p < .001$) but not by participant group ($F(2, 83) = 1.12$, $p = .331$). *Post-hoc* analyses revealed overall quicker decision times for White than for Black faces (Table 6.2), with a significant difference for White participants ($t(86) = 7.70$, $p < .001$, $d = 1.09$) and Coloured participants ($t(86) = 4.59$, $p < .001$, $d = .85$), but not for Black participants ($t(86) = 1.29$, $p = 200$, $d = .14$).

6.3.4 Detection performance in post-test and field description

Assignment within the four alternatives

Overall, the attribution within the four groups of presence/absence of the targets was fairly balanced (Table 6.3).

Table 6.3 – *Assignment of participants within the four field conditions of presence/absence of the targets.*

	Full sample (N = 192)	Final sample (N = 166)
Both targets present	44	37
None of the target present	92	76
Only Black target present	30	29
Only White target present	26	24

Note. Full sample is 192 instead of 196 here since four participants did not completed the field task due to collection failures during a previous task.

Localization and activity of the targets

The location of the targets inside the library was collected over the course of the study. In total, the area had 142 seats. Overall, the targets sat down in 47 different seats over the sessions (34.5%; Figure 6.5). One seat was occupied for a maximum of eight times across the study (M = 2.70, SD = 2.37) and four times in a row (M = 2.17, SD = .80). They mostly sat alone (84% of the time), but also next to friends (9.79%) or next to the other confederate (5.59%).Their time was spent mostly working on a laptop (53.15% of the time), being on their phone (16.08%), drawing (17.48%), being on both their laptop and phone (4.90%), or other activities included listening to music or thinking (8.39%).

Extent of occupation of the area

The extent of occupation of the area during tasks was recorded by both myself/RA and by the confederates when present. The area was 60.25% occupied on average (SD = 17.88%, percentages are the number of people in the area divided by the total number of seats), with a minimum of 5% and a maximum of 100%. Because I did not have any exclusion criteria regarding this data and since there was a wide range of possibilities with most of the values

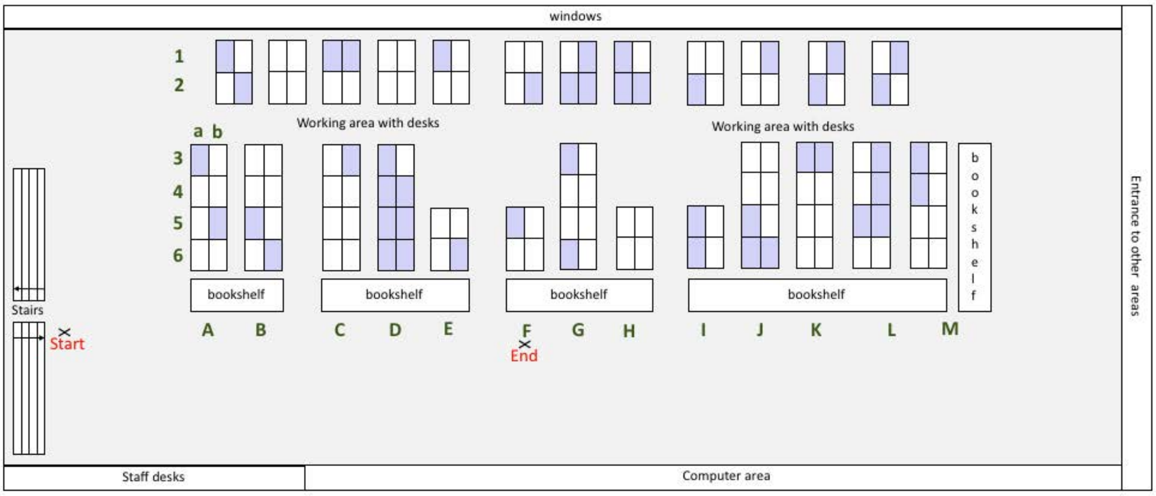

Figure 6.5 – Map of the confederates' location over the study. Colorized desks were used at least once during the study.

contained between 50% and 74.10%, this offered a good frame for a field study. In relation to the high range of variability, I added this information as a random effect into my analyses while studying the effect of training. Variability within the persons present into the area (i.e., group for instance) has not been recorded since it was too time consuming and especially not stable across time for each participant[30].

Detection accuracy, hits and false alarms during the post-test - Field task analyses

In order to explore the effect of training, a mixed logistic model was constructed to predict the accuracy of the field decision. The model is more complex than the previous ones, and includes: the main effects of participant group, target group, training, confidence and decision time, and presence/absence of the target with participants and the extent of occupation of the area as random effects. The model resulted in two significant effects: the main effect of confidence ($\chi^2(1) = 10.75$, $p < .001$) and the main effect of presence/absence of the target ($\chi^2(1) = 29.06$, $p < .001$). The other variables showed no significant effect on accuracy (Table 6.4).

---

[30] The other students present in the library were sometimes arriving of leaving during the experiment.

Table 6.4 – *Mixed logistic regression coefficient table, participants and extent of occupation of the area as random effect, and accuracy as dependent variable.*

	χ^2	*df*	*p*
Presence	29.06	1	**<.001**
Participant group	3.41	2	.182
Target group	1.89	1	.169
Training	.61	1	.433
Confidence	10.75	1	**<.001**
Decision latency	.01	1	.909

Note. Significant *p*-values are in **bold**

While looking at the odds ratios (Figure 6.6) it seems clear that the accuracy is not different between trained and untrained groups also training resulted on no main effect on accuracy. However, there is a great different for presence/absence of the target, resulting on a greater accuracy for target present (82%) conditions than target absent conditions (45%). Training did not improved accuracy, but presence/absence of the targets influenced it.

Since SDT measures could not be computed for this task because only two decisions had to be taken, *z*-scores were used to explore the difference in accuracy rate between the two targets, from each participant group in each condition. Assuming that a significant different means the presence of an OGB, trained White participants still display an OGB in target present condition, however, not in the other groups. An OGB was not present either for Black participants, while they did in pre-test (Table 6.5).

To better understand and locate the difference between target present and absent conditions, the nature of the answer was explored. An absent-target condition included two possibilities: a false alarm or a correct rejection; while a present-target condition included three

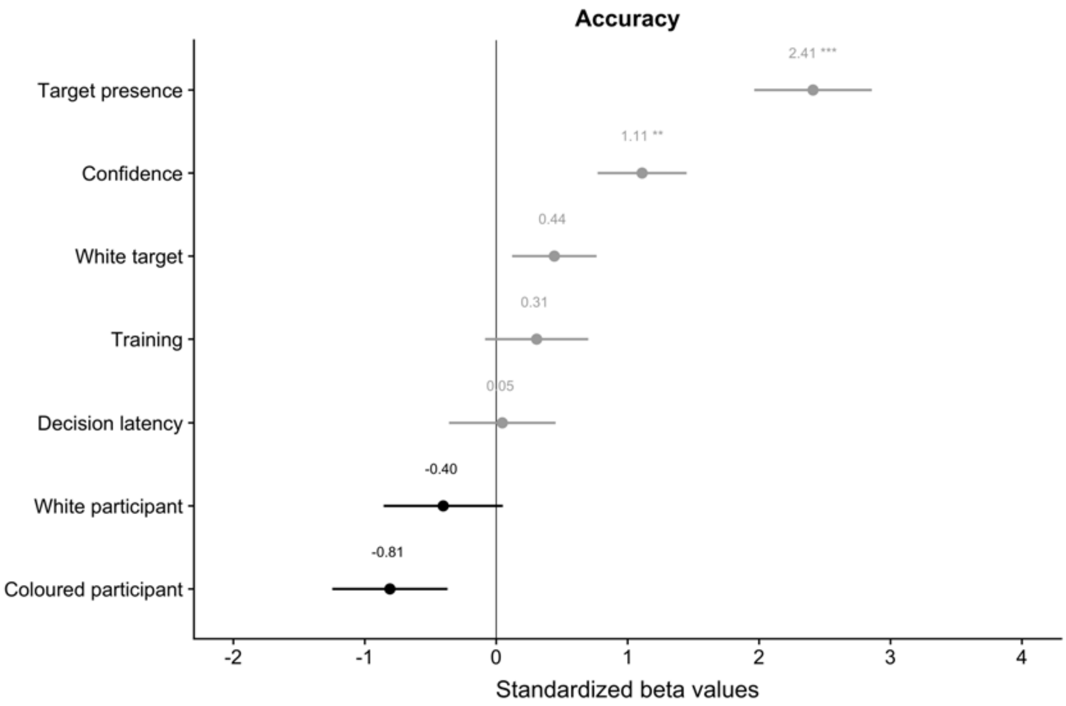

Figure 6.6 – Forest-plot of standardized beta odds ratio from the mixed logistic regression model with participants and extent of occupation of the area as random effect, and accuracy as dependent variable. Horizontal lines are standard errors.
$^{**}p < .01$; $^{***}p < .001$

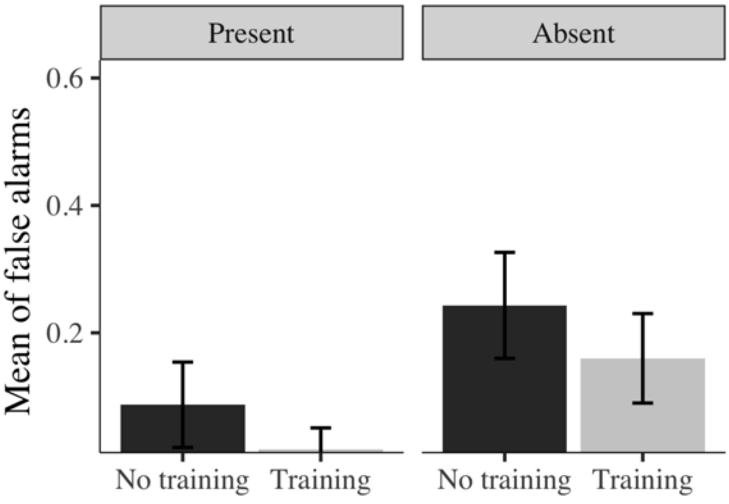

Figure 6.7 – Mean of false alarms for each training condition (training; no training) and target presence condition (presence; absence). I bars are 95% confidence intervals.

Table 6.5 – *Number of participants and accuracy rate (proportion) of the field decision regarding targets (Black; White) for each participant group (Black; White; Coloured), presence condition (presence; absence) and training condition (training; no training). Within-subject z-score difference between the targets (Black; White) are also presented.*

		Part. group	Stim. Group	*n*	Accuracy	z-score	*p*
Training	Present	Black	Black	15	.87	.498	.619
			White	17	.88		
		White	Black	21	.81	2.147	**.032**
			White	22	1		
		Coloured	Black	15	.73	.245	.806
			White	16	.69		
	Absent	Black	Black	12	.58	.095	.924
			White	10	.60		
		White	Black	14	.36	.528	.597
			White	13	.46		
		Coloured	Black	5	.40	.474	.635
			White	4	.25		
No training	Present	Black	Black	18	.89	.365	.715
			White	20	.85		
		White	Black	16	.75	1.519	.129
			White	17	.94		
		Coloured	Black	14	.57	.961	.336
			White	12	.75		
	Absent	Black	Black	14	.50	.408	.683
			White	12	.42		
		White	Black	17	.29	.896	.370
			White	16	.44		
		Coloured	Black	2	1	.937	.349
			White	6	.67		

Note. Significant *p*-values are in **bold**

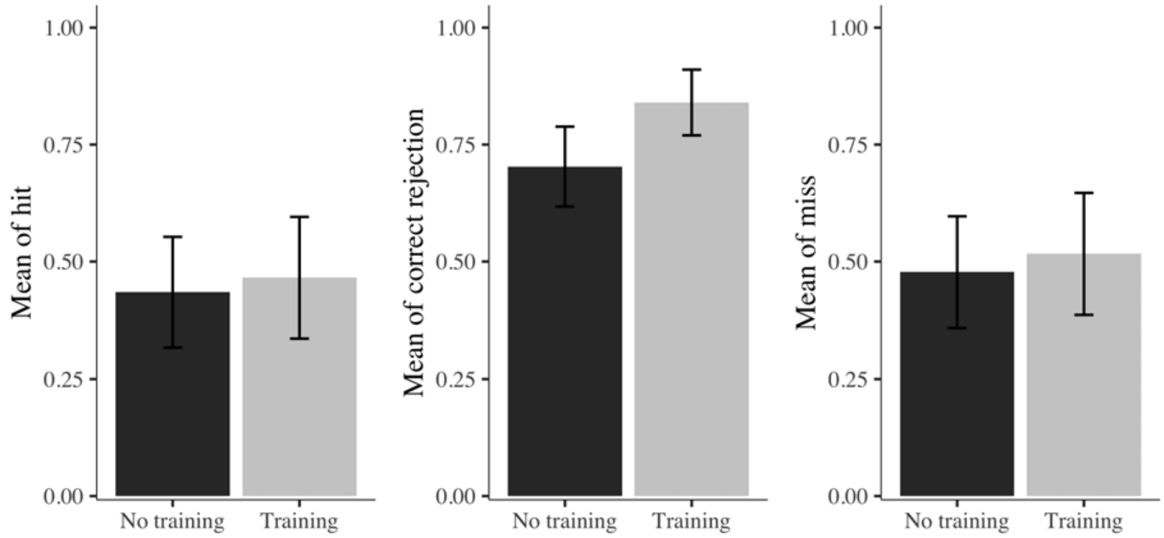

Figure 6.8 – Mean of hit, miss and correct rejection for each training condition (training; no training) and target presence condition: left and right are target present condition outcomes and middle is target absent condition outcomes. I bars are 95% confidence intervals.

possibilities: a hit, a false alarm or a miss. Overall, more false identifications were made when the target was absent, even if also occurred when the target was present. That being said, participants were less likely to make a false identification and therefore were more likely to make a correct rejection when the target was absent (Figure 6.7).

Looking at hits and misses, the two possible additional answers in the case of a target present situation, there seemed to be no differences between training and no training even if these decisions were more frequent than false alarms and fairly equally taken (Figure 6.8). Correct rejections seemed to be the most common decision, and trained participants were more likely to make a correct rejection in a target absent situation than untrained participants.

Considering these results, one should be careful not to draw conclusions that are too direct since there are no further analyses done to support or reject these observations. Indeed, once broken down by participant group and target group (Table 6.5), it was not relevant to conduct any further analysis than reporting the descriptive data (Table 6.6).

Table 6.6 – *Mean and standard deviation of hit, false alarms, miss and correct rejection across both targets (Black; White) for each participant group (Black; White; Coloured), training condition (training; no training) and target presence condition (presence; absence).*

		Target present					
		Hit		False alarm		Miss	
Stim.	Part.	Training	No training	Training	No training	Training	No training
group	group	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Black	Black	.58 (.51)	.50 (.52)	.00 (.00)	.07 (.27)	.42 (.51)	.43 (.51)
	White	.36 (.50)	.29 (.47)	.00 (.00)	.12 (.33)	.64 (.50)	.59 (.51)
	Coloured	.40 (.55)	.50 (.58)	.00 (.00)	.25 (.50)	.60 (.55)	.25 (.50)
White	Black	.60 (.52)	.42 (.51)	.10 (.32)	.17 (.39)	.30 (.48)	.42 (.51)
	White	.46 (.52)	.44 (.51)	.00 (.00)	.00 (.00)	.54 (.52)	.56 (.51)
	Coloured	.25 (.50)	.67 (.52)	.00 (.00)	.00 (.00)	.75 (.50)	.33 (.52)

		Target absent			
		Correct rejection		False alarm	
Stim.	Part.	Training	No training	Training	No training
group	group	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Black	Black	.87 (.35)	.89 (.32)	.13 (.35)	.11 (.32)
	White	.81 (.40)	.52 (.51)	.19 (.40)	.37 (.50)
	Coloured	.73 (.46)	.57 (.51)	.27 (.46)	.43 (.51)
White	Black	.88 (.33)	.85 (.37)	.12 (.33)	.15 (.37)
	White	1.00 (.00)	.67 (.48)	.00 (.00)	.20 (.41)
	Coloured	.69 (.48)	.75 (.45)	.31 (.48)	.25 (.45)

6.3.5 Confidence in the field task

Confidence given after making recognition decisions (from 0 to 100% certainty) has been shown in the literature as being a predictor of accuracy[31]. In this study, confidence was correlated with accuracy (r = .31, t(390) = 6.48, p < .001, 95%CI [.22, .40]). Participants also display a significantly higher confidence for own-group rather than other-group targets for both Black and White participants. This has been shown by the mixed linear regression run on the prediction of confidence rate using participant group and target group as main and interaction effects, training as a main effect, and participant as a random effect along with presence/absence of the target as a second random effect. Results showed significant effects of participant group ($\chi^2(2)$ = 9.43, p = .010) and of the interaction between participant group and target group ($\chi^2(2)$ = 23.54, p < .001). However, no significant main effect was present for the target group ($\chi^2(1)$ = 1.41, p = .235) and training ($\chi^2(1)$ = .64, p = .425), suggesting that, while an expected effect was found, training did not influence confidence level in the participant's decision. Regardless of training, *post-hoc* analyses confirmed the tendency: Black participants reported a significantly higher confidence when taking a decision for the Black target (M = 88.54, SD = 16.16) than for the White target (M = 80.05, SD = 20.71; β = 8.29, t(205) = 2.83, p = .005, d = .46) and White participants reported a significantly lower confidence when taking a decision for the Black target (M = 72.17, SD = 15.60) than for the White target (M = 82.40, SD = 15.47; β = -10.23, t(205) = -4.05, p < .001, d = .66). Coloured participants did not display any differences in confidence level for any of the group targets

---

[31] Unfortunately, I did not pre-register the hypothesis on confidence, but I expected to find a higher confidence for own-group targets than other-group targets and a higher confidence for trained than untrained participants. Indeed, This relationship is often found in the literature (Meissner & Brigham, 2001; Wright et al., 2003, 2001).

(M_{bk} = 76.39, SD_{bk} = 19.52; M_{wh} = 76.79, SD_{wh} = 20.21; β = -.40, $t(205)$ = -.11, p = .916, d = .02).

6.3.6 Decision time in the field

On average, participants took 3 minutes and 45 seconds (205.56 seconds; SD = 174 sec., min = 74 sec., max = 600 sec.) to make decision in the field task. Decision time is not a predictor of accuracy regarding the model presented above, however, it correlated with the extent of occupation of the area (r = .19, $t(390)$ = 3.74, $p < .002$, 95%CI [.09, .28]) which is not surprising. Since decision time was an overall measure for both decisions (participants were looking for the two targets at the same time), it is not possible to explore the decision latency according to the own-group or other-group variables. Looking at decision time between the trained and the untrained group with a mixed linear model including participant and the extent of occupation of the area as random effect, there is no effect of training on the decision time ($F(1, 392)$ = 30.25, p = .500).

6.4 Discussion and conclusion

The aim of this study was to test whether an OGB would be present in a field task where participants had to spot targets within a crowd, and whether the designed multiple-view matching training would help to achieve this task equally for own-group and other-group targets. In the pre-test, participants presented a high accuracy performance (around 91%) when tested with a frontal-to-profile matching task, although this was lower than in Study 5, which is in line with that a matching task across views increases difficulty (Bindemann et al., 2013; McKone, 2008; Stephan & Caine, 2007). In addition, White participants took significantly more trials to complete the training task involving Black faces rather than White faces, while Black and Coloured participants did not present any differences for either stimulus group. In Study 5, there were no differences found in the number of trials taken, which, when considered with the results of Study 6, suggests that difficulty is also higher during training while involves multiple views rather than degradation while keeping view constant.

As in the previous studies, White participants presented an OGB in the pre-test. For the first time, an OGB was also observed for Black participants, even if the effect size is not as large as for White participants (d = .20). This finding of an OGB could be due to the sample size (N = 63, compared to 40 in Study 5 and 20 in Study 4), or the difficulty created from the frontal-to-profile matching task, suggesting that Black participants only display an OGB relative to a task difficulty, or both. This suggestion is also supported by the finding in Study 5 that there were no differences in decision time during the pre-test, while in the present study, both White and Coloured participants presented a significantly higher decision time for Black faces than for White faces, while Black participants did not show any differences. In fact, Black participants were more accurate, but not faster, in their decisions involving Black faces rather than White faces. Coloured participants still did not display a higher discrimination performance for either of their other-group targets, although this is the only group which presented a significant conservative bias, but this was only for Black faces. By comparison, all participants in Study 5 presented a tendency to be more liberal towards Black faces in the pre-test, while no response bias was present toward White faces. The absence of any response bias could be a result of task difficulty, since participants were more careful in completing the task. These differences in discrimination performance and response bias therefore suggest that the nature of the task requires different strategies. In addition, the completion of the field task was well received by participants, and they showed enthusiasm completing it. This might have affected their motivation to find the targets, which could explain why overall, the OGB was absent in this task as participants may not have truly performed differently for any of the targets.

In the post-test, White participants presented an OGB; namely a significantly higher accuracy in their decision for the own-group target than for the other-group target when the target was present, but not when the target was absent, and that was seen regardless of training condition. Black participants did not present any OGB during the field task, as they performed similarly in identifying and rejecting the presence of both targets. Coloured participants did not present any significant differences between targets. The effect of the presence or absence of the target in participants' decisions also revealed more false alarms in target-absent than target-present conditions, even if the instructions clearly stated that the targets could be present or absent from the library at the moment of the study. Some efforts were also made to control the conditions of the field task, reducing the likelihood of biasing factors: Targets were instructed to vary their location and their activity in order to have their faces exposed at different angles and at varying visibility across the study. These instructions were adhered to well, and results from the field situation records show acceptable variability to reduce field-variability associated biases. In addition, eight research assistants were involved in the research, and the variability added from their own way of conducting the study would have been spread over the conditions, while the fact that they were trained and equally instructed offering a good standardization across participants. Collecting such information was crucial, and it should be recorded or controlled for in any further studies of this kind, not only for face detection or face recognition, but also for eyewitness testimony.

The results showed that training did not improve performance for a matching field task. However, it might have been useful to test the effect of training in a more controlled environment, and to ask participants from an additional group to complete the task from pre-test in the post-test. Further studies should address the question of an immediate effect of a multiple-view matching training in the laboratory[32]. Beside the difficulties and limits of this study, it confirmed the importance of using present and absent target conditions, and the interesting contribution of a field study. Crucially, the OGB was found in only one condition (i.e., trained, target-present) for White participants, while all White participants displayed the OGB prior to training and in every other study of this thesis where tested in a laboratory/artificial environment with pictures. It would also be of interest to conduct this research using a memory, rather than matching, paradigm.

---

[32] Because of the context and the organisation of the thesis, it was not possible at the time to go back to South Africa to conduct an additional data collection with this purpose. Since it does not really make sense to compare results without using the same population, this could therefore be considered for future projects.

Section III

General discussion

General Discussion

The present thesis, shifting away from any social contact while focusing on the development of specific and perceptual processing, aimed to remove the OGB as a result of training. As reminder, the two main objectives were to involve a visual individuation through the focus on what makes a face distinguishable from the others, and to develop a set of generalizable skills that could be applied to novel faces. That is, this research included a total of 398 participants in person and 181 participants online[33], across two countries, and from four different groups (French White, South African White, Black and Coloured.). Results mainly indicated no training effect. The OGB was mostly found in White participants, and this was seen in every study. Black participants revealed an OGB only in one type of task, and in studies with a larger sample size.

