HAL
open science

Utilisation d'une mesure indirecte du stéréotype : apport théorique des modèles multinomiaux

Rémi Courset

To cite this version:

Rémi Courset. Utilisation d'une mesure indirecte du stéréotype : apport théorique des modèles multinomiaux. Psychologie. Université Grenoble Alpes [2020-..], 2021. Français. NNT : 2021GRALS003 . tel-03255199

HAL Id: tel-03255199

https://theses.hal.science/tel-03255199

Submitted on 9 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : PCN - Sciences cognitives, psychologie et neurocognition

Arrêté ministériel : 25 mai 2016

Présentée par

Rémi COURSET

Thèse dirigée par **Dominique MULLER**, Professeur,
Université Grenoble Alpes et
Codirigée par **Marco PERUGINI**, Professeur,
Université de Milan-Bicocca

préparée au sein du **Laboratoire Interuniversitaire de Psychologie / Personnalité, Cognition, Changement Social** (LIP/PC2S)
dans **l'École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement** (EDISCE)

Utilisation d'une mesure indirecte du stéréotype : apport théorique des modèles multinomiaux

Thèse soutenue publiquement le **12 janvier 2021**,

devant le jury composé de :

M. Olivier CORNEILLE
Professeur, Université catholique de Louvain, Examinateur
Mme Jessica MANGE
Maîtresse de conférences – HDR, Université de Caen, Rapportrice
M. Dominique MULLER
Professeur, Université Grenoble Alpes, Directeur de thèse
M. Marco PERUGINI
Professeur, Université de Milan-Bicocca, Directeur de thèse
M. François RIC
Professeur, Université de Bordeaux, Examinateur – Président du jury
Mme Odile ROHMER
Professeure, Université de Strasbourg, Rapportrice

Remerciements

Arrivé au moment tant attendu des remerciements liés à ces travaux de thèse, un doute m'envahit. Tant de personnes ont contribué, de près ou de loin, à faire de ce travail ce qu'il est aujourd'hui. Vais-je oublier quelqu'un ? Vais-je faire des déçus ? Vais-je être à la hauteur de ce que vous avez toutes et tous pu m'apporter ? La tâche est rude, mais je vais essayer de m'y atteler au mieux.

Je voudrais en premier lieu remercier les membres du jury qui ont accepté d'évaluer ce manuscrit. Jessica Mange, Odile Rohmer, merci d'avoir accepté aussi rapidement d'être rapportrices pour ce travail et de m'avoir permis de rédiger mon manuscrit de thèse dans des conditions optimales. Olivier Corneille, François Ric, merci d'avoir accepté d'examiner mon travail, en espérant que celui-ci soit agréable à lire. Il me tarde de continuer nos débats commencés par le passé. À tous les quatre, encore merci. Merci également à l'EDISCE et plus particulièrement Orélie Garzena et Anne Guérin-Dugué de m'avoir permis de réaliser ma thèse dans les meilleures conditions.

Bien sûr, dans ces remerciements je pense d'abord à ma famille la plus proche. Merci pour votre soutien sans faille. Depuis toujours, vous m'avez suivi dans ce que je faisais sans jamais douter de mes capacités. Si je suis devenu la personne que je suis aujourd'hui, c'est en premier lieu grâce à vous. J'ai beaucoup de chance d'avoir pu grandir et évoluer au sein d'un tel environnement familial. Peut-être que je ne le montre pas assez, peut-être que je ne suis pas assez expressif dans toute l'affection que je vous porte, mais sachez que je mesure pleinement la chance que j'ai de vous avoir à mes côtés.

Je voudrais ensuite remercier ceux que je connais (presque) depuis toujours. Bocquet, Erik, Etienne, (Joris), Thomas, sans oublier bien sûr Aude, Marine et maintenant Luka, grandir avec vous et savoir que je peux compter aujourd'hui encore sur chacun d'entre vous est juste parfait. Déjà 25ans d'amitié (plus ou moins), sans aucun doute vous êtes l'une de mes plus grandes fiertés. À ce petit monde, j'ajoute bien sûr Vincent avec qui j'ai partagé énormément de bons moments, de nos années de collège, de lycée, nos différents voyages ou le fait d'avoir eu la chance d'être ton témoin de mariage. On n'a pas besoin de se voir beaucoup pour savoir que l'on peut être présent l'un pour l'autre et il me tarde de revenir vous voir Katja, Nathanael et toi.

En écrivant ces remerciements, je me rends un peu plus compte de la chance que j'ai d'être entouré comme je le suis. Arrivé en L2 Psycho, j'avais peur de me retrouver seul et de ne pas réussir à reconstruire ce que j'avais pu avoir par le passé. Dire que je me trompais relève de l'euphémisme. C'est de vous dont je parle, les Psychopotes : Adrien, Alizée, Céline, Chris, Fifou, Jonath, Laura, Lola, Lucie, Lucile, Marielle et Rèm. On dit souvent que les années fac sont inoubliables, que leurs souvenirs restent parmi les meilleurs d'une vie. Je ne peux qu'être d'accord avec cela. Merci d'avoir été là, que ce soit sur les bancs d'un amphi, le canap d'un de nos apparts ou littéralement collés au Barberousse… Aux remerciements associés à ces années d'études, il me faut aussi penser à tous celles et ceux qui ont partagé ces années de Licence et de Master (ClemClem, je t'oublie pas tu sais) et tout particulièrement à Yoan, mon binôme de M2. C'est bon Yoan, tu peux être rassuré, tu as toute ta place dans mes remerciements.

Parmi toutes ces rencontres et amitiés qui ont su se créer sur les bancs de l'université, il y en a une qui me rend encore plus fier. Une amitié qui partage aujourd'hui ma vie. Lola, comme tu le sais déjà je n'aurais pas pu rêver mieux que de tomber amoureux de ma meilleure amie. Tu m'as supporté avant cette thèse, tu m'as supporté pendant cette thèse et j'espère vraiment que tu me supporteras encore longtemps après. Tu m'as vu souffrir ces derniers mois, mais tu as répondu présent à chaque fois qu'il le fallait. Je te l'ai déjà souvent répété, mais merci d'être exactement la personne que tu es. Bien entendu, tu n'es pas entrée dans ma vie toute seule. Ma belle-famille (première fois que j'emploie ce terme), vous êtes toutes et tous absolument parfaits ! Merci encore pour l'accueil que vous m'avez fait.

Dans ce déroulé au final très chronologique, j'en arrive à toutes les personnes qui ont gravité plus directement autour de ce travail de thèse. Si je me suis senti aussi bien pendant ces trois années,

c'est aussi parce que j'ai eu la chance d'être entouré de doctorants et post-doctorants exceptionnels. À vous celles et ceux ayant transité par le bureau D213, D214, D215 ou les bureaux dédiés aux post-doctorants, merci d'avoir été d'un soutien indéfectible. Que ce soit nos repas souvent animés en salle de pause, nos pauses café au rond ou nos (trop) nombreux apéros au Martin's ou ailleurs, j'ai eu la chance de partager avec vous certains moments particulièrement mémorables. Une petite mention à Soufian qui, du haut de ses huit années de thèse, a toujours été un modèle pour moi, que ce soit dans l'inclusion de chaque personne au sein du LIP ou plus généralement dans le maintien d'une atmosphère joviale à longueur de journée. Promis, je vais continuer à aller courir (un peu). Je penssse que sss'est le bon moment pour mensssionner Lison. Neuroud, tu me respectes pas, on est d'accord là-dessus, mais je te kiffe tellement ! Quand on a commencé ces longues sessions à manipuler ces gobelets, je n'aurais pas pu m'imaginer une seule seconde la place que tu allais t'octroyer parmi mes amis les plus proches. Même si tu n'es plus au labo, que tu aies réussi à devenir psychologue du travail, on est restés super proches et j'espère vraiment que ça va continuer. Je pense aussi à Rick Klein, my boy, j'ai tellement hâte de pouvoir réentendre l'un de tes « oui, je comprends ! » autour d'une petite bière (un galopin..).

J'en arrive à mes deux acolytes que je ne peux dissocier tant ils ont été l'un et l'autre d'une importance capitale tout au long de ces trois années. Cédric, d'abord, qui m'a accompagné depuis le tout début sur les bancs de l'université et que j'ai eu la chance de rejoindre dans cette aventure qu'est le doctorat. Tu sais toute l'estime et l'admiration que je te porte au quotidien et cela n'a fait que se renforcer au cours du temps. Toutes ces conférences passées ensemble, ces moments hors et à l'intérieur du labo, toutes ces heures passées à douter, à se rassurer, à se soutenir, mais aussi à s'embrouiller, parfois, n'ont fait que renforcer notre amitié. Kevin, ensuite. Comment te définir… Je pense que toutes les personnes qui te connaissent seront d'accord avec moi pour dire que tu es un personnage unique. Toujours de bonne humeur, toujours avide de cet humour décalé, tu as fait la joie de l'ensemble du labo (doucement par contre en soirée…). Mais te définir seulement par rapport à tes traits d'humour serait réducteur. T'es beaucoup plus que ça. Tu as su être toujours (et j'insiste sur le « toujours ») être présent quand il le fallait, pour prendre du temps pour moi, me rassurer et remonter un peu l'estime que je pouvais avoir de moi-même. A tous les deux, merci. Au passage, je ne me fais aucun souci pour la relève du bureau D213. Yoann, Olivier, vous avez été une belle découverte et c'est une grande chance pour le labo de vous avoir encore pendant un petit moment (vraiment) !

Dans le désordre, je voudrais aussi remercier un ensemble de personnes qui ont été présentes tout au long de cette thèse. Déjà, l'ensemble des enseignants-chercheurs du LIP/PC2S et plus particulièrement celles et ceux que j'ai pu côtoyer au cours des ateliers de recherches, des réunions d'axes ou plus simplement dans la vie quotidienne du labo. Celine Baeyens, Catherine Bortolon, Arnaud Carré, Florian Delmas, Aurélien Graton, Jean-Louis Monestès, Hans Rocha IJzerman, Annique Smeding et tous les autres, merci pour nos échanges, nos rires, mais aussi vos conseils, vos critiques et surtout, votre présence. On ne vous le dit peut-être pas assez, mais votre proximité et le temps que vous accordez à l'ensemble des doctorants et doctorantes contribuent à faire du LIP ce qu'il est aujourd'hui : un laboratoire où l'on se sent bien. Au sein du laboratoire, je voudrais enfin remercier deux personnes sans qui rien ne pourrait fonctionner. Anne et Nico vous avez été parfaits tout au long de ces années. Vous répondez toujours présent, vous avez toujours résolu sans broncher les différents problèmes que je pouvais avoir et je ne suis certainement pas le premier ni le dernier à le dire, mais merci pour toute l'énergie que vous dépensez dans votre travail. Je n'oublie également pas tous les étudiants et étudiantes qui ont participé à l'ensemble des expérimentations constituant cette thèse. Que ce soit celles et ceux ayant participé aux réunions d'équipe ou plus directement les stagiaires que j'ai eu la chance d'avoir tout au long de ces années, vous avez fait de ce travail ce qu'il est aujourd'hui.

Je sors du LIP pour faire un détour par le LaRAC et remercier Cécile Nurra avec qui j'ai passé d'excellents (mais trop courts) moments tout au long de ces années et qui a su trouver les mots quand cela était nécessaire. Je n'oublie pas ici Margot, certainement le futur un peu plus lointain de la recherche qui va, j'en suis sûr, faire des ravages partout où elle va passer ! Natacha Boissicat et Camille

Sanrey, merci également à vous pour ces moments que l'on a pu passer ensemble. Après le LIP et le LaRAC, il me reste le LPNC. Pour moi, les doctorants du LIP ne vont pas sans les doctorants du LPNC. J'ai eu la chance de faire ma thèse à un moment où la frontière entre nos deux labos n'existait pour nous tout simplement pas. Le meilleur exemple ? Les Furies. Oui, les Furies me manquent… Mention spéciale à un ancien doctorant et une ancienne doctorante du LPNC, Amélie (Taïga) et Brice Beffara-Bret ! Avec vous, on ira, on saura, sauver notre existence, pour refaire un monde sans dangers !!

Je sors enfin de mon expérience Grenobloise (en fait, pas vraiment) pour remercier tous les anciens et anciennes du LIP/PC2S. Jennifer Bastart, Marie-Pierre Fayant, Anthony Lantian, Marine Rougier, Elisa Sarda, Oulmann Zerhouni et d'autres, vous avez été des exemples à suivre pour nous permettre nous, les petits nouveaux, d'avoir l'opportunité un jour de pouvoir entreprendre ce que vous avez réussi. Par le prisme de chacun d'entre vous, on peut voir toute l'importance que peut avoir le LIP dans nos vies et à quel point celui-ci est une famille dont on ne sort au final jamais réellement. Ensuite, comment ne pas remercier celui avec qui j'ai partagé tant de Skype, celui qui m'aura permis de parler de foot durant ma thèse (merci). Luc, je suis désolé, tes 43 clubs n'arrivent pas à la cheville du meilleur club de France. Allez l'OM ! Enfin, un remerciement certainement trop court à tous celles et ceux avec qui j'ai pu partager de merveilleux moments à Milan ou Davis lors de mes deux in-doc, ainsi qu'à tous celles et ceux dont j'ai croisé la route lors des différentes conférences qui ont émaillé ces trois années. J'ai hâte de pouvoir revenir vous embêter un peu ! I would especially like to thank Jeff Sherman, for allowing me to spend some time with his team at UCDavis, as well as Christoph Klauer for the many exchanges we had and for allowing me to present my work at the University of Freiburg.

Bien sûr, je vais finir mes remerciements par mes deux directeurs de thèse Marco Perugini et Dominique Muller. First, for you Marco, I am going to switch to English. I am sorry Marco, you would certainly prefer that I speak Italian, but I am definitely not born to speak Italian (despite my five years of learning it behind me). Before my thesis, we did not know each other. And yet you agreed to be my supervisor and to believe in me and my capacities. With you, I have always felt supported. Not on the front line, not present on a daily basis, but present every time it was needed. Our many exchanges have all been really instructive and I hope I have followed some of the many advices you have been able to give me. Thank you again for welcoming me to Milan, for taking time for me, and for being more optimistic about my work than I will ever be! Enfin, ces dernières lignes ne suffiront certainement pas à retranscrire toute la gratitude que je peux avoir envers Dominique Muller. Dominique, je ne vais pas être très original dans mes remerciements. Je suis sûr que si l'on regardait les remerciements de tes anciens doctorants, on pourrait retrouver exactement les mêmes thèmes. Mais c'est diagnostique de quelque chose. Cela montre à quel point tu es investi pour tes doctorants et à quels points ceux-ci sont conscients de la chance qu'ils ont eue de faire leur thèse avec toi. Comme tu le sais, je ne voulais pas faire une thèse à n'importe quel prix. Je voulais uniquement faire une thèse avec toi et je ne m'étais pas trompé : je n'aurais pas pu rêver meilleur directeur. Dans ma formation professionnelle, déjà, tu m'as apporté plus que je ne l'aurais espéré, toujours pédagogue pour m'expliquer certains concepts clés dont tu savais qu'ils pourraient m'être utiles dans mon futur professionnel. Tu as été patient, tu as toujours trouvé du temps pour réfléchir à mes recherches ou m'aider dans mes enseignements. Mais si tu es autant apprécié de tes doctorants, et même je dirais de l'ensemble des doctorants du labo, c'est aussi et surtout pour ton humanité. Je pense qu'avoir un directeur capable d'une telle empathie est quelque chose de particulièrement rare. Définitivement, tu es un modèle pour moi et pour l'ensemble des doctorants. Quelqu'un qui s'investit avant tout pour les autres. Et je ne t'en remercierais jamais assez. (PS : Désolé pour les virgules…)

Ces remerciements sont au final certainement trop longs. Mais ils sont avant tout le reflet de la chance que j'ai d'être entouré par autant de personnes formidables. Ce mot « chance », je l'ai employé plus d'une dizaine de fois et ce n'est pas un hasard. Je mesure pleinement la chance que j'ai de vous avoir dans ma vie. J'aurais aimé prendre plus de temps pour vous le redire, mais une dernière fois : merci.

Résumé

Les mesures indirectes occupent une place centrale dans l'étude des stéréotypes et préjugés individuels. Pourtant, ces mesures sont aujourd'hui critiquées : contrairement à ce qui était avancé, celles-ci seraient le reflet d'une multitude de processus couvrant des degrés d'automaticité divers. Les modèles multinomiaux, modèles mathématiques couplés à ce type de mesures, ont l'avantage de démêler la contribution spécifique des principaux processus à l'origine de la réponse individuelle. L'objectif de ce travail était de montrer l'intérêt théorique que peut avoir l'utilisation des modèles multinomiaux dans l'étude des stéréotypes et préjugés. À travers l'exemple d'une mesure indirecte du stéréotype liant les individus maghrébins à une menace, nous avons défendu la thèse selon laquelle les processus et l'organisation des processus à l'origine de la réponse individuelle pouvaient varier en fonction d'un contexte particulier et du type de participants prenant part à la mesure. Notre première partie empirique ne nous a pas permis de confirmer la possibilité qu'un contexte d'attentats puisse influencer de manière contraire deux des processus à l'origine de la réponse. Dans notre deuxième partie empirique, nous avons montré que la relation conditionnelle entre processus pouvait être liée à certaines composantes de la tâche, mais aussi et surtout à certains facteurs individuels. Enfin, notre dernière partie a mis en lumière le caractère reproductible de nos effets avec d'autres stimuli et auprès des participants cibles du stéréotype, mais pas dans un contexte géographique différent. En somme, ce travail de thèse était pour nous l'opportunité de mettre en avant l'importance des précautions à prendre dans l'interprétation des résultats issus de mesures indirectes et l'apport théorique et empirique des modèles multinomiaux dans ces interprétations.

Mots clés : mesures indirectes, stéréotypes, modèles multinomiaux, cognition sociale

Abstract

Indirect measures are central to the study of individual stereotypes and prejudice. However, these measures are not without criticism: contrary to what was previously argued, these measures are now thought to reflect multiple processes covering varying degrees of automaticity. Multinomial models, mathematical models coupled with indirect measures, have the benefit of disentangling the specific contribution of the main processes behind the individual response. The objective of this work was to show the theoretical value that multinomial models may have on stereotyping and prejudice studies. Through the example of an indirect measure of the stereotype linking people from North Africa to a threat, we argued that the processes and the conditional relationship between the processes behind the individual response could vary according to a specific context and the type of participants taking part in the measure. Our first empirical section did not allow us to confirm the possibility that a terrorist attack context could influence in an opposite way two of the processes behind the participant response. In our second empirical section, we showed that the conditional relationship between processes could be linked to some task components, but also and above all to some individual factors. Finally, our last part highlighted the replicability of our effects with other stimuli and with the participants linked to the stereotype, but not in a different geographical context. In sum, this work was an opportunity for us to highlight the need to be cautious when interpreting results from indirect measures as well as the theoretical and empirical contribution of multinomial models in these interpretations.

Keywords: indirect measures, stereotypes, multinomial models, social cognition

Table des matières

Avant-propos _____ *11*

Chapitre 1 – Mesurer les attitudes raciales : considérations théoriques _____ *15*

1.1 Stéréotype, préjugé et discrimination : trois composantes de l'attitude raciale _ **16**

1.1.1 Modèle tripartite de l'attitude _____ 16

1.1.2 Le stéréotype _____ 17

1.1.3 Le préjugé _____ 20

1.1.4 La discrimination _____ 22

1.1.5 Relation entre stéréotype, préjugé et discrimination _____ 24

1.2 La mesure des attitudes raciales _____ **26**

1.2.1 Les mesures précoces : contenu de croyances et personnalité _____ 26

1.2.2 Problèmes liés aux mesures directes : l'exemple du biais de désirabilité sociale _____ 29

1.2.3 L'apport de la psychologie cognitive : les critères d'automaticité _____ 31

1.2.4 Les modèles duaux du traitement de l'information _____ 34

1.2.5 Mesures indirectes des attitudes raciales _____ 38

1.2.6 Limites des mesures indirectes _____ 42

1.3 L'apport des modèles multinomiaux dans l'étude des attitudes raciales _____ **46**

1.3.1 Du PDP aux modèles multinomiaux _____ 46

1.3.2 Le fonctionnement des modèles multinomiaux : l'exemple du Quad Model _____ 52

1.3.3 Modèles multinomiaux dans le champ des attitudes raciales _____ 59

1.4 La Stereotype Misperception Task (SMT) _____ **61**

1.4.1 Les processus à l'origine de la stéréotypisation _____ 62

1.4.2 SMT : structure de la tâche et modèles associés _____ 63

1.4.3 Stéréotype de menace : le cas des Maghrébins _____ 68

1.5 Étude 1 (préliminaire) : réplication des résultats originaux _____ **70**

1.5.1 Méthode _____ 70

1.5.2 Résultats et discussion _____ 73

1.6 Une thèse en trois parties autour de la mesure du stéréotype _____ **76**

Chapitre 2 – Impact des attentats sur la perception des Maghrébins _____ *79*

2.1 L'impact des attentats, ce que nous savons _____ **79**

2.2 Utilisation de mesures indirectes, une absence de résultats ? 81

2.3 Étude 2 : utilisation de la SMT suite à un rappel d'attentats 84

2.3.1 Méthode 85

2.3.2 Résultats 87

2.3.3 Discussion 91

2.4 Étude 3 : utilisation de la SMT dans un contexte d'attentats 93

2.4.1 Méthode 94

2.4.2 Résultats 95

2.4.3 Discussion 100

2.5 Étude 4 : application d'un Quad Model à des données récoltées au moment des attentats 101

2.5.1 Méthode 101

2.5.2 Résultats 103

2.5.3 Discussion 110

2.6 Discussion du Chapitre 2 : l'amalgame Maghrébin-terroriste existe-t-il ? 113

Chapitre 3 – Rôle des facteurs liés à la tâche et des facteurs liés aux participants dans la relation conditionnelle entre processus 117

3.1 Comparaison de modèles multinomiaux dans la littérature 118

3.2 SAC-dominant vs. D-dominant 120

3.3 Étude 5 : SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc 124

3.3.1 Méthode 125

3.3.2 Résultats 127

3.3.3 Discussion 131

3.4 Étude 6 : manipulation de la perception de l'amorce et de la cible 133

3.4.1 Méthode 134

3.4.2 Résultats 136

3.4.3 Discussion 143

3.5 Étude 7 : mesure directe et relation conditionnelle entre processus 146

3.5.1 Méthode 147

3.5.2 Résultats 148

3.5.3 Discussion 153

3.6 Discussion du Chapitre 3 : aller plus loin dans nos précautions 156

Chapitre 4 – Que trouve-t-on derrière l'effet SMT ? _____ *160*

4.1 L'effet SMT : un effet robuste ? _____ **160**

4.2 Effet SMT : exemplaire versus catégorie _____ **163**

4.3 Étude 8 : impact des exemplaires vs. impact de la catégorie _____ **167**

4.3.1 Méthode _____ 168

4.3.2 Résultats _____ 170

4.3.3 Discussion _____ 174

4.4 L'effet SMT est-il généralisable à d'autres populations ? _____ **175**

4.4.1 Généralisabilité de l'effet SMT au sein d'un autre environnement culturel _____ 176

4.4.2 Généralisabilité de l'effet SMT au sein du groupe cible du stéréotype _____ 178

4.4.3 Vue d'ensemble des Études 9 et 10 _____ 183

4.5 Étude 9 : SMT dans un autre environnement culturel, réplication aux États-Unis _____ **184**

4.5.1 Méthode _____ 184

4.5.2 Résultats _____ 185

4.5.3 Discussion _____ 194

4.6 Étude 10 : analyse intégrative des données de participants de culture maghrébine _____ **198**

4.6.1 Méthode et résultats _____ 199

4.6.2 Discussion _____ 202

4.7 Discussion du Chapitre 4 : l'effet SMT est-il généralisable ? _____ **206**

Chapitre 5 – Discussion générale _____ *209*

5.1 Rappel des résultats : mise en évidence de la multiplicité des processus _____ **209**

5.2 Que retenir de ces études ? _____ **212**

5.2.1 Une prise en compte encore trop partielle du caractère multi-processuel des mesures indirectes _____ 213

5.2.2 De nouvelles avancées dans l'utilisation des modèles multinomiaux _____ 217

5.3 Nécessité d'une théorisation forte préalable à l'utilisation des modèles multinomiaux _____ **221**

5.3.1 Réactivation du stéréotype ou d'un préjugé négatif ? _____ 221

5.3.2 La SMT est-elle une mesure directe ? _____ 223

5.3.3 Indice de fiabilité de la SMT et de validité des modèles associés _____ 225

5.3.4 Et si l'activation du stéréotype était inversée ? 230
5.3.5 Choix d'un modèle à quatre paramètres ou d'un modèle à six paramètres 234

5.4 Pour conclure : des perspectives futures dans l'utilisation de ces modèles 237

***Bibliographie* 239**

***Index des figures* 263**

***Index des tableaux* 267**

***Annexes* 268**

Annexe 1 : inductions expérimentales du contexte (Étude 3) 268
Condition Attentat 268
Condition Contrôle 269

Annexe 2 : description des résultats pour le modèle D-dominant (Étude 5) 271

Annexe 3 : tableaux de corrélations entre les paramètres du modèle et les mesures directes (Études 6 et 7) 273
Étude 6 273
Étude 7 274

Annexe 4 : Échelle de Préjugés Généralisés (Études 3, 6 et 7) 275

Annexe 5 : échelles et sous-échelles de motivation à réduire l'expression de ses préjugés (Étude 7) 276
Échelle de motivation à contrôler ses réactions teintées de préjugé (MCPR) 276
Échelle de motivation interne et externe à répondre sans préjugés (MRWP) 277

Annexe 6 : amorces (prénoms) maghrébins, français et neutres (Étude 8) 278

Annexe 7 : réanalyse des étapes de validation des modèles SAC-dominant et D-dominant (Krieglmeyer & Sherman, 2012 ; méthode *latent-trait*) 279

Avant-propos

Imaginez-vous, jeune chercheur en psychologie voulant participer aux recherches actuelles traitant des stéréotypes raciaux. Votre intérêt se porte sur les stéréotypes associés aux individus d'origine maghrébine. Vous savez qu'étudier ces stéréotypes n'est pas chose aisée. Vous n'êtes pas naïf, recueillir directement l'avis des individus par questionnaire est particulièrement risqué. En effet, vous avez vu tout au long de votre cursus que les réponses aux questionnaires pouvaient être le reflet de multiples biais. Parmi d'autres, les participants pourraient cacher leur attitude négative envers ces populations stigmatisées. Le contexte actuel est propice à cela. Combien de fois a-t-on entendu dans les médias de ne pas faire d'amalgame, que les maghrébins n'étaient pas des terroristes, qu'ils étaient tout autant victime des attentats et qu'il fallait faire attention à ne pas discriminer davantage ces individus. Pourtant, c'est justement ce contexte qui vous a poussé à étudier ces stéréotypes, et en particulier celui liant les populations maghrébines à une menace potentielle. Certes, les participants pourraient modifier leur réponse pour se conformer à la pression sociale ou à certaines de leurs valeurs, mais qu'en est-il réellement de leur stéréotype, ces croyances ou connaissances ancrées dans leur esprit ? Le contexte actuel n'est-il pas propice à renforcer cette tendance à percevoir les Maghrébins comme une menace ? Le questionnaire étant trop risqué, il vous faut une mesure différente pour investiguer la réactivation de ce stéréotype, une mesure qui ne serait pas contaminée par un quelconque biais externe.

Vous savez que les psychologues sociaux utilisent aujourd'hui un autre type de mesures, des mesures permettant de contourner l'influence volontaire du participant. Vous avez lu que ces mesures étaient révolutionnaires, elles permettraient d'avoir accès à la « vraie » attitude, celle qui existe réellement à l'intérieur de l'esprit des individus. À vos yeux, cet autre type de mesures serait idéal pour votre objet d'étude. Vous en êtes sûr, vos participants n'auront pas conscience de ce qui influence leur réponse au sein de cette tâche. De toute façon, même s'ils en ont conscience, ils ne pourront pas contrôler leur réponse, la tâche est faite pour ça. En êtes-vous certain ? Étudiant en psychologie, vous avez vous-même maintes fois passé ce type de tâches. Dire que vous aviez une idée de ce qui était mesuré relève de l'euphémisme. Souvenez-vous, n'avez-vous jamais essayé de contrôler vos réponses ? Vous n'êtes plus très sûr. Au final, vous vous demandez : existe-t-il réellement un moyen d'investiguer cette question ?

Comme nous le verrons au Chapitre 1, les psychologues sociaux se sont très tôt intéressés aux stéréotypes et préjugés qui peuvent être à l'origine d'un comportement discriminatoire. En conséquence, différents types de mesures ont émergé et évolué au fil du temps permettant l'apparition de nouvelles problématiques et facilitant l'avancée des connaissances sur ces objets de recherche. Ainsi, les questionnaires papier mesurant les stéréotypes et préjugés partagent maintenant l'affiche avec des tâches informatisées, popularisées par l'apparition de nouveaux modèles théoriques. Ces modèles font la part belle aux processus qualifiés d'automatiques, ainsi qu'à leur opposition ou interaction avec des processus non automatiques. En conséquence, de nombreuses tâches informatisées sont aujourd'hui disponibles pour essayer de capter au plus près ces processus automatiques.

Ces dernières années, des critiques ont émergées quant à l'utilisation de ces mesures et aux conclusions qu'elles nous permettent de tirer. L'une de ces critiques concerne leur capacité à capter des processus automatiques sans influence d'un quelconque contrôle de la part du participant. Ce travail de thèse s'inscrit dans le prolongement de ces critiques. En prenant l'exemple d'une mesure indirecte du stéréotype, nous montrerons l'importance des précautions à prendre dans l'utilisation de ce type de tâches. Disposons-nous des bons outils pour mesurer ce que nous voulons mesurer ? Plus précisément, les outils que nous avons en notre possession mesurent-ils réellement ce que nous pensons qu'ils mesurent ? Tout au long de ce manuscrit, nous nous intéresserons aux processus cognitifs à l'œuvre au sein des mesures indirectes et montrerons qu'une investigation systématique de ces processus peut nourrir notre réflexion quant à la situation particulière qu'est la passation d'une tâche de psychologie sociale. Nous nous poserons ainsi les questions suivantes : quels processus ces tâches permettent-elles de mesurer ? Quel peut être l'impact d'un contexte particulier sur ces processus ? Le type de participants ou de stimuli que nous utilisons peut-il influencer l'organisation de ces processus ? Et enfin, est-on sûr que les résultats de ces mesures reflètent effectivement notre résonnement théorique ? Au final, une thèse générale a pu émerger de l'ensemble de ces questions : nous pensons que les processus et l'organisation des processus à l'origine de la réponse individuelle peuvent varier en fonction d'un contexte particulier et du type de participants prenant part à la mesure indirecte du stéréotype. Globalement, nous montrerons qu'en utilisant une tâche adéquate couplée à un modèle mathématique particulier, nous pouvons apporter des éléments de réponse à l'ensemble de ces questions.

Le Chapitre 1 aura pour objectif de revenir sur l'étude des stéréotypes au cours du temps, ainsi que sur les outils à la disposition des psychologues sociaux pour les mesurer.

Nous élargirons volontairement notre champ théorique aux concepts de préjugé et de discrimination, concepts indissociables du stéréotype. Ainsi, ces trois concepts seront regroupés sous le terme d'attitude raciale. Après avoir illustré les premières mesures de ces trois composantes de l'attitude, nous nous centrerons sur les mesures indirectes à disposition, ainsi que leurs critiques ces dernières années. En conséquence, nous aborderons en détail le fonctionnement d'un type de modèles en particulier, les modèles multinomiaux, et l'intérêt que représentent ces modèles pour répondre à ces critiques. Enfin, nous mettrons en lumière une mesure du stéréotype, la *Stereotype Misperception Task* (SMT ; Krieglmeyer & Sherman, 2012), qui nous accompagnera tout au long de ce manuscrit. Cette mesure nous permettra d'étudier un type de population particulièrement discriminé en France, la population maghrébine. Nous investiguerons en détail le stéréotype qui peut lier un individu maghrébin à une menace. Trois grands axes, représentés par trois chapitres, vont nous permettre d'explorer différentes questions quant à l'utilisation de la SMT pour étudier le stéréotype de menace associé à cette population.

Le Chapitre 2 nous permettra d'explorer l'impact d'un contexte d'attentats sur notre perception des individus d'origine maghrébine. Plus particulièrement, ce chapitre sera pour nous l'occasion d'aborder la nécessité de prendre en compte à la fois l'impact du contexte sur la réponse du participant, mais aussi et surtout l'impact de ce contexte sur les processus aboutissant à cette réponse. Nous tenterons de mettre en évidence le fait qu'un contexte particulier comme un contexte d'attentats peut influencer de manière contraire deux des processus jouant un rôle dans la réponse du participant. Globalement, ce premier chapitre empirique sera l'occasion d'illustrer l'importance, pour le chercheur, d'avoir conscience non seulement des processus à l'origine de la réponse du participant, mais aussi et surtout de l'influence de nos manipulations sur ces processus.

Dans le Chapitre 3 nous mettrons en évidence l'importance des facteurs liés aux participants et des facteurs liés à la tâche dans la relation conditionnelle entre les processus à l'œuvre dans la SMT. Nous montrerons qu'en fonction des stimuli choisis, mais aussi du type de participant prenant part à la tâche, le processus principal qui guide la réponse du participant peut varier. Ainsi, ce troisième chapitre aura pour but de souligner l'importance non seulement de la prise en compte de ces différents processus, mais aussi de la relation conditionnelle que ces processus entretiennent et de leur importance relative pour aboutir à la réponse du participant.

Enfin, le Chapitre 4 sera consacré au caractère généralisable de nos effets. Nous nous attarderons ainsi sur l'effet classiquement obtenu dans un paradigme SMT et sur les réflexions théoriques quant à l'origine de cet effet. Dans un premier temps, nous laisserons de côté les processus cognitifs sous-tendant la réponse du participant pour nous intéresser aux stimuli que nous avons utilisés. Nous tenterons de voir si la réponse du participant reflète effectivement une réactivation des catégories d'intérêt et des stéréotypes associés ou plus simplement du type de stimuli en eux-mêmes. Dans un second temps, nous questionnerons le caractère reproductible de notre effet auprès d'un échantillon issu d'un contexte géographique différent ainsi qu'auprès d'individus issus du groupe habituellement associé au stéréotype. Nous montrerons ainsi la nécessité d'une prise en considération du caractère multi-processuel des réponses dans la contribution théorique liée à cette littérature.