Define groups in France and South Africa

One of the challenges in studying the OGB concerns the categorization used to define participants and stimuli, and the diversity of the population and group representativeness (in terms of number, and socioeconomic status) of countries like South Africa.

To categorize people according to their group, I used two different measures depending on the country of data collection: ‘race’ in South Africa, and phototype in France. However, I initially did not consider the possibility of using race for my first study conducted in South

---

[33] A total of 621 participants in person, and 230 participants online, were recruited before exclusion due to incomplete data or specific inclusion criteria.

Africa (Study 4), so I relied on phototype to categorize my participants, which revealed some limitations that have already been considered in the discussion section of the concerned study (see p.160). From these observations, I was curious about how could correlate 'race' and phototype, and chose to record both in the studies that followed, by using 'race' categorization only to categorize participants. Within the 359 participants[34] self-declared as Black, White or Coloured, a correlation analysis was conducted between the self-declared group and the categorized group from phototype (i.e., White are I, II and III; IV is Coloured; V and VI are Black). Results revealed a high correlation (ρ = .86, p < .001). Considering the importance of correctly categorizing participants when studying the OGB, even though this correlation is high, it still shows that not all categories overlap. Looking at the exact distribution of each group within each phototype, it is clear that the phototype categorization can result in erroneous categorization, and that it is not perfectly adapted to collect information such as 'race' (Figure 6.1). Indeed, only 55% of participants of phototype IV are self-identified as Coloured, whereas 21% are self-identified as Black, and 24% are self-identified as White. By comparison, amongst participants of phototypes V and VI, assumed to be Black, 24% are self-identified as Coloured. Phototype I, II and III, assumed to be White, are more accurate. However, phototype II also included 2% self-identified as Coloured, and 1% of self-identified as Black and 9% self-identified as Coloured. This was already observed in Study 2 from the removal of participants categorized into these three phototype groups but with Caribbeans or Maghrebian ancestors.

---

[34] Data collected during studies 'X', 5 and 6, including 271 women, M_{age} = 19.72, SD_{age} = 2.17.

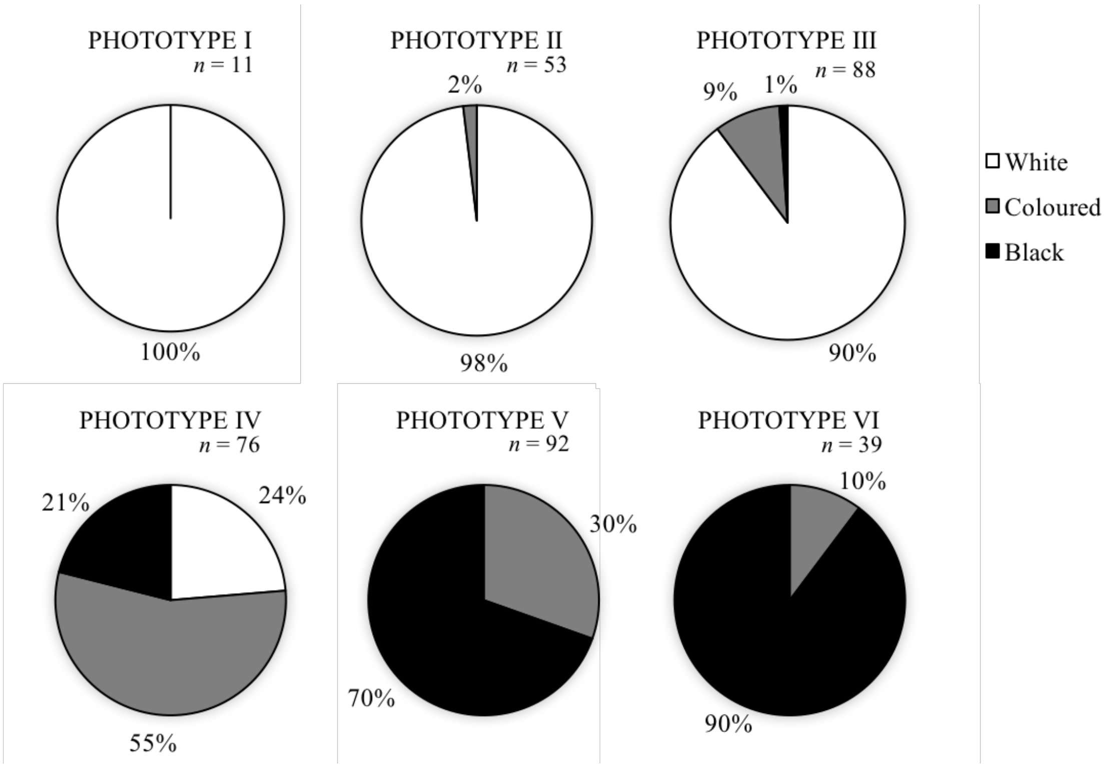

Figure 6.1 – Distribution of participant 'race' (Black; White; Coloured) in each of the phototype (I to VI, see Fitzpatrick categorization p.32).

Although interesting, these observations suggest that my results and conclusions in Study 4 are likely to be biased because of the categorization method. Moreover, these results highlight the limits and boundaries of such a categorization method, especially when considering groups as Coloured or Maghrebian.

There is a crucial distinction to make between phototype and 'race'. While phototype is related to the physical appearance of an individual, 'race' relies more on a social category related to sense of belonging and identity of an individual to a group. While the former is purely descriptive, and could even be made by an external observer, the latter cannot be concluded from inference, and relates to individual identity. Considering the OGB, phototype and 'race' could be informative, and even though they are different, they are complementary. Indeed, arising from cognitive (i.e., visual) and social (i.e., contact) variables, the OGB relies on both physical appearance and group belonging as cognitive and social effects are intertwined (Hugenberg et al., 2010). Considering the importance of visual and social information, there is no perfect solution for categorizing people in order to study the OGB. It would be of some interest to record both phototype and 'race', and eventually keep only participants where the two corresponded. However, if that is a solution, the phototype scale has to be developed and adapted to address the non-exhaustiveness of the phototype scale.

An alternative solution, that I would consider more ethical, is based on visual cues rather than descriptive cues, and would be to ask participants to self-identity themselves on the basis of pictures. Indeed, a consequent number of appropriate groups would be targeted according to the country of the study (e.g., White, Black, Coloured, Indian, Asian and Mixed-race for South Africa; White, Black, Maghrebian, Caribbeans for France), and 6 to 8 pictures would be presented for each group, offering some variability of the physical appearance of people in the group. No physical description or group terminology would be given to any of the groups, and participants would simply have to self-identity as belonging to one or the other group. Of course, this would have to be tested, and many difficulties could arise in the creation of such categorization process, but it would be of some interest considering the sensitivity of racial terms in France, and the fact that even though 'race' can be asked in South Africa, this is starting to be questioned as well. Indeed, more and more students choose not to disclose this information when asked in their registration forms, with an increase of students doing so from 7% in 2013 to 17% in 2018 (https://www.uct.ac.za/main/about/finance/annual-statements).

Along with the classification of participants group, classification of stimuli group should also be more systematic, and self-reported by the models while photographed, as in UCT2007 database. This would allow researchers to select stimuli and participants on the same categorization basis. Nevertheless, since the OGB is a perceptual and social bias, how observers would categorize faces seems to be more important and relevant to explore than how the model consider and self-declare him/herself.

The own-group bias in pre-test

Over the six studies conducted in the present thesis, the presence of the OGB was dependant on participant group. Indeed, OGBs were found every time for White participants, regardless of the country of the study (Figure 6.2). These results are in line with previous observations on the presence of an OGB in White participants when using Black and White stimuli, both in France (Bataille, 2018; Brunet, 2017, 2018) and South Africa (Chiroro et al., 2008; Goodman et al., 2007; Seutloali, 2014; Wright et al., 2001, 2003). However, a study conducted by Derbyshire (2018) did not reveal any OGB for White participants, whom were also students at the University of Cape Town. Although these results are not surprising for France, and even though it was expected for South Africa due to previous studies, it is surprising considering the demography and racial minority position of White people in South African. In fact, it reveals that exposure to many Black faces on a daily basis is not sufficient to protect White people from the OGB.

For Black participants, only participants in Study 6 presented an OGB in pre-test. Considering the effect sizes of the three studies with Black participants (Figure 6.3), this cannot

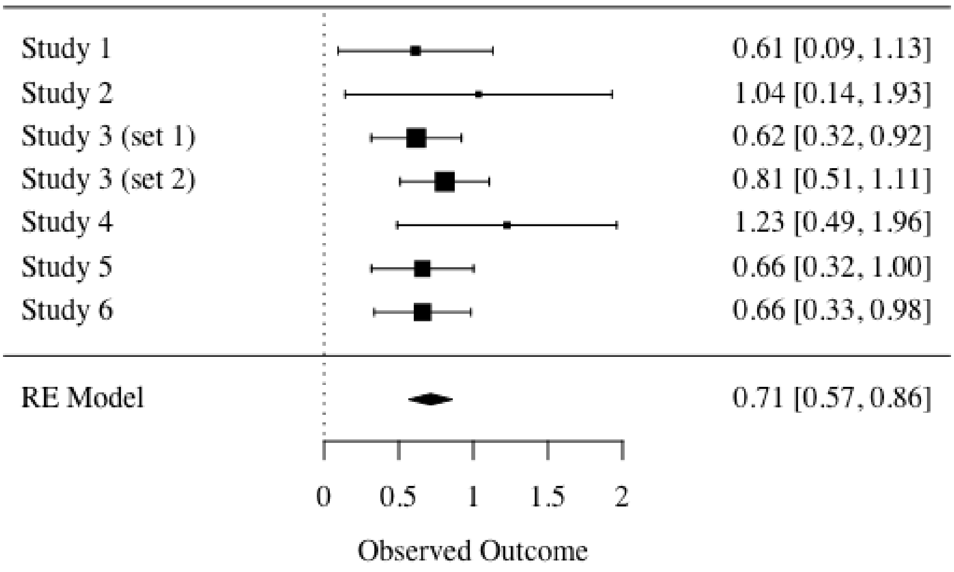

Figure 6.2 – Summary of the random effects meta-analysis of the OGB results from the seven studies including White participants. Mean differences of discrimination performance for each of the stimulus group, and the standard deviation of the differences were computed. The resulting average Hedges' g is .71 (95%CI = [.57; .86]). Together, these studies reveal the presence of a strong OGB in my studies for White participants. Study 1 to 3 are studies with French White participants, Studies 4 to 6 are studies with White South African participants. Sample sizes were 30, 11, 88, 93, 17, 69 and 77 participants, respectively.

be entirely explained by the increased sample size relative to the two other studies. Previous studies also demonstrated inconsistent results, sometimes finding an OGB (Chiroro et al., 2008; Derbyshire, 2018; Wright et al., 2001), and sometimes not (Sadozai et al., 2018), with some even showing a tendency to better recognize other-group (i.e., White) over own-group faces (Seutloali, 2014; Wright et al., 2003). Nevertheless, from the demography of the country (see p.34), I expected to find an OGB in Black participants that would reveal a significantly higher discrimination performance for Black rather than White faces.

Considering the demography of UCT, exposure to White faces is greater than in the general population, which could be a possible explanation. However, exposure is not sufficient for White participants, so it is unlikely that it would be for Black participants. Considering that

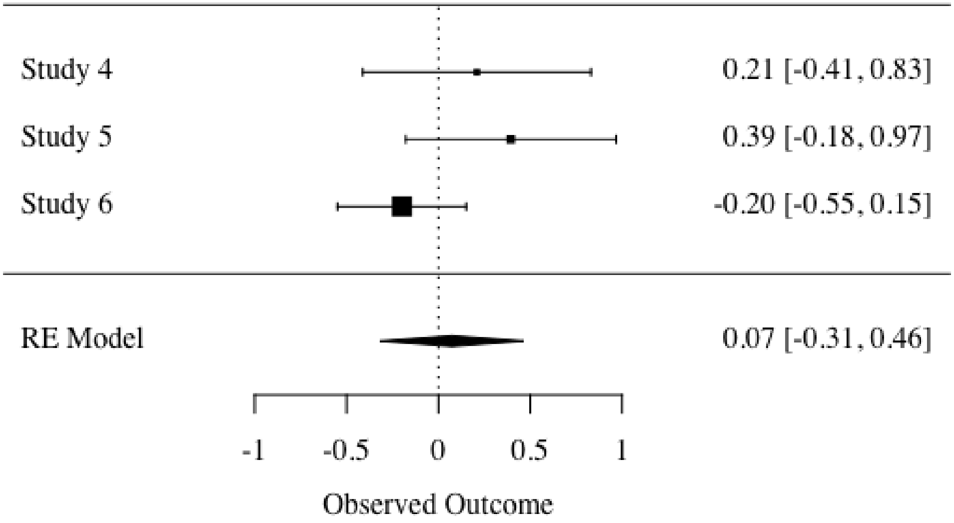

Figure 6.3 – Summary of the random effects meta-analysis of the OGB results from the three studies including Black participants. Mean differences of discrimination performance for each of the stimulus group, and the standard deviation of the differences were computed. The resulting average Hedges' *g* is .07 (95%CI = [-.31; 46]). Together, these studies suggest an absence of an OGB for Black participants in my studies. Sample sizes were 20, 40 and 59 participants respectively.

contact, and not only exposure, related to the absence of OGB among their Black participants, Wright et al. (2003) demonstrated an interaction between contact and recognition performance such that the higher the reported contact by Black participants, the more accurate they were in the recognition task. Since students at UCT come from across the country, Seutloali (2014) considered the relationship between the hometown demography of the participants, which is considered as an objective indication of contact (usually recorded from questionnaires). A significant correlation between hometown demography and OGB magnitude was observed, and the percentage of own-group people in the hometown explained approximately 15% of the variance of the OGB. This supports the idea of contact, and especially of long-term contact. In fact, McKone et al. (2019) concluded from their study (with White and Asian in Australia), that contact in childhood accounts for more than contact during adulthood.

However, this does not fully explain the differences in the presence of the OGB between Black and White participants. Socioeconomic status, and power status, might have to be considered as well. In fact, White people are not only the majority amongst students, but also among Professors. For instance, in 2017, 47% of the academic staff (e.g., lecturers, professors) were White, meaning they were the most represented group to which students are exposed to during the completion of their university education (https://www.uct.ac.za/main/annual-statements). That is, Black students have more interest, and motivation, to individuate and recognize White people, than White students do. The History heritage (i.e., colonialism, slavery, and Apartheid), and past and present social climate might be an additional explanation of White people being well represented in highly considered positions such as professors. For instance, in the United States of America, where Black and White people also have different socioeconomic and power status from colonialism onward, studies observed a moderation of the OGB for White participants when manipulating status, such that Black faces associated with a higher status were better recognized than those associated with a lower status (Shriver & Hugenberg, 2010). This might explain why Black participants performed better for other group faces compared to White participants, because White faces are more likely to have a 'powerful' position in the society.

Overall, Black participants are less likely than White participants to present an OGB (Meissner & Brigham, 2001). For more recent examples, testing Black, White, Hispanic and Asian children on Black, White, Hispanic and Asian stimuli, Gross (2014) found an OGB for all four groups, observing a better recognition performance for own-group than for other-groups overall. However, studying the same groups with adult participants, Gross (2009) found that when separating other-groups, Black participants performed equally for both Black and White faces. Crucially, bi-racial participants from mixed White American and African American parents also performed equally for Black faces and Whites faces, but worse for Asian faces, therefore presenting an OGB for both Black and White faces (Goodman et al., 2007). By contrast, bi-cultural Latino American participants performed better for the primed group regardless of the own-group or other-group race of faces, suggesting that they can perform equally for both groups (Marsh et al., 2016).

Unlike Black and White participants, Coloured participants did not have an opportunity to demonstrate an OGB as the stimuli in my studies were from other groups. They were thus included as a ‘control group’, in the sense that they would have a similar recognition performance for both Black and White faces, assuming similar contact with these groups. In her study, conducted with Coloured participants and stimuli, Seutloali (2014) found dissimilar results, namely a better recognition of White than Black faces. Gross (2009, 2014) also found that Hispanic and Asian participants performed better for White than Black faces. However, similarly to my results, Teitelbaum and Geiselman (1997) found that Latino and Asian participants presented no differences in discrimination performance for Black and White faces.

Modern South African society is changing, moving toward greater diversity. Black students are more numerous nowadays than years ago, and are still increasing toward a greater representativeness of the South African population. However, the number of Coloured students at UCT is still lower than the two other groups, while Coloured is the second most represented group in Cape Town. With each year, the demography continues to change, and with this inter-racial contact does as well. The roles of numerous variables in the OGB makes it something that takes roots in a broader fields than Psychology, including history, geography, politics, sociology and anthropology.

The training effect

The OGB was still present in most of my studies at post-test. Indeed, only Study 5 revealed an absence of the OGB in trained participants in the post-test who had OGB in the pre-test, and for untrained participants in the post-test. In this study, participants were trained to better perform a specific task: Matching an original image to pixelated pictures. It has already been demonstrated that specific instructions or tasks can improve face matching performance for own-group faces (Bindemann & Sandford, 2011; Towler et al., 2019; White, Burton, et al., 2014), but not for other-group faces (Megreya & Bindemann, 2018). My study therefore suggests that firstly, it is possible to increase performance for both own-group and other-group faces, and secondly, such an improvement can be made as a result of training and not due to specific instruction during the completion of the task. However, this has only been demonstrated for White participants, who were the only group who presented an OGB in the pre-test.

Overall, my studies confirmed that simple exposure to many faces, in this case as photographs, is not sufficient to remove the OGB. On the basis of the categorization-individuation model (Hugenberg et al., 2010), I assumed that South African participants would already have sufficient experience to be able to individuate faces when motivated to do so. However, considering the training tasks developed in my studies, individuation invoked by specific visual processing (i.e., focusing on individual face features in order to succeed at the tasks) might be of a different nature than individuation, often created from an association with a semantic information. Another point is that I did not induced any further motivation (beside study 1) than specifying to the participants that they were asked to complete the tasks as fast and as accurately as possible. On that point, the overall similar accuracy in finding/rejecting the presence of the two targets in Study 6 might have been created as a result of motivation that my participants shared after the task. Given the lack of social meaning of my tasks (in terms of individuation and motivation), the objective was simply to develop a better representation of faces in memory (Valentine, 1991; Valentine et al., 2016), and to that end one should be exposed to more faces and should develop an efficient way to discriminate and recognize faces. Training was suppose to address this in three different ways: first, exposing participants to quite a high number of faces, then, supporting them in finding the right dimensions to look at, in order to build a good representation of other-group faces and finally, providing insightful feedback. However, I think the training had a few weakness that might have limited the development of a well elaborated face space for other group faces. Indeed, the simple exposure is not sufficient, and not all the tasks permitted participants to focus on the right dimensions, as study 5 and 6 encouraged configural processing while studies 2 and 4 suffered from the lack of feedback and involved perhaps processing strategies that were too broad. That is, it is of some interest to work on the development of more efficient training tasks.

In previous training studies, the majority used individuation training with face-label association tasks, and observed that such a task is efficient at improving other-group performance for novel faces (Elliott et al., 1973; Goldstein & Chance, 1985; Lebrecht et al., 2009; Matthews & Mondloch, 2018; McGugin et al., 2011; Stahl, 2010; Tanaka & Pierce, 2009). The two training tasks from previous studies which used more visual processing served as a basis to develop my training task. Replicating Hills and Lewis' (2006) study, I found the opposite of their findings, namely an increase rather than a removal of the OGB. Lavrakas et al. (1976) also developed a training regimen during which participants focused on internal parts of faces, and found a significantly better performance for their White participants for Black faces after a single session of training, even if this training effect disappeared one week after the post-test. They attributed their results to the fact that participants would have paid more attention to individual features than group feature (i.e., skin colour, see Levin, 2000). However, this study was conducted in 1976 in the United States, in a less diversified context than studies conducted in the present thesis.

Since the training I developed mostly took place within a single session of about 15 to 30 minutes, during which they were presented an average of 100 pictures, it might not have been sufficient to create a strong learning effect, which is considered independently of the nature of the task and the occasional small sample size. This is supported by the fact that training effects have been found in studies using multiple sessions, and a greater number of pictures than I did (e.g., Goldstein & Chance, 1985; Lebrecht et al., 2009; McGugin et al., 2011). However, the lack of strong findings in my studies opens to question about the fact that exposure to relatively few faces, and for a limited amount of time, could not create enough variability to develop expertise, that is, not being representative enough of the variability present within a group.

To conclude, the acquisition of expertise for other-group faces would require more than exposure and contact, namely explicit and time-intensive training (Tanaka, Heptonstall, & Hagen, 2013; McKone et al., 2019), as it also does for own-group faces where training showed even less optimistic results (Hussain et al., 2009; Young & Burton, 2018; Yovel et al., 2012). That is, a lot of work still needs to be done regarding training, and one should consider the limitations and results from the present thesis as baseline.

General conclusion and perspectives

The work completed in the present thesis did not reveal the expected effect that training would remove the OGB. Since implications and consequences of the OGB can be serious, it is important to conduct research on this topic, and to explore to what extent training would have positive effects on discrimination and recognition performance during a generalization task, while addressing the limitations raised within the present work. Also, the effect of training found for Study 5 is promising for training professionals. Indeed, it shows that while it did not improve own-group discrimination performance, it did for other-group faces and thereby removed the OGB.

Although one limitation of this thesis was that there was not a thorough follow-up from one study to the next. However, considering the small number of different training tasks that have already been tested, several more or less promising training regimen have been tested throughout my thesis, that one could use to conduct further studies. One should consider the results and methodology as a basis for further studies while addressing the acknowledged limitations. In opposition to studies on other-group contact (Pettigrew & Tropp, 2006), the question of face training has received limited empirical support (Hole & Bourne, 2010), and very few researchers have focused on the development of a training regimen, particularly when considering the removal of the OGB. However, this could be due to the publication bias (Franco, Malhotra, & Simonovits, 2015), resulting in unpublished non-significant results. Indeed, researchers might have worked on training, but are not visible, and without Open

Science practice (Franco et al., 2015), no pre-print or registration is made available. In this regard, and given my growing support of Open Science practice, I registered two studies (Study 5 and Study 6) during my thesis to make them available, even if I had unintentionally omitted some information. Nevertheless, none of my data are in open access since this usage was not specified in any of the consent forms signed by my participants. This is considered for the future.

Perspectives

Considering the results and design used in this thesis, some aspects should be kept in mind and carried out in future studies, and these are not limited to training studies.

First, studying a South African population is of a great interest for research in the OGB. In fact, this population is diverse, and even when limited to the University of Cape Town, three main groups of people are available. In further studies conducted with this population, one should consider adding pictures of Coloured people when studying the OGB. In addition, South Africa is a fertile place for increasing inter-group contact, and inclusion of minorities is in constant progress. Research regarding the OGB should also involve people from different fields of research due to its social meaning. Alongside social psychology, including point of view from other fields of research such as sociology or politics would help to explore and understand the underlying social and contextual causes, and consequences, of the OGB in different countries.

Then, considering training tasks, further studies should explore different tasks than those already explored, or should adapt them into multiple session regimen utilizing exposure to a greater variability of other-group faces (i.e., many pictures), at different times. In particular, it would be of great interest to focus on the training task used in Study 6 (i.e., multiple view face matching), and to test its effect in a laboratory task as post-test, rather than as a field task.

Alternatively, for individuation training, one should also consider testing long-term effects in addition to short-term effects, since this was only studied once (Goldstein & Chance, 1985). Indeed, observing first a short-term effect is an important finding, but one should assess for long-term effects as well.

Field studies, even though time consuming to prepare and difficult to standardize, should be more widespread as well. By better approximating real life experiences, they allow us to understand and consider different processing and are more informative of what could be expected in a natural environment, and therefore offer better ecological validity than laboratory-based tests. These tasks allow other information, such as gait, voice and body shape, to be considered as well, and are closer to eyewitness research than face recognition research.

Finally, this thesis has to be contextualised in the project it is a part of. Indeed, one aspect of the MisIdentification Contact Project aimed to develop a Serious Game to improve face discrimination and recognition performance, especially regarding other-group faces. To create such a game, and to implement efficient training tasks, further studies have to be conducted on the designed training tasks, and adapted for inclusion in the game.

References, appendices and French summation

References

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 997–1010. doi: 10.1037/0278-7393.25.4.997

Andrews, S., Jenkins, R., Cursiter, H., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *Quarterly Journal of Experimental Psychology*, *68*(10), 2041–2050. doi: 10.1080/17470218.2014.1003949

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., Tanaka, J. W., & Lee, K. (2013). Developmental origins of the other-race effect. *Current Directions in Psychological Science*, *22*(3), 173–178. doi: 10.1177/0963721412474459

Arizpe, J., Kravitz, D. J., Walsh, V., Yovel, G., & Baker, C. I. (2016). Differences in looking at own- and other-race faces are subtle and analysis-dependent: An account of discrepant reports. *PLoS ONE*, *11*(2), e0148253. doi: 10.1371/journal.pone.0148253

Bartlett, J., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology*, *25*(3), 281–316. doi: 10.1006/cogp.1993.1007

Bataille, A. (2018). *Does social context (really) modulate the cross-race effect in face recognition? A test of the social categorisation hypothesis as an explanation of the cross-race effect* (Unpublished master's thesis). Université de Toulouse, Université de Toulouse: France.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. doi: 10.18637/jss.v067.i01

Bertone, A., Mélen, M., Py, J., & Somat, A. (1995). *Témoins sous influences: Recherches de* *psychologie sociale et cognitive*. Presses universitaires de Grenoble.

Bindemann, M., Attard, J., Leach, A., & Johnston, R. A. (2013). The effect of image pixelation on unfamiliar-face matching. *Applied Cognitive Psychology*, *27*(6), 707–717. doi: 10.1002/acp.2970

Bindemann, M., & Sandford, A. (2011). Me, myself, and I: Different recognition rates for three photo-IDs of the same person. *Perception*, *40*(5), 625–627. doi: 10.1068/p7008

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE*, *3*(8), e3022. doi: 10.1371/journal.pone.0003022

Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology*, *23*, 1933–1967. doi: 10.1111/j.1559-1816.1993.tb01074.x

Brigham, J. C. (2008). The role of race and racial prejudice in recognizing other people. In *Motivational aspects of prejudice and racism* (pp. 68–110). Springer New York. doi: 10.1007/978-0-387-73233-6_4

Brigham, J. C., Bennett, L. B., Meissner, C. A., & Mitchell, T. L. (2007). The influence of race on eyewitness memory. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, vol. 2. memory for people* (pp. 257—281). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of eyewitness identification in a field setting. *Journal of Personality and Social Psychology*, *42*(4), 673–681. doi: 10.1037/0022-3514.42.4.673

Brigham, J. C., & Malpass, R. S. (1985). The role of experience and contact in the recognition of faces of own- and other-race persons. *Journal of Social Issues*, *41*(3), 139–155. doi: 10.1111/j.1540-4560.1985.tb01133.x

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, *73*(1), 105–116. doi: 10.1111/j.2044-8295.1982.tb01795.x

Bruce, V. (1994). Stability from variation: The case of face recognition the m.d. vernon memorial lecture. *The Quarterly Journal of Experimental Psychology Section A*, *47*(1), 5–28. doi: 10.1080/14640749408401141

Bruce, V. (2008). Remembering faces. In *The visual world in memory* (pp. 78–100). Routledge.

Brunet, M. (2017). *Le contact social et son impact sur les processus visuels et attentionnels dan le cadre du biais inter-ethnique* (Unpublished master's thesis). Université de Toulouse, Université de Toulouse: France.

Brunet, M. (2018). *The impact of social contact on own-group bias: a study of visual and attention processes* (Unpublished master's thesis). Université de Toulouse, Université de Toulouse: France.

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, *66*(8), 1467–1485. doi: 10.1080/17470218.2013.800125

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42*(1), 286–291. doi: 10.3758/BRM.42.1.286

Byatt, G., & Rhodes, G. (2004). Identification of own-race and other-race faces: Implications for the representation of race in face space. *Psychonomic Bulletin & Review*, *11*(4), 735–741. doi: 10.3758/bf03196628

Cavazos, J. G., Noyes, E., & OToole, A. J. (2019). Learning context and the other-race effect: Strategies for improving face recognition. *Vision Research*, *157*, 169–183. doi: 10.1016/j.visres.2018.03.003

Chance, J., Goldstein, A. G., & McBride, L. (1975). Differential experience and recognition memory for faces. *The Journal of Social Psychology*, *97*(2), 243–253. doi: 10.1080/00224545.1975.9923344

Chen, W., & Liu, C. H. (2009). Transfer between pose and expression training in face recognition. *Vision Research*, *49*(3), 368–373. doi: 10.1016/j.visres.2008.11.003

Chiroro, P. M., Tredoux, C. G., Radaelli, S., & Meissner, C. A. (2008). Recognizing faces across continents: The effect of within-race variations on the own-race bias in face recognition. *Psychonomic Bulletin & Review*, *15*(6), 1089–1092. doi: 10.3758/pbr.15.6.1089

Chiroro, P. M., & 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. doi: 10.1080/14640749508401421

Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*, *31*(8), 985–994. doi: 10.1068/p3335

Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual Cognition*, *11*(7), 857–869. doi: 10.1080/13506280444000021

Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology*, *17*(1), 97–116. doi: 10.1080/09541440340000439

Collishaw, S. M., & Hole, G. J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, *29*(8), 893–909. doi: 10.1068/p2949

Corcos, A. F. (2016). *The myth of human races*. Wheatmark, Inc.

Cross, J. F., Cross, J., & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception & Psychophysics*, *10*(6), 393–396. doi: 10.3758/bf03210319

Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology*, *23*(4), 482–505. doi: 10.1002/acp.1490

De Heering, A., De Liedekerke, C., Deboni, M., & Rossion, B. (2010). The role of experience during childhood in shaping the other-race effect. *Developmental Science*, *13*(1), 181–187. doi: 10.1111/j.1467-7687.2009.00876.x

Derbyshire, D. (2018). *Investigating the effects of race-based exclusion and prejudice on the own-race bias in facial recognition*. (Unpublished honours' thesis). University of Cape Town, University of Cape Town, South Africa.

de Viviés, X., Kelly, D., Cordier, V., & Pascalis, O. (2010). Reconnaissance des visages d'un autre groupe ethnique : éclairage d'une approche développementale. *Psychologie Française*, *55*(3), 243–257. doi: 10.1016/j.psfr.2010.07.001

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. doi: 10.1037/0096-3445.115.2.107

Dolzycka, D., Herzmann, G., Sommer, W., & Wilhelm, O. (2014). Can training enhance face cognition abilities in middle-aged adults? *PLoS ONE*, *9*(3), e90249. doi: 10.1371/journal.pone.0090249

Dowsett, A. J., Sandford, A., & Burton, A. M. (2016). Face learning with multiple images leads to fast acquisition of familiarity for specific individuals. *Quarterly Journal of Experimental Psychology*, *69*(1), 1–10. doi: 10.1080/17470218.2015.1017513

Duchaine, B., & Nakayama, K. (2006). The cambridge face memory test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585. doi: 10.1016/j.neuropsychologia.2005.07.001

Elliott, E. S., Wills, E. J., & Goldstein, A. G. (1973). The effects of discrimination training on the recognition of white and oriental faces. *Bulletin of the Psychonomic Society*, *2*(2), 71–73. doi: 10.3758/bf03327717

Ellis, H. D., Deregowski, J. B., & Shepherd, J. W. (1975). Description of White and Black faces by White and Black subjects. *International Journal of Psychology*, *10*(2), 119–123. doi: 10.1080/00207597508247325

Estudillo, A. J., & Bindemann, M. (2014). Generalization across view in face memory and face matching. *i-Perception*, *5*(7), 589–601. doi: 10.1068/i0669

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using gpower 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. doi: 10.3758/brm.41.4.1149

Feingold, G. A. (1914). The influence of environment on identification of persons and things. *Journal of the American Institute of Criminal Law and Criminology*, *5*(1), 39. doi: 10.2307/1133283

Fitzpatrick, T. B. (1988). The validity and practicality of sun-reactive skin types I through VI. *Archives of Dermatology*, *124*(6), 869. doi: 10.1001/archderm.1988.01670060015008

Fletcher, K. I., Butavicius, M. A., & Lee, M. D. (2008). Attention to internal face features in unfamiliar face matching. *British Journal of Psychology*, *99*(3), 379–394. doi: 10.1348/000712607x235872

Fluehr-Lobban, C. (2018). *Race and racism: An introduction*. Rowman & Littlefield.

Fox, J., Friendly, M., & Weisberg, S. (2013). Hypothesis tests for multivariate linear models using the car package. *The R Journal*, *5*(1), 39. doi: 10.32614/rj-2013-004

Franco, A., Malhotra, N., & Simonovits, G. (2015). Under-reporting in psychology experiments. *Social Psychological and Personality Science*, *7*(1), 8–12. doi: 10.1177/1948550615598377

Fysh, M. C., & Bindemann, M. (2017). Forensic face matching: A review. *Face processing: Systems, disorders and cultural differences*, *4*(6), 1–20. doi: 10.1098/rsos.170249

Goldberg, D. T. (2006). Racial europeanization. *Ethnic and Racial Studies*, *29*(2), 331–364. doi: 10.1080/01419870500465611

Goldinger, S. D., He, Y., & Papesh, M. H. (2009). Deficits in cross-race face learning: Insights from eye movements and pupillometry. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1105–1122. doi: 10.1037/a0016548

Goldstein, A. G., & Chance, J. E. (1985). Effects of training on japanese face recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society*, *23*(3), 211–214. doi: 10.3758/bf03329829

Goodman, G. S., Sayfan, L., Lee, J. S., Sandhei, M., Walle-Olsen, A., Magnussen, S., ... Arredondo, P. (2007). The development of memory for own- and other-race faces. *Journal of Experimental Child Psychology*, *98*(4), 233–242. doi: 10.1016/j.jecp.2007.08.004

Gross, T. F. (2009). Own-ethnicity bias in the recognition of Black, East Asian, Hispanic, and White faces. *Basic and Applied Social Psychology*, *31*(2), 128–135. doi: 10.1080/01973530902880381

Gross, T. F. (2014). Face recognition and own-ethnicity bias in Black, East/Southeast Asian, Hispanic, and White children. *Asian American Journal of Psychology*, *5*(3), 181–189. doi: 10.1037/a0034082

Hajji, S. (2018). *Reinvestigation of implicit prejudice and intergroup contact effects on the other-ethnicity bias* (Unpublished master's thesis). Université de Toulouse, Université de Toulouse: France.

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. doi: 10.1348/000712607x199981

Hancock, P. J., Bruce, V., & Burton, A. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*(9), 330–337. doi: 10.1016/s1364-6613(00)01519-9

Hayward, W. G., Favelle, S. K., Oxner, M., Chu, M. H., & Lam, S. M. (2017). The other-race effect in face learning: Using naturalistic images to investigate face ethnicity effects in a learning paradigm. *Quarterly Journal of Experimental Psychology*, *70*(5), 890–896. doi: 10.1080/17470218.2016.1146781

Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition*, *106*(2), 1017–1027. doi: 10.1016/j.cognition.2007.04.002

Heisz, J. J., & Shore, D. I. (2008). More efficient scanning for familiar faces. *Journal of Vision*, *8*(1), 9. doi: 10.1167/8.1.9

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, *33*(1), 98–106. doi: 10.3758/bf03195300

Herlitz, A., & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition*, *21*(9-10), 1306–1336. doi: 10.1080/13506285.2013.823140

Heron-Delaney, M., Anzures, G., Herbert, J. S., Quinn, P. C., Slater, A. M., Tanaka, J. W., … Pascalis, O. (2011). Perceptual training prevents the emergence of the other race effect during infancy. *PLoS ONE*, *6*(5), e19858. doi: 10.1371/journal.pone.0019858

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013). Removing the own-race bias in face recognition by attentional shift using fixation crosses to diagnostic features: An eye-tracking study. *Visual Cognition*, *21*(7), 876–898. doi: 10.1080/13506285.2013.834016

Hills, P. J., & Lewis, M. B. (2006). Reducing the own-race bias in face recognition by shifting attention. *Quarterly Journal of Experimental Psychology*, *59*(6), 996–1002. doi: 10.1080/17470210600654750

Hills, P. J., & Lewis, M. B. (2011). Reducing the own-race bias in face recognition by attentional shift using fixation crosses preceding the lower half of a face. *Visual Cognition*, *19*(3), 313–339. doi: 10.1080/13506285.2010.528250

Hills, P. J., & Lewis, M. B. (2018). The development of face expertise: Evidence for a qualitative change in processing. *Cognitive Development*, *48*, 1–18. doi: 10.1016/j.cogdev.2018.05.003

Hills, P. J., & Pake, J. M. (2013). Eye-tracking the own-race bias in face recognition: Revealing the perceptual and socio-cognitive mechanisms. *Cognition*, *129*(3), 586–597. doi: 10.1016/j.cognition.2013.08.012

Hole, G., & Bourne, V. (2010). *Face processing: Psychological, neuropsychological, and applied perspectives*. Oxford University Press.

Howells, T. H. (1938). A study of ability to recognize faces. *The Journal of Abnormal and*

Social Psychology, *33*(1), 124–127. doi: 10.1037/h0062942

Hsiao, J. H.-W., & Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychological Science*, *19*(10), 998–1006. doi: 10.1111/j.1467-9280.2008.02191.x

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

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology*, *43*(2), 334–340. doi: 10.1016/j.jesp.2006.02.010

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. doi: 10.1037/a0020463

Hussain, Z., Sekuler, A. B., & Bennett, P. J. (2009). Perceptual learning modifies inversion effects for faces and textures. *Vision Research*, *49*(18), 2273–2284. doi: 10.1016/j.visres.2009.06.014

Jenkins, R., White, D., Montfort, X. V., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–323. doi: 10.1016/j.cognition.2011.08.001

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577–596. doi: 10.1080/09658210902976969

Kelly, D. J., Liu, S., Ge, L., Quinn, P. C., Slater, A. M., Lee, K., … Pascalis, O. (2007). Cross-race preferences for same-race faces extend beyond the african versus caucasian contrast in 3-month-old infants. *Infancy*, *11*(1), 87–95. doi: 10.1207/s15327078in1101_4

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy. *Psychological Science*, *18*(12), 1084–1089. doi:

10.1111/j.1467-9280.2007.02029.x

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., … Pascalis, O. (2005). Three-month-old, but not newborns, prefer own-race faces. *Developmental Science*, *8*(6), F31–F36. doi: 10.1111/j.1467-7687.2005.0434a.x

Kemp, R. I., Caon, A., Howard, M., & Brooks, K. R. (2016). Improving unfamiliar face matching by masking the external facial features. *Applied Cognitive Psychology*, *30*(4), 622–627. doi: 10.1002/acp.3239

Kemp, R. I., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *11*(3), 211–222. doi: doi.org/10.1002/(SICI)1099-0720(199706)11:3<211::AID-ACP430>3.0.CO;2-O

Kokje, E., Bindemann, M., & Megreya, A. M. (2018). Cross-race correlations in the abilities to match unfamiliar faces. *Acta Psychologica*, *185*, 13–21. doi: 10.1016/j.actpsy.2018.01.006

Kramer, R. S. S., & Reynolds, M. G. (2018). Unfamiliar face matching with frontal and profile views. *Perception*, *47*(4), 414–431. doi: 10.1177/0301006618756809

Kramer, R. S. S., Ritchie, K. L., & Burton, A. M. (2015). Viewers extract the mean from images of the same person: A route to face learning. *Journal of Vision*, *15*(4), 1–9. doi: 10.1167/15.4.1

Kramer, R. S. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition*, *172*, 46–58. doi: 10.1016/j.cognition.2017.12.005

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). doi: 10.18637/

jss.v082.i13

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00863

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud faces database. *Cognition & Emotion*, *24*(8), 1377–1388. doi: 10.1080/02699930903485076

Laurence, S., & Mondloch, C. J. (2016). That's my teacher! Children's ability to recognize personally familiar and unfamiliar faces improves with age. *Journal of Experimental Child Psychology*, *143*, 123–138. doi: 10.1016/j.jecp.2015.09.030

Laurence, S., Zhou, X., & Mondloch, C. J. (2016). The flip side of the other-race coin: They all look different to me. *British Journal of Psychology*, *107*(2), 374–388. doi: 10.1111/bjop.12147

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. doi: 10.3758/bf03208285

Lebrecht, S., Pierce, L. J., Tarr, M. J., & Tanaka, J. W. (2009). Perceptual other-race training reduces implicit racial bias. *PLoS ONE*, *4*(1), e4215. doi: 10.1371/journal.pone.0004215

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. *Nature*, *410*(6831), 890–890. doi: 10.1038/35073749

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1). doi: 10.18637/jss.v069.i01

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. doi: 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. doi: 10.1037/0096-3445.129.4.559

Little, A. C., DeBruine, L. M., Jones, B. C., & Waitt, C. (2008). Category contingent aftereffects for faces of different races, ages and species. *Cognition*, *106*(3), 1537–1547. doi: 10.1016/j.cognition.2007.06.008

Liu, C. H., Bhuiyan, M. A.-A., Ward, J., & Sui, J. (2009). Transfer between pose and illumination training in face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 939–947. doi: 10.1037/a0013710

Liu, C. H., Chen, W., & Ward, J. (2015). Effect of exposure to facial expression variation in face learning and recognition. *Psychological Research*, *79*(6), 1042–1053. doi: 10.1007/s00426-014-0627-8

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 77–100. doi: 10.1037/0096-1523.34.1.77

Lüdecke, D. (2019a). esc: Effect size computation for meta analysis [Computer software manual]. (R package version 0.5.0) doi: 10.5281/zenodo.1249218

Lüdecke, D. (2019b). sjplot: Data visualization for statistics in social science [Computer software manual]. (R package version 2.6.3) doi: 10.5281/zenodo.1308157

Ma, Y., Yang, S., & Han, S. (2011). Attitudes influence implicit racial face categorization in a perceptual task. *Group Processes & Intergroup Relations*, *14*(6), 887–899. doi: 10.1177/1368430211409962

MacLin, O. H., MacLin, M. K., & Malpass, R. S. (2001). Race, arousal, attention, exposure and delay: An examination of factors moderating face recognition. *Psychology, Public Policy, and Law*, *7*(1), 134–152. doi: 10.1037/1076-8971.7.1.134

MacLin, O. H., & Malpass, R. S. (2003). The ambiguous-race face illusion. *Perception*, *32*(2), 249–252. doi: 10.1068/p5046

Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. *The Journal of Open Source Software*, *3*(22), 470. doi: 10.21105/joss.00470

Malpass, R. S. (1974). Racial bias in eyewitness identification? *Proceedings of the Division of Personality and Society Psychology*, *1*(1), 42–44. doi: 10.1177/014616727400100115

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

Malpass, R. S., Lavigueur, H., & Weldon, D. E. (1973). Verbal and visual training in face recognition. *Perception & Psychophysics*, *14*(2), 285–292. doi: 10.3758/bf03212392

Marsh, B. U., Pezdek, K., & Ozery, D. H. (2016). The cross-race effect in face recognition memory by bicultural individuals. *Acta Psychologica*, *169*, 38–44. doi: 10.1016/j.actpsy.2016.05.003

Matthews, C. M., Davis, E. E., & Mondloch, C. J. (2018). Getting to know you: The development of mechanisms underlying face learning. *Journal of Experimental Child Psychology*, *167*, 295–313. doi: 10.1016/j.jecp.2017.10.012

Matthews, C. M., & Mondloch, C. J. (2018). Improving identity matching of newly encountered faces: Effects of multi-image training. *Journal of Applied Research in Memory and Cognition*, *7*(2), 280–290. doi: 10.1016/j.jarmac.2017.10.005

McGugin, R. W., Tanaka, J. W., Lebrecht, S., Tarr, M. J., & Gauthier, I. (2011). Race-specific perceptual discrimination improvement following short individuation training with faces. *Cognitive Science*, *35*(2), 330–347. doi: 10.1111/j.1551-6709.2010.01148.x

McKone, E. (2008). Configural processing and face viewpoint. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 310–327. doi: 10.1037/0096-1523.34.2.310

McKone, E., Brewer, J. L., MacPherson, S., Rhodes, G., & Hayward, W. G. (2007). Familiar other-race faces show normal holistic processing and are robust to perceptual stress. *Perception*, *36*(2), 224–248. doi: 10.1068/p5499

McKone, E., Wan, L., Pidcock, M., Crookes, K., Reynolds, K., Dawel, A., … Fiorentini, C. (2019). A critical period for faces: Other-race face recognition is improved by childhood but not adult social contact. *Scientific Reports*, *9*(1). doi: 10.1038/s41598-019-49202-0

Megreya, A. M. (2018). Feature-to-feature comparison and holistic processing in unfamiliar face matching. *PeerJ*, *6:e4437*. doi: 10.7717/peerj.4437

Megreya, A. M., & Bindemann, M. (2018). Feature instructions improve face-matching accuracy. *PLoS ONE*, *13*(3), e0193455. doi: 10.1371/journal.pone.0193455

Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied*, *14*(4), 364–372. doi: 10.1037/a0013464

Megreya, A. M., White, D., & Burton, A. M. (2011). The other-race effect does not rely on memory: Evidence from a matching task. *Quarterly Journal of Experimental Psychology*, *64*(8), 1473–1483. doi: 10.1080/17470218.2011.575228

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. doi: 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. doi: 10.1002/acp.1097

Menon, N., White, D., & Kemp, R. I. (2015). Variation in photos of the same face drives improvements in identity verification. *Perception*, *44*(11), 1332–1341. doi: 10.1177/0301006615599902

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

Michel, C., Rossion, B., Han, J., Chung, C.-S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science*, *17*(7), 608–615. doi: 10.1111/j.1467-9280.2006.01752.x

Mukudi, P. B., & Hills, P. J. (2019). The combined influence of the own-age, -gender, and -ethnicity biases on face recognition. *Acta Psychologica*, *194*, 1–6. doi: 10.1016/j.actpsy.2019.01.009

Murphy, J., Ipser, A., Gaigg, S. B., & Cook, R. (2015). Exemplar variance supports robust learning of facial identity. *Journal of Experimental Psychology: Human Perception and*

Performance, *41*(3), 577–581. doi: 10.1037/xhp0000049

Ng, A. H., Steele, J. R., & Sasaki, J. Y. (2016). Will you remember me? Cultural differences in own-group face recognition biases. *Journal of Experimental Social Psychology*, *64*, 21–26. doi: 10.1016/j.jesp.2016.01.003

Ng, W.-J., & Lindsay, R. C. (1994). Cross-race facial recognition. *Journal of Cross-Cultural Psychology*, *25*(2), 217–232. doi: 10.1177/0022022194252004

Paterson, H. M., Luppino, D., Calderwood, C., MacDougall, H. G., Taubert, J., & Kemp, R. I. (2017). Can training improve eyewitness identification? The effect of internal feature focus on memory for faces. *Psychology, Crime & Law*, *23*(10), 927–945. doi: 10.1080/1068316x.2017.1346099

Pettigrew, T. F. (1997). Generalized intergroup contact effects on prejudice. *Personality and Social Psychology Bulletin*, *23*(2), 173–185. doi: 10.1177/0146167297232006

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. doi: 10.1037/0022-3514.90.5.751

Platz, S. J., & Hosch, H. M. (1988). Cross-racial/ethnic eyewitness identification: A field study. *Journal of Applied Social Psychology*, *18*(11), 972–984. doi: 10.1111/j.1559-1816.1988.tb01187.x

Proietti, V., Laurence, S., Matthews, C. M., Zhou, X., & Mondloch, C. J. (2019). Attending to identity cues reduces the own-age but not the own-race recognition advantage. *Vision Research*, *157*, 184–191. doi: 10.1016/j.visres.2017.11.010

Publication manual of the american psychological association (6th ed.). (2006). American Psychological Association.