D'une manière générale, ce travail de thèse sera pour nous l'occasion de mettre en évidence l'importance d'une prise en compte précise des principaux processus sous-tendant la réponse individuelle au sein d'une mesure indirecte. Par le biais de ces trois chapitres, nous insisterons sur un ensemble de réflexions théoriques et méthodologiques souvent mises de côté dans nos recherches qui, pourtant, peuvent avoir de l'importance dans l'interprétation de nos résultats.

Chapitre 1 – Mesurer les attitudes raciales : considérations théoriques

S'intéresser aux stéréotypes et au meilleur moyen pour les mesurer nécessite plus largement d'aborder le concept d'attitude. Toutefois, faire un historique de l'ensemble des mesures d'attitudes ainsi que de leur évolution serait hors de propos. Au sein de ce manuscrit de thèse, nous nous centrerons uniquement sur les attitudes que les individus peuvent avoir envers les membres de groupes dits « ethniques ». La notion de groupe ethnique n'étant pas définie juridiquement en France, nous utiliserons cette terminologie à visée opératoire. Par membre d'un groupe ethnique, nous considérons tout individu pouvant être catégorisé comme faisant partie d'un groupe issu de l'immigration africaine, antillaise (individus de couleur noire) ou nord-africaine (individus maghrébins). Nous nous centrerons sur la catégorisation subjective, l'appartenance ethnique pouvant être réelle (e.g., parents d'origine maghrébine), mais aussi simplement supposée (e.g., individu ayant la peau mate). La littérature anglophone utilise le terme « racial attitude » (e.g., Fazio et al., 1995) pour faire référence aux attitudes dont l'objet attitudinal est un groupe ethnique particulier ou un individu membre de ce groupe. En conséquence, et en adéquation avec les termes employés dans la littérature, nous utiliserons tout au long de ce manuscrit le terme d' « attitude raciale ». Nous insistons une fois encore sur le fait que le terme « raciale » ne présuppose aucune séparation biologique en termes de race. Au contraire, et comme nous pouvons le retrouver parfois dans la langue française (e.g., « discrimination raciale »), ce terme est défini présentement par sa notion de construit social, c'est-à-dire de catégorisation subjective, et dans le cas présent abusive, sur une base ethnique (e.g., « les Maghrébins »). Au sein de ce manuscrit, nous nous intéresserons en particulier aux individus catégorisés comme étant d'origine maghrébine. Suivant en cela certains travaux traitant de la labélisation des groupes en France, nous parlerons du groupe des Maghrébins pour référer aux individus français d'origine maghrébine (Harabi, 2018). À l'opposé, nous parlerons du groupe des Blancs pour parler des individus français d'origine européenne. De la même manière, nous mentionnerons régulièrement des études s'intéressant aux individus de couleur noire. La littérature américaine utilise aujourd'hui le label « Afro-américains » pour désigner les individus de couleur noire, ce label s'opposant au label « Européo-américains » pour désigner les individus de couleur blanche. Ces labels n'ayant à nos yeux que peu de sens quand appliqués aux populations françaises, nous préférerons des labels plus classiques et parlerons d'individus noirs ou d'individus blancs.

1.1 Stéréotype, préjugé et discrimination : trois composantes de l'attitude raciale

1.1.1 Modèle tripartite de l'attitude

Il nous semble tout d'abord important de clarifier ce que nous entendons par attitude. Bien qu'il existe une multitude de définitions du concept d'attitude, une grande partie de ces définitions s'accorde aujourd'hui sur son aspect évaluatif. Parmi les plus connus, Eagly et Chaiken (1993) définissent l'attitude comme « une tendance psychologique qui serait exprimée par l'évaluation d'une entité particulière avec un certain degré de faveur ou de défaveur » (p. 1). Cependant, dans le but d'englober l'ensemble des dimensions constituant l'attitude raciale, nous préférons adopter une définition plus générale du concept d'attitude. Nous suivrons en cela la définition première d'Allport (1935) qui définissait l'attitude comme « un état mental et neural de préparation, organisé à travers l'expérience, exerçant une influence directrice et dynamique sur les réponses de l'individu envers tous les objets et situations avec lesquelles il est lié » (p. 810). Dans la conception classique de l'attitude (e.g., Rosenberg & Hovland, 1960), celle-ci peut être divisée en trois composantes : la composante cognitive, la composante affective et la composante comportementale. La composante cognitive correspondrait aux croyances et connaissances liées à un objet attitudinal. La composante affective correspondrait pour sa part aux affects et émotions liés à cet objet. Enfin, la composante comportementale regrouperait les tendances à l'action et comportements en rapport à l'objet attitudinal. L'attitude d'un individu peut donc être la résultante d'une, deux, ou de l'ensemble de ces trois composantes.

Lorsque l'on parle spécifiquement d'attitude raciale, ces trois composantes peuvent être décrites selon une autre terminologie (e.g., Correll et al., 2010 ; Dovidio et al., 1996 ; Fiske, 1998). La dimension cognitive de l'attitude correspondrait au stéréotype (e.g., « les Maghrébins sont menaçants »). La dimension affective de l'attitude correspondrait au préjugé (e.g., « je n'aime pas les Maghrébins »). Enfin, la dimension comportementale de l'attitude correspondrait à la discrimination (e.g., « j'évite de m'approcher de cet individu maghrébin »). Ainsi, mesurer l'attitude raciale d'un individu envers un groupe ou un individu membre de ce groupe reviendrait à mesurer son stéréotype, son préjugé ou bien la discrimination qu'il pourrait entretenir envers cet individu. Avant de détailler l'évolution des mesures des

attitudes raciales, il nous parait nécessaire de revenir plus en détail sur la définition de ces trois composantes.

1.1.2 Le stéréotype

Il convient d'abord de signaler que, encore aujourd'hui, définir ce qu'est un stéréotype n'est pas chose aisée. En effet, malgré le nombre important de définitions que nous pouvons trouver dans la littérature (plusieurs centaines ; Stangor, 2016), il n'existe à ce jour aucun consensus sur ce qu'est ou compose un stéréotype. Ainsi, en nous basant sur certaines définitions existantes, nous tâcherons en quelques lignes de dresser un panorama rapide des caractéristiques importantes d'un stéréotype.

D'un point de vue épistémologique, le terme stéréotype apparait au XVIIIe siècle. Il est à la base un terme d'imprimerie, repris en psychiatrie, censé appuyer sur la notion de rigidité, de fixité (Légal & Delouvée, 2015). Le terme stéréotype comme nous l'entendons aujourd'hui a été en premier lieu utilisé par l'éditorialiste Lippmann en 1922. Lippmann définissait le stéréotype comme des « images dans notre tête », la plupart du temps erronées, concernant des groupes ou individus. Pour Lippmann, ces images dans notre tête avaient pour rôle de simplifier le réel et de pouvoir permettre à l'individu de faire face à la complexité du monde qui l'entoure. Un peu plus tard, Katz et Braly (1935) ont défini le stéréotype comme « une impression fixe correspondant très peu aux faits qu'il prétend représenter et résultant en premier lieu de notre définition, et en second lieu de notre observation » (p. 181). Encore plus tard, Brigham (1971a) définit le stéréotype comme « une généralisation à propos d'un groupe ethnique concernant l'attribution d'un trait, et qui est considérée comme injustifiée par un observateur » (p. 31).

Ces premières descriptions ont en commun de définir le stéréotype par son critère de véracité ou d'exactitude. Pour ces auteurs, le stéréotype serait une vision erronée de la réalité. Au final, peu de travaux ont testé empiriquement l'exactitude d'un stéréotype (e.g., Judd & Park, 1993 ; Swim, 1994). De ces travaux, il en ressort souvent la notion de « noyau de vérité », hypothèse selon laquelle la plupart des stéréotypes sont basés sur une réalité empirique partielle ou minime (e.g., Schneider, 2004 ; Swim, 1994). Cependant, savoir si cette réalité empirique serait sous-estimée ou surestimée n'est pas encore tranchée (Stangor, 2016 ; Swim, 1994). De plus, il semble qu'il y ait une différence entre l'exactitude des

stéréotypes que l'on attribue à son endogroupe par rapport à ceux attribués à un exogroupe, ces derniers étant plus sensibles à une exagération (Judd & Park, 1993). Globalement, il semble néanmoins que ce soit davantage le processus de surgénéralisation qui peut amener à un jugement incorrect (Fiske, 1989). Ainsi, même si le groupe possède effectivement la caractéristique qu'on lui attribue, le fait d'appliquer cette caractéristique sans distinction à toutes les individualités du groupe peut provoquer cette inexactitude (Judd & Park, 1993 ; Stangor, 2016).

Les recherches en cognition sociale ont permis de remettre en question la perception que l'on pouvait avoir du stéréotype. Comme nous l'avons montré ci-dessus, bon nombre d'auteurs de travaux précoces voyaient dans le processus de formation d'un stéréotype quelque chose de mauvais, de négatif (voir Dovidio et al., 1996 ; Schneider, 2004). La vision des chercheurs en cognition sociale est différente. En cognition sociale, la représentation mentale constituant le stéréotype peut être décrite comme étant de l'ordre du schéma cognitif (Hilton & von Hippel, 1996). Selon cette approche, un schéma serait une structure cognitive représentant des connaissances organisées sur un concept (Fiske & Taylor, 2013). Nous pourrions ainsi avoir le schéma des Maghrébins, qui représenterait en mémoire toutes les connaissances que nous pourrions avoir des Maghrébins. Ainsi, par le biais d'un phénomène de catégorisation mentale, nous rassemblerions des connaissances abstraites par rapport au groupe ainsi qu'aux membres de ce groupe (Dovidio et al., 1996 ; Macrae & Bodenhausen, 2000, 2001). Plus particulièrement, les schémas mentaux que nous aurions contiendraient non seulement les attributs propres au groupe, mais aussi la relation entre ces attributs (Fiske & Taylor, 2013). Parmi ces attributs nous pouvons retrouver des connaissances générales sur le groupe, allant de caractéristiques physiques (e.g., les personnes âgées ont les cheveux blancs), aux traits de personnalités (e.g., les personnes âgées sont réfléchies) ou bien ses comportements prototypiques (e.g., les personnes âgées dorment beaucoup).

Comme nous le savons maintenant, notre système cognitif a une capacité limitée (e.g., Fiske & Taylor, 2013 ; Moskowitz, 2005 ; Ric & Muller, 2017). Ainsi, et rejoignant en cela la définition initiale de Lippmann (1922), catégoriser sous forme abstraite permettrait de faire face à la multitude d'informations qui nous arrive au quotidien et ainsi de faciliter le traitement de l'information. Bien que ce point soit aujourd'hui partagé par de nombreux chercheurs, il reste cependant la question de ce que l'on entend par représentation abstraite, celle-ci pouvant à la fois référer à un schéma, comme dit précédemment, mais aussi à un

réseau associatif entre attributs, ou même une représentation abstraite du prototype de l'individu (Hilton & von Hippel, 1996). De manière générale, la littérature utilise régulièrement la notion d'informations sémantiques pour parler du stéréotype, en opposition au contenu affectif du préjugé (Dovidio et al., 1996).

Aborder ce processus de catégorisation mentale nous permet également d'aborder la valence des connaissances stockées, et donc la valence du stéréotype. Bien que certains chercheurs insistent sur la valence globalement négative d'un stéréotype, en particulier dans le cas des attitudes raciales (e.g., Judd & Park, 1993 ; Stangor, 2016), le stéréotype a la particularité de pouvoir être à la fois de valence positive (e.g., « les personnes de couleur noire sont sportives »), comme négative (e.g., « les personnes de couleur noire sont menaçantes » ; Bodenhausen & Richeson, 2010 ; Fiske et al., 2002 ; Krieglmeyer & Sherman, 2012). En outre, il semblerait une fois encore que les stéréotypes associés à notre endogroupe soient majoritairement positifs, là où ceux associés aux exogroupes soient majoritairement négatifs (Hilton & von Hippel, 1996 ; Schneider, 2004).

Les définitions actuelles des stéréotypes se veulent beaucoup plus larges et consensuelles. Celles-ci n'abordent que très peu l'aspect d'exactitude ou la valence du stéréotype. Par exemple, Hilton et von Hippel (1996) décrivent le stéréotype comme « des croyances à propos des caractéristiques, attributs et comportements des membres de certains groupes » (p. 240). De la même manière, Leyens et collaborateurs (1996) définissent le stéréotype comme « des croyances partagées concernant les caractéristiques personnelles, généralement des traits de personnalité, mais souvent aussi des comportements, d'un groupe de personnes » (p. 24). Cette définition soulève toutefois un autre point de désaccord entre chercheurs. Est-ce qu'un stéréotype peut être individuel (i.e., propre aux croyances de l'individu) ou doit-il être partagé par un ensemble d'individus (i.e., profils de croyances partagées au sein d'une communauté ; Leyens et al., 1996) ? En cela, bon nombre de chercheurs s'accordent aujourd'hui sur le fait qu'un stéréotype a un fondement social (e.g., amis, familles) et culturel (e.g., Légal & Delouvée, 2015 ; Leyens et al., 1996 ; Payne et al., 2017 ; mais voir Judd & Park, 1993, pour un avis contraire). En complément, il convient de souligner qu'il peut y avoir connaissance du stéréotype sans toutefois adhésion au stéréotype (e.g., Arkes & Tetlock, 2004 ; Devine, 1989 ; Payne et al., 2017). Ainsi, nous pouvons avoir connaissance du stéréotype « les femmes sont mauvaises en mathématiques », stocker cette connaissance dans notre représentation mentale des femmes, sans toutefois adhérer à cette idée. Nous reprendrons ce point au Chapitre 4.

Pour finir, il est important de séparer stéréotype et processus de stéréotypisation (Leyens et al., 1996). Là où le stéréotype représente un contenu mental, le processus de stéréotypisation représente pour sa part l'utilisation de ce contenu mental. Ainsi, le processus de stéréotypisation consiste, en rapport à des individus cibles, à « leur appliquer un jugement —stéréotypique — qui rend ces individus interchangeables avec les autres membres de leur catégorie » (Leyens et al., 1996, p. 24). On retrouve ici l'aspect de généralisation abordé précédemment. Par exemple, Emilie est une bibliothécaire. Utiliser le processus de stéréotypisation revient à dire que, comme toutes les bibliothécaires, Emilie est forcément intelligente, studieuse, n'aime pas faire la fête, etc. Nous reviendrons sur ce processus de stéréotypisation lorsque nous aborderons les mesures des stéréotypes. Mais pour l'instant, il convient de définir ce que constitue la composante affective de l'attitude raciale, à savoir le préjugé.

1.1.3 **Le préjugé**

Définir ce qu'est un préjugé peut s'avérer problématique au regard des autres définitions auxquelles nous nous sommes intéressés jusqu'à présent. Nous tâcherons de faire un rapide état des lieux des définitions existantes afin dans un second temps de nous centrer sur la définition qui, en rapport à notre objet d'étude, nous semble la plus pertinente.

La plus célèbre contribution à la littérature sur les préjugés est l'œuvre d'Allport qui, dans son livre the Nature of Prejudice (1954), donne l'une des premières définitions du concept de préjugé. Allport définissait le préjugé comme très proche de son sens littéral, à savoir de « pré-jugement ». Pour lui, la seule différence entre ces deux notions était que le pré-jugement, à l'inverse du préjugé, pouvait être modifié par de nouvelles connaissances (Fishbein, 2002). Pour Allport, le préjugé se définissait donc comme « une antipathie basée sur une généralisation rigide et erronée. Il peut être ressenti ou exprimé. Il peut être dirigé envers un groupe dans son ensemble ou envers un individu parce qu'il est membre de ce groupe » (p. 9). Suivant cette première définition, mais élargissant celle-ci, la définition du préjugé la plus communément utilisée de nos jours est la suivante : un préjugé est une attitude négative envers un groupe social ou les membres de ce groupe (Eagly & Diekman, 2005 ; Stangor, 2016). Défini en ces termes, le préjugé peut prendre d'autres appellations en fonction du groupe cible. Ainsi, il n'est pas rare de voir le préjugé comme synonyme de racisme, de

sexisme, d'antisémitisme ou d'homophobie (e.g., Brown, 2011 ; Dovidio et al., 2010 ; Schneider, 2004). Nous reviendrons plus tard sur la notion de négativité du préjugé, mais voyons d'abord en quoi définir le préjugé de cette manière peut s'avérer problématique.