Quinn, P. C., Lee, K., & Pascalis, O. (2018). Perception of face race by infants: Five developmental changes. *Child Development Perspectives*, *12*(3), 204–209. doi: 10.1111/cdep.12286

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Revelle, W. (2017). psych: Procedure for personality and psychological research [Computer software manual]. Evanston, IL: Northwestern University. (R package version 1.8.12)

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

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. doi: 10.1037/a0025750

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology*, *70*(5), 897–905. doi: 10.1080/17470218.2015.1136656

Rodriguez, J., Bortfeld, H., & Gutierrez-Osuna, R. (2008, sep). Reducing the other-race through caricatures. *2008 8th IEEE International Conference in Automatic Face & Gesture Recognition*, 1—5. doi: 10.1109/AFDR.2008.4813398

Sadozai, A. K., Kempen, K., Tredoux, C., & Robbins, R. A. (2018). Can we look past people's race? The effect of combining race and a non-racial group affiliation on holistic processing. *Quarterly Journal of Experimental Psychology*, *72*(3), 557–569. doi: 10.1177/1747021818760482

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

Science, *16*(6), 440—444. doi: 10.1111/j.0956-7976.2005.01554.x

Sekiguchi, T. (2011). Individual differences in face memory and eye fixation patterns during face learning. *Acta Psychologica*, *137*(1), 1–9. doi: 10.1016/j.actpsy.2011.01.014

Seutloali, G. M. (2014). *Own-race bias in facial recognition amongst Black, Coloured and White participants*. (Unpublished honours' thesis). University of Cape Town, University of Cape Town, South Africa.

Shepherd, J. W., & Deregowski, J. B. (1981). Races and faces - a comparison of the responses of africans and europeans to faces of the same and different races. *British Journal of Social Psychology*, *20*(2), 125–133. doi: 10.1111/j.2044-8309.1981.tb00485.x

Shriver, E. R., & Hugenberg, K. (2010). Power, individuation, and the cross-race recognition deficit. *Journal of Experimental Social Psychology*, *46*(5), 767–774. doi: 10.1016/j.jesp.2010.03.014

Singmann, H., Kellen, D., & Klauer, K. C. (2013). Investigating the other-race effect of germans towards turks and arabs using multinomial processing tree models. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*(35).

Slone, A. E., Brigham, J. C., & Meissner, C. A. (2000). Social and cognitive factors affecting the own-race bias in Whites. *Basic and Applied Social Psychology*, *22*(2), 71–84. doi: 10.1207/s15324834basp2202_1

Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law*, *7*(1), 36–97. doi: 10.1037/1076-8971.7.1.36

Sporer, S. L., & Horry, R. (2011). Recognizing faces from ethnic in-groups and out-groups: Importance of outer face features and effects of retention interval. *Applied Cognitive Psychology*, *25*(3), 424–431. doi: 10.1002/acp.1709

Stahl, J. (2010). *Determinants of the own-race bias and neural correlates of own- and other-race face processing* (Unpublished master's thesis). University of Jena, Germany.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. doi: 10.3758/bf03207704

Stephan, B. C. M., & Caine, D. (2007). What is in a view? The role of featural information in the recognition of unfamiliar faces across viewpoint transformation. *Perception*, *36*(2), 189–198. doi: 10.1068/p5627

Sugden, N. A., & Marquis, A. R. (2017). Meta-analytic review of the development of face discrimination in infancy: Face race, face gender, infant age, and methodology moderate face discrimination. *Psychological Bulletin*, *143*(11), 1201–1244. doi: 10.1037/bul0000116

Tan, C. B. Y., Stephen, I. D., Whitehead, R., & Sheppard, E. (2012). You look familiar: How malaysian chinese recognize faces. *PLoS ONE*, *7*(1), e29714. doi: 10.1371/journal.pone.0029714

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. doi: 10.1080/14640749308401045

Tanaka, J. W., Heptonstall, B., & Hagen, S. (2013). Perceptual expertise and the plasticity of other-race face recognition. *Visual Cognition*, *21*(9-10), 1183–1201. doi: 10.1080/13506285.2013.826315

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), B1–B9.

doi: 10.1016/j.cognition.2003.09.011

Tanaka, J. W., & Pierce, L. J. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(1), 122–131. doi: 10.3758/cabn.9.1.122

Tanaka, J. W., & Simonyi, D. (2016). The “parts and wholes” of face recognition: A review of the literature. *Quarterly Journal of Experimental Psychology*, *69*(10), 1876–1889. doi: 10.1080/17470218.2016.1146780

Teitelbaum, S., & Geiselman, R. E. (1997). Observer mood and cross-racial recognition of faces. *Journal of Cross-Cultural Psychology*, *28*(1), 93–106. doi: 10.1177/0022022197281006

Templeton, A. R. (2013). Evolution and notions of human race. In *How evolution shapes our lives* (pp. 346–361). Princeton University Press. doi: 10.2307/j.ctv7h0s6j.26

Towler, A., Kemp, R. I., Burton, A. M., Dunn, J. D., Wayne, T., Moreton, R., & White, D. (2019). Do professional facial image comparison training courses work? *PLoS ONE*, *14*(2), e0211037. doi: 10.1371/journal.pone.0211037

Towler, A., White, D., & Kemp, R. I. (2014). Evaluating training methods for facial image comparison: The face shape strategy does not work. *Perception*, *43*(2-3), 214–218. doi: 10.1068/p7676

Tredoux, C., Nunez, D., Oxtoby, O., & Prag, B. (2006). An evaluation of id: An eigen-face based construction system: reviewed article. *South African Computer Journal*, 90–97.

Tuttenberg, S. C., & Wiese, H. (2019). Learning own- and other-race facial identities from natural variability. *Quarterly Journal of Experimental Psychology*, 174702181985984. doi: 10.1177/1747021819859840

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. doi: 10.1080/14640749108400966

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. doi: 10.1080/14640749208401305

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *Quarterly Journal of Experimental Psychology*, *69*(10), 1996–2019. doi: 10.1080/17470218.2014.990392

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3). doi: 10.18637/jss.v036.i03

Walker, P. M., & Hewstone, M. (2008). The influence of social factors and implicit racial bias on a generalized own-race effect. *Applied Cognitive Psychology*, *22*(4), 441–453. doi: 10.1002/acp.1382

Walker, P. M., & Tanaka, J. W. (2003). An encoding advantage for own-race versus other-race faces. *Perception*, *32*(9), 1117–1125. doi: 10.1068/p5098

Wallraven, C., Whittingstall, L., & Bulthoff, H. H. (2013). Learning to recognize face shapes through serial exploration. *Experimental Brain Research*, *226*(4), 513–523. doi: 10.1007/s00221-013-3463-y

White, D., Burton, A. M., Jenkins, R., & Kemp, R. I. (2014). Redesigning photo-ID to improve unfamiliar face matching performance. *Journal of Experimental Psychology: Applied*, *20*(2), 166–173. doi: 10.1037/xap0000009

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2013). Feedback training for facial image comparison. *Psychonomic Bulletin & Review*, *21*(1), 100–106. doi: 10.3758/s13423-013-0475-3

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PLoS ONE*, *9*(8), e103510. doi: 10.1371/journal.pone.0103510

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1814), 20151292. doi: 10.1098/rspb.2015.1292

Wiese, H., Kaufmann, J. M., & Schweinberger, S. R. (2014). The neural signature of the own-race bias: Evidence from event-related potentials. *Cerebral Cortex*, *24*, 826–835. doi: 10.1093/cercor/bhs369

Wilson, J. P., Bernstein, M. J., & Hugenberg, K. (2016). A synthetic perspective on the own-race bias in eyewitness identification. *Advances in Psychology and Law*, 241–270. doi: 10.1007/978-3-319-43083-6_8

Wittwer, T., Tredoux, C. G., Py, J., & Paubel, P.-V. (2019). Training participants to focus on critical facial features does not decrease own-group bias. *Frontiers in Psychology*, *10*. doi: 10.3389/fpsyg.2019.02081

Woodhead, M. M., Baddeley, A. D., & Simmonds, D. C. V. (1979). On training people to recognize faces. *Ergonomics*, *22*(3), 333–343. doi: 10.1080/00140137908924617

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology, Public Policy, and Law*, *7*(1), 119–133. doi: 10.1037/1076-8971.7.1.119

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own-race bias for face recognition in South Africa and England. *Applied Cognitive Psychology*, *17*(3), 365–373. doi: 10.1002/acp.898

Xiao, W. S., Fu, G., Quinn, P. C., Qin, J., Tanaka, J. W., Pascalis, O., & Lee, K. (2015). Individuation training with other-race faces reduces preschoolers' implicit racial bias: A link between perceptual and social representation of faces in children. *Developmental Science*, *18*(4), 655–663. doi: 10.1111/desc.12241

Young, A. W., & Burton, A. M. (2017). Recognizing faces. *Current Directions in Psychological Science*, *26*(3), 212–217. doi: 10.1177/0963721416688114

Young, A. W., & Burton, A. M. (2018). Are we face experts? *Trends in Cognitive Sciences*, *22*(2), 100–110. doi: 10.1016/j.tics.2017.11.007

Yovel, G., Halsband, K., Pelleg, M., Farkash, N., Gal, B., & Goshen-Gottstein, Y. (2012). Can massive but passive exposure to faces contribute to face recognition abilities? *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 285–289. doi: 10.1037/a0027077

Zhou, X., & Mondloch, C. J. (2016). Recognizing "Bella Swan" and "Hermione Granger": No own-race advantage in recognizing photos of famous faces. *Perception*, *45*(12), 1426–1429. doi: 10.1177/0301006616662046

Appendices

UNIVERSITY OF CAPE TOWN

Department of Psychology

University of Cape Town Rondebosch 7701 South Africa
Telephone (021) 650 3417
Fax No. (021) 650 4104

02 May 2018

Tania Wittwer
Department of Psychology
University of Cape Town
Rondebosch 7701

Dear Tania

I am pleased to inform you that ethical clearance has been given by an Ethics Review Committee of the Faculty of Humanities for your study, *The own-group bias in face recognition: improving accuracy with training.* The reference number is PSY2018 -017.

I wish you all the best for your study.

Yours sincerely

Lauren Wild (PhD)
Associate Professor
Chair: Ethics Review Committee

University of Cape Town
ΨPSYCHOLOGY DEPARTMENT
Upper Campus
Rondebosch

Université
Fédérale
Toulouse
Midi-Pyrénées

Toulouse, le mercredi 3 avril 2019

A l'attention de PY Jacques

CER : Comité d'Ethique sur les Recherches

Objet : Avis de la commission du 19/03/2019

Numéro d'enregistrement : 2019-147

Titre du projet soumis : **Entraînement à la reconnaissance des visages, étude sur eye-tracker**

Porteur de projet : PY Jacques, laboratoire CLLE-LTC, UTJJ

Monsieur,

Compte tenu des éléments fournis dans votre demande, le Comité d'Ethique pour les Recherches de l'Université de Toulouse émet l'avis suivant : **Avis Favorable.**

Nous rappelons, par ailleurs, qu'il relève de la responsabilité des chercheurs de se conformer à leurs obligations légales notamment en ce qui concerne les aspects d'homologation du lieu de recherche ou RGPD : Règlement Général sur la Protection des Données.

Nous restons à votre disposition pour toute question.

Les membres du bureau CER.

Pr Maria Teresa Munoz Sastre

Mª Teresa

Rémi Capa

CER - Université Fédérale Toulouse Midi-Pyrénées
Département Recherche, Doctorat et Valorisation
41, Allées Jules Guesde - CS 61321 - 31013 Toulouse CEDEX 6 - Tél. : 05 61 10 80 30
Courriel : bureau-cerni@univ-toulouse.fr

CONSENT FORM

The investigator of this research is Tania Wittwer, PhD, and it is supervised by Pr. Colin Tredoux (UCT) and Pr. Jacques Py (University of Toulouse – France). The present study is funded by the French project Chaire d'Attractivité « MisIdentificationContact 2016-19 ». For any query, please contact Tania Wittwer (*tania.wittwer@univ-tlse2.fr*).

The aim of this experiment is to study face recognition, and especially the effect of training on face recognition performance. You will complete three tasks during this experiment: the first and the last ones are recognition tasks during which you will look at several faces and then try to remembered them. The tasks between are training tasks. The detailed instructions will be given to you precisely before each task, often directly on the screen. The experiment will be conducted in the lab for *5 sessions* on from the 30th of July, to the 31st of September. Your presence is required for every session, and the 5 points will only be given to you if you complete all the sessions.

Collected data in the context of this experiment are strictly confidential in nature and your name will not be available to the public or used in any part of the final report. All data will be collected anonymously, and kept on a secure place only accessible to scientific managers of the study. Results from this experiment, which could be available to you on simple request, could be share on Conferences in written or oral form or/and be published in conference act or/and scientific journals or reviews.

Furthermore, there are no foreseeable risks in participate to this experiment, and you should not experience any psychological distress nor stress and no harm should befall you. You can stop the experiment at any moment without given any justification and without any repercussions, and your data will be removed from the data collection.

I hereby confirm that the study in which I voluntary has been explained to me and that I hereby give my informed consent to participate to the study. I understand that my participation is voluntary and that I may withdraw my participation at any stage, without repercussions or penalty.

Place and date: ___________

Full name and student number of the participant: ____________

Full name and signature of the researcher: Tania Wittwer

Electronically signed, recorded in two exemplars, one for each part.

DEBRIEF FORM

You took part of an experiment conducted on training and face recognition, and I am grateful for your volunteering. This study was a part of my PhD named "*The own-group bias in face recognition: improving accuracy with training*" which is a collaboration between Universities of Cape Town and Toulouse (France).

Theoretical background of the Research

In the field of forensic psychology, lot of studies highlight a deficit in people recognition called the *other-race bias* (or other-group bias, OGB). The OGB is defined as the difficulty of recognise other-group faces while own-group faces are better remembered.

This OGB effect is known as one of the strongest in forensic psychology, and a meta-analysis found that across 39 scientific papers, the OGB actually explains 15% of the total variance across studies (Meissner & Brigham, 2001). This effect is known to be universal, shared by many group in different countries (Brigham, Bennett, Meissner, & Mitchel, 2007) and not due to perceptible features of groups themselves (Tanaka & Pierce, 2009) and their distinctiveness (Ellis, 1975) but more due to differences in encoding faces (Arizpe, Kravitz, Walsh, Yovel, & Baker, 2016; Hills & Pake, 2013). Nevertheless, the OGB is dependent on the demography of each country or region, and is more pronounced towards minority groups than majority groups (Wright, Boyd, & Tredoux, 2003) even residential segregation can moderate this tendency (Bataille & Hajji, 2017). In field situations, the OGB has serious implications and increases the risk of misidentification of minorities. According to The Innocence Project (2017), 70% of the mistaken identifications in the United States of 351 processed cases were of someone from a minority. However, because the OGB is acquired and dependent on a particular visual and social daily environment, it is possible that it might be decreased or entirely erased in the right circumstances as this bias appears early in life (see Anzures, Quinn, Pascalis, Slater, Tanaka, & Lee, 2013).

Several theories tend to explain the apparition of OGB in terms of visual expertise (see Brigham, Bennet, Meissner & Mitchel, 2007; Michel, Caldara, & Rossion, 2006), social contact (for a review see Pettigrew & Tropp, 2006) or individuation/categorization (for a review see Wilson, Hugenberg, & Bernstein, 2013). From all those theories, increasing visual exposure though positive interactions with motivation to pay attention to other-race people should decrease OGB. Thus, acting on those in training sessions is a way to improve other-race faces recognition, and several studies already shown interesting and promising results (Elliott, Willis, & Goldstein, 1973; Goldstein & Chance, 1985; Hills and Lewis, 2006; Hills & Lewis, 2011; Hills, Cooper, & Pake, 2013; Lebrecht, Pierce, Tarr, & Tanaka, 2009; Paterson et al, 2017).

Explanation and hypothesis of the present study

The experiment you took part of is only one among others I conduct for my thesis. The explanation bellow is applied to all my studies, but each one have its own specifications and hypothesis. In my studies, I attempted to build training tasks to erase the OGB with sometimes exploration of the modification of visual patterns of attention with eye-movement recording, matching task between faces from multi-views, recognition from different views, or field matching/recognition tasks.

As you might have noticed, you saw both Black and White people during the experiment. According to your own-group, you might have the OGB against one or both of the stimuli groups. Thus, the first part of each experiment measures the presence and robustness of the OGB, and because the purpose is to erase it, I expect to find it in every study and for each participant. The third task (after training) has the same goal and thus, I can compare the two OGB rates to observe how it is – or not, modified after and according to the training task. Generally, I expect a reduction of the OGB in each study, meaning after every kind of training I propose.

Then, depending on the study you took part in, I might explore for example how a training task during which participants have to focus on certain parts of a face can modify their visual patterns of attention to faces after the training, and how this can have an impact on recognition ability for other-group faces. I also explore how encoding a face from multiple views (i.e., frontal, three-quarter and profile) can increase face recognition accuracy and decrease the OGB, and how these new abilities developed in the lab can be transposed to a field task. I also explore long-term effects of training.

If you are interested in more explanations of the study you participated in, or in other experiments I have conducted or will conduct for my thesis, or more generally on my topic, you can take a look at my proposal at the following link: https://tinyurl.com/tania-w-proposal

For any questions or concerns about my research, or if you are interested in a global report on my results, please contact me by email at: tania.wittwer@univ-tlse2.fr.

Tania Wittwer,
PhD Candidate
Department of Psychology,
University of Cape Town

References

Arizpe, J., Kravitz, D. J. Walsh, V., Yovel, G., & Baker, C. I. (2016). Differences in Looking at Own- and Other-Group Faces Are Subtle and Analysis-Dependent: An Account of Discrepant Reports. *PLoS ONE, 11*(2), e0148253. doi:10.1371/journal.pone.0148253

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., Tanaka, J. W., & Lee, K. (2013). Developmental Origins of the Other-Group Effect. *Current Directions in Psychological Science*, *22*(3), 173?178. doi:10.1177/0963721412474459

Bataille, A., & Hajji, S. (2017). Le Biais de Reconnaissance Inter-Ethnie : une Méta-Analyse de 47 ans de Recherche (mémoire de maîtrise). Université Toulouse Jean Jaurès: Toulouse.

Brigham, J. C., Bennett, L. B., Meissner, C. A., & Mitchell, T. L. (2007). The Influence of Group on Eyewitness Memory. In R. Lindsay, D. Ross, J. Read, & M. Toglia (Eds.), *Handbook of Eyewitness Psychology* (pp. 257?281). Mahwah, NJ: Lawrence Erlbaum & Associates.

Elliott, E. S., Wills, E. J., & Goldstein, A. G. (1973). The Effects of Discrimination Training on the Recognition of White and Oriental Faces. *Bulletin of the Psychonomic Society*, *2*(2), 71?73. doi:10.3758/bf03327717

Ellis, H. D. (1975). Recognizing Faces. *British Journal of Psychology*, *66*(4), 409?426. doi:10.1111/j.2044-8295.1975.tb01477.x

Goldstein, A. G., & Chance, J. E. (1985). Effects of Training on Japanese Face Recognition: Reduction of the Other-group Effect. *Bulletin of the Psychonomic Society*, *23*(3), 211?214. doi:10.3758/bf03329829

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013). Removing the Own-group Bias in Face Recognition by Attentional Shift Using Fixation Crosses to Diagnostic Features: An Eyetracking Study. *Visual Cognition*, *21*(7), 876 ?898. doi:10.1080/13506285.2013.834016

Hills, P. J., & Lewis, M. B. (2006). Short Article: Reducing the Own-Group Bias in Face Recognition by Shifting Attention. *Quarterly Journal of Experimental Psychology, 59*(6), 996 ?1002. doi :10.1080/17470210600654750

Hills, P. J., & Pake, J. M. (2013). Eye-tracking The Own-group Bias in Face Recognition: Revealing the Perceptual and Socio-cognitive Mechanisms. *Cognition, 129*(3), 586 ?597. doi:10.1016/j.cognition.2013.08.012

Lebrecht, S., Pierce, L. J., Tarr, M. J., & Tanaka, J. W. (2009). Perceptual Other-Group Training Reduces Implicit Racial Bias. *PLoS ONE, 4*(1), e4215. doi:10.1371/journal.pone.0004215

Meissner, C. A., & Brigham, J. C. (2001). Thirty Years of Investigating the Own-group Bias in Memory for Faces: A Meta-Analytic Review. *Psychology, Public Policy, and Law, 7*(1), 3?35. doi :10.1037/1076-8971.7.1.3

Michel, C., Caldara, R., & Rossion, B. (2006). Same-group Faces are Perceived More Holistically than Other-group faces. *Visual Cognition, 14*(1), 55 ?73. doi:10.1080/13506280500158761

Paterson, H. M., Luppino, D., Calderwood, C., MacDougall, H. G., Taubert, J., & Kemp, R. I. (2017). Can Training Improve Eyewitness Identification? The Effect of Internal Feature Focus on Memory for Faces. *Psychology, Crime & Law, 23*(10), 927 ?945. doi:10.1080/1068316x.2017.1346099

Pettigrew, T. F., & Tropp, L. R. (2006). A Meta-Analytic Test of Intergroup Contact Theory. *Journal of Personality and Social Psychology, 90*(5), 751?783. doi:10.1037/0022-3514.90.5.751

Tanaka, J. W., & Pierce, L. J. (2009). The neural plasticity of other-group face recognition. *Cognitive, Affective, & Behavioral Neuroscience, 9*(1), 122 ?131. doi:10.3758/cabn.9.1.122

The Innocence Project (2017). Retrieved from https ://25years.innocenceproject.org

Wilson, J. P., Hugenberg, K., & Bernstein, M. J. (2013). The Cross-Group Effect and Eyewitness Identification: How to Improve Recognition and Reduce Decision Errors in Eyewitness Situations. *Social Issues and Policy Review, 7*(1), 83 ?113. doi:10.1111/j.1751-2409.2012.01044.x

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial Contact and the Owngroup Bias for Face Recognition in South Africa and England. *Applied Cognitive Psychology, 17*(3), 365 ?373. doi:10.1002/acp.898

Observation Grid

Participant ID	Day/Hour	% of people	Condition	People	Identification decision	Confidence *(0-100%)*	Decision latency (min, sec)	Comments
			TA+TA TP+TP TA+TPw TA+TPb	M	ID Rejection			
				I	ID Rejection			
			TA+TA TP+TP TA+TPw TA+TPb	M	ID Rejection			
				I	ID Rejection			
			TA+TA TP+TP TA+TPw TA+TPb	M	ID Rejection			
				I	ID Rejection			
			TA+TA TP+TP TA+TPw TA+TPb	M	ID Rejection			
				I	ID Rejection			
			TA+TA TP+TP TA+TPw TA+TPb	M	ID Rejection			
				I	ID Rejection			
			TA+TA TP+TP TA+TPw TA+TPb	M	ID Rejection			
				I	ID Rejection			
			TA+TA TP+TP TA+TPw TA+TPb	M	ID Rejection			
				I	ID Rejection			
			TA+TA TP+TP TA+TPw TA+TPb	M	ID Rejection			
				I	ID Rejection			

Résumé français du contenu de la thèse

Le biais intergroupe dans le traitement des visages : l'effet de l'entraînement sur la performance de reconnaissance

Contexte de la présente recherche

La présente thèse a été effectuée dans le cadre d'une convention de cotutelle entre l'Université de Toulouse Jean Jaurès (UT2J, Toulouse, France), et de l'Université de Cape Town (University of Cape Town, UCT, Le Cap, Afrique du Sud), sous la direction du Professeur Jacques Py et du Professeur Colin G. Tredoux. Cette recherche s'inscrit dans le projet de Chaire d'attractivité 'MisIdentification Contact'. Les données des études menées lors de cette thèse ayant été recueillies dans les deux pays, et le sujet de recherche incluant des mots ayant une connotation sensible, des précisions sur les termes utilisés sont nécessaires et ce, notamment cas ils portent en leur sens une sensibilité différente selon la langue, et le pays dans lequel ces termes sont utilisés. De ce fait, les termes sont adaptés dans ce résumé afin de rester fidèle à la fois aux règles éthiques, mais également au nom des populations étudiées. Pour rapporter les résultats d'études antérieures, ainsi que pour décrire les participants des études menées en France, les termes 'personne typée d'origine européenne' et 'personne typée d'origine africaine' sont utilisés. Pour décrire les études menées en Afrique du Sud, les termes 'White', 'Black' et 'Coloured[2]' sont utilisés, afin qu'une traduction ne dénature pas le groupe dans lequel les participants se sont identifiés. Ces termes sont utilisés à la fois pour les participants,

---

[2] Les 'Coloured' sont principalement les descendants du peuple indigène des Khoi-San (vivant en Afrique du Sud depuis des centaines d'années) ou d'esclaves d'origine Indonésienne (présents depuis environ 300 ans sur le territoire).

mais également pour les stimuli, qui sont des photographies majoritairement recueillies en Afrique du Sud. Toutefois, les termes 'personnes d'origine européenne' et 'White' décrivent une population similaire considérant le biais intergroupe.

Avant toutes choses, il est important de situer le contexte démographique de chacun des pays dans lesquels les expérimentations de cette thèse se sont déroulées. Alors que l'Institut National de la Statistiques et des Etudes Economiques (INSEE) n'est pas autorisé à recueillir des données ethniques en France, cette information est demandée de manière quasiment systématique en Afrique du Sud, que ce soit dans les documents administratifs quelconques, ou recueillie par l'organisme similaire à l'INSEE: Stats South Africa. Ces données révèlent que les 'Black' sont majoritaires à l'échelle du pays, mais à l'échelle de Cape Town, ils ne sont pas beaucoup plus représentés que les 'Coloured'. A l'université de Cape Town, le groupe White, majoritaire pendant des années, a été rejoint en 2017, et dépassé en 2018, par les 'Black'. Les 'Coloured', quant à eux, sont sous-représentés à l'université de Cape Town, alors même qu'ils sont presque autant présents que les 'Black', à l'échelle de la ville. Ces spécificités démographiques ont été à la base de la décision d'étudier ces trois groupes de population. Toutefois, l'Afrique du Sud, tout comme la France, sont des pays diversifiés et composés de personnes appartenance également à d'autres groupes ethniques.

Éléments théoriques

En observant une rue bondée, la terrasse d'un café, celle d'un restaurant, ou encore les tables d'une bibliothèque universitaire à Toulouse ou à Cape Town, la notion de diversité est très saillante pour caractériser la population qui s'y trouve. Loin d'être le cas il y a des dizaines, voire des centaines d'années, l'Histoire - et notamment le colonialisme, la mondialisation, ou encore le tourisme pour ne prendre que ces exemples, sont autant de facteurs qui ont contribué au développement de la démographie de nos pays. Par conséquent, les populations se sont mélangées, diversifiées et la majorité des sociétés modernes sont aujourd'hui multiculturelles, multi-religieuses et multi-ethniques. Contrairement à des pays plus homogènes, cette diversité nous amène à fréquemment rencontrer des personnes de différents groupes[3]. Alors que l'exposition à une telle diversité est positive sur de bien nombreux points, des conséquences inattendues sont apparues : nous sommes amenés à, et devons être capables de, discriminer (c'est-à-dire., différencier), et reconnaître des personnes de groupes différents, et plus particulièrement, différent du nôtre (c'est-à-dire., exogroupes). Reconnaître un visage, tâche d'ores et déjà difficile en elle-même, l'est d'autant plus lorsqu'il s'agit de reconnaître des personnes appartenant à un exogroupe. Les conséquences de cette difficulté peuvent être plus ou moins sérieuses. Cette difficulté peut nous amener à confondre deux camarades de classe, saluer la mauvaise personne, ou encore se retrouver en difficulté lorsqu'il s'agit de différencier les membres d'une équipe de Basketball que l'on ne connait pas, situations pouvant être inconfortables, et dans certains cas, conduire à des propos inappropriés, voire discriminatoires. Pour illustrer ce propos, et dans le but d'éveiller les consciences sur les conséquences de ce problème, le photographe Peter De Vito (https://www.peterdevito.com) a photographié des personnes Asiatiques, arborant sur une partie de leur visage un autocollant sur lequel il était écrit "we all look the same" (« on se ressemble tous »). Les photographies sont accompagnées de témoignages des modèles, à propos de leurs expériences, majoritairement négatives, en relation avec cette célèbre phrase. La difficulté à discriminer et reconnaître des person-

---

[3] Le terme de 'groupe' est utilisé en remplacement des termes 'race' ou 'ethnie' habituellement utilisés dans ce domaine de recherche.

nes issues d'un exogroupe se retrouve également lors de tâches impliquant de vérifier qu'une photographie sur un passeport correspond bien à la personne détentrice du passeport, ou dans l'augmentation des erreurs d'identifications judiciaires. Selon l'association The Innocence Projet (https://www.innocenceproject.org), qui œuvre depuis 27 ans aux Etats-Unis pour aider des personnes incarcérées à prouver leur innocence, 367 personnes ont été innocentées grâce à des preuves ADN. Parmi ces affaires, 69% impliquaient des erreurs de témoignage oculaire, dont 42% concernait des cas d'identifications intergroupes, dans lesquels le témoin et le suspect appartenaient à deux groupes différents. Ces situations montrent l'importance de la compréhension des processus impliqués dans la reconnaissance des visages, et plus particulièrement dans des cas intergroupes.

La reconnaissance des visages et le biais intergroupe

Un visage est composé de propriétés relationnelles de premier ordre et de second ordre (Diamond & Carey, 1986). Les éléments de premier ordre sont ceux qui composent un visage et qui le rendent reconnaissable comme tel : la disposition de deux yeux, au-dessus d'un nez, lui-même au-dessus d'une bouche. Les éléments de second ordre, quant à eux, concernent les relations que ces traits de visages entretiennent les uns avec les autres, définissant la configuration unique de chacun des visages : la distance entre les yeux, la configuration générale, etc. Puisqu'un visage est facilement reconnu comme tel, et parce que nous sommes habitués et performants à reconnaître des visages familiers, il est souvent considéré - à tort, que nous sommes experts en reconnaissance des visages. En réalité, nous ne sommes expert qu'en ce qui concerne la reconnaissance des visages familiers, et présentons des compétences de reconnaissance altérées dans le traitement de visages non familiers (Young & Burton, 2018, 2017).

De par l'exposition répétée et le contact prolongé avec certains visages, et ce parfois depuis notre plus tendre enfance, nous devenons experts dans la reconnaissance des visages des personnes constituant notre cercle social immédiat (ex. membres de la famille, amis). C'est en bénéficiant d'une exposition répétée à ces visages, et plus particulièrement dans différentes conditions à la fois contextuelles (luminosité, distance, pose) ou propres au visage (émotion, coupe de cheveux, pilosité faciale, lunettes, etc.), que la reconnaissance des visages familiers est meilleure que pour les visages non familiers (Bruce, 1994; Burton, 2013; Johnston & Edmonds, 2009). En effet, l'exposition à des visages sous différentes variations permet de développer une bonne représentation en mémoire de ces visages, nécessaire pour en améliorer leur récupération par ailleurs (Andrews et al., 2015). Contrairement aux visages familiers, les visages non-familiers sont bien souvent encodés sous une variation unique, limitant ainsi leur représentation en mémoire ainsi que leur récupération, ainsi dépendante des conditions d'encodage (Bruce, 1982; P. J. Hancock et al., 2000). Les visages d'exogroupes bénéficient également de l'effet positif de la familiarité. Ainsi, l'effet de la familiarité sur la représentation en mémoire et la récupération est telle qu'elle outrepasse l'effet de l'appartenance à un exogroupe. De fait, les visages familiers bénéficieront d'une meilleure discrimination et reconnaissance que les visages non familiers, et ce, indépendamment de la relation intergroupe (Laurence et al., 2016; Zhou & Mondloch, 2016).

Contrairement aux visages familiers, la discrimination et la reconnaissance des visages non familiers appartenant à des exogroupes est encore plus difficile que pour les visages non-familiers de l'endogroupe. Ce phénomène, portant le nom de biais intergroupe (own-group bias; OGB), est bien souvent décrit comme un amoindrissement de la reconnaissance des visages. L'OGB est, en fait, le résultat d'une adaptation à l'environnement visuel spécifique d'un observateur. En effet, dès leur plus jeune âge, les enfants sont principalement confrontés à des personnes de leur endogroupe (Heron-Delaney et al., 2011). Ainsi, alors qu'ils présentent une préférence visuelle pour les visages de leur endogroupe plutôt que ceux d'un exogroupe dès l'âge de 3 mois (Kelly et al., 2005), une difficulté lors de la reconnaissance des visages de l'exogroupe n'apparaît qu'à partir de 6 mois (Kelly, Liu, et al., 2007), résultant en un OGB (c'est-à-dire une grande difficulté) déjà robuste à partir de 9 mois (Anzures et al., 2013; Kelly, Quinn, et al., 2007). La nature acquise de l'OGB est également visible considérant le cas d'enfants ayant été adoptés par une famille appartenant à un groupe différent du leur. En effet, ces enfants présentent bien souvent un OGB à l'encontre de leur endogroupe biologique, et donc, dans la même direction que celui de leur famille adoptive. La cause de cet OGB 'inversé' est très probablement due à l'environnement dans lequel ces enfants évoluent, ayant comme conséquence le développement de leurs capacités à discriminer et reconnaître des personnes de leur endogroupe adoptif. Des études sur ces enfants ont mis en avant qu'ils sont tout aussi capables de discriminer et de reconnaître des visages du groupe de leurs parents biologiques que du groupe de leurs parents adoptifs (De Heering et al., 2010), voire qu'ils présentent une meilleure performance pour les visages appartenant au groupe de leurs parents adoptifs, plutôt que ceux du groupe de leurs parents biologiques (de Viviés et al., 2010; Sangrigoli et al., 2005). Plus généralement, les enfants présentent un OGB de plus faible magnitude que celui des adultes (Goodman et al., 2007). Ces résultats soutiennent, d'une part, l'idée selon laquelle l'OGB est une conséquence de l'homogénéité de l'environnement visuel dans lequel les enfants grandissent et, d'autre part, du fait que l'OGB se développe dans le temps (Hills & Lewis, 2018; McKone et al., 2019). L'apparition et la consistance de l'OGB est alors relatif à un manque d'exposition à des personnes d'exogroupes, alors même que l'exposition à des personnes de leur endogroupe est quotidienne.

L'OGB se caractérise donc par une meilleure performance de discrimination et de reconnaissance des visages de l'endogroupe, au détriment de celle des visages d'exogroupes. Démontré expérimentalement pour la première fois en 1969 par Malpass and Kravitz, l'OGB a dès lors fait l'objet de nombreuses études. Une méta-analyse de 39 études (Meissner & Brigham, 2001) a montré que les visages de l'endogroupe sont 1.4 fois plus à même d'être correctement reconnus par rapport aux visages d'un exogroupe, qui eux, ont 1.56 fois plus de (mal)chance de faire l'objet d'une reconnaissance erronée. L'OGB se traduit également par un taux élevé de fausses alarmes (c'est-à-dire., de reconnaissances erronées) concernant l'exogroupe par rapport à l'endogroupe, alors que les reconnaissances correctes restent plutôt stables pour les deux groupes (Bertone et al., 1995; Meissner et al., 2005; Meissner & Brigham, 2001; M. G. Rhodes & Anastasi, 2012; Slone et al., 2000). Au-delà de ces caractéristiques, l'OGB serait universel, et partagé par de nombreux groupes, au sein de nombreux pays.

Les processus visuels de traitement des visages et l'OGB

Lors du traitement d'un visage, un observateur peut avoir recours à deux processus : un processus parcellaire, ou un processus configural. Le processus parcellaire est le fait de traiter un visage trait par trait, alors que le processus configural le considère dans son ensemble, à savoir en considérant également la position des traits les uns par rapport aux autres. Ces deux processus peuvent être utilisés indépendamment, ou en combinaison pour un encodage plus robuste (Collishaw & Hole, 2000; Hayward et al., 2008; Tanaka & Simonyi, 2016). Le processus configural résulte néanmoins en une meilleure performance que le processus parcellaire (Tanaka & Farah, 1993), et est principalement associé au traitement des visages familiers, ou de ceux de l'endogroupe. A contrario, le traitement parcellaire est plus fréquemment utilisé lors du traitement de visages appartenant à un exogroupe (Byatt & Rhodes, 2004; Le Grand et al., 2001; G. Rhodes et al., 2009). Bien que donnant lieu à de moins bons résultats que le traitement configural, le traitement parcellaire peut être relativement efficace lorsqu'il est dirigé vers les parties critiques (c'est-à-dire., contenant le plus de variabilité au sein d'un groupe donné) des visages. Traiter les caractéristiques permettant de différencier les visages les uns des autres au sein d'un groupe montre que le traitement parcellaire peut s'avérer utile dans la discrimination des visages, lorsque ce traitement s'appuie sur les caractéristiques diagnostiques du groupe en question.

Le modèle multidimensionnel de représentation des visages en mémoire

Selon le *multidimensionnal face space model* (MDFS; Valentine, 1991; Valentine & Endo, 1992; Valentine et al., 2016), les visages sont représentés en mémoire dans un espace multidimensionnel. La reconnaissance d'un visage dépendrait ainsi de sa localisation dans cet espace. Un visage est encodé et représenté en mémoire à la fois en fonction de ses caractéristiques propres mais aussi par rapport aux autres visages déjà représentés dans l'espace multidimensionnel. Afin de visualiser le modèle, on pourrait imaginer deux lignes, qui s'entrecoupent en leur origine, traduisant chacune toutes les variations d'une dimension (c'est-à-dire., caractéristiques d'un trait du visage, par exemple, la forme des yeux, ou le blanc des yeux). L'aire entourant l'intersection de ces dimensions correspond à la tendance centrale, à savoir la zone dans laquelle la majorité des visages sont représentés puisqu'ils partagent des variations similaires des dimensions. Ainsi, chaque nouveau visage encodé est représenté comme un point unique par rapport à ses propres variations dans chacune des dimensions. Même s'il est plus facile de se faire une idée du modèle en ne parlant que de deux dimensions, le modèle propose que chacune des caractéristiques d'un trait du visage se traduise en une dimension particulière. Prenant de la distance par rapport à la tendance centrale, les visages atypiques sont eux, représentés selon une distance plus relative en fonction de leur propre caractéristique distinctive, et de ce fait, sont plus facilement reconnaissables au sein de tous. Considérant l'OGB dans le présent modèle, les visages d'un exogroupe sont représentés à la fois comme étant un visage atypique puisqu'une des caractéristiques est saillante pour les regrouper, comme la couleur de la peau par exemple[4], mais en l'absence de dimensions adaptées aux spécificités phénotypiques de ces visages, ils sont représentés en clusters. De ce fait, il est plus difficile de les différencier et les distinguer les uns des autres. Il existe deux versions du MDFS (Figure 6.2) : un modèle basé sur la norme, et un modèle basé sur les exemplaires. Le modèle basé sur la norme soutient que les visages sont encodés relativement à une norme (un prototype), et c'est la différenciation avec cette norme qui joue un rôle dans la représentation des nouveaux visages dans l'espace des visages.

---

[4] Voir le modèle de catégorisation perceptuelle, dans lequel les visages d'exogroupe sont catégorisés en fonction de la couleur de peau, trait saillant, plutôt qu'individualisés comme le sont les visages de l'endogroupe (Levin, 1996, 2000). Une alternative de ce modèle, se distinguant du MDFS, est le modèle de catégorisation sociale (Sporer, 2001) qui indique que les visages sont individualisés ou catégorisés en fonction d'une caractéristique d'appartenance sociale (par exemple, l'université), résultant en une meilleure reconnaissance des personnes de l'endogroupe social plutôt que ceux de l'exogroupe social, renvoyant l'effet du groupe 'ethnique' au second plan.

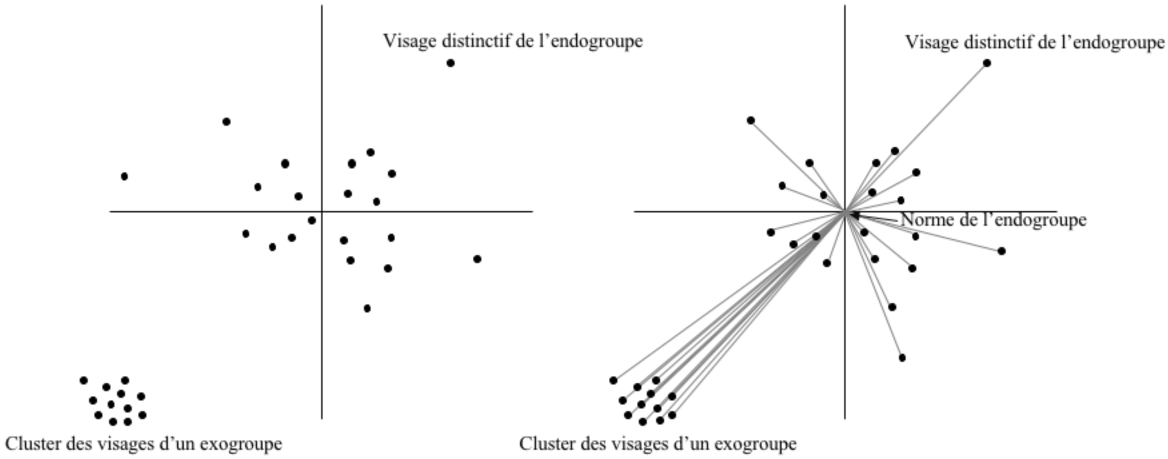

Figure 6.2 – Représentation schématique du *multidimensional face space model* de Valentine, incluant la représentation des visages de l'endogroupe et d'un exogroupe. A gauche, le modèle basé sur les exemplaires, dans lequel tous les visages sont représentés les uns relativements aux autres, et les visages de l'endogroupe sont représentés en cluster distictif puisque représentés relativement à des dimensions inappropriées. A droite, le modèle basé sur la source selon lequel chaque visage est représenté en fonction du prototype de l'endogroupe (la norme), les visages de l'exogroupe étant représentés tous dans la même direction et en cluster, en l'absence de leur norme propre.

Le modèle basé sur les exemplaires soutient que tous les visages sont encodés en fonction de leurs relations les uns avec les autres, et non pas par rapport à une norme unique. Le second modèle est préféré lorsqu'on parle de l'OGB. En effet, selon ce dernier, l'exposition à de nombreux visages permettrait d'affiner les dimensions et d'en créer des nouvelles, plus appropriées à la discrimination des membres du groupe concerné. De fait, en multipliant les interrelations des visages les uns avec les autres, le modèle s'affine jusqu'à permettre, à terme, de développer un espace avec des dimensions adaptées permettant une meilleure représentation et ainsi, une meilleure récupération. L'existence de multiples sous-espaces a également été considérée (Little et al., 2008), chacun d'entre eux pouvant être dédié à une catégorie spécifique, comme le groupe 'ethnique', ou le genre. Ces sous-espaces pourraient être inclus dans un espace globale, ou être des sous-espaces indépendants vers lesquels nous serions redirigés en fonction de la motivation (voir Valentine et al., 2016). Au lieu d'être représentés relativement à tous les autres visages, les nouveaux visages seraient tout d'abord 'triés' en fonction du groupe auquel ils appartiennent, et seraient ensuite codés dans l'espace pertinent.

Le contact avec des visages d'un groupe donné au détriment d'autres groupes résulte donc en une représentation pertinente des visages en mémoire, alors même que l'absence, ou le plus faible contact avec des visages d'exogroupes ne permet pas le développement d'une représentation en mémoire suffisamment fine des groupes considérés. En effet, plus on a de contact (au sens perceptif du terme) avec des personnes d'un groupe donné, plus la représentation en mémoire est affinée et robuste, et plus la discrimination et/ou la reconnaissance est facile et correcte. Toutefois, le contact strictement visuel, bien qu'il soit difficile à isoler du contact social, ne semble pas être suffisant pour éliminer l'OGB. Au-delà donc d'une simple exposition, l'interaction avec des personnes d'exogroupes a également un certain intérêt. Plus un observateur est en contact avec des membres d'exogroupes, meilleures sont ses compétences de discrimination et de reconnaissance (Goldstein & Chance, 1985; Michel, Caldara, & Rossion, 2006). Le contact explique, en fait, 2% de la variance de l'OGB (Meissner & Brigham, 2001). Pour illustrer l'importance du contact, des études menées avec des enfants Zimbabwéens fréquentant une école mixte (Chiroro & Valentine, 1995) ou des expatriés Asiatiques vivant en Allemagne depuis 22 mois en moyenne (Wiese et al., 2014) ont révélé une meilleure reconnaissance des visages d'un exogroupe typé d'origine européenne, par rapport à des participants présentant un plus faible taux de contact. Le contact apparaît alors comme un élément important dans le développement et le maintien de l'OGB. À l'inverse, et puisque l'OGB se développe au fur et à mesure de l'exposition et du contact, il apparaît comme possible de pouvoir le réduire, voire l'éliminer, en manipulant le contact avec des exogroupes.