Le préjugé conceptualisé comme une attitude, et suivant en cela le modèle tripartite de l'attitude (e.g., Rosenberg & Hovland, 1960), il peut être vu comme ayant trois composantes : une composante cognitive, une composante affective, et une composante comportementale (Correll et al., 2010 ; Dovidio et al., 2010). Ainsi, en plus de l'aspect affectif caractéristique du préjugé (e.g., Allport, 1954), le préjugé est ici décrit comme englobant à la fois le concept de stéréotype (dimension cognitive) et celui de discrimination (dimension comportementale). Cela peut de ce fait créer une confusion, le préjugé pouvant être perçu comme attitude dans son ensemble, mais aussi comme composante de l'attitude. En conséquence, et suivant en cela certains auteurs (Correll et al., 2010 ; Fiske, 1998), nous préférons une définition du préjugé se concentrant sur sa dimension affective ou émotionnelle. Parmi d'autres, nous pouvons citer les définitions de Schneider (2004), le préjugé étant vu comme « l'ensemble des réactions affectives que nous avons envers des individus en fonction de leur appartenance à une catégorie » (p. 27) ou celle de Correll et collaborateurs (2010) définissant le préjugé comme « une réponse affective ou évaluative valencée (positive ou négative) envers une catégorie sociale et ses membres. » (p. 46). Cette distinction plus précise entre stéréotype, préjugé et discrimination nous parait importante pour la suite de ce document. En effet, cela nous permettra de séparer plus distinctement les différents types de mesures s'intéressant au contenu sémantique de l'attitude raciale (stéréotype), du contenu affectif de l'attitude raciale (préjugé ; Correll et al., 2010).

Par de nombreux aspects, stéréotypes et préjugés partagent des points de convergence. Comme nous venons de le voir, le préjugé, comme le stéréotype, peut être valencé positivement comme négativement. En cognition sociale cela peut s'expliquer par le fait que tous deux peuvent être vus comme reflétant le fonctionnement normal de notre catégorisation mentale (Schneider, 2004). Malgré tout, il semble que la littérature sur les attitudes raciales traite majoritairement les préjugés valencés négativement (Stangor, 2016). Pour certains auteurs, la littérature se concentre sur les préjugés négatifs du fait des conséquences plus importantes que ceux-ci peuvent entrainer (e.g., discrimination ; Kite & Whitley Jr, 2016). Il est à noter enfin que savoir si l'aspect affectif de notre attitude est, comme nos croyances, stocké en mémoire (Fazio, 2007) ou reconstruit *on the spot* à partir de celles-ci (Schwarz, 2007) est aujourd'hui encore sujet à débat.

Un autre point de débat concerne la question de l'origine individuelle ou culturelle des préjugés. Pour certains auteurs, la différence entre stéréotype et préjugé viendrait du fait que le stéréotype serait culturellement partagé, là où la composante affective du préjugé serait davantage personnelle (Devine, 1989). Cependant, comme nous le verrons plus tard, cette distinction reflète plus l'adhésion ou non au contenu réactivé que le contenu lui-même, et reflèterait donc davantage le type de processus à l'œuvre (automatique vs. non automatique ; Fiske, 2000). Le débat de l'origine individuelle du préjugé rejoint en réalité celui de la composante personnelle ou extra-personnelle de l'attitude (e.g., Arkes & Tetlock, 2004 ; Banaji et al., 2004 ; Jost, 2019 ; Payne et al., 2017 ; Rougier et al., 2020). En accord avec une vision culturelle du préjugé, de nombreux travaux empiriques ont mis en évidence l'aspect partagé du préjugé chez un groupe de culture ou d'environnement social similaire (voir Arkes & Tetlock, 2004 ; Payne et al., 2017), et ce, même chez les individus membres du groupe cible du préjugé (e.g., Ashburn-Nardo et al., 2003 ; Nosek et al., 2002 ; Richeson et al., 2005 ; Rudman et al., 2002). Globalement, bien que nos ressentis affectifs puissent être issus de notre propre expérience, le fondement social et culturel du préjugé est assez bien établi (e.g., Payne et al., 2017) Comme pour le stéréotype, nous reprendrons ce débat au Chapitre 4.

1.1.4 La discrimination

À l'inverse du stéréotype et du préjugé, nous trouvons un relatif consensus en psychologie quant à la définition du concept de discrimination. Parmi ces définitions, nous pouvons en retenir deux qui, à notre sens, regroupent ses principales caractéristiques. Schneider (2004) définit la discrimination comme « l'utilisation injustifiée d'informations catégorielles pour porter des jugements sur d'autres individus » (p. 29). Pour Correll et collaborateurs (2010), la discrimination serait « un comportement à l'égard des membres de la catégorie qui est conséquent pour leurs résultats et qui leur est adressé non pas en raison d'un quelconque mérite particulier ou d'une quelconque réciprocité, mais simplement parce qu'ils se trouvent être membres de cette catégorie » (p. 46).

Ainsi, la discrimination est avant tout caractérisée par son aspect comportemental ou de tendance comportementale. De ce fait, contrairement au stéréotype et au préjugé, elle peut être directement observable. Tout type de comportement pouvant être discriminatoire (Stroebe & Insko, 1989), celle-ci peut prendre de nombreuses formes, allant de simples

comportements d'évitements (e.g., Word et al., 1974) à des comportements d'agression verbale ou physique (Bodenhausen & Richeson, 2010). En psychologie, les études existantes se sont principalement centrées sur les comportements de discrimination interpersonnels (e.g., Benson et al., 1976 ; Gaertner & Bickman, 1971 ; Word et al., 1974). Cependant, celle-ci peut également être organisationnelle ou institutionnelle (e.g., proportion plus faible de femmes dans des postes hauts placés), voire même culturelle (e.g., sous-représentation de personnes noires tenant des rôles principaux au cinéma ; Kite & Whitley Jr, 2016 ; Schneider, 2004).

Tout comme le stéréotype et le préjugé, la discrimination peut être à la fois positive et négative (Dovidio et al., 2010). Toutefois, en psychologie, et particulièrement quand on s'intéresse aux relations intergroupes, la discrimination est décrite comme un comportement avant tout négatif (Dovidio et al., 2010 ; Fishbein, 2002 ; Kite & Whitley Jr, 2016). Il est à noter que cette négativité peut être le fruit à la fois d'une valorisation de notre groupe social ou bien d'une dévalorisation d'un groupe social différent (Schneider, 2004). Ainsi, un recruteur pourra embaucher un homme parce qu'il se sent plus à l'aise avec des hommes ou parce qu'il juge les femmes non compétentes. Encore plus que par sa négativité, la discrimination est associée à un comportement injustifié, non mérité (Schneider, 2004). Le comportement discriminatoire serait de ce fait non objectif. Par conséquent, ce qui constitue ou non une discrimination peut ne pas faire consensus, un comportement justifié pour un individu pouvant ne pas l'être pour un autre individu (e.g., embaucher un mannequin sur un critère physique ou un critère d'âge ; Al Ramiah et al., 2010 ; Correll et al., 2010). À un niveau interindividuel, il peut donc exister un flou sur ce qui constitue ou non une discrimination. C'est la raison pour laquelle une définition légale de la discrimination existe, celle-ci ayant pour but de séparer les comportements pouvant être objectivement justifiés, de ceux qualifiés de discriminatoires.

À un niveau interindividuel, la discrimination est donc vue comme un comportement biaisé (Dovidio et al., 2010). L'information issue de nos catégories mentales va prendre le pas sur l'information objective pour guider nos actions. De cette façon, nos croyances quant à un groupe social (i.e., nos stéréotypes), nos affects et émotions envers ce groupe social (i.e., préjugé) ou la combinaison des deux vont pouvoir guider nos comportements face à un individu membre de ce groupe. Nous voyons ici un premier lien entre stéréotype, préjugé et discrimination. En réalité, le lien entre ces trois concepts est un peu plus complexe.

1.1.5 Relation entre stéréotype, préjugé et discrimination

Bien qu'il n'y ait que peu de travaux qui aient théorisé le lien entre stéréotype, préjugé, et discrimination (e.g., Cottrell & Neuberg, 2005 ; Cuddy et al., 2007), la relation entre ces trois composantes est souvent décrite comme complexe et multifactorielle (Schneider, 2004). En cognition sociale, la littérature met la plupart du temps en avant une relation causale ou temporelle (e.g., Bodenhausen & Richeson, 2010 ; Schneider, 2004). Suivant en cela certaines théories classiques des attitudes, comme par exemple les théories de l'action raisonnée et du comportement planifié (Ajzen, 1991 ; Ajzen & Fishbein, 2000), les stéréotypes négatifs que nous aurions envers un groupe donneraient naissance à notre préjugé envers ce groupe, celui-ci pouvant aboutir à un comportement négatif envers ce groupe ou l'un de ses membres. En d'autres termes, l'affect négatif que l'on aurait par rapport au groupe (e.g., « je n'aime pas les Maghrébins ») serait tout d'abord déterminé par la valence globale des croyances réactivées (e.g., « les Maghrébins sont menaçants »). C'est cet affect qui guiderait nos tendances comportementales et aboutirait à nos comportements potentiellement discriminatoires (e.g., « je ne vais pas embaucher cet individu d'origine maghrébine »). Dans ce modèle classique, le stéréotype causerait donc le préjugé, qui causerait la discrimination (e.g., Bodenhausen & Richeson, 2010 ; Dovidio et al., 1996 ; Park & Judd, 2005 ; Schneider, 2004).

Le *Behavior from Intergroup Affect and Stereotype map* (i.e., BIAS map ; Cuddy et al., 2007) fais partie des rares modèles ayant testé empiriquement la relation causale entre ces trois composantes. Ce modèle est basé sur le *Stereotype Content Model* (SCM ; Fiske et al., 2002) qui a pour but de classer différents groupes sociaux selon les dimensions stéréotypées de chaleur et de compétence. Par exemple, le groupe des « Blancs » serait perçu comme chaleureux et compétent, alors que le groupe des « Arabes » serait lui perçu comme moyennement chaleureux et compétent (Fiske et al., 2002). Pour Fiske et collaborateurs (2002), ces différentes combinaisons vont aboutir à une émotion associée au groupe. Les groupes hauts sur les dimensions de chaleur et compétence vont être associés à des émotions positives (i.e., admiration, respect). À l'inverse, les groupes bas sur ces dimensions vont être associés à des émotions négatives (i.e., mépris). Dans leur BIAS Map, Cuddy et collaborateurs ont lié le contenu de ces stéréotypes (et leurs réactions émotionnelles associées) à deux dimensions comportementales : la dimension facilitateur-nuisible et la dimension actif-passif. Par exemple, ces auteurs ont montré que les groupes perçus comme haut sur les dimensions de chaleur et compétence, donc associés à des émotions positives, allaient être sujets à des

comportements globalement facilitateurs (à la fois actifs et passifs). À l'inverse, les groupes bas sur ces dimensions, donc associés à des émotions négatives, vont subir des comportements globalement nuisibles (là encore, à la fois actifs et passifs). De manière générale, Cuddy et collaborateurs ont confirmé, par le biais de cette étude, le lien causal entre stéréotype, préjugé et discrimination. Plus particulièrement, ils ont mis en évidence un lien direct entre préjugé et discrimination, ainsi qu'un lien indirect entre stéréotype et discrimination, celui-ci étant médiatisé par le préjugé.

Un point particulièrement remis en cause par rapport à la relation entre stéréotype, préjugé et discrimination est cette vision unilatérale de la relation qu'entretiennent ces trois composantes (Bodenhausen & Richeson, 2010 ; Dovidio et al., 1996 ; Park & Judd, 2005 ; Schneider, 2004). Par exemple, un grand nombre d'auteurs ont mis en évidence le rôle de rationalisation que pouvait en réalité avoir le stéréotype (Allport, 1954 ; Dovidio et al., 1996 ; Jost & Banaji, 1994). Pour Jost et Banaji (1994), le stéréotype aurait pour fonction idéologique de légitimer la domination de certains groupes sur d'autres. Selon la théorie de la justification du système, le stéréotype viendrait donner un sens à nos préjugé et comportements (e.g., « si les Maghrébins sont aujourd'hui encore discriminés, c'est parce qu'ils sont menaçants et sont un danger pour la société »). Le stéréotype ne causerait donc pas préjugé et discrimination, mais devient une conséquence de ceux-ci. De la même manière, d'autres théories s'appuient sur ce rôle de justification ou de légitimation du stéréotype, comme la théorie de la croyance en un monde juste (e.g., Dion & Dion, 1987). Selon cette théorie, nous serions motivés à maintenir comme croyance que nous méritons ce qui nous arrive. Ainsi, le stéréotype pourrait fonctionner ici comme maintien de cette croyance (e.g., « si les Noirs gagnent moins d'argent, c'est qu'ils le méritent, car ils sont fainéants »).

Enfin, une dernière critique de la relation supposée entre stéréotype, préjugé et discrimination est la corrélation relativement faible entre ces trois composantes de l'attitude raciale ($r < .30$; e.g., Dovidio et al., 1996 ; Park & Judd, 2005) et l'instabilité de cette relation (Ajzen & Fishbein, 2000). Pour certains auteurs, cette faible corrélation illustre le fait que stéréotype, préjugé et discrimination peuvent, dans certains cas, ne pas être liés directement (Schneider, 2004). D'autre part, il est également mis en avant l'existence de nombreux modérateurs (e.g., pression sociale, motivation ; Devine, 1989 ; Fazio, 1990) pouvant influer sur le lien direct entre ces variables. Pour Ajzen et Fishbein (2000), la faible corrélation entre ces variables serait davantage le reflet d'un problème de mesure. Ces auteurs pointent deux problématiques quant à ces mesures. La première problématique invoque le principe de

correspondance, point qui sera abordé plus loin dans ce chapitre, ainsi qu'au Chapitre 4. La seconde rejoint une problématique plus générale de la mesure des attitudes : l'impact des processus automatiques et des processus non automatiques au sein de la mesure (e.g., De Houwer & Moors, 2010). Ces critères d'automaticité, nous n'en avons pas parlé jusqu'à présent. Pourtant, comme nous allons l'illustrer maintenant, ceux-ci ont dicté et dictent encore l'évolution des mesures des attitudes raciales en psychologie.

1.2 La mesure des attitudes raciales

Comme nous allons le voir, les mesures des attitudes raciales se sont développées et ont évoluées en fonction des considérations théoriques associées. La psychologie cognitive a joué un rôle important au cours des années 80, permettant l'éclosion de mesures centrées sur les processus. Bien que notre intérêt théorique se porte avant tout sur les mesures associées aux stéréotypes, nous aborderons à la fois les mesures des stéréotypes et des préjugés, ces deux types de mesures ayant connu une trajectoire relativement similaire au cours du temps. Nous laisserons volontairement de côté les mesures centrées sur la discrimination, plus complexes à mettre en place et donc minoritaires dans l'étude des attitudes raciales en psychologie sociale.

1.2.1 Les mesures précoces : contenu de croyances et personnalité

En parallèle du développement théorique des notions de stéréotype et préjugé, les chercheurs se sont intéressés à la meilleure façon de les mesurer. Les premières mesures relatives aux stéréotypes se sont centrées sur leur contenu (Brigham, 1971a ; pour une revue de ces premières mesures). L'une des plus célèbres mesures du contenu des stéréotypes provient des travaux de Katz et Braly (1933). Ces auteurs demandèrent à des étudiants de Princeton de choisir parmi une liste d'adjectifs ceux caractérisant le mieux divers groupes cibles. Certains groupes ont été décrits de manière plutôt positive. Par exemple, 78% des participants ont décrit les Allemands comme ayant un esprit scientifique, 53% des participants ont décrit les Italiens comme artistiques, et 48% des participants ont décrit les Japonais comme intelligents. À l'inverse, d'autres groupes ont sans surprise été associés à des traits

négatifs. Les Noirs ont été associés par 75% des participants à l'adjectif « paresseux » et les Turques par 54% des participants à l'adjectif « cruels ». Afin d'observer l'évolution culturelle des stéréotypes au cours du temps, cette étude a été répliquée à deux reprises, là encore avec des étudiants de Princeton. Ainsi, Gilbert (1951), puis Karlins et collaborateurs (1969) ont mis en évidence une diminution progressive du pourcentage d'individus associés à ces traits. Nous reprendrons ce point par la suite.

En réponse aux divergences théoriques sur la définition même du stéréotype, d'autres mesures ont été développées. Brigham avait une vision du stéréotype comme croyance non pas culturelle, mais individuelle. Dans ses études, Brigham (1971b, 1973) a demandé à des participants de donner leur propre estimation du pourcentage des membres d'un groupe qui selon eux possédaient un trait particulier (e.g., pourcentage des Afro-américains qui étaient paresseux). Brigham (1973) a également demandé aux participants de donner l'estimation minimum et maximum acceptable qu'un autre individu pourrait avoir quant à ce pourcentage. Par le biais de ces deux mesures, ce chercheur voulait ainsi d'une part, déterminer les traits stéréotypiques d'un groupe (i.e., estimations au pourcentage élevé), et d'autre part séparer le contenu des croyances personnelles du consensus général (i.e., écart entre l'estimation individuelle et celle communément admise).

Enfin, McCauley et Stitt (1978) ont intégré à leur mesure l'aspect discriminant du stéréotype, c'est-à-dire l'aspect permettant de différencier un groupe d'un autre (e.g., les Afro-américains sont paresseux, mais pas les individus de couleur blanche). En se basant sur le théorème de Bayes, ces auteurs ont élaboré ce qu'ils appellent le « diagnostic ratio ». Ce diagnostic ratio est un indice calculé par rapport au pourcentage d'individus d'un groupe A possédant un trait divisé par le pourcentage d'individus d'un groupe B (ou de la population générale) possédant ce même trait. Si le diagnostic ratio est supérieur à 1, alors le trait est stéréotypique du groupe A. Par exemple, si 30% des individus afro-américains sont décrits comme paresseux, mais seulement 10% des individus blancs (diagnostic ratio de 3), le trait paresseux sera décrit comme stéréotypique des individus afro-américains.