Un cadre explicatif de l'OGB : le CIM

Considérant le principe d'individualisation, qui se traduit par le fait de traiter un visage par ses spécificités propres, Hugenberg et al. (2010) ont développé un modèle intégratif : le modèle de catégorisation-individualisation (Categorization-Individuation Model, CIM). Ce modèle s'appuie sur l'interaction de trois concepts : la catégorisation, l'expérience, et la motivation. La *catégorisation* est l'activation automatique d'une catégorisation sociale au détriment de l'individualisation des membres de cette catégorie, ayant pour conséquence une confusion entre les différents membres de cette catégorie. La catégorisation fait échos aux modèles de Levin (2000) et de Sporer (2001). L'*expérience* renvoie au fait d'avoir été exposé et confronté à des personnes d'exogroupes, donnant lieu à des expériences d'individualisation par le passé. La *motivation* se traduit par l'apparition d'une attention sélective, définissant l'intérêt et la pertinence de la mobilisation d'un processus d'individualisation. Le CIM part du principe que, si les observateurs ont *a priori* les capacités cognitives et visuelles pour traiter les visages d'exogroupes, de par leur expérience et exposition à ces visages, alors induire une motivation à les individualiser contrerait la catégorisation mise en place automatiquement, améliorant alors la discrimination et la reconnaissance de ces visages. En considérant ce modèle, le développement d'une expérience perceptive, dans le cas où elle n'existerait pas *a priori*, est primordial afin ensuite d'induire de l'individualisation par le biais de la motivation.

Les études sur l'entraînement

Compte tenu de l'importance de la reconnaissance des visages dans la vie quotidienne de tout un chacun, et plus particulièrement les conséquences que peuvent avoir les erreurs de reconnaissance des visages d'exogroupes, un champ de recherche s'est constitué sur la compréhension et le développement de la discrimination et de la reconnaissance des visages. Les recherches sur ce sujet se distinguent de l'apprentissage d'identités spécifiques via l'exposition multiple à des visages dans le but de développer la familiarité, d'une part, et de l'entraînement au développement de compétences perceptives généralisables, d'autre part. En effet, un ensemble d'études a démontré que, puisque la familiarité est développée en réponse à de multiples confrontations à un même visage et sous différentes conditions (Burton, 2013), demander à des participants d'apprendre des visages spécifiques à l'aide de tâches les exposant à des variations de ces identités pourrait améliorer la reconnaissance de ces mêmes identités sous une différente variation (Hussain et al., 2009; Longmore et al., 2008; Matthews et al., 2018; Menon et al., 2015). Des études ont effectivement montré que plus les participants sont exposés à des variations différentes, plus leur performance est élevée (Clutterbuck & Johnston, 2002, 2004, 2005; Dowsett et al., 2016; Laurence & Mondloch, 2016), et ce que l'exposition ait été induite de façon active (par une tâche de pratique particulière) ou de façon passive, par une simple exposition (Andrews et al., 2015). En réalité, les participants semblent construire une bonne représention de ces visages, indépendamment de leur volonté de le faire (Kramer et al., 2015; Menon et al., 2015). Alors même que ces études ont été menées avec des participants et stimuli du même groupe, d'autres études utilisant des paradigmes intergroupes ont mis en avant un effet positif de ce type de pratique sur la reconnaissance d'identités spécifiques, même lorsque ces dernières appartiennent à un exogroupe (McKone et al., 2007; Tuttenberg & Wiese, 2019), bien que d'autres études aient montré que la performance augmente globalement, l'OGB reste présent (Cavazos et al., 2019; Hayward et al., 2017; Proietti et al., 2019). Toutes ces études démontrent qu'il est possible d'apprendre des identités spécifiques, ce qui est cohérent avec l'idée selon laquelle les visages deviennent familiers au fur et à mesure des expositions. Toutefois, il n'est pas question de généralisation à de nouvelles identités dans ces études, dont l'objectif n'était pas d'améliorer la discrimination et la reconnaissance de visages d'un groupe particulier de manière générale, mais de visages individuels.

Contrairement à ces études, dans une autre ligne de recherche, des auteurs ont dirigé l'attention visuelle d'observateurs sur des parties critiques et pertinentes des visages pour en améliorer la discrimination et la reconnaissance. Partant du principe que la partie inférieure du visage est plus pertinente pour discriminer et reconnaître des visages typés d'origine africaine, et la partie supérieure pour reconnaître des visages typés d'origine européenne (Ellis et al., 1975; Shepherd & Deregowski, 1981), des auteurs (Hills & Lewis, 2011) ont exploré les effets d'une attention dirigée vers ces parties-là des visages en utilisant des croix de fixation. Les participants, tous typés d'origine européenne, reconnaissaient mieux les visages typés d'origine africaine plutôt que les visages typés d'origine européenne lorsque la croix de fixation se situait dans la partie inférieure des visages, plutôt que dans la partie supérieure. L'inverse était observé pour les croix de fixations localisées sur la partie supérieure des visages, favorisant la reconnaissance des visages typés d'origine européenne par rapport aux visages typés d'origine africaine. Cette étude a été répliquée avec des participants typés d'origine africaine (Hills et al., 2013). Ces deux études ont permis l'élimination de l'OGB par la redirection des patterns visuels automatiques vers les parties critiques des visages de chacun des groupes. Ces études montrent l'effet positif de la focalisation pertinente sur l'élimination de l'OGB, mais n'ont, elles n'ont plus, pas eu pour objectif de tester la persistance de ces patterns lors d'une tâche de généralisation à de nouveaux visages jamais présentés au préalable, et plus particulièrement une fois que les participants ont recours à une exploration visuelle libre.

Peu d'études ont également cherché à construire un entraînement qui amènerait les participants à développer des compétences spécifiques dans le traitement des visages afin d'éliminer l'OGB dans une tâche de généralisation. Relativement aux présents objectifs de recherche, un total de 10 études publiées dans neuf articles ont mis en place une méthodologie similaire et surtout, testant ainsi un effet de l'entraînement. Parmi ces études, la majorité a eu recours à une tâche d'individualisation de visages appartenant à un exogroupe, à travers l'association d'images et d'étiquettes, dans le but d'induire l'individuation des visages de ce groupe dans une tâche de généralisation (Elliott et al., 1973; Goldstein & Chance, 1985; Matthews & Mondloch, 2018; McGugin et al., 2011; Stahl, 2010; Tanaka & Pierce, 2009). Considérant les résultats des études dirigeant le traitement des visages vers les parties critiques de ces derniers, en fonction de leur groupe, d'autres études ont utilisé la modification de certaines caractéristiques des visages dans le but de diriger les participants à se focaliser sur ces parties spécifiques (Hills & Lewis, 2006; Lavrakas et al., 1976). Toutes ces études ont montré que l'entraînement améliorait la reconnaissance des visages indépendamment du groupe. Bien que ce ne soit pas systématiquement testé, l'entraînement a également bien souvent amélioré la reconnaissance des visages de l'exogroupe, provoquant parfois une élimination de l'OGB. Ces études montrent qu'un entraînement, parfois relativement court (session unique), a un effet positif sur la reconnaissance des visages, et en particulier la reconnaissance des visages d'exogroupes lorsque les participants ont été entraînés à les individualiser, ou à les traiter en se focalisant sur leurs caractéristiques critiques. Ces études s'étant principalement focalisées sur un processus d'individuation à travers une association entre une photographie et une étiquette, il semble intéressant de s'attarder sur le développement d'un processus d'individualisation plus visuel, en poursuivant les travaux ayant montré qu'un traitement spécifique des visages d'exogroupes permettaient d'améliorer leur reconnaissance.

Considérant ces études, les travaux inclus dans cette thèse ont eu pour objectif d'induire une individualisation perceptive à travers un entraînement à la focalisation sur les traits qui rendent un visage distinctif et différenciable des aux autres membres d'un même groupe (en pariculier d'un exogroupe). Ces entraînements avaient pour ambition d'éliminer l'OGB à travers le développement et/ou l'amélioration de stratégies visuelles spécifiques, et généralisables.

Mesures et analyses

Afin de mesurer l'OGB, les mesures issues de la théorie de la détection du signal (SDT; Stanislaw & Todorov, 1999) ont été utilisées, et plus particulièrement la mesure non-paramétrique de performance de discrimination (A')[5]. La performance de discrimination, lors d'une tâche de reconnaissance par exemple, traduit la capacité à différencier un signal (c'est-à-dire une photographie présentée en encodage), du bruit (c'est-à-dire d'une image présentée uniquement en reconnaissance). Afin d'analyser l'effet de l'entraînement, des modèles linéaires généraux ou mixtes ont été utilisés, afin de prédire l'effet des différentes variables (groupe des participants, groupes des stimuli, entraînement) sur la performance de discrimination. Toutes les analyses ont été effectuées avec R (R Core Team, 2017)[6].

---

[5] Compte-tenu de la nature de ce résumé et de l'objectif principal de ma thèse, seule cette mesure sera rapportée ici.

[6] Pour le détail des packages, se référer p.98.

Étude 1 : Entraîner des participants à se focaliser sur des caractéristiques critiques des visages n'élimine pas le biais inter-ethnique[7]

Des études ont montré que pour encoder un visage typé d'origine européenne de façon efficace, un observateur devrait davantage se focaliser sur la moitié supérieure d'un visage, alors que pour les visages typés d'origine africaine, il faudrait plus porter son attention sur la partie inférieure du visage (Hills & Lewis, 2006, 2011; Hills & Pake, 2013). Dans cette étude, des participants typés d'origine européenne ont été amenés à se focaliser sur la partie inférieure des visages présentés, dans le but de réduire l'OGB.

Hypothèses. Dans cette étude, les participants étant typés d'origine européenne, il était attendu (1) qu'ils se focalisent de manière spontanée sur le haut du visage avant l'entraînement quels que soient les stimuli (endogroupe ; exogroupe) ; (2) qu'ils se focalisent plus sur la partie inférieure plutôt que supérieure du visage après l'entraînement plutôt qu'avant ; (3) qu'ils présentent un OGB avant l'entraînement ; (4) que la diminution de l'OGB après l'entraînement soit expliquée par le changement de l'exploration visuelle des visages.

Echantillon. Trente participants typés d'origine européenne (22 femmes, M_{age} = 22.73, SD_{age} = 4.91) recrutés sur le campus de l'Université Toulouse Jean Jaurès.

Design. Cette étude suit un plan expérimental factoriel intra-sujet incluant deux variables : le moment de la mesure de l'OGB (avant ; après l'entraînement) et le groupe des stimuli (endogroupe ; exogroupe).

---

[7] Cette étude a fait l'objet d'une publication chez Frontiers, voir Wittwer et al., 2019

Matériel. Les photographies de 140 jeunes hommes différents (70 de chaque groupe) présentant une expression neutre ont été utilisées dans cette étude. Parmi ces photographies, 40 étaient utilisées en pré-test, et 40 en post-test. Pour la tâche d'entraînement, 60 essais ont été construits (30 pour chaque groupe). Dans chacun de ces essais, et donc pour chacune des 60 photographies, six dérivations ont été construites à l'aide d'un programme de synthèse de visages (ID ; Tredoux et al., 2006), parmi lesquelles était aléatoirement désigné le visage cible. Un essai présentait donc six visages : le visage cible et cinq dérivations de ce dernier. Une duplication de l'image-cible était également présentée à côté des six, de fait indiquant aux participants le visage à retrouver. Pour chaque essai, soit le nez (n = 20), la bouche (n = 20), ou les deux (n = 20) ont été modifiés pour construire les dérivations. La tâche d'entraînement a été présentée à l'aide du logiciel de conception d'étude E-prime 3.0 (Psychology Software Tools, Pittsburgh, PA). Les tâches de pré-test et post-test ont été présentées à l'aide de l'interface Experiment Center 3.6 (SMI, Teltow, Allemagne). Les mouvements oculaires ont été enregistrés avec l'outil SMI RED250 mobile (SMI, Teltow, Allemagne).

Procédure. L'étude se découpait en trois phases : deux tâches de reconnaissance et la tâche d'entraînement. Les mouvements oculaires étaient uniquement enregistrés lors des deux tâches de reconnaissance. Pour chacune des tâches de reconnaissance, 20 visages étaient présentés pendant trois secondes lors de l'encodage (10 de chaque groupe), avec une croix de fixation (placée au milieu de l'écran) entre chaque visage pour une durée d'une seconde. Après une tâche distractrice de 5 minutes, les 20 visages déjà présentés étaient présentés à nouveau, mélangés à 20 nouveaux visages, constituant ainsi la phase de reconnaissance. Durant la phase d'encodage, les visages étaient présentés de face, alors que dans la phase de reconnaissance, ils étaient présentés de trois-quarts. Entre le pré-test et le post-test, les participants complétaient la tâche d'entraînement. Pour ce faire, il leur était d'abord précisé l'importance de se focaliser sur la moitié inférieure d'un visage pour en améliorer la reconnaissance, notamment car cette partie détient des informations cruciales sur la configuration du visage. Ensuite, les participants complétaient les 60 essais, dans un ordre aléatoire. Il leur était demandé de retrouver, parmi les six images, laquelle était la duplication du visage désigné comme visage-cible. Un feedback correctif leur était présenté suite à leur décision, encadrant en bleu la duplication de la cible, et en rouge le visage incorrectement choisi, le cas échéant. Suite à l'entraînement, les participants complétaient la tâche de reconnaissance post-test. Le calibrage de l'oculomètre était fait avant le pré-test, et avant le post-test uniquement car les mouvements oculaires n'étaient pas enregistrés pendant l'entraînement.

Mesures et résultats. Les visages étaient découpés en deux aires d'intérêt (AOI) : partie inférieure (c'est à dire le bas du visage : nez, bouche, mâchoire, etc.), et partie supérieure (c'est-à-dire le haut du visage : yeux, sourcils, front, etc.).

L'entraînement a eu l'effet escompté sur l'exploration visuelle des participants : après l'entraînement (par rapport à avant), les participants passaient plus de temps à regarder la partie supérieure des visages après, plutôt qu'avant, l'entraînement (M_{aprs} = 2.95, SD_{aprs} = .21 ; M_{avnt} = 2.29, SD_{avnt} = .27; β = -.93, $t(182) = -13.16$, $p < .001$, $d = 2.75$). Notons qu'après (par rapport à avant), les participants dirigeaient plus et plus vite leur première fixation vers la partie inférieure du visage que vers la partie supérieure. Ces résultats confirment le changement de stratégie visuelle mise en place par les participants suite à l'entraînement. L'OGB a été observé en pré-test, avec une meilleure performance de reconnaissance pour les visages de l'endogroupe (M = .75, SD = .14) plutôt que pour ceux de l'exogroupe (M = .66, SD = .15; β = -.09, $t(93)$ = -2.29, $p < .025$, d = .62). Toutefois, l'OGB reste présent suite à l'entraînement (M_{endo} = .77, SD_{endo} = .13 ; M_{exo} = .53, SD_{exo} = .17 ; β = -.24, $t(93)$ = -6.39, $p < .001$, d = 1.60). Alors que la performance n'est pas significativement différente entre le pré-test et le post-test entre pour les visages de l'endogroupe (β = -.03, $t(93)$ = .67, p = .505, d = .74), la performance pour les visages de l'exogroupe a diminué de façon significative suite à l'entraînement (β = .13, $t(93)$ = 3.43, $p < .001$, d = .81). Cette différence concerne principalement le taux de fausses reconnaissances, significativement plus élevé pour les visages de l'exogroupe plutôt que pour ceux de l'endogroupe, avant l'entraînement.

Discussion et conclusion. Dans cette étude, j'ai confirmé qu'entraîner les participants à se focaliser sur une partie du visage (ici, la partie inférieure) résulte en une plus grande focalisation sur cette partie des visages. Toutefois, l'entraînement a également pour conséquence une augmentation de l'OGB, contrairement à ce qui était attendu. Ce dernier résultat est contraire à ce qui a été trouvé par ailleurs (Hills & Lewis, 2006). Les raisons pouvant expliquer cette différence de résultats ainsi que l'augmentation de l'OGB sont les suivantes. Premièrement, il est possible que les participants aient été conscients du fait que leur performance était meilleure dans la tâche d'entraînement pour les visages de l'exogroupe plutôt que ceux de l'endogroupe[8]. En effet, de par le feedback utilisé lors de l'entraînement, les participants ont eu un retour direct sur leur performance, meilleure pour les visages de l'exogroupe. De fait, ils ont pu présenter une plus grande motivation à utiliser la focalisation sur la partie inférieure du visage pour les stimuli de l'exogroupe. Toutefois, réduisant le processus parcellaire d'ores

---

[8] (M_{exo} = .75, SD_{exo} = .16; M_{endo} = .36, SD_{endo} = .11)

et déjà restreint (puisque se focalisant sur les traits du visages de façon indépendante les uns des autres), la reconnaissance s'en est trouvée impactée.

Étude 2 : Un entraînement longitudinal n'élimine pas le biais inter-ethnique chez des participants français typés d'origine européenne

Dans cette étude, un entraînement longitudinal a été créé dans le but d'amener les participants à se focaliser sur différentes parties des visages. Cette étude, menée durant cinq semaines, comprenait trois sessions d'entraînement, chacune ayant un but différent, modélisé par les consignes. Les consignes utilisées lors des séances d'entraînement se sont fondées sur des études ayant montré que la reconnaissance est meilleure lorsqu'un observateur (1) se focalise sur le bas du visage plutôt que le haut du visage pour les visages de l'exogroupe (typés d'origine africaine ; Hills & Lewis, 2006, 2011; Hills et al., 2013), (2) se focalise sur l'intérieur plutôt que l'extérieur des visages, peu importe leur groupe (Fletcher et al., 2008; Kemp et al., 2016; Paterson et al., 2017), ou encore (3) utilise un processus configural plutôt que parcellaire pour traiter les visages (Michel, Rossion, et al., 2006; Sadozai et al., 2018; Tanaka & Simonyi, 2016).

Hypothèses. Dans cette étude, il était attendu de retrouver (1) un OGB avant l'entraînement ; (2) une réduction, voire une élimination de l'OGB suite à l'entraînement ; (3) de meilleures performances des participants lorsqu'ils se focalisent : (i) sur le bas plutôt que le haut des visages ; (ii) sur l'intérieur plutôt que l'extérieur des visages ; (iii) de façon configurale plutôt que parcellaire.

Echantillon. L'échantillon final comprend 11 femmes typées d'origine européenne (M_{age} = 20.45, SD_{age} = 0.52), recrutées lors d'un cours de psychologie sociale appliquée à l'Université de Toulouse Jean Jaurès.

Design. Cette étude suit un plan expérimental factoriel avec une variable intra-sujet : groupe des stimuli (endogroupe ; exogroupe).

Matériel. Les photographies d'un total de 200 hommes différents (100 de chaque groupe) ont été utilisées. Les vêtements étaient visibles, le fond n'était pas standardisé, et le modèle photographié de face. Durant chacune des sessions, 40 identités différentes étaient utilisées : 20 cibles (présentées en encodage, et en reconnaissance), et 20 distracteurs (présentés en reconnaissance). La présentation des visages se faisait en groupe, à l'aide d'une présentation Microsoft Office Power Point. Des feuilles de réponses individuelles étaient distribuées aux participants : une contenant des contours de visages afin que les participants puissent entourer/mettre une croix sur l'endroit des visages qu'ils regardent pendant la phase d'encodage, une deuxième contenant une grille de réponse (oui/non) pour la phase de reconnaissance, et une troisième lors des entraînements contenant les consignes spécifiques et un exemple illustratif.

Procédure. Chaque session était administrée en début de cours, pour un total de 10 à 25 minutes, et était principalement constituée d'une tâche de mémoire des visages : encodage de 20 visages présentés automatiquement (10 secs.), et reconnaissance de ces 20 visages parmi 40, également présentés automatiquement. Lors de la phase d'encodage des trois sessions d'entraînement, les participants devaient se focaliser sur certaines parties spécifiques des visages lors de l'encodage (haut versus bas en session 2 ; intérieur versus extérieur en session 3 ; de façon configurale versus parcellaire en session 4). Afin de vérifier le respect de la consigne, il leur était demandé, en utilisant la feuille avec les contours de visage, d'entourer, ou de mettre une croix, sur la partie du visage qu'ils jugeaient la plus importante à encoder pour pouvoir mieux reconnaître chacun des visages, en fonction de leur consigne de focalisation (ex. un participant dans la consigne 'bas du visage' devrait alors n'avoir que des cercles ou des croix dans les parties inférieures des contours de visages). Suite à une tâche distractrice de 5 minutes (anagrammes, mots-mêlées, sept différences, etc.), la phase de reconnaissance était proposée. Il était alors demandé aux participants de répondre, en entourant 'oui' ou 'non' sur leur feuille de réponse, à la question « Avez-vous vu cette personne pendant la première phase ? » et ce, pour chacun des visages.

Résultats. Conformément à nos attentes, un OBG a été observé avant l'entraînement. En effet, les participants ont significativement mieux reconnu les visages de l'endogroupe (*M* = .93, *SD* = .06), que ceux de l'exogroupe (*M* = .80, *SD* = .161 ; β = -.131, $t(109) = -3.09$, $p = .003$, $d = 1.18$). Cet OGB a également été retrouvé dans les autres sessions, sauf dans la session 3. Suite à l'entraînement, les participants présentaient toujours un OGB ($M_{endo} = .88$, $SD_{endo} = .11$; $M_{exo} = .77$, $SD_{exo} = 16$; $\beta = -.110$, $t(109) = -2.59$, $p = .011$, $d = .81$), résultant en une absence d'effet de l'entraînement sur l'OGB.

Concernant les différences entre consignes, la taille de l'échantillon ($N = 11$) ne permetant que des analyses limitées, aucune conclusion sur une meilleure performance après l'une ou l'autre des consignes d'entraînement n'a pû être tirée.

Discussion et conclusion. La présente étude a permis, une fois de plus, de confirmer l'existence de l'OGB chez des participants français typés d'origine européenne. Toutefois, l'entraînement administré n'a pas permis d'éliminer ce biais. La taille de l'échantillon ainsi que l'absence d'un groupe 'contrôle' (un problème dans les collectes de données a eu pour conséquence une impossibilité d'inclure le groupe 'contrôle') peuvent avoir limité les interprétations. Toutefois, les résultats ayant été répliqués dans l'Étude 4, leurs implications seront discutées dans la discussion de cette dernière étude (p.308).

Étude 3 : Mesurer l'OGB à l'aide de différentes séries de photographies

Dans les deux premières études, les photographies utilisées étaient organisées en deux séries, ensuite administrées soit en pré-test, soit en post-test. Leur présentation n'étant pas contrebalancée, j'ai voulu vérifier que les séries ne variaient pas en terme de 'difficulté' afin d'éviter un biais entre les deux. Une étude en ligne, menée en France, a permis de tester l'effet de la série sur l'OGB.

Hypothèses. Dans cette étude, il était attendu que les participants confrontés aux deux séries présentent un OGB.