Là où les premières mesures des stéréotypes se concentraient sur leur contenu, les premières mesures des préjugés se sont centrées sur la mise en évidence des groupes victimes d'évaluations négatives. Parmi ces mesures, l'une des plus connues est l'échelle de distance sociale de Bogardus (1947) qui se base sur une échelle cumulative de Guttman et qui permet de mesurer la proximité sociale que des individus accepteraient d'avoir avec des membres de divers groupes sociaux. Lambert (1952) a par exemple voulu estimer la proximité

que des participants français accepteraient d'avoir face à des individus allemands, juifs ou noirs. Les participants devaient répondre par oui ou par non quant à leur degré d'acceptation à des propositions allant de « complètement exclu de mon pays » à « comme époux, uni par les liens du mariage ». Entre autres, Lambert a mis en évidence que les participants plus jeunes avaient une plus faible distance sociale avec les Allemands que les participants plus âgés ou que les femmes avaient une distance sociale plus forte envers les Noirs que les hommes. Comme pour le stéréotype, ce type d'échelle a permis de suivre l'évolution au cours du temps de ce préjugé, Lambert ayant lui-même comparé ses résultats aux résultats de Borgadus de 1928 et ceux de Hartley de 1938 (Lambert, 1952).

Dans un contexte post-seconde guerre mondiale, un influant courant de pensée s'est intéressé à la meilleure compréhension des différences individuelles pouvant mener au racisme et à l'antisémitisme. Adorno et collaborateurs, dans leur livre *The Authoritarian Personality* (1950), ont ainsi tenté de mettre en lumière la personnalité propice à l'utilisation de préjugés. Ils ont ainsi développé plusieurs échelles de Likert, chacune d'entre elles ayant été modifiée et améliorée au cours du temps. Nous pouvons par exemple citer l'échelle d'antisémitisme (échelle A-S ; e.g., « Le problème lorsqu'on laisse entrer des Juifs dans un quartier agréable, c'est qu'ils lui donnent progressivement une atmosphère typiquement juive »), d'ethnocentrisme (échelle E ; e.g., « Si les Noirs vivent mal, c'est surtout parce qu'ils sont naturellement paresseux, ignorants et sans maîtrise de soi ») ou de fascisme (échelle F ; e.g., « l'obéissance et le respect envers l'autorité sont les vertus les plus importantes à enseigner aux enfants »). Adorno et collaborateurs, influencés par les théories psychodynamiques de l'époque, avançaient une explication de la personnalité autoritaire liée à l'enfance. Toutefois, ces auteurs ont aussi pu mettre en évidence le fait que ce type de personnalité n'était lié ni à l'intelligence ni à la religion, mais propice à l'intolérance, un faible raisonnement abstrait et à une surgénéralisation.

Les mesures citées jusqu'à présent ont en commun deux caractéristiques. La première est qu'elles ont toutes influencé les mesures actuelles des stéréotypes et préjugés. Encore aujourd'hui, nous retrouvons des mesures du contenu du stéréotype, comme par exemple des mesures utilisant le modèle SCM de Fiske et collaborateurs (2002) décrit plus haut (e.g., Rohmer & Louvet, 2018). Au niveau du préjugé, l'échelle de distance sociale de Bogardus a influencé des mesures de ressenti plus actuel, comme le *feeling thermometer* (e.g., Haddock et al., 1993). Enfin, les échelles de préjugé d'Adorno ont pour leur part contribué à la création

d'autres échelles grandement utilisées comme la *Right Wing Authoritarian* (RWA ; Altemeyer, 1988) ou la *Modern Racism Scale* (McConahay, 1986).

La seconde caractéristique de ces mesures est qu'elles peuvent toutes être regroupées sous le terme de mesures directes. Effectivement, par le biais de ces mesures, le chercheur demande directement à l'individu ses croyances, ressentis ou comportements en rapport à l'objet d'intérêt (e.g., les Maghrébins). Les mesures directes ont deux caractéristiques (De Houwer & Moors, 2010). Le résultat de la mesure (i.e., la réponse) est issu d'un auto-report de la part du participant (1) et cet auto-report concerne directement l'attribue cible de la mesure (2). Bien qu'il existe des exceptions, les mesures directes consistent la plupart du temps en des questionnaires ou échelles de mesure (e.g., échelle de Likert) et demandent une réponse sans limites de temps imparti. La grande majorité des études avant les années 80 utilisaient des mesures directes pour étudier les stéréotypes et préjugés présents dans la société. Toutefois, comme nous allons le voir maintenant, ces mesures ont reçu de nombreuses critiques (e.g., Gawronski & Hahn, 2019).

1.2.2 Problèmes liés aux mesures directes : l'exemple du biais de désirabilité sociale

Dans les années 70, les chercheurs américains se trouvent face à un paradoxe. D'un côté, plusieurs sondages d'opinion montrent une diminution progressive des évaluations négatives des populations noires depuis le milieu du XXe siècle (e.g., Taylor et al., 1978). Pour certains auteurs, cela reflèterait une diminution des préjugés raciaux au sein de la société américaine (e.g., Poskocil, 1977). D'un autre côté, les actes racistes diminuent eux beaucoup plus lentement (voir Arkes & Tetlock, 2004 ; Gawronski & Creighton, 2013). Ainsi, d'autres auteurs remettent en doute l'affirmation d'une diminution des attitudes raciales négatives envers les Afro-américains (e.g., Crosby et al., 1980 ; McConahay, 1986 ; Sigall & Page, 1971). Pour ces auteurs, cette différence entre paroles et actes serait davantage le reflet d'une évolution des attitudes raciales. La société devenant de plus en plus égalitaire, le racisme deviendrait lui plus latent ; il deviendrait moins accepté socialement d'avoir une attitude négative envers les populations afro-américaines.

Crosby et collaborateurs (1980) ont effectué une revue de la littérature des différentes mesures comportementales, c'est-à-dire ici discriminatoires, pouvant laisser supposer une perpétuation des actes racistes aux États-Unis. Parmi d'autres mesures comportementales, il a

ainsi été montré qu'un individu était plus disposé à aider une personne au téléphone quand celle-ci était blanche plutôt que noire (Gaertner & Bickman, 1971). De même, un individu allait davantage renvoyer une lettre égarée quand l'expéditeur était blanc que lorsqu'il était noir (Benson et al., 1976). Enfin, un participant se faisant passer pour un recruteur lors d'un entretien d'embauche fictif allait avoir davantage de comportements de proximité quand le recruté potentiel était blanc que lorsqu'il était noir (Word et al., 1974). En somme, Crosby et collaborateurs ont conclu de cette revue de littérature que les données effectives à disposition ne reflétaient pas les sondages réalisés à cette époque. En d'autres termes, il y avait effectivement une différence entre l'opinion auto-rapportée des Américains, issue des mesures directes, et leur comportement effectif.

Pour les auteurs de cette époque et encore aujourd'hui, il est reproché aux mesures directes d'être soumises à divers biais d'auto-présentation, et particulièrement au biais de désirabilité sociale (e.g., Crosby et al., 1980 ; Sigall & Page, 1971). Le biais de désirabilité sociale consiste en la tendance pour l'individu à répondre de manière socialement acceptable ou désirable, au détriment de ses croyances ou ressenti personnel dans le but de se faire voir sous un jour favorable (Crowne & Marlowe, 1960). L'une des expériences les plus connues mettant en évidence l'existence de ce biais est l'expérience de Sigall et Page (1971), celle-ci reprenant les différentes réplications de l'étude princeps de Katz et Braly (1933).

Comme nous l'avons décrit plus haut, les réplications successives de l'étude de Katz et Braly (1933) ont montré une diminution importante des traits négatifs associés à certains groupes habituellement discriminés (Gilbert, 1951 ; Karlins et al., 1969). Pour reprendre notre exemple précédent, là où en 1933, 75% des participants associaient les Noirs au trait « paresseux », ils ne sont plus que 31% en 1951 et seulement 26% en 1969. À l'inverse, un trait positif, comme le trait « musical », a vu son association augmenter, passant de 26% en 1933, à 33% en 1951 et enfin 47% en 1969. Sigall et Page (1971) ont voulu tester expérimentalement si, comme ils en faisaient l'hypothèse, cette diminution du stéréotype au cours du temps venait plus de l'impact d'un biais de désirabilité sociale plutôt que d'une réelle diminution du stéréotype dans la société. Leur manipulation expérimentale consistait à faire croire à la moitié des participants qu'ils disposaient d'un appareillage permettant de détecter leur attitude réelle, poussant donc ces derniers à répondre le plus honnêtement possible (et ainsi, supprimer le biais de désirabilité sociale). Les résultats sont allés dans le sens de l'hypothèse des auteurs : les individus noirs ont été plus négativement stéréotypés en

condition expérimentale qu'en condition contrôle. À l'inverse, les individus blancs ont été plus positivement stéréotypés en condition expérimentale qu'en condition contrôle.

D'une manière générale, cette étude montre l'une des limites des mesures directes : ces mesures sont soumises à un contrôle de la part du participant et peuvent ne pas refléter le contenu de croyances ou affectif réactivé (e.g., Fazio et al., 1986, 1995). Nous venons de voir que les participants pouvaient contrôler leur réponse pour se conformer à leur environnement social. Toutefois, d'autres études ont montré que les participants pouvaient également contrôler leurs réponses pour que celles-ci correspondent plus fortement à leurs croyances personnelles (Devine, 1989). Globalement, nous voyons que des processus intermédiaires peuvent jouer un rôle dans la réponse finale des participants. En cela, les années 80 ont été un tournant dans l'étude des attitudes raciales. Les chercheurs en cognition sociale sont ainsi passés de l'étude d'un contenu en mémoire ou d'un type de personnalité, à l'étude des processus pouvant aboutir aux évaluations et jugements individuels (Dovidio et al., 2010 ; Schneider, 2004). En réalité, ce passage de l'étude du contenu à l'étude des processus a coïncidé avec la révolution cognitive et l'intérêt croissant pour un type de processus particulier : les processus automatiques.

1.2.3 L'apport de la psychologie cognitive : les critères d'automaticité

Nous venons d'exposer l'une des critiques les plus fréquentes quant aux mesures directes. Il nous semble toutefois important de mentionner qu'en parallèle, ces mesures vont être critiquées pour d'autres aspects, et notamment pour leur incapacité à capter croyances et évaluations prenant racine en dehors d'un traitement conscient (Greenwald & Banaji, 1995). En somme, de nombreux reproches vont émerger quant à l'incapacité qu'ont les mesures directes à capter des processus pouvant être qualifiés d'automatiques.

Bien que l'on puisse remonter au début de la psychologie pour trouver trace des premières mentions de ce type de processus (e.g., James, 1890), de nombreux auteurs ont souligné le fait que l'intérêt réel pour les processus automatiques a émergé avec l'essor de la psychologie cognitive (e.g., Gawronski & Creighton, 2013 ; Monteith et al., 2013 ; Payne & Gawronski, 2010). Les études sur l'attention sélective et la mémoire à court terme ont ainsi permis de suggérer une dichotomie entre processus qualifiés d'automatiques et processus qualifiés de contrôlés (Neely, 1977 ; Posner & Snyder, 2004 ; Shiffrin & Schneider, 1984 ;

Treisman, 1969). Les auteurs de ces études définissent les processus automatiques comme rapides, parallèles, difficiles à supprimer volontairement, demandant peu d'attention, peu de ressources et illimités en capacité. À l'inverse, les processus contrôlés sont eux définis comme étant lents, sériels, initiés volontairement, demandant de l'attention et des ressources, et limités en capacités (voir aussi Payne & Gawronski, 2010). En compléments de ces premiers critères d'automaticité, les travaux sur la mémoire implicite (e.g., Jacoby, 1991) ont pour leur part mis en évidence la présence de processus prenant forme en dehors du champ de la conscience et ont de ce fait permis une distinction entre processus conscients et processus non conscients (e.g., Greenwald & Banaji, 1995).

De nos jours, les critères d'automaticités recueillant le plus large consensus sont aux nombres de quatre (i.e., les quatre cavaliers de l'automaticité ; Bargh, 1994). Ainsi, un processus sera qualifié d'automatique s'il est non conscient, non intentionnel, incontrôlable et/ou efficient (Bargh, 1994 ; Melnikoff & Bargh, 2018). Un processus sera qualifié de non conscient d'une part si l'individu n'a pas conscience de la présence du stimulus, mais aussi de l'effet de ce stimulus sur son jugement ou bien de la relation de cause à effet entre stimulus et jugement. Par exemple, amorcer de manière subliminale (i.e., en dessous du seuil de la conscience) des mots associés au terrorisme orienterait notre attention vers des visages maghrébins (Horry & Wright, 2009). Ensuite, un processus sera non intentionnel s'il n'est pas initié volontairement, que ce soit directement ou par un but externe. Le phénomène de catégorisation d'un visage, par exemple, interviendrait sans que l'individu n'ait l'intention de catégoriser les stimuli présentés (Ito & Tomelleri, 2017 ; Ito & Urland, 2005). De la même manière, un processus sera qualifié d'incontrôlable s'il ne peut être stoppé après avoir été initié ou bien si l'individu ne peut pas modifier ou altérer sa réponse. Comme nous le verrons par la suite, à l'inverse de son application, la réactivation d'un stéréotype est souvent décrite comme incontrôlable (e.g., Krieglmeyer & Sherman, 2012). Enfin, un processus sera efficient s'il ne consomme pas ou très peu de ressources cognitives. Par exemple, la catégorisation sociale serait un processus ne nécessitant que très peu de ressources, libérant celles-ci pour effectuer une double tâche (Macrae et al., 1994).

Bien que ces critères d'automaticité soient aujourd'hui largement repris dans la littérature (pour une revue de la littérature, voir Moors, 2016), il nous parait important de prendre quelques instants pour mentionner les nombreuses critiques existantes à leur sujet. Tout d'abord, comme les travaux de psychologie cognitive ont pu le montrer, la distinction entre ce qui constitue ou non un critère d'automaticité peut être questionnable (e.g.,

existence ou non d'un critère de rapidité ; Moors, 2016 ; Moors & De Houwer, 2006). De plus, à l'intérieur même d'un critère d'automaticité, il peut exister des divergences sur la manière dont celui-ci peut opérer (Melnikoff & Bargh, 2018). En conséquence, la présence ou non d'un critère d'automaticité ne serait au final que très rarement binaire (e.g., efficient vs. non efficient), mais plus souvent graduelle (e.g., processus demandant plus ou moins de ressources ; Melnikoff & Bargh, 2018 ; Moors, 2016).

La critique la plus importante relative aux critères d'automaticités concerne le problème de leur alignement (Corneille & Hütter, 2020 ; Melnikoff & Bargh, 2018 ; Moors, 2016 ; Moors & De Houwer, 2006). Pour un même processus, et en admettant une dichotomie présence/absence de chaque critère, nous avons 16 possibilités de combinaisons possibles entre ces différents critères. Par exemple, un processus peut potentiellement être conscient et contrôlable (i.e., non automatique), mais non intentionnel et efficient (i.e., automatique). Poussant le raisonnement plus loin, un processus peut donc être décrit comme automatique, quand bien même il ne recouvre qu'une seule composante d'automaticité. En réalité, la probabilité qu'un alignement parfait des critères d'automaticité soit la norme, c'est-à-dire la probabilité que l'ensemble des processus soient entièrement automatiques ou entièrement non automatiques n'a jamais été testée empiriquement et serait difficilement testable (Melnikoff & Bargh, 2018). En définitive et pour bon nombre d'auteurs, très peu de processus seraient totalement automatiques ou totalement non automatiques (e.g., Gawronski et al., 2014 ; Ric & Muller, 2017). C'est la raison pour laquelle, lorsque nous aborderons les critères d'automaticité, nous parlerons de processus davantage automatiques ou davantage non automatiques (Ric & Muller, 2017) ou tâcherons de définir au mieux à quel(s) critère(s) précisément nous faisons référence (Corneille & Hütter, 2020 ; Moors & De Houwer, 2006).

Malgré les critiques précoces sur l'alignement des différents critères d'automaticité (Bargh, 1994), l'intérêt pour les processus automatiques s'est accompagné d'un important développement théorique relatif aux modes de traitement de l'information au sein de la littérature en cognition sociale en général, et celle relative aux attitudes raciales en particulier.

1.2.4 Les modèles duaux du traitement de l'information

À la suite de l'attrait grandissant en cognition sociale pour l'étude des processus mentaux, les chercheurs ont tenté de modéliser la manière dont ceux-ci pouvaient aboutir aux jugements, évaluations et comportements individuels. La littérature s'est principalement développée autour d'un type de modèles bien particulier, appelés modèles duaux (pour une revue de la littérature voir Evans, 2008 ; Gawronski & Creighton, 2013). Comme leur nom l'indique, l'idée principale des modèles duaux est que les processus mentaux à l'origine des jugements, évaluations et comportements individuels peuvent être divisés en deux catégories distinctes (Gawronski et al., 2014). De nombreux labels ont été utilisés pour nommer ces catégories, celles-ci pouvant entre autres être décrites comme implicite et explicite, automatique et contrôlé, non consciente et consciente, ou bien heuristique et analytique (Evans & Stanovich, 2013). Nous préférerons un label plus neutre, Type 1 vs. Type 2, ceci nous permettant de rester agnostique quant au type de processus ou critères d'automaticités inclus dans chaque catégorie (e.g., Evans & Stanovich, 2013 ; Kahneman, 2011). En réalité, ce nombre important de labels reflète avant tout une forte hétérogénéité quant aux processus pouvant être inclus au sein de ces catégories. En effet, les processus de Type 1 peuvent régulièrement être décrits comme inconscients, rapides, automatiques, associatifs ou heuristiques alors que les processus de Type 2 peuvent être eux décrits comme conscients, lents, contrôlables, propositionnels et délibératifs (e.g., Evans, 2008). D'une manière globale, la principale distinction entre les processus de Type 1 et de Type 2 se situe au niveau de leur degré d'automaticité, les processus de Type 1 recouvrant les processus davantage automatiques et ceux de Type 2 les processus davantage non automatiques (Gawronski et al., 2014). Il est toutefois à signaler que, bien que la littérature distingue ces deux types de processus par le biais de leurs conditions opérantes, c'est-à-dire les conditions selon lesquels un processus opère (i.e., de manière automatique ou non), le but de ces modèles est avant tout de décrire les principes opérants de ces processus, c'est-à-dire les mécanismes mentaux permettant d'aboutir à la réponse individuelle (Gawronski et al., 2014 ; Sherman et al., 2008).

La littérature traitant des modèles duaux recouvre de nombreux domaines de la psychologie sociale. Certains des modèles précurseurs étaient spécifiques à un domaine donné, comme les modèles s'intéressant au champ de la persuasion (e.g., Chaiken, 1987 ; Petty & Cacioppo, 1986) ou de l'attribution (e.g., Trope, 1986). D'autres modèles actuels se veulent plus généraux et intégratifs, et sont ainsi construits indépendamment d'un domaine

spécifique (e.g., Gawronski & Bodenhausen, 2006 ; Smith & DeCoster, 2000 ; Strack & Deutsch, 2004). Dans les paragraphes qui vont suivre, nous allons nous intéresser à certains des modèles pouvant être appliqués directement (i.e., modèles spécifiques) ou indirectement (i.e., modèles généraux) au champ des attitudes raciales.

Comme nous l'avons vu précédemment, Devine (1989) différenciait stéréotype et préjugé par leur aspect culturel ou personnel. Dans son modèle de la dissociation, Devine a théorisé la façon dont nos croyances en la véracité d'un stéréotype pouvaient moduler notre attitude par rapport aux membres d'un groupe cible. Pour Devine, nos stéréotypes culturels et croyances personnelles proviendraient de deux structures cognitives différentes. Les stéréotypes seraient des connaissances acquises depuis la petite enfance par le biais de notre environnement culturel. Ces croyances ayant été développées très précocement dans le développement, elles auraient été intériorisées et renforcées au cours du temps. À l'inverse, les croyances personnelles (i.e., croyances sur la véracité ou non du stéréotype) seraient plus récentes, moins bien ancrées en mémoire et de ce fait, moins accessibles que le stéréotype culturel. Par conséquent, face à une personne de couleur noire, quel que soit notre niveau de préjugé (nos croyances personnelles), notre stéréotype s'activerait automatiquement (i.e., de manière non intentionnelle). En revanche, nos croyances personnelles joueraient un rôle dans la manière dont nous exercerions un contrôle par rapport à ce stéréotype. En effet, seuls les individus n'ayant pas ou peu de préjugés observeraient un conflit entre le stéréotype culturel réactivé et leurs croyances personnelles. Par le biais de processus contrôlés, ces individus inhiberaient le stéréotype réactivé au profit de leurs croyances personnelles plus égalitaires (Devine, 1989).

Nous voyons ici que le modèle de la dissociation fait des prédictions opposées à ce que nous avons vu lorsque nous avons traité du biais de désirabilité sociale. En effet, Devine fait la prédiction que notre stéréotype réactivé automatiquement serait issu d'une influence culturelle et pourrait être inhibé par nos croyances personnelles. À l'inverse, comme nous l'avons vu précédemment, certains auteurs faisaient l'hypothèse que la réactivation de nos connaissances personnelles du stéréotype pouvait être inhibée par la pression culturelle environnante (e.g., Crosby et al., 1980 ; Sigall & Page, 1971). C'est ce second type de prédiction qu'a repris Fazio (1990 ; Fazio & Olson, 2014) dans son modèle *Motivation and Opportunity as DEterminants* (MODE). Pour Fazio, notre attitude (à prendre ici comme association mentale entre un objet et l'évaluation de cet objet) pourrait avoir un impact direct ou indirect sur notre comportement. Si cette association mentale est suffisamment

importante (e.g., évaluation négative des individus noirs), celle-ci pourrait être réactivée de manière automatique en présence de l'objet d'attitude. Pour Fazio, si nous n'avons pas la motivation ou l'opportunité (i.e., les ressources) nécessaires, notre comportement serait directement guidé par l'attitude préalablement réactivée. Néanmoins, si nous avons la motivation et l'opportunité suffisantes, nous pourrions nous engager dans un traitement plus approfondi (i.e., délibéré) de la situation. En conséquence, l'impact de l'attitude réactivée sur notre comportement pourrait être modulé par des caractéristiques propres à la situation, comme l'influence des normes sociales en vigueur (e.g., biais de désirabilité sociale).

Les modèles duaux traitant de la formation d'impression peuvent également être pertinents lorsque l'on s'intéresse aux attitudes raciales (e.g., Brewer, 1988 ; Fiske & Neuberg, 1990). Ces modèles proposent que l'impression que l'on se forme sur autrui provienne en priorité et de manière automatique d'informations saillantes (i.e., processus d'identification ou de catégorisation). Ce ne serait que dans un second temps que nous procéderions de manière volontaire à une analyse plus détaillée (i.e., personnalisation ou individualisation). Pour le modèle du continuum, la manière dont nous nous formons une impression sur autrui prendrait la forme d'un continuum (Fiske & Neuberg, 1990)[1]. À l'un des pôles du continuum, l'impression que nous avons d'autrui serait basée sur ces informations catégorielles saillantes (e.g., âge, couleur de peau). Cette première impression serait le fruit de processus automatiques (i.e., efficients et non intentionnelles). Dans un second temps, si nous avons là encore la motivation et les ressources nécessaires, cette impression initiale pourrait être affinée par l'ajout d'attributs individuels (i.e., propre à la cible). Si ces attributs confirment la catégorisation initiale, celle-ci prédominera. Si à l'inverse ces attributs sont contraires à la catégorisation initiale et si une recatégorisation n'est pas possible, l'impression finale se rapprochera du pôle opposé du continuum, à savoir basée sur des informations individuelles.

Les modèles décrits ci-dessus étaient spécifiques à des domaines précis. Néanmoins, comme nous l'avons spécifié plus haut, à partir des années 2000 certains auteurs ont proposé des modèles duaux plus généraux, mais pouvant également être appliqués au champ des attitudes raciales. L'un des modèles, si ce n'est le modèle le plus influent aujourd'hui, est *l'Associative-Propositional Evaluation model* (APE ; Gawronski & Bodenhausen, 2006, 2007,

[1] Ce modèle décrivant un continuum entre processus plutôt que deux grands types de processus, il pourrait ne pas être inclue parmi les modèles duaux. Toutefois, les deux grands pôles du continuum correspondent aux processus de Type 1 et de Type 2 vus précédemment. C'est pourquoi, il nous semble pertinent de l'inclure dans le cadre des modèles duaux.

2011, 2014, 2018). Comme d'autres modèles similaires (e.g., *Reflective-Impulsive model* ; Deutsch & Strack, 2006 ; Strack & Deutsch, 2004), le modèle APE distingue deux types de processus : un processus associatif (i.e., associations en mémoire entre plusieurs concepts reliés sémantiquement) et un processus propositionnel (i.e., propositions décrivant la relation particulière entre plusieurs concepts). Le processus associatif reflèterait un ensemble d'associations mentales entre concepts formées par le biais de simples cooccurrences spatio-temporelles. Par exemple, un individu voyant régulièrement dans les médias des individus d'origine maghrébine décrits comme menaçants associera en mémoire le concept « Maghrébins » et le concept « menace ». Plus la cooccurrence entre ces concepts sera importante, plus l'association sera forte. Le processus propositionnel reflèterait pour sa part un ensemble de propositions, stockées elles aussi sous forme associatives, préalablement validées et sensibles à la relation entre concepts (Gawronski & Bodenhausen, 2018). Par exemple, un individu sensible à la cause anti-raciale pourra stocker en mémoire une proposition du type « les Maghrébins ne sont pas menaçants ». Selon Gawronski et Bodenhausen, la réactivation d'un concept entrainerait automatiquement la réactivation des concepts reliés dans le réseau associatif. Ainsi, un individu confronté à un stimulus caractéristique du groupe des Maghrébins (e.g., visage prototypique maghrébin) réactivera automatiquement (i.e., de manière non intentionnelle) en mémoire le concept Maghrébin et les concepts reliés (e.g., « menace »). L'ensemble des concepts activés aboutiront à une réaction affective globale associée au stimulus (appelée par les auteurs « attitude implicite »), indépendante de tout critère de vérité. Le processus propositionnel, lui, aurait pour rôle de valider ou non le contenu réactivé par le système associatif. Ainsi, la réaction affective générée par le système associatif va être transformée en proposition (e.g., « je n'aime pas les Maghrébins ») et sera comparée à d'autres propositions également activées (e.g., « il est mal de discriminer les Maghrébins », « j'aime les Maghrébins »). Si cette proposition est compatible avec les autres propositions réactivées, elle pourra être acceptée comme base de jugement (appelée par les auteurs « attitude explicite »). Si au contraire la proposition est incompatible avec les autres propositions réactivées (comme dans notre exemple), le système propositionnel aura la charge de réduire cette dissonance cognitive. Une des solutions pourra être de rejeter la proposition issue de la réaction affective réactivée (e.g., « j'aime les Maghrébins ») ou bien de rejeter les autres propositions réactivées (e.g., « les Maghrébins ne sont pas un groupe discriminé », « je ne suis pas sûr d'aimer les Maghrébins »).

L'ensemble des modèles que nous venons de décrire suivent au final la même logique. Les processus de Type 1, de par certaines de leurs composantes d'automaticité, seront

réactivés prioritairement (e.g., associations mentales, catégorisation, stéréotype culturel). Si l'individu n'a pas la motivation nécessaire et les ressources suffisantes ou bien s'il est en accord avec le contenu réactivé, celui-ci va être validé et utilisé par l'individu dans son jugement, son évaluation ou son comportement. À l'inverse, si l'individu dispose de la motivation nécessaire, de ressources suffisantes et s'il n'est pas en accord avec ce contenu, la réponse de l'individu reflètera davantage les processus de Type 2. Les mesures classiques des attitudes raciales étant décrites comme reflétant majoritairement des processus de Type 2 (e.g., Sigall & Page, 1971), de nouvelles mesures ont dû être créées. Ainsi, afin de capter plus directement le contenu automatiquement réactivé (i.e., processus de Type 1), des mesures indirectes ont été proposées.

1.2.5 Mesures indirectes des attitudes raciales

Nous l'avons vu précédemment, une mesure peut être qualifiée de directe si le résultat de la mesure est issu d'un auto-report et que cet auto-report concerne l'objet cible de la mesure (De Houwer & Moors, 2010). À l'inverse, une mesure indirecte, comme son nom l'indique, aura pour but d'inférer les croyances, ressentis ou comportements individuels de manière détournée. Ainsi, une mesure peut être définie comme indirecte quand (1) la réponse de l'individu n'est pas basée sur un auto-report (e.g., performance dans une tâche) ou bien (2) quand celle-ci est basée sur un auto-report mais relatif à un attribut autre que l'objet d'intérêt (e.g., jugement d'une cible précédée par l'amorce d'intérêt ; Gawronski & De Houwer, 2014 ; Gawronski & Hahn, 2019). Bien que des exceptions existent (e.g., IAT Papier-Crayon ; Lemm et al., 2008), les mesures indirectes prennent la plupart du temps la forme de tâches informatisées (Correll et al., 2010). Développées dans les années 80 et s'inspirant des tâches existantes en psychologie cognitive (e.g., Neely, 1977), leur popularité n'a cessé de s'accroitre depuis. Cet attrait se retrouve dans l'étude des attitudes raciales, les mesures indirectes y étant grandement utilisées, car censées entre autres contourner l'influence des biais d'auto-présentation (e.g., Devine, 1989 ; Fazio & Olson, 2014 ; Gawronski & Bodenhausen, 2006 ; Greenwald et al., 1998). Nous avons choisi de décrire certaines de ces mesures (e.g., AMP ; Payne et al., 2005), car structurellement proches de la mesure principale utilisée dans nos travaux (i.e., SMT) ou parce que nous avons directement utilisé cette mesure dans certains de nos travaux (i.e., IAT ; Greenwald et al., 1998). De ce fait, nous laisserons volontairement de côté d'autres mesures pourtant pertinentes dans l'étude des attitudes raciales (e.g., mesures

d'approche-évitement, Rougier et al., 2018 ; biais du tireur, Mange et al., 2012), car décrites plus loin dans le manuscrit ou moins utiles pour la suite de notre exposé.

La mesure indirecte la plus connue à ce jour est l'IAT (*Implicit Association Task* ; Greenwald et al., 1998 ; Nosek et al., 2002, 2007). L'IAT, dans sa version classique, est censée refléter la préférence des individus pour un groupe (e.g., les personnes de couleur noire) comparativement à un autre groupe (e.g., les personnes de couleur blanche ; pour une interprétation alternative, voir Rothermund & Wentura, 2004). Dans le cas des attitudes raciales, elle constitue donc une mesure de préjugés. Le participant va avoir pour tâche de catégoriser le plus rapidement possible deux types de stimuli apparaissant à l'écran. D'une part, il doit catégoriser des mots positifs (Bons) ou négatifs (Mauvais) selon leur valence. Une clé de réponse (i.e., une touche du clavier) est associée aux mots positifs, l'autre clé de réponse étant associée aux mots négatifs. D'autre part, le participant doit catégoriser des photos d'individus blancs ou noirs selon leur origine ethnique. Une clé de réponse est associée à la catégorie des individus de couleur blanche, l'autre clé de réponse étant associée aux individus de couleur noire. L'IAT est divisée en plusieurs blocs (la plupart du temps 5 ou 7 blocs ; Greenwald et al., 2003). Dans les blocs dits d'entrainement, le participant doit catégoriser un seul type de stimuli à la fois (i.e., mots ou visages). À l'inverse, dans les blocs dits critiques le participant doit effectuer, au sein du même bloc, les deux types de catégorisation en utilisant les mêmes clés de réponse. Dans le bloc compatible, les mots positifs et visages blancs partagent une clé de réponse, les mots négatifs et visages noirs partageant l'autre clé de réponse. Dans le bloc incompatible, les mots positifs et visages noirs sont associés à une même clé de réponse, les mots négatifs et visages blancs étant associés à l'autre clé de réponse. L'idée sous-jacente de l'IAT est que le participant devrait être plus rapide quand deux concepts associés en mémoire partagent une même clé de réponse (i.e., bloc compatible) que l'inverse (i.e., bloc incompatible).

De nombreuses variantes de l'IAT existent (pour une revue de la littérature, voir (Gawronski & De Houwer, 2014 ; Gawronski & Hahn, 2019). Les modifications par rapport à la tâche originale peuvent être minimes, des versions de l'IAT pouvant par exemple comporter une unique catégorie, c'est-à-dire uniquement la catégorie des Noirs ou la catégorie des Blancs (SC-IAT ; Karpinski & Steinman, 2006). D'autres mesures s'inspirent de la structure de l'IAT, mais l'appliquent à d'autres types de tâches, comme une tâche de Stroop (*Extrinsic Affective Simon Task*, EAST ; De Houwer, 2003) ou une tâche de *Go/No-Go* (*The Go/No-Go Association Task*, GNAT ; Nosek & Banaji, 2001). Dans cette dernière, le participant doit aller

appuyer sur un bouton (*go*) à l'apparition de certains stimuli (e.g., visage noir, mot négatif) et au contraire ne pas appuyer sur ce bouton (*no go*) en à l'apparition d'autres stimuli (e.g., visage blanc, mot positif). Comme pour la version classique de l'IAT, la GNAT comprend un type de bloc compatible (e.g., appuyer sur le bouton à l'apparition d'un visage noir ou d'un mot négatif) ainsi qu'un type de bloc incompatible (e.g., appuyer sur le bouton à l'apparition d'un visage noir ou d'un mot positif). Toutefois, et à la différence de l'IAT, les scores sont calculés à partir du taux d'erreur, nécessitant dès lors la présence d'une fenêtre de réponse.

L'IAT peut aussi être adapté pour capter non plus l'aspect évaluatif, mais le contenu sémantique de l'attitude et peut donc constituer une mesure du stéréotype. Dans des tâches de ce type, les mots positifs ou négatifs sont simplement remplacés par des mots constitutifs du stéréotype des catégories étudiées. Par exemple, Amodio et Devine (2006) ont utilisé des mots reflétant soit le sport (attribut stéréotypique des individus de couleur noire), soit l'intelligence (attribut stéréotypique des individus de couleur blanche). Comme pour l'IAT classique, l'idée derrière ce type d'IAT est que les individus seront plus rapides dans des blocs où les attributs stéréotypiques des catégories partagent une même clé de réponse avec les visages de cette catégorie (e.g., mots liés au sport et visages noirs) que l'inverse (e.g., mots liés à l'intelligence et visages noirs).