Echantillon. L'échantillon final contient 88 participants pour la série 1 (72 femmes, M_{age} = 22.19, SD_{age} = 5.77) et 93 participants pour la série 2 (82 femmes, M = 22.58, SD = 7.26). Tous les participants inclus sont français typés d'origine européenne.

Design. Cette étude suit un plan expérimental factoriel mixte, avec la série (Série 1 ; Série 2) en variable inter-sujets et le groupe des stimuli (endogroupe ; exogroupe) en variable intra-sujets.

Matériel. Les photographies de 80 hommes différents ont été utilisées dans cette étude : 20 cibles, et 20 distracteurs et ce, dans chacune des deux séries. Les cibles étaient présentées dans la phase d'encodage, et dans la phase de reconnaissance. Les photographies étaient prises de face, le modèle avait une expression neutre, les vêtements étaient visibles, et le fond n'était pas uniformisé. En somme, les photographies étaient identiques à celles utilisées dans l'Étude 2. La complétion de la tâche se faisait en ligne, via la plateforme Qualtrics (Qualtric, Provo, UT).

Procédure. La procédure est un paradigme de reconnaissance standard. Les participants devaient regarder et être attentifs aux visages allant leur être présentés. Dans la phase d'encodage, 20 photographies étaient présentées pendant 3 secondes chacune, avec un intervalle d'une seconde entre deux photographies. Suite à une tâche de complétion d'anagrammes de 5 minutes, les 20 photographies étaient présentées à nouveaux, mélangées aux 20 photographies des distracteurs. Le temps n'était pas limité, et pour chacune des photographies, les participants devaient répondre à la question « Avez-vous vu cette personne pendant la première phase ? », en cliquant sur 'oui' ou 'non'.

Résultats. Cette étude a permis de mettre en avant la présence d'un OGB dans chacun des deux séries : série 1 (M_{endo} = .91, SD_{endo} = .07 ; M_{exo} = .86, SD_{exo} = .09 ; β = .05, $t(89)$ = -5.27, $p < .001$, d = .58) et série 2 (M_{endo} = .92, SD_{endo} = .07 ; M_{exo} = .85, SD_{exo} = .10 ; β = .07, $t(89)$ = -6.76, $p < .001$, d = .82), ainsi qu'une absence de différence significative entre les deux sets, suggérant ainsi qu'ils permettent tous deux de mesurer l'OGB.

Discussion et conclusion. Cette étude a permis de mettre en avant que les deux séries de photographies utilisés dans les deux études précédentes permettaient toutes deux de détecter un OGB. De fait, l'absence d'effet d'entraînement dans les deux études précédentes n'est pas dû à une difficulté inhérente aux séries. Dans les études suivantes, j'ai ajouté un contrebalancement entre les séries utilisées en pré-test et en post-test, afin d'éviter tout éventuel biais.

Étude 4 : Un entraînement longitudinal n'élimine pas le biais inter-ethnique chez des participants sud-africains

Considérant les limites évoquées dans l'Étude 2, la présente étude est une réadaptation de l'étude d'entraînement longitudinal, avec une population différente. Les groupes inclus dans cette étude sont des personnes 'Black', ''White' et 'Coloured' d'Afrique du Sud.

Hypothèses. Dans cette étude, il était attendu que (1) les participants 'Black' et 'White' présentent un OGB avant l'entraînement ; (2) les participants 'Coloured' présentent une performance similaire pour les deux groupes de stimuli ; (3) les participants entrainés aient une meilleure performance que les participants non-entraînés ; (4) les participants aient une meilleure performance après l'entraînement (par rapport à avant), résultant en une réduction, voire en une élimination, de l'OGB.

Echantillon. L'échantillon final inclut 51 participants (41 femmes, 20 'Black', 17 'White', 14 'Coloured' ; M_{age} = 20.63, SD_{age} = 3.24), recrutés à l'Université de Cape Town.

Design. Cette étude suit un plan expérimental factoriel mixte, avec l'entraînement (avec ; sans) en variable inter-sujets, le groupe des stimuli ('Black' ; 'White') et temps de mesure (pré-test ; post-test) en variables intra-sujet.

Matériel. Les photographies de 200 hommes ont été utilisées dans cette étude. Les 80 photographies utilisées pour les tâches de pré-test et de post-test sont identiques à celles utilisées dans l'Étude 2. Pour chacune des identités utilisées, la photographie correspondante et prise de trois-quarts a également été utilisée. Les vêtements ont été masqués, et le fond uniformisé. Lors des sessions d'entraînement, les photographies ont subi les mêmes modifications (vêtements masqués, et fond uniformisé), mais étaient également proposées dans une version ayant subi une rotation verticale, et présentées en nuance de gris (en remplacement de la présentation de trois-quarts). L'étude a été administrée individuellement avec la plateforme Qualtrics (Qualtrics, Provo, UT).

Procédure. La procédure était identique à cette utilisée pour l'Étude 2, avec quatre différences : (1) les passations étaient effectuées sur ordinateur et de façon individuelles, permettant ainsi une aléatoirisation à la fois des séries, mais également des photographies au sein de chacune des séries ; (2) un groupe contrôle complétait la même tâche que celle de pré-test et post-test pendant les trois sessions pendant lesquelles les autres participants s'entraînaient, impliquant ainsi les mêmes stimuli, mais sans les consignes ; (3) pendant les sessions d'entraînement, le temps de présentation des stimuli était géré manuellement (plutôt que pendant 10 secondes) afin de laisser le temps aux participants d'encercler/mettre une croix sur les contours des visages, sans que le temps soit perçu comme étant une contrainte ; (4) toutes les consignes et le recueil des réponses étaient présentés et enregistrés via Qualtrics. Les participants choisissaient leurs créneaux, et se présentaient le même jour à la même heure toutes les semaines, pendant cinq semaines. En cas d'empêchement, il leur était possible de décaler leur participation de plus ou moins 24h par rapport au rendez-vous initial.

Résultats. Dans cette étude, un OGB a été observé chez les participants ‘White’, avec une meilleure reconnaissance pour les stimuli de leur endogroupe (M = .84, SD = .10) plutôt que de leur exogroupe (M = .68, SD = .15 ; β = -.16, t(54) = -3.51, p < .001, d = 1.28), mais pas pour les participants ‘Black’ (M_{bk} = .74; SD_{bk} = .15; M_{wh} = .77; SD_{wh} = .13; β = -.04, t(54) = -.92, p = .360, d = .21). Les participants ‘Coloured’ n’ont également présenté aucune différence significative entre les deux groupes de stimuli (M_{bk} = .73; SD_{bk} = .16; M_{wh} = .77; SD_{wh} = .13; β = -.04, t(54) = -.82, p = .42, d = .26). Suite à l’entraînement, l’OGB reste présent pour les participants ‘White’ ayant suivi l’entraînement (M_{wh} = .77, SD_{wh} = .30; M_{bk} = .67, SD_{bk} = .14; β = -.150, t(55) = -2.607, p = .012, d = .45), ainsi que ceux n’ayant pas été entraînés (M_{wh} = .80, SD_{wh} = .09; M_{bk} = .66, SD_{bk} = .14; β = -.114, t(55) = -2.18, p = .034, d = 1.13). Globalement, l’entraînement n’a donc pas permis d’éliminer l’OGB observé en pré-test chez ces participants. Il n’y a également pas de différences significatives de performance entre les différents types d’entraînement.

Discussion et conclusion. Cette étude a permis de répliquer l’Étude 2, en répondant aux limites soulevées dans la discussion de cette dernière, et en l’administrant à des participants différents, dans un pays différent. Il est tout d’abord intéressant de voir que l’OGB a été trouvé pour les participants ‘White’, mais pas pour les participants des deux autres groupes. L’absence d’OGB chez les participants ‘Black’ avait déjà été démontrée dans d’autres études menées en Afrique du Sud (Chiroro et al., 2008; Wright et al., 2001). Toutefois, des recherches menées à l’Université de Cape Town ont trouvé des résultats hétérogènes, observant parfois un OGB chez ce groupe (Derbyshire, 2018), parfois non (Seutloali, 2014; Wright et al., 2003). Suite à l’entraînement, le biais est toujours présent, suggérant une absence d’effet d'entraînement sur l'OGB. Cette observation peut avoir été affectée par un manque de puissance statistique compte tenu des tailles des sous-échantillons, résultat d'un très fort taux d'exclusion des participants lors de cette étude. En effet, seuls 50% de l'échantillon initial a complété les cinq sessions, pré-requis pourtant, pour analyser les résultats.

Les résultats sont donc différents de ceux attendus et de ceux rapportés dans la littérature. D'autres études ayant utilisé un entraînement longitudinal (Goldstein & Chance, 1985; Lebrecht et al., 2009; Matthews & Mondloch, 2018; McGugin et al., 2011; Stahl, 2010; Tanaka & Pierce, 2009), impliquaient entre trois et sept sessions sur des jours consécutifs ou dans un total de deux semaines, alors même que les études 2 et 4 nécessitaient un délai de six à sept jours sur un total de cinq semaines. De plus, dans ces études les participants étaient amenés à compléter la même tâche durant toutes les sessions, alors que dans les études 2 et 4, il leur était demandé d'effectuer une tâche certes similaire, mais en suivant des consignes différentes lors de chacune des sessions. Ces deux différences peuvent avoir eu des effets délétères sur l'apprentissage en réduisant, par exemple, la consolidation. De plus, l'absence de feedback peut également avoir eu un rôle à jouer dans ces résultats, les feedbacks étant importants pour les processus d'apprentissage par entraînement (Estudillo & Bindemann, 2014; Hussain et al., 2009; White et al., 2013). En effet, n'ayant eu aucun retour sur leur performance suite à l'application des consignes, les participants n'ont pas pu cibler laquelle des stratégies mises en place leur ayant éventuellement permis d'avoir une meilleure performance de reconnaissance. Finalement, le contenu même de l'entraînement peut ne pas avoir été optimal pour le but recherché. Alors qu'il fait sens, à la lumière des théories avancées dans la littérature, d'amener les participants à mobiliser un traitement configural (plutôt que parcellaire) ou à se focaliser sur l'intérieur (plutôt que l'extérieur) des visages indépendamment du groupe, il n'en est pas de même pour la troisième consigne. En effet, se focaliser sur la partie inférieure, ou supérieure, des visages est dépendant de la nature du visage, à savoir s'il appartient à l'endogroupe ou à l'exogroupe (Ellis et al., 1975; Hills et al., 2013; Shepherd & Deregowski, 1981). Ainsi, cette étude n'a, une fois encore, pas permis de mettre en avant un effet positif de l'entraînement sur l'OGB.

Étude 5 : Entraîner des participants sur une tâche visuelle spécifique élimine l'OGB[9]

Compte-tenu du fait que des tâches d'appariement sont souvent effectuées par des officiers de police ou de contrôle aux frontières, et que les images utilisées peuvent parfois être de faible qualité (ex. vidéo-surveillance), il est intéressant de proposer et d'évaluer un entraînement permettant d'améliorer la performance lors de ces tâches, car elle est généralement relativement faible (Fysh & Bindemann, 2017) en particulier dans des situations intergroupes (Kokje et al., 2018; Megreya et al., 2011). La présente étude s'est appuyée sur le travail de Bindemann et al. (2013) qui ont conduit des expériences faisant varier à la fois la qualité (en modifiant la résolution des images) et la taille de leurs stimuli. Ils ont constaté que la performance est fortement influencée par la résolution de l'image, avec une meilleure performance pour les paires constituées de deux photographies originales plutôt que d'une originale et d'une pixélisée (étude 1). Ils ont également observé que la réduction de la taille de l'image pouvait partiellement inverser l'effet délétère de la pixellisation (étude 2), en particulier lorsque la taille de l'image pixellisée, par rapport à l'image originale, était réduite (étude 3). Cependant, la pixellisation et la pose (frontale par rapport au profil) ont toutes deux des effets principaux délétères, mais

---

[9] Cette étude a été pré-enregistrée sur la plateforme Open Science Framework : `osf.io/kg8hd`

aucun effet d'interaction (étude 4).

Étant donné que l'exactitude diminue en fonction de la détérioration de la résolution, entraîner des participants en utilisant des images de différents niveaux de pixellisation les amèneraient à avoir une meilleure compréhension de l'extraction des traits du visage à partir d'une image pixélisée et donc une meilleure performance d'appariement lors d'une tâche de généralisation. Cette tâche a alors été transposée à une situation intergroupe dans cette étude.

Hypothèses. Dans cette étude, il était attendu que (1) les participants 'Black' et 'White' présentent un OGB avant l'entraînement ; (2) les participants 'Coloured' présentent une performance similaire pour les deux groupes de stimuli ; (3) les participants entraînés aient une meilleure performance que les participants non entraînés ; (4) les participants aient une meilleure performance suite à l'entraînement, relativement à avant, résultant en une réduction, voire élimination, de l'OGB.

Echantillon. L'échantillon final comprend 140 participants (108 femmes, 40 'Black', 69 'White', 31 'Coloured' ; $M_{age} = 19.30$, $SD_{age} = 1.49$). Les participants ont été recrutés à l'Université de Cape Town.

Design. Cette étude suit un plan expérimental factoriel mixte, avec l'entraînement (avec ; sans) en variable inter-sujets, le groupe des stimuli ('Black' ; 'White') et le moment de la mesure (pré-test ; post-test) en variables intra-sujet.

Matériel. Les photographies de 308 hommes différents ont été utilisées. Toutes les images étaient présentées de face, en nuances de gris. Chacune des identités étaient présentées en deux versions lors des tâches pré-test et post-test : version originale et version pixélisée ; et en trois versions lors de la tâche d'entraînement : version originale, version pixélisée intermédiaire, version hautement pixélisée. Les paires ont été créées sur la base du partage de caractéristiques faciales similaires (âge perçu, couleurs des cheveux, teint de la peau, etc.). Le niveau de difficulté des paires présentées lors de la tâche d'entraînement augmentait au fur et à mesure du temps. Plus il y avait de caractéristiques communes aux paires, plus la difficulté était élevée. Toutes les tâches ont été construites et présentées avec le logiciel E-Prime 3.0 (Psychology Software Tools, Pittsburg, PA).

Procédure. La présente étude se découpait en trois phases : deux tâches d'appariement comme mesures pré-test et post-test, et la tâche d'entraînement entre les deux mesures. Lors des tâches de pré-test et de post-test, les participants devaient déterminer si les deux images représentaient la même personne. Dans chacune de ces deux tâches, un total de 100 paires (50 présentant la même personne, 50 présentant deux personnes différentes) étaient présentées. Dans chacune des paires, l'une était présentée dans sa version originale, et l'autre dans sa version pixélisée. La tâche d'entraînement consistait à apparier les photographies de façon progressive, en deux étapes. La première étape consistait à apparier une identité présentée dans sa version originale à sa version pixélisée intermédiaire parmi six images. Une fois trouvée, les participants devaient ensuite identifier la version hautement pixélisée à partir de la version pixélisée intermédiaire, toujours parmi six possibilités. Des feedbacks correctifs étaient donnés aux participants après chaque choix, leurs précisant quel était le choix correct en cas de réponse incorrecte. Un groupe contrôle (groupe n'ayant pas suivi d'entraînement) jouait pendant 5 minutes à un jeu dans lequel ils devaient faire courir un avatar sur une piste d'athlétisme. Suite à cette tâche, ils devaient répondre à des questions sur le jeu. La tâche était d'une durée approximativement similaire à la tâche d'entraînement.

Résultats. Dans cette étude, un OGB a encore une fois été observé pour les participants 'White', présentant une meilleure performance pour les stimuli de leur endogroupe (M = .95, SD = .04) plutôt que ceux de leur exogroupe (M = .92, SD = .05; β = -.01, $t(136)$ = -2.61, p = .010, d = .67). Une fois encore, il n'y a de différences ni pour les participants 'Black' (M_{bk} = .93, SD_{bk} = .05; M_{wh} = .94, SD_{wh} = .05; β = -.01, $t(136)$ = -.69, p = .492, d = .20), ni pour les participants 'Coloured' (M_{bk} = .94, SD_{bk} = .04; M_{wh} = .93, SD_{wh} = .04; β = .00, $t(136)$ = .27, p = .791, d = .50).

Suite à l'entraînement, uniquement les participants 'White' n'ayant pas été entraînés présentaient encore un OGB (M_{wh} = .93; SD_{wh} = .05; M_{bk} = .91, SD_{bk} = .07; β = -.02, $t(139)$ = -2.68, p = .008, d = .33), alors que ceux ayant suivis un entraînement présentaient une performance non significativement différente pour les deux groupes de stimuli (M_{wh} = .95, SD_{wh} = .04; M_{bk} = .94, SD_{bk} = .04; β = -.00, $t(139)$ = -.38, p = .705, d = .25). Ces résultats suggèrent que l'entraînement a éliminé l'OGB chez ces participants. La performance d'appariement chez ce même groupe ('White') était globalement plus élevée chez les participants entraînés que chez les participants non entraînés, et ce, significativement plus pour les visages de l'exogroupe, résultant en une performance significativement meilleure dans la reconnaissance des visages de l'exogroupe pour les participants entraînés (M = .95, SD = .04) plutôt que non entraînés (M = .91, SD = .07 ; β = -.03, $t(198)$ = -2.82, p = .005, d = 1.14).

Discussion et conclusion. Dans cette étude, bien que la performance globale ait été relativement élevée, l'entraînement proposé a permis d'éliminer l'OGB présent en pré-test chez les participants 'White', encore une fois étant les seuls à présenter ce biais en pré-test. L'élimination de l'OGB chez ces derniers semble être due à une diminution des fausses reconnaissances pour les visages de l'exogroupe, alors que le taux de reconnaissances correctes est resté stable. L'entraînement utilisé dans cette étude semble avoir permis de développer des compétences spécifiques à la tâche, en améliorant la performance d'appariement lorsque les deux visages présentés ne sont pas de qualité identique. Toutefois, il est possible que ces compétences restent spécifiques à ce type de tâche, n'étant donc potentiellement pas transférable à une tâche d'appariement sans dégradation des images.

Étude 6 : Entraîner des participants à se représenter les visages en trois dimensions n'améliore pas la performance dans une tâche écologique de détection[10]

Partant de l'observation qu'une tâche d'appariement peut également être effectuée sur la base d'images non dégradées, et notamment entre une photographie et un individu en face-à-face, la présente étude visait à proposer une tâche d'entraînement ayant pour but d'améliorer la détection d'un individu au sein d'une foule. Peu de recherches ont étudié ce type de tâche dans un paradigme intergroupe. Toutefois, certaines études ont observé un OGB lors de tâches en milieu écologique, faisant participer des caissiers (Brigham et al., 1982; Platz & Hosch, 1988) ou encore les passants d'un centre commercial (Wright et al., 2001). Toutefois, ces études n'avaient pas pour but de développer une tâche d'entraînement permettant l'amélioration des performances.

---

[10] Cette étude a été pré-enregistrée sur la plateforme Open Science Framework : `osf.io/59jfu`

La tâche d'entraînement développée dans la présente étude avait donc pour objectif d'aider les participants à extraire une représentation en trois dimensions des visages présentés, suivant le principe de rotation mentale. L'entraînement consistait à apparier des photographies d'une même identité vue sous plusieurs angles (de face, de trois-quarts et de profil).

Hypothèses. Dans cette étude, il était attendu que (1) les participants 'Black' et 'White' présentent un OGB avant l'entraînement ; (2) les participants 'Coloured' présentent une performance similaire pour les deux groupes de stimuli ; (3) les participants entraînés aient une meilleure performance lors de la tâche de détection, par rapport aux participants non entraînés ; (4) les participants aient une confiance plus élevée en leur décision concernant la cible de son endogroupe plutôt que celle de son exogroupe (hors participants 'Coloured').

Echantillon. L'échantillon final contient 166 participants (123 femmes, 59 'Black', 69 'White', 38 'Coloured', M_{age} = 19.65, SD_{age} = 2.40). Les participants ont été recrutés à l'Université de Cape Town. Huit assistants de recherche ont également été recrutés afin d'aider lors des différentes tâches, ainsi que deux complices (hommes) dont le rôle étaient de jouer le rôle des 'cibles' à retrouver lors de la tâche de détection en milieu écologique.

Design. Cette étude suit un plan expérimental factoriel mixte, avec l'entraînement (avec ; sans) en variable inter-sujets, le groupe des stimuli ('Black' ; 'White'), le groupe des cibles ('Black' ; 'White') et le moment de la mesure (pré-test ; post-test) en variables intra-sujet. Quatre configurations étaient possibles : les cibles étaient soit les deux présentes, les deux absentes, ou l'une était présente mais l'autre absente (et inversement).

Matériel Les photographies de 208 hommes ont été utilisées lors de cette étude. Pour la tâche de pré-test, 50 des 100 paires utilisées lors de l'Étude 5 ont été sélectionnées aléatoirement. Chacune des paires présentait la même photographie originale (de face, en nuances de gris), ainsi qu'une vue de profil (toujours en nuance de gris), en remplacement de la photographie pixélisée comme dans l'Étude 5. Pour l'entraînement, les mêmes identités que celles utilisées dans l'Étude 5 étaient présentées, toutefois, les deux niveaux de pixellisation ont été remplacés par une vue de trois-quarts, et une vue de profil. Pour la tâche de post-test (tâche écologique), les photographies des deux complices ont été imprimées sur du papier glacé, en format 10x14 cm. Les tâches effectuées en laboratoire ont été construites et présentées sur E-prime 3.0 (Psychology Software Tools, Pittsburg, PA). La tâche écologique était effectuée dans une partie (définie en amont) de la bibliothèque universitaire principale du campus, comprenant un total de 142 espaces individuels de travail. Un questionnaire était à compléter par les complices sur la plateforme Qualtrics (Qualtrics, Provo, UT), et une grille d'observation a également été imprimée, permettant à l'observateur de la tâche écologique de recueillir: la date, l'heure, le niveau d'occupation de l'espace utilisé lors de la tâche (par rapport au nombre total de sièges), la présence/absence des cibles, la décision des participants (identification ou rejet), le temps de décision, et la certitude des décisions.

Procédure. Cette étude se déroulait en trois phases différentes : le pré-test, l'entraînement/ non-entraînement, et le post-test. Durant la tâche de pré-test, étaient présentées aux participants des paires de photographies. Pour chacune des paires, il était demandé aux participants de dire si les deux images de chaque paire étaient des images de la même personne, ou de deux personnes différentes. La tâche d'entraînement suivait le même principe d'appariement par étape que l'Étude 5. L'image originale (de face) était présentée, et les participants avaient pour objectif de retrouver, parmi six images de trois-quarts, celle présentant la même identité que l'image cible. Ensuite, il leur était demandé, à partir de l'image de trois-quarts, de retrouver parmi les six images de profil, l'image correspondante. Les participants du groupe sans entraînement complétaient la même tâche que ceux de l'Étude 5 (jeu et questionnaire). Pour le post-test, il était expliqué aux participants qu'ils allaient compléter une tâche de terrain, à la bibliothèque. Les consignes leur étaient présentées, et les participants devaient les répéter afin de s'assurer de leur bonne compréhension. Une fois sur place, les photographies des deux cibles leur étaient données, et ils disposaient d'un maximum de 10 minutes pour prendre une décision quant à la présence ou l'absence de chacune des deux cibles. Une fois leur décision prise, elles étaient recueillies.