Bien que l'IAT soit à ce jour l'une des mesures des attitudes raciales les plus populaires, d'autres mesures restent pertinentes. Un nombre important de ces mesures se base sur la procédure dite d'amorçage séquentiel (Wentura & Degner, 2010). Dans ces paradigmes, le participant est exposé à deux types de stimuli présentés de façon succincte et l'un à la suite de l'autre. La tâche du participant est de répondre par rapport au second stimulus présenté (la cible), sans tenir compte du premier stimulus (l'amorce). L'intérêt de ce type de tâche est que, bien que le participant ne traite pas directement l'amorce, celle-ci va avoir une influence sur le jugement de la cible. Plus particulièrement, l'idée derrière ce type de tâche est que la réponse du participant devrait être facilitée quand la cible est congruente avec le contenu mental réactivé par l'amorce et au contraire inhibée quand la cible est non congruente avec le contenu mental réactivé.

Ce type de tâche a été utilisé pour mesurer le préjugé individuel. Par exemple, dans l'*Evaluative Priming Task* (EPT ; Fazio et al., 1986, 1995), le participant doit classer des mots positifs ou négatifs selon leur valence. Ces mots cibles sont précédés d'amorces relatives aux catégories sociales d'intérêts (e.g., visages noirs vs. blancs). Si le participant dispose d'une évaluation plus négative des personnes de couleur noire par rapport aux personnes de

couleur blanche, il va être plus rapide (ou faire moins d'erreurs) quand un mot négatif sera précédé d'un visage de couleur noire et/ou quand un mot positif sera précédé d'un visage de couleur blanche. À l'inverse de l'IAT, une amorce neutre peut également être incluse dans la tâche (Gawronski & De Houwer, 2014 ; Wittenbrink, 2007). Inclure une amorce neutre permet de s'affranchir de la comparaison d'une catégorie par rapport à l'autre, en comparant directement l'impact de chaque type d'amorce à l'impact d'une amorce neutre. Par exemple, un temps de réponse plus rapide pour détecter des mots cibles négatifs précédés d'un visage noir par rapport à une amorce neutre sera diagnostic d'une évaluation négative des individus noirs.

Les tâches d'amorçages séquentiels peuvent également être utilisées dans l'étude des stéréotypes (e.g., *Semantic Priming Procedure* ; Banaji & Hardin, 1996 ; *Lexical Decision Task* ; Wittenbrink et al., 1997). Dans la *Lexical Decision Task* (Wittenbrink et al., 1997), le participant va devoir juger si les cibles présentées sont des mots ou des non-mots. Comme pour l'EPT, ces cibles sont précédées d'amorces relatives aux catégories Noirs et Blancs, mais correspondent ici directement aux labels des catégories (i.e., « noir » et « blanc »). En plus des non-mots, les mots cibles sont soit neutres (e.g., table), soit font partie du stéréotype des Noirs (e.g., sportif) ou des Blancs (e.g., intelligent). Là encore, des amorces neutres peuvent être incluses pour s'affranchir de la comparaison entre les catégories. Si les participants ont effectivement des stéréotypes envers les Noirs ou les Blancs, ils devraient être plus rapides quand le mot cible (e.g., sportif) est congruent avec l'amorce (e.g., noir) que l'inverse (e.g., blanc).

Pour finir avec les tâches basées sur le paradigme d'amorçage séquentiel, il nous semble important de citer l'*Affective Misattribution Procedure* (AMP ; Payne et al., 2005), l'une des tâches les plus populaires à ce jour dans l'étude des préjugés. À l'inverse des autres tâches décrites précédemment, l'AMP ne s'appuie pas sur un principe de congruence ou d'incongruence entre le contenu mental réactivé par l'amorce et le jugement de la cible. Dans l'AMP, il n'y a pas de bonnes ou de mauvaises réponses. En effet, le participant a pour consigne de porter un jugement affectif (i.e., plaisant / déplaisant) sur un ensemble de cibles neutres (i.e., idéogrammes chinois). Néanmoins, comme pour les autres tâches d'amorçage, ces idéogrammes chinois sont précédés de stimuli relatifs aux catégories des Noirs ou des Blancs (e.g., visages noirs ou blancs). L'AMP se base sur le principe de fausse attribution affective (Payne et al., 2005, 2010), c'est-à-dire que, selon ce principe, la valence activée par un premier stimulus devrait être faussement attribuée au second. Plus particulièrement, l'idée

derrière cette tâche est que le participant va faussement attribuer l'affect activé par l'amorce à son ressenti évoqué par l'idéogramme cible. Le participant devrait donc trouver un idéogramme déplaisant quand celui-ci est précédé d'une amorce réactivant un affect négatif (e.g., visage noir). À l'inverse, il devrait trouver un idéogramme plaisant quand celui-ci est précédé d'une amorce réactivant un affect positif (e.g., visage blanc).

Les mesures indirectes, et notamment celles présentées ci-dessus, ont révolutionné le monde de la recherche concernant les attitudes raciales. Aujourd'hui, ces mesures sont devenues le principal moyen pour contourner l'impact de processus de Type 2 dans la réponse du participant. La référence aux « biais implicites » (i.e., biais résultants de processus automatiques) n'a jamais été aussi élevée (Gawronski, 2019 ; Mitchell & Tetlock, 2017). Pourtant, ces mesures ne sont pas immunisées contre d'éventuelles remises en question. En effet, comme nous avons pu le voir avec les mesures directes, les mesures indirectes ont été grandement critiquées ces dernières années.

1.2.6 **Limites des mesures indirectes**

La littérature s'étoffant peu à peu d'un ensemble de mesures indirectes, des chercheurs se sont intéressés à la corrélation que ces mesures pouvaient entretenir avec des mesures directes. Plusieurs méta-analyses ont mis en évidence une corrélation relativement faible entre ces deux types de mesures ($r < .25$; e.g., Cameron et al., 2012 ; Hofmann et al., 2005 ; mais voir Nosek & Hansen, 2008 ; Ratliff et al., 2020, pour des exemples de corrélations plus importantes). Doit-on s'attendre à une corrélation plus élevée ? Moins élevée ? Comme nous allons le voir maintenant, cette corrélation relativement faible entre mesures indirectes et directes a soulevé divers questionnements et tentatives de réponses aboutissant à certaines critiques des mesures indirectes.

Pour expliquer la faible corrélation entre mesures indirectes et directes, certaines raisons avancées sont purement méthodologiques. Certains auteurs ont ainsi mis en avant la faible consistance interne des mesures indirectes, cela ayant pour conséquences une sous-estimation de la corrélation réelle entre ces mesures. Effectivement, bien que certaines mesures indirectes puissent avoir une relativement bonne fiabilité (e.g., AMP, IAT), d'autres sont souvent critiquées pour leur faible fiabilité (e.g., EPT ; Gawronski & De Houwer, 2014 ; Gawronski & Hahn, 2019). En outre, comme nous l'avons évoqué plus haut et comme nous en

parlerons au Chapitre 4, cette faible corrélation peut également s'expliquer par le principe de correspondance entre les mesures (Ajzen & Fishbein, 1977), largement ignoré dans le champ des attitudes utilisant des mesures indirectes (Gawronski, 2019). Pour résumer très succinctement l'idée derrière ce principe et pour prendre l'exemple d'une correspondance quant au contenu de la mesure, une mesure directe relative au groupe dans son ensemble devrait être plus fortement corrélée à une mesure indirecte elle-même relative au groupe (i.e., mesure centrée sur la catégorie) qu'à une mesure indirecte relative aux membres de ce groupe (i.e., mesure centrée sur les exemplaires de cette catégorie).

Néanmoins, les mesures indirectes et directes étant censées capter des types de processus différents (Type 1 vs. Type 2), il pourrait sembler logique que la corrélation entre ces mesures soit relativement faible. Pour certains auteurs, cette faible corrélation serait même la preuve que ces mesures captent des processus distincts (e.g., Fazio & Olson, 2014). Tout d'abord, il a souvent été avancé que les mesures indirectes permettraient de refléter des processus non conscients, processus que les mesures directes ne permettraient pas de capter (e.g., Greenwald & Banaji, 1995 ; Rudman et al., 1999). La corrélation entre mesures indirectes et directes serait faible, car le contenu capté par les mesures indirectes ne pourrait pas être consciemment rapporté par le participant répondant à des mesures directes (Nosek & Hansen, 2008 ; Nosek et al., 2007). Cependant, bien que la structure des mesures indirectes n'impose pas au participant d'avoir connaissance de ses attitudes, cela ne signifie pas pour autant que celui-ci n'y a pas accès (Gawronski, 2019 ; Hahn & Gawronski, 2019). Des études empiriques récentes viennent ainsi remettre en cause l'aspect inconscient de l'attitude au sein des mesures indirectes (Hahn & Gawronski, 2019 ; Hahn et al., 2014). Dans leur étude, Hahn et collaborateurs (2014) ont demandé à des participants de prédire leur score à plusieurs IAT relatifs à différents groupes (Noirs, Asiatiques, Hispaniques, Célébrités, Enfants). Par la suite, les participants ont passé les IAT dont ils avaient prédit leur score. Les auteurs ont montré que les participants prédisaient correctement leurs scores relatifs aux différents IAT (corrélation médiane de .65), remettant effectivement en cause le postulat voulant que les mesures indirectes soient imperméables aux processus conscients.

Dans la même lignée, et comme nous avons pu le voir auparavant, certains auteurs avancent que les mesures indirectes captent le contenu mental réactivé automatiquement (processus de Type 1), indépendamment d'un quelconque contrôle ou d'une quelconque validation de la part du participant (processus de Type 2 ; Devine, 1989 ; Fazio, 1990 ; Gawronski & Bodenhausen, 2006). Pour ces auteurs et particulièrement dans le champ des

attitudes raciales, la réponse issue des mesures indirectes aurait l'avantage de ne pas être contaminée ou modifiée volontairement par des influences externes (e.g., biais de désirabilité sociale) ou internes (e.g., validation ou non du contenu réactivé). Pourtant, plusieurs études ont montré que les participants pouvaient modifier volontairement leurs réponses au sein même de différentes mesures indirectes (e.g., Fiedler & Bluemke, 2005 ; Langner et al., 2010 ; Lowery et al., 2001). Au cours de trois expériences et utilisant un IAT Allemand-Turc, Fiedler et Bluemke (2005) ont demandé à des participants allemands de répondre de telle manière à ce qu'un biais anti-Turcs n'apparaisse pas. Ces auteurs ont ainsi montré que des participants familiers avec la procédure IAT pouvaient contrôler leurs réponses pour faire disparaitre le biais anti-Turcs qu'ils avaient eu précédemment. Plus précisément, ces participants n'avaient qu'à ralentir dans les blocs compatibles afin de faire disparaitre la différence de temps de réponse entre les blocs incompatibles et compatibles. Plus intéressants encore, des experts ont été incapables de détecter si ces participants avaient modifié volontairement leurs réponses ou non. Bien que ces résultats concernent uniquement des participants à qui il a été demandé explicitement de contrôler leur réponse, cette étude constitue l'une des preuves empiriques permettant la remise en cause de l'association classiquement faite entre mesure indirecte et processus non contrôlables.

Pour résumer, les réponses issues des mesures indirectes, que l'on pensait être le reflet de processus de Type 1, peuvent au moins en partie refléter des processus de Type 2 (e.g., contrôle de la part du participant). Ce point illustre une réflexion plus générale quant aux mesures indirectes et directes. Les mesures indirectes comme les mesures directes ne sont pas *process pure* (e.g., Hütter & Klauer, 2016 ; Jacoby, 1991 ; Klauer et al., 2011 ; Sherman et al., 2008). Ces deux types de mesures sont le reflet d'une combinaison entre différents processus et notamment des processus pouvant être décrits comme davantage de Type 1 et des processus pouvant être décrits comme davantage de Type 2. En d'autres termes, ces deux types de mesures sont sensibles à des processus automatiques et non automatiques. Cela constitue une première illustration du risque que nous pouvons avoir à faire le raccourci entre un type de mesure (i.e., indirecte vs. directe) et un type de processus (i.e., automatique vs. non automatique). Mais il existe une autre caractéristique de ces mesures qu'il nous faut prendre en compte.

Plus haut, nous avons évoqué le principe de correspondance entre les mesures et avons donné l'exemple d'une correspondance relative au contenu de ces mesures (Ajzen & Fishbein, 1977). Toutefois, ce principe de correspondance peut également s'appliquer au

niveau de leur structure (Payne et al., 2008). Effectivement, il ne faut pas négliger la possibilité que les réponses des participants au sein de ces mesures puissent refléter non pas le type de processus étudié, mais des caractéristiques différentes de ces mesures (e.g., Hütter & Klauer, 2016 ; Payne et al., 2008). Payne et collaborateurs (2008) ont ainsi montré que lorsque la structure d'une mesure directe d'attitudes raciales se rapprochait de la structure d'une mesure indirecte (i.e., AMP), la corrélation entre ces deux types de mesures augmentait.

Afin d'illustrer la différence de structure pouvant exister entre ces mesures, attardons-nous sur deux de ces mesures. Tout d'abord, prenons une mesure directe d'attitude raciale que nous développerons par la suite, l'échelle de Préjugés Généralisés (Dambrun & Guimond, 2001). Dans cette échelle, le participant doit donner son degré d'accord à plusieurs propositions parmi lesquelles « Il est insensé d'attribuer aux Algériens et aux Marocains les problèmes économiques de la France » ou « Les Français devraient être prioritaires en matière de logement ». Pour répondre à cette échelle, le participant ne dispose pas de limite de temps imparti. Il doit cependant interpréter à la fois l'échelle de réponse proposée (allant ici de 1 = « Pas du tout d'accord », à 7 = « Tout à fait d'accord ») et les différents items composant le questionnaire (et notamment les items contre-trait). À l'opposée, prenons une mesure indirecte de préjugé, l'IAT-Maghrébin/Blanc. Comme nous l'avons vu, dans les blocs critiques il est demandé de catégoriser à la fois des stimuli Maghrébins et Blancs, ainsi que des mots positifs et négatifs. À l'inverse de la mesure directe décrite ci-dessus, le participant doit ici répondre le plus rapidement possible. De plus, bien que le participant n'ait pas à interpréter d'échelle de réponse, pour chaque essai il va devoir détecter qu'elle sera la bonne réponse (e.g., est-ce qu'il m'est demandé de catégoriser ce stimulus Maghrébin sur la même clé de réponse qu'un stimulus positif ou sur la même clé de réponse qu'un stimulus négatif). Ainsi, les réponses des participants au sein de ces deux types de mesures peuvent ne pas refléter uniquement les processus cognitifs d'intérêts (e.g., processus plus ou moins contrôlés), mais également des caractéristiques différentes de ces mesures qui n'ont aucun rapport avec les processus d'intérêt (e.g., Krieglmeyer & Sherman, 2012).

Nous venons de l'évoquer, utiliser des mesures indirectes pour étudier les stéréotypes et préjugés individuels peut s'avérer problématique. D'une part, les réponses des participants peuvent refléter des processus de Type 1, mais aussi des processus de Type 2 et notamment des processus contrôlés (e.g., Fiedler & Bluemke, 2005). D'autre part, les réponses des participants peuvent également refléter des caractéristiques spécifiques aux tâches et qui ne sont pas liées aux processus d'intérêts (e.g., Hütter & Klauer, 2016). Toutefois, des solutions

existent. Parmi ces solutions, des modèles mathématiques, appelés modèles multinomiaux ont l'avantage de pouvoir contourner ces différents problèmes. Nous allons montrer que, grâce à ces modèles, les chercheurs peuvent désormais étudier au sein d'une même tâche les processus automatiques et non automatiques à l'origine des attitudes individuelles.

1.3 L'apport des modèles multinomiaux dans l'étude des attitudes raciales

Le but de cette partie n'est pas de faire un historique des modèles multinomiaux comme nous avons pu le faire pour les mesures indirectes ou directes. Ainsi, bien que les modèles multinomiaux aient été popularisés avant les années 90 (e.g., Riefer & Batchelder, 1988), nous allons laisser ces modèles de côté pour nous concentrer tout d'abord sur une procédure développée par Jacoby (1991 ; Jacoby et al., 1993) appelée *Process Dissociation Procedure* (PDP). Comme nous allons le voir à présent, le PDP a joué un rôle majeur dans le développement des modèles multinomiaux actuels dans le champ des attitudes.

1.3.1 Du PDP aux modèles multinomiaux

Dans le but de décrire le plus clairement possible le fonctionnement des modèles multinomiaux, il nous faut nous arrêter sur le fonctionnement du PDP. À l'origine, le PDP avait pour rôle de séparer, au sein d'une tâche de mémoire, les processus issus de la mémoire implicite des processus issus de la mémoire explicite. En effet, Jacoby avait mis en lumière la possibilité que des processus conscients, issus de la mémoire explicite, puissent influencer les performances des participants dans des tâches indirectes (i.e., tests de mémoire implicite), et que dans le même temps des processus non conscients, issus de la mémoire implicite, puissent influencer leurs performances dans des tâches plus directes (i.e., tests de mémoire explicite ; Jacoby, 1991 ; Payne & Bishara, 2009).