Résultats. Tout comme dans les études précédentes, un OGB a été observé chez les participants 'White', présentant ainsi une meilleure performance pour les stimuli de leur endogroupe (M = .92, SD = .04) que ceux de leur exogroupe (M = .89, SD = .05 ; β = -.03, $t(170)$ = -4.20, p < .001, d = .67). Pour la première fois, un OGB a également été observé chez les participants 'Black', qui ont mis en avant une meilleure performance pour les stimuli de leur endogroupe (M = .91, SD = .05) plutôt que ceux de leur exogroupe (M = .90, SD = .05 ; β = .01, $t(170)$ = 2.57, p = .011, d = .20). Les participants 'Coloured' n'ont présenté, une fois encore, aucune différence significative dans leur performance de reconnaissance pour les membres des deux groupes de stimuli (M_{bk} = .89, SD_{bk} = .06 ; M_{wh} = .91, SD_{wh} = .04 ; β = -.01, $t(170)$ = -1.46, p = .147, d = .40).

Suite à l'entraînement, la tâche de terrain a mis en avant des résultats intéressants. De manière générale, l'exactitude était plutôt faible et, indépendamment du groupe des participants et de celui des cibles, les décisions étaient significativement plus exactes lorsque les cibles étaient présentes (82%) que lorsqu'elles étaient absentes (45%), révélant également un plus grand nombre d'erreur lorsque les cibles étaient absentes plutôt que présentes. Un OGB a également été observé, de manière inattendue, pour les participants 'White' ayant été entraînés (M_{wh} = 100%, M_{bk} = 81% ; z = 2.147, p = .032). Aucun autre OGB n'a été observé en post-test. Concernant la confiance en leur décision, les participants 'Black' et 'White' ont fait part d'une plus grande confiance en leur décision lorsque celle-ci concernait la cible de leur endogroupe (respectivement 89% et 82%) plutôt que la cible de leur exogroupe (respectivement 80% et 72%), alors que les participants 'Coloured' n'ont présenté aucune différence entre les deux groupes. De plus, pour l'ensemble des participants, confiance et exactitude sont modérément corrélés (r = .31, $t(390)$ = 6.48, p < .001, 95%CI [.22, .40]).

Discussion et conclusion. L'objectif de cette étude était double. D'une part, il s'agissait de voir si un OGB serait présent dans une tâche de détection en milieu écologique, tout comme il a été démontré auparavant. D'autre part, il s'agissait de voir comment une tâche d'entraînement fondée sur l'idée de rotation mentale permettrait aux participants de mieux reconnaître un individu dans une tâche écologique, c'est-à-dire reconnaitre un individu en trois dimensions à partir d'une photographie en deux dimensions.

Dans cette étude, un OGB a été observé en pré-test chez les participants 'White', mais également, chez les participants 'Black'. Lors de la tâche écologique (post-test), seuls les participants ‘White’ ayant suivis un entraînement présentaient toujours un OGB. Indépendamment de l’OGB, les participants avaient un taux plus élevé d’identifications erronées lorsque les cibles étaient absentes plutôt que lorsqu’elle étaient présentes durant la tâche. Il est important de noter que les conditions de complétion de cette tâche écologique ont été contrôlées le plus possible : il était demandé aux cibles de varier leur positionnement au sein de l’espace dédié, ainsi que leur activité, afin d’induire le plus de variabilité possible en termes d’exposition de leurs visages. De plus, le taux d’occupation de l’espace dédié (nombre de sièges occupés par rapport au nombre total de sièges) a été relevé, et inclus dans les analyses statistiques afin de prendre en compte les variations inter-participants. Finalement, cette étude n’a pas permis de mettre en avant un effet de l’entraînement. Il conviendrait néanmoins de tester l’effet de ce type d’entraînement sur une tâche similaire à celle utilisée en pré-test, et donc, en laboratoire.

Discussion générale et conclusion

La présente thèse avait pour objectif d’utiliser l’entraînement dans le but d’éliminer l’OGB. Un total de 398 participants en présentiel ainsi que 181 participants en ligne a été nécessaire[11]. Les participations se sont effectuées dans deux pays différents, incluant des participants de groupes différents. Les résultats principaux indiquent d’une part, qu’un OGB a été observé pour les participants français typés d’origine européenne et les participants ‘White’, dans toutes les études et d’autre part qu’un OGB n’a été détecté chez les participants ‘Black’ que dans l’Étude 6. Toutefois, les différentes tâches d’entraînement testées n’ont pas permis d’induire une meilleure performance en post-test, Étude 5 mise à part.

---

[11] Un total de 621 participants en présentiel, et de 230 participants en ligne a été recruté avant l’application des critères d’inclusion/exclusion.

Catégoriser les individus dans les groupes étudiés. Un des challenges de cette thèse concernait les catégorisations utilisées pour décrire les participants et les stimuli, compte-tenu du fait que catégoriser les participants (et les stimuli) au sein de groupes est ce qui permet d'étudier l'OGB. En effet, il est nécessaire de savoir dans quel groupe les participants se situent, afin d'étudier s'ils présentent une meilleure performance de reconnaissance envers les membres de leur groupe plutôt qu'envers les membres d'un autre groupe. Différentes mesures ont été utilisées pour catégoriser les participants, selon les pays. En Afrique du Sud, les termes anglais 'race' ainsi que 'Black', 'White' et 'Coloured' ont été utilisés lorsqu'il était demandé aux participants de s'auto-identifier à un groupe. En France, il a été plus difficile de définir une catégorisation, compte tenu du contexte social, culturel et légal. Une échelle médicale de tolérance au soleil (phototype ; Fitzpatrick, 1988) ainsi que les pays ou continents de naissance des proches (parents, grands-parents) ont été utilisés à défaut. Lors de la collecte des données de l'Étude 4, première des études menées en Afrique du Sud, la catégorisation par phototype plutôt que par groupe ethnique avait été utilisée, avant de se rendre compte qu'elle n'était pas optimale pour cette population. De fait, et puisqu'il reste, pour le moment, encore acceptable dans ce pays d'utiliser des termes raciaux, la méthode de catégorisation a été modifiée et adaptée pour les études 5 et 6. Toutefois, afin d'étudier de manière plus précise les deux systèmes de catégorisation, il a été décidé de recueillir les données des deux méthodes afin de comparer les deux types de catégorisation et d'explorer dans quelle mesure les catégories sont identiques, ou non, compte tenu de la méthode utilisée. Incluant 359 participants sur trois études (271 femmes, M_{age} = 19.72, SD_{age} = 2.17), une forte corrélation a été observée (ρ = .86, $p <$ .001) entre les catégories ethniques auto-rapportées ('Black', 'White', et 'Coloured'), et celles basées sur les phototypes (I, II, III pour 'White' ; IV pour 'Coloured' ; V et VI pour 'Black').

Bien que la corrélation soit forte, il semblerait qu'il y ait des différences importantes à prendre en compte entre les deux méthodes, résultant en des participants classés dans des groupes différents en fonction de ces méthodes. En effet, lorsque l'on s'attarde sur la distribution des trois groupes au seins des six phototypes, on se rend compte qu'il y a des confusions (Figure 6.3). Considérant les participants 'Coloured', alors qu'il était attendu qu'ils s'identifient au phototype IV, ils s'identifient à tous les phototypes (sauf le I) dans des proportions plus ou moins grandes. Pour ne prendre qu'un exemple, le phototype V, sensé correspondre au groupe 'Black', contient 30% de personnes s'identifiant également au groupe 'Coloured'. Utiliser la catégorisation par phototypes en Afrique du Sud pose question. Il reste possible d'utiliser une catégorisation ethnique, bien que celle-ci soit à questionner d'un point de vue sociétal et légal.

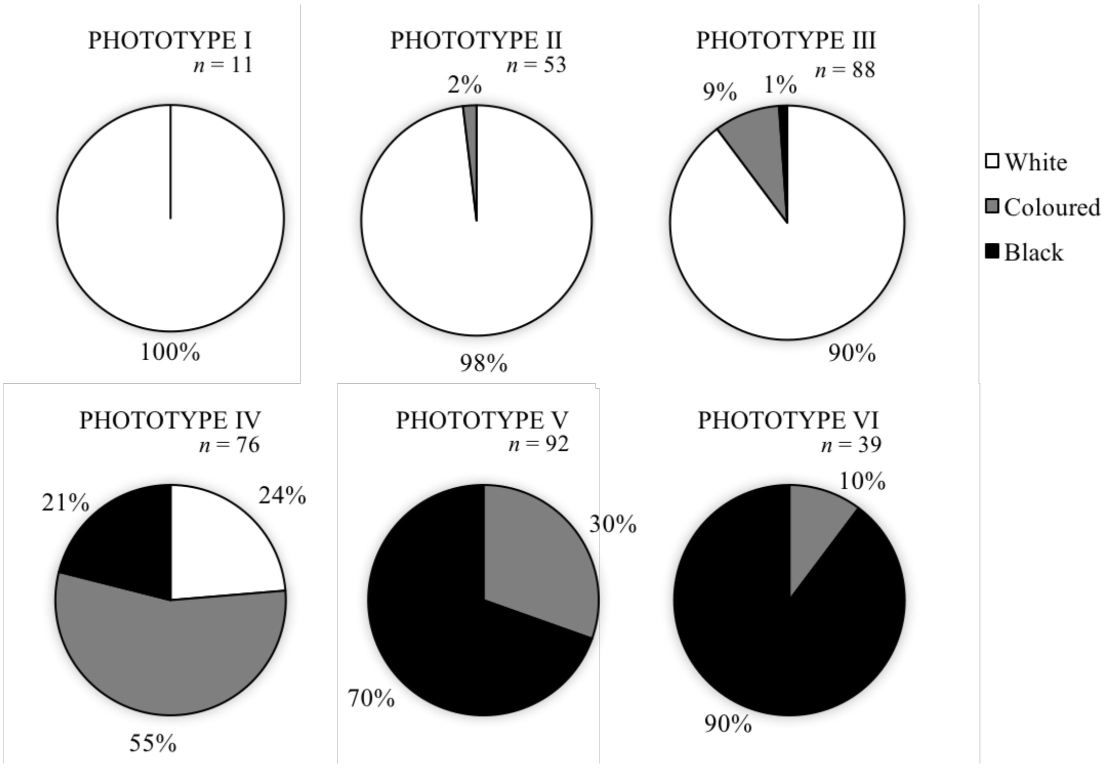

Figure 6.3 – Distribution de l'ethnie des participants (Black; White; Coloured) en fonction des phototypes (I à VI, voir la classification de Fitzpatrick p.32).

Conceptuellement, le phototype et l'ethnie renvoient à des dimensions différentes, de fait, l'absence d'une corrélation parfaite est cohérente, bien que cette indépendance des termes n'ait été questionné que suite à l'étude 4. Alors que le phototype relève d'une description purement physique, l'ethnie relève d'une appartenance à une catégorie sociale et identitaire. Contrairement au phototype pouvant être catégorisé objectivement, l'appartenance à un groupe ethnique est nécessairement auto-rapporté, impliquant une dimension plus identitaire que physique, bien que majoritairement fortement corrélé. Les deux informations peuvent alors être complémentaires, d'autant plus que l'OGB est à la fois une conséquence de variables cognitives (notamment visuelles), mais également sociales (particulièrement le contact). Il serait ainsi pertinent de travailler sur le développement d'une nouvelle méthode de catégorisation.

La présence d'un OGB en pré-test. À travers les études de cette thèse, une observation est apparue comme constante : la présence d'un OGB chez les participants français typés d'origine européenne, et les participants 'White' sud-africains (Figure 6.4). Ces résultats correspondent globalement à ce qui a été observé dans des études précédentes avec l'utilisation des mêmes deux groupes de stimuli, à la fois en France (Brunet, 2017, 2018), mais aussi en Afrique du Sud (Chiroro et al., 2008; Goodman et al., 2007; ?, ?; Wright et al., 2001, 2003). Toutefois, une autre étude menée en 2018 par Derbyshire n'a pas mis en avant d'OGB chez ce groupe de participants.

A l'inverse, cinq des études menées dans cette thèse n'ont pas démontré la présence d'un OGB chez les 'Black' (Figure 6.5), observation consistante avec des résultats récents (Sadozai

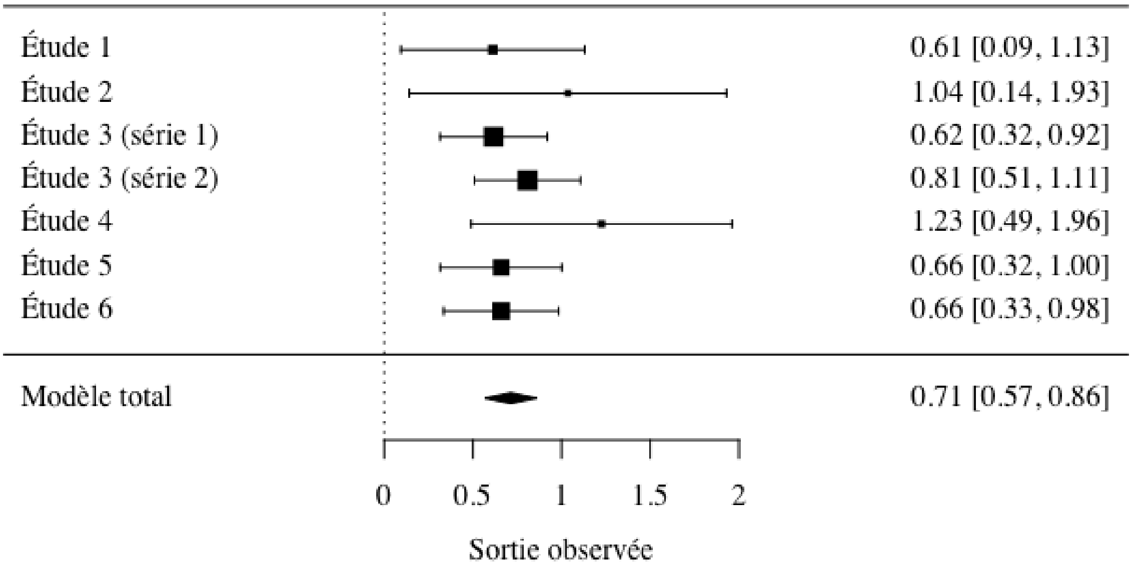

Figure 6.4 – Récapitulatif de la méta-analyse de la présence de l'OGB à travers les études concernant les participants typés d'origine européenne ou 'White'. Ces études révèlent ensemble la présence d'un OGB avec une taille d'effet globale importante (Hedges' g = .71, 95%IC = [.57; .86]). Les études 1 à 3 ont été effectuées en France, les études 4 à 6 ont été effectuées en Afrique du Sud. La taille des échantillons était respectivement de 30, 11, 88, 93, 17, 69 et 77 participants.

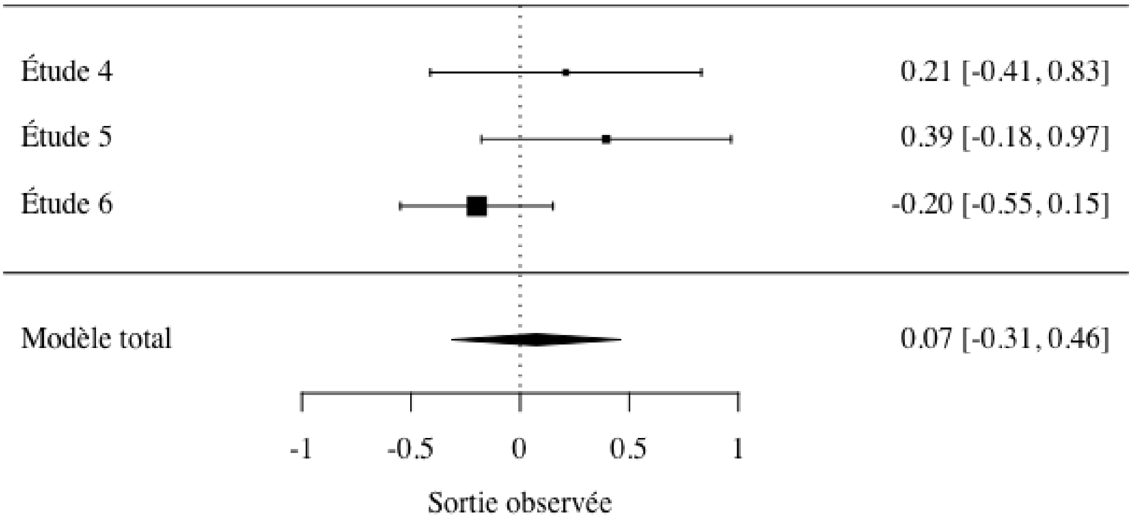

Figure 6.5 – Récapitulatif de la méta-analyse de la présence de l'OGB à travers les études concernant les participants 'Black'. Ces études révèlent ensemble l'absence d'un OGB avec une taille d'effet globale faible (Hedges' g = .07, 95%IC = [-.31; 46]). Les études ont toutes été effectuées en Afrique du Sud. La taille des échantillons était respectivement de 20, 40 et 59 participants.

et al., 2018). Néanmoins, la littérature a également mis en évidence un OGB chez les ‘Black’, tout comme dans l’Étude 6 (Chiroro et al., 2008; Derbyshire, 2018; Wright et al., 2001). Parfois, une meilleure performance de reconnaissance pour les personnes de leur exogroupe plutôt que de leur endogroupe a même été mis en avant (Seutloali, 2014; Wright et al., 2003), tendance également observée dans les études 4 et 5. Au regard du contexte démographique du pays, il était pourtant attendu que les ‘Black’ présentent un OGB, et ce notamment de par leur présence majoritaire dans le pays. Toutefois, et bien qu’ils restent majoritaires à Cape Town, suivis de peu par les ‘Coloured’ et avec une minorité de ‘White’, les étudiants ‘Black’ étaient en plus faible nombre que les étudiants ‘White’ à l’Université de Cape Town et ce, jusqu’en 2018. La sur-représentativité du groupe ‘White’ à l’université par rapport au reste du pays, ainsi que l’héritage historique et notamment l’Apartheid ou encore la question des statuts économiques et du pouvoir (contexte hiérarchique), peuvent être des explications à ces observations concernant l’OGB, induisant ainsi à la fois du contact, mais également et surtout, un plus grand intérêt à individualiser ce groupe (Pettigrew & Tropp, 2006; Shriver & Hugenberg, 2010). Concernant les ‘Coloured’, aucune des études n’a permis de démontrer une quelconque meilleure performance pour l’un ou l’autre des deux groupes de stimuli, tout deux des exogroupes. Pourtant, dans son étude incluant des participants ‘Coloured’, Seutloali (2014) a mis en avant une meilleure performance de reconnaissance de ces participants pour les ‘White’ plutôt que les ‘Black’.

L’absence générale d’effet de l’entraînement. Alors qu’il était attendu d’observer une élimination de l’OGB comme conséquence positive de l’entraînement, seule l’Étude 5 a permis de soutenir cette hypothèse. Toutefois, dans cette étude, les participants étaient entraînés à améliorer leur performance dans une tâche spécifique : l'appariement de visages présentés de manière plus ou moins dégradée (manipulé par le niveau de pixellisation des photographies). Cette étude suggère alors qu'il est possible d'augmenter les performances pour les stimuli des deux groupes (endogroupe et exogroupe), et qu'un entraînement de ce type permet une augmentation des performances pour de nouveaux visages.

De manière générale, les études menées dans cette thèse suggèrent qu'une simple exposition à de nombreux visages n'est pas suffisante pour éliminer l'OGB. Pourtant, considérant le modèle d'individualisation-catégorisation (Hugenberg et al., 2010), il était également supposé que les participants sud-africains bénéficient d'ores et déjà d'une expérience d'individuation préalable suffisante pour être capable d'induire de l'individuation lorsque motivés à le faire. Il est dès lors possible que les tâches développées n'aient pas induit suffisamment de motivation, alors même que ce n'était pas nécessairement leur objectif. L'objectif des tâches d'entraînements développées dans cette thèse consistait à développer des compétences visuelles. En effet, pour développer une meilleure représentation en mémoire des visages d'exogroupes (Valentine, 1991; Valentine et al., 2016), un observateur devrait être exposé à de nombreux visages, et devrait disposer d'un moyen efficace pour discriminer et reconnaître les visages. L'entraînement était considéré comme une réponse à ces besoins. Toutefois, l'absence d'effet d'entraînement peut être expliquée de différentes manières. Il est possible que les tâches d'entraînement développées dans ce travail n'aient pas permis de créer des processus et/ou compétences suffisamment robustes, efficaces ou même pertinentes dans le but d'éliminer l'OGB, et ce, notamment lors d'une tâche de généralisation à de nouveaux visages.

La majorité des études publiées a utilisé un processus d'individualisation permettant ainsi de réduire, voire d'éliminer l'OGB (Elliott et al., 1973; Goldstein & Chance, 1985; Lebrecht et al., 2009; Matthews & Mondloch, 2018; McGugin et al., 2011; Stahl, 2010; Tanaka & Pierce, 2009). Les deux études n'induisant pas d'individualisation ont cherché à éliminer l'OGB avec des méthodes basées sur la modification des processus visuels. Les tâches d'entraînement développées dans cette thèse reposent sur le même principe. En répliquant le principe sous-jacent à l'entraînement de l'une d'elles (Hills & Lewis, 2006), l'opposé de leurs résultats, soit une augmentation de la magnitude de l'OGB, a été observé. Dans une autre étude (Lavrakas et al., 1976), une amélioration de la performance de reconnaissance intergroupe a été observée, lorsque les participants se focalisaient sur la partie interne des visages, suite à un bref entraînement. Cet effet a été observé immédiatement après l'entraînement, mais pas après un délai d'une semaine. Cette dernière observation pourrait expliquer l'absence d'effet de l'entraînement observée dans les études 2 et 4. En effet, le post-test intervenait une semaine après la dernière session d'entraînement, induisant dès lors un certain délai. Il est donc possible que de tels résultats ne soient observables que dans une tâche de généralisation réalisée immédiatement après l'entraînement.

Finalement, il est possible que les tâches d'entraînement développées dans cette thèse n'aient pas été suffisamment difficiles, n'aient pas impliqué suffisamment de stimuli, ou suffisamment de sessions afin de développer des compétences pouvant être généralisables. De plus, la taille des échantillons utilisés est également une limite n'ayant parfois pas pu permettre de conclure de façon plus tranchée sur une absence réelle d'effet de l'entraînement. L'acquisition d'une réelle expertise de discrimination et de reconnaissance pour des membres d’exogroupes nécessiterait alors une plus grande exposition, ainsi que plus de contacts, lors de tâches d’entraînement plus longues, plus intensives et explicites. Toutefois, ce travail de thèse permet d’ores et déjà d’étayer le champs des possibles, et d’apporter des réponses importantes et intéressantes.

Peu d’études ont considéré l’utilité d’un entraînement pour développer des compétences de discrimination et de reconnaissance des visages, et notamment lorsque ceux-ci appartiennent à des exogroupes. Il est donc d’un grand intérêt de continuer les recherches en ce sens, et plus particulièrement compte tenu des conséquences négatives que l’OGB peut avoir dans la vie quotidienne de tout un chacun, et plus particulièrement dans certains milieux professionnels comme les professions judiciaires (police, agents de contrôle au frontières, etc.).