L'utilisation du PDP nécessite la présence, au sein d'une même tâche, de deux types de phases (ou deux types d'essais) appelées phase d'inclusion et phase d'exclusion. Dans les études de Jacoby (e.g., Jacoby, 1991) la phase d'inclusion est construite de telle sorte que les processus issus des mémoires explicite et implicite donnent des réponses identiques. À l'inverse, dans la phase d'exclusion les processus issus des mémoires explicite et implicite

devraient donner des réponses différentes. La particularité du PDP est qu'en comparant algébriquement la proportion de réponses issues de la phase d'inclusion avec celle issue de la phase d'exclusion, il est possible de connaitre la contribution spécifique des processus provenant de la mémoire implicite ainsi que de ceux provenant de la mémoire explicite. Il nous semble important d'insister sur le fait que le PDP ne se base pas sur des temps de réponse, mais sur les réponses catégorielles des participants (ou taux d'erreur). Afin d'illustrer le fonctionnement général du PDP, évoquons l'une de ses applications en psychologie sociale.

Le PDP a été grandement repris et utilisé en psychologie sociale pour séparer les processus automatiques des processus non automatiques (pour une revue de littérature, voir Payne & Bishara, 2009). Prenons l'exemple de la *Weapon Identification Task* (WIT ; Payne, 2001), autre tâche indirecte utilisée en cognition sociale. Les participants doivent, au sein de cette tâche, détecter le plus rapidement possible si des objets (cibles) présentés sur un écran d'ordinateur sont des outils ou des armes. Reprenant en cela les tâches d'amorçages vues précédemment (e.g., EPT ; Fazio et al., 1995), les objets cibles sont précédés d'amorces de visages noirs ou de visages blancs, le tout présenté très rapidement. Là encore, l'idée derrière la WIT est que la présentation d'une amorce de visage noir va avoir pour conséquence la réactivation du stéréotype liant les individus noirs aux armes et ainsi influencer la réponse des participants, ceux-ci étant plus rapides et/ou faisant moins d'erreurs pour détecter les armes quand celles-ci sont précédées de visages noirs par rapport à des visages blancs.

Comme nous l'avons évoqué précédemment, le PDP va nous permettre de séparer les processus automatiques des processus non automatiques, qualifiés ici de processus contrôlés. Les processus contrôlés (C) sont ici définis comme la capacité du participant à discriminer de manière intentionnelle la cible (outil vs. arme). À l'inverse, les processus automatiques (A) sont définis comme les processus influençant la réponse du participant quand les processus contrôlés ne sont pas à l'œuvre (Payne & Bishara, 2009). Comme nous venons de l'évoquer, utiliser le PDP nécessite la présence de deux types d'essais, raison pour laquelle la WIT se compose d'essais congruents (i.e., phase d'inclusion) mélangés aléatoirement avec des essais incongruents (i.e., phase d'exclusion). Décrivons ces deux types d'essais.

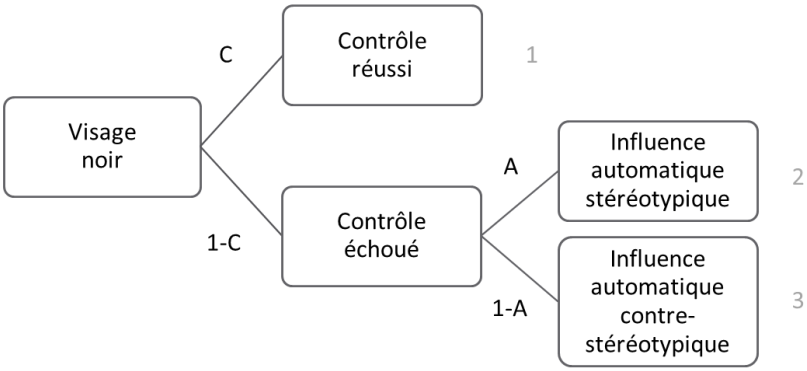
Dans les essais congruents, les processus automatiques et les processus contrôlés aboutissent à une même réponse. Prenons l'exemple d'un essai où l'amorce est un visage noir et la cible une arme. Pour Payne (2001), les processus automatiques (i.e., association stéréotypique entre les Noirs et les armes) et les processus contrôlés (i.e., identification correcte de l'arme) amèneront le participant à répondre « arme ». D'un point de vue

algébrique, la probabilité de répondre « arme » dans un essai congruent est donc égale à la probabilité d'identifier correctement l'arme (i.e., « C ») plus la probabilité conditionnelle d'être influencé par l'association stéréotypique Noirs/armes quand l'identification échoue (i.e., « (1-C)*A »). En résumé, P(arme/congruent) = C + (1-C)*A.

À l'inverse, les essais incongruents sont des essais où les processus automatiques et les processus contrôlés aboutissent à une réponse différente. Prenons l'exemple d'un essai où l'amorce est un visage noir, mais où la cible est un outil. Les processus automatiques (i.e., association stéréotypique entre les Noirs et les armes) vont orienter la réponse du participant vers la réponse « arme ». À l'opposée, les processus contrôlés (i.e., identification correcte de l'outil) vont orienter la réponse du participant vers la réponse « outil ». La probabilité de répondre « arme » dans un essai incongruent est donc uniquement égale à la probabilité conditionnelle d'être influencé par l'association stéréotypique quand l'identification échoue (i.e., « (1-C)*A »). De ce fait, P(arme/incongruent) = (1-C)*A. À partir de ces deux équations, la probabilité que la réponse du participant soit influencée par les processus automatiques et contrôlés peut être directement estimée algébriquement. Ici, la probabilité que les processus contrôlés soient à l'œuvre est donc : C = P(arme/congruent) – P(arme/incongruent). En remplaçant les différents termes de ces équations, la probabilité que les processus automatiques soient à l'œuvre est : A = P(arme/incongruent)/(1-C).

Nous venons de représenter algébriquement la prédiction du PDP quant à la contribution des différents paramètres (i.e., A et C) en fonction des différents types d'essais (i.e., congruents vs. incongruents). Ceci constitue une première possibilité de représentation des paramètres au sein du PDP. Mais nous pouvons également représenter les différentes prédictions du PDP forme d'arbre (Figure 1). Pour expliquer le fonctionnement de l'arbre, nous allons reprendre le même exemple que celui que nous avons utilisé. Tout à gauche de l'arbre, nous avons le type d'amorce (ici, un visage noir). Ensuite, nous avons les différentes « branches » de l'arbre. Les différents rectangles de l'arbre correspondent aux processus cognitifs (représentés ici par des paramètres) aboutissant à la réponse du participant. Les paramètres ayant une ligne menant à eux sont conditionnels au paramètre précédent. Il est important de signaler qu'un paramètre précédent un autre au sein de cet arbre n'est en aucun cas le reflet d'une précédence temporelle du processus, mais bien d'une influence prioritaire pour aboutir à la réponse du participant. Enfin, le tableau de droite représente les « feuilles de l'arbre » c'est-à-dire ici les prédictions du modèle en termes de réponses correctes (+) et

incorrectes (-) en fonction du type d'essai (i.e., congruent vs. incongruent) et de la contribution des différents paramètres (i.e., branches produisant ces réponses).



Congruent (arme)	Incongruent (outil)
+	+
+	-
-	+

Figure 1. Représentation en arbre de la *Process Dissociation Procedure* (PDP). À gauche, le type d'amorce (e.g., visage noir). Au milieu, les processus cognitifs aboutissant à la réponse du participant (i.e., C = Contrôle ; A = Automatique). À droite, les réponses prédites par le modèle en fonction du type d'essai (i.e., Congruent = Visage noir/arme ; Incongruent = Visage noir/outil). Les réponses correctes sont représentées par le signe « + » et les réponses incorrectes par le signe « - ». Les numéros (1 à 3) correspondent aux numéros donnés aux branches de l'arbre.

Pour illustrer le fonctionnement de cet arbre, reprenons nos deux exemples précédents, à savoir l'exemple d'un essai congruent où l'amorce est un visage noir et la cible une arme, et d'un essai incongruent où l'amorce est un visage noir et la cible un outil. Commençons par un essai congruent. Pour rappel, dans ce type d'essai, donner la réponse « arme » est correct, et correspond donc ici aux réponses « + » de l'arbre. Sur la première branche de l'arbre, nous voyons que le seul paramètre représenté est le paramètre « C » et que celui-ci mène à une réponse correcte (+). Cela signifie que si le participant a identifié la bonne réponse, c'est-à-dire si le processus contrôlé (C) est à l'œuvre, le participant va répondre correctement « arme ». Sur la deuxième branche de l'arbre, nous voyons que le premier paramètre représenté est « 1-C », ce qui signifie l'absence de contrôle, puis le paramètre « A », qui signifie l'influence de l'association stéréotypique Noirs/armes, le tout menant à une réponse correcte (+). Comme nous l'avons vu précédemment, cela signifie que

si l'identification de la réponse correcte échoue (1-C), mais qu'il y a influence de l'association stéréotypique, le participant va répondre correctement « arme ». Pour résumer, la probabilité ici de répondre arme dans un essai congruent est égale à la première branche de l'arbre, plus la deuxième branche de l'arbre. Nous retrouvons donc bien l'équation décrite précédemment : P(arme/congruent) = C + (1-C)*A. Nous pouvons utiliser le même raisonnement pour un essai incongruent. Pour rappel, répondre « arme » dans ce type d'essai équivaut à donner une réponse incorrecte, correspondant aux réponses « - » de l'arbre. Dans cette condition, la deuxième branche de l'arbre aboutit à une réponse incorrecte. En effet, dans ce type d'essai, si l'identification de la bonne réponse échoue (1-C), mais que le participant est influencé par l'association stéréotypique Noirs/armes (A), le participant va répondre de manière incorrecte « arme ». Encore une fois, nous retrouvons l'équation que nous avons décrite précédemment, P(arme/incongruent) = (1-C)*A.

Ce modèle est appelé C-dominant, car la réponse du participant va être conditionnée en priorité par les processus contrôlés. En d'autres termes, comme nous venons de le voir, ce modèle prédit que si le participant a identifié correctement la cible, celui-ci va répondre en fonction de cette identification, et ce, même si son stéréotype est activé. Toutefois, un modèle inverse peut exister, c'est-à-dire un modèle où les processus contrôlés n'influenceraient la réponse du participant que lorsque les processus automatiques ne sont pas à l'œuvre (i.e., modèle A-dominant). En d'autres termes, le modèle A-dominant prédit que si le participant a activé son stéréotype, il va répondre en fonction de ce stéréotype, et ce, même s'il a identifié la bonne réponse. Quel modèle choisir ? En réalité, ce qui va orienter le chercheur vers l'utilisation d'un modèle C-dominant ou d'un modèle A-dominant est avant tout lié à son raisonnement théorique. La plupart du temps le modèle C-dominant fera théoriquement plus de sens (Bishara & Payne, 2009 ; Payne & Bishara, 2009). Toutefois, dans de rares exceptions (e.g., tâche de Stroop ; Lindsay & Jacoby, 1994) le modèle A-dominant prédominera. Nous insistons sur ce point : dans la version classique du PDP, il est impossible d'avoir un indice objectif, autre qu'un choix théorique, permettant la sélection du modèle décrivant au mieux la contribution des processus (Hütter & Klauer, 2016). Ce point constitue une première critique majeure de la procédure classique du PDP. En outre, comme nous allons le voir maintenant, une autre critique a entrainé la remise en cause de ces modèles.

Le PDP, bien que constituant une procédure relativement accessible, a été critiqué, car jugé trop simpliste dans la description des processus à l'œuvre (e.g., Buchner et al., 1995 ; Hütter & Klauer, 2016 ; Klauer & Meiser, 2000). Illustrons tout d'abord ce point avec les

processus automatiques. Pour rappel, dans leur définition classique, les processus automatiques vont être ceux qui vont guider la réponse du participant quand les processus contrôlés ne sont pas à l'œuvre (Jacoby, 1991 ; Payne, 2001). Cela constitue une définition plus large que ce que nous avons pu voir jusqu'à présent. Nous avons en effet pu voir que, dans le champ des attitudes raciales, les processus automatiques étaient décrits comme étant le reflet d'un contenu stéréotypique ou affectif réactivé (e.g., Devine, 1989). En conséquence, nous pourrions imaginer que le paramètre d'automaticité (A) du PDP renvoie directement à la réactivation de ce contenu stéréotypique ou affectif (Payne, 2001). Mais que se passe-t-il si le stéréotype n'est pas activé ? Pour certains auteurs (e.g., Buchner et al., 1995), la version classique du modèle ignore la possibilité que le participant puisse « deviner » la bonne réponse et les biais qui peuvent être associés. Par biais associés, nous entendons la tendance à favoriser de manière non consciente un type de réponse par rapport à l'autre. Par exemple, Nisbett et Wilson (1977) ont montré que des participants devant choisir une paire de collants posée sur une table parmi quatre pairs identiques avaient un biais systématique à préférer des items placés à droite de la table quand bien même ils n'avaient pas conscience de ce biais. Ramenons cela à notre exemple précédent. Si les processus contrôlés échouent, il est possible qu'un participant réponde « arme », non pas parce que son stéréotype a été activé, mais tout simplement parce que celui-ci doit fournir une réponse. Cette réponse peut être totalement aléatoire (50% arme, 50% outil), mais peut aussi avoir été guidée par une impression (vérifiée ou non) de voir globalement plus d'armes que d'outils ou par l'impression inverse.

Nous pouvons émettre le même type de critiques quant aux processus contrôlés. Comme nous l'avons évoqué précédemment, les processus contrôlés vont être définis comme la capacité du participant à discriminer de manière intentionnelle la cible présentée (outil vs. arme ; Payne, 2001). La notion de contrôle correspondrait ici simplement à une détection de la bonne réponse (i.e., extraction de l'information issue de la tâche afin de fournir une réponse adéquate ; Sherman et al., 2008). De manière surprenante, cela laisse totalement de côté la notion de contrôle ou d'inhibition du contenu mental réactivé comme certains modèles duaux peuvent l'évoquer (e.g., Devine, 1989 ; Fazio, 1990). Ainsi, pour bon nombre d'auteurs (e.g., Conrey et al., 2005 ; Sherman, 2006 ; Sherman et al., 2008), ces deux définitions de la notion de contrôle (i.e., détection ou inhibition) peuvent aboutir à des réponses différentes et doivent de ce fait être prises en compte. En résumé, comme nous venons de le montrer à la fois pour le paramètre d'automaticité comme pour le paramètre de contrôle, ces deux paramètres ne semblent pas suffisants pour décrire avec précision les processus à l'œuvre au sein d'une mesure indirecte. Or, pour permettre une résolution algébrique de la valeur des

différents paramètres, le PDP ne peut inclure plus de deux paramètres. Une fois encore, les chercheurs ont dû se tourner vers un autre type d'outils : les modèles multinomiaux.

1.3.2 Le fonctionnement des modèles multinomiaux : l'exemple du Quad Model

Les modèles multinomiaux sont des modèles mathématiques qui, combinés à une tâche indirecte, permettent de démêler l'impact des processus cognitifs à l'œuvre au sein de la tâche (pour une revue de la littérature, voir Batchelder & Riefer, 1999 ; Hütter & Klauer, 2016). Tout comme le PDP, les modèles multinomiaux exigent l'utilisation de données catégorielles (e.g., correct/incorrect), c'est-à-dire d'une tâche où chaque réponse du participant est incluse dans une seule et unique catégorie de réponses (Klauer et al., 2011 ; mais voir Klauer & Kellen, 2018, pour des modèles multinomiaux basés sur les temps de réponse). Le principal atout des modèles multinomiaux est la possibilité d'inclure dans le modèle plus de deux paramètres mathématiques, ces paramètres représentant là encore des processus cognitifs spécifiques (e.g., activation du stéréotype, application du stéréotype). Le choix des paramètres à inclure au sein de ces modèles se fait une fois encore de manière théorique (Calanchini, 2020 ; Calanchini et al., 2018 ; Hütter & Klauer, 2016 ; Sherman, 2006 ; Sherman et al., 2008, 2014). Il est primordial d'inclure l'ensemble des principaux processus à l'œuvre au sein de la tâche, tout en respectant certaines contraintes inhérentes à ce type de modèle (e.g., ne pas avoir plus de paramètres que de conditions au sein de la tâche ; Klauer et al., 2011). De plus, il est également nécessaire de spécifier théoriquement la relation conditionnelle supposée entre ces paramètres. Comme pour le PDP, cette relation conditionnelle spécifie quel paramètre sera prioritaire pour aboutir à la réponse du participant et quels paramètres joueront un rôle si le premier paramètre n'est pas à l'œuvre. Encore une fois, cette priorité n'est pas temporelle, l'ensemble des paramètres pouvant être activés en même temps (Calanchini, 2020 ; Calanchini et al., 2018). Comme pour le PDP, la valeur de chaque paramètre sera estimée sur une échelle de probabilité (de 0 à 1) représentant la contribution spécifique de chaque processus (Calanchini, 2020).

Afin d'illustrer plus clairement le fonctionnement des modèles multinomiaux, nous allons prendre l'exemple du *Quadruple Process Model* (Quad Model ; Conrey et al., 2005 ; Sherman, 2006), modèle classiquement associé à l'IAT. Bien que l'IAT soit majoritairement une mesure du préjugé et non pas du stéréotype, le Quad Model est à ce jour le modèle

multinomial le plus utilisé en cognition sociale dans le champ des attitudes raciales. Il nous semble donc pertinent de présenter ce modèle pour illustrer le fonctionnement général des modèles multinomiaux. Il est tout d'abord à souligner que le Quad Model ne se base pas sur les temps de réponse classiquement mesurés dans l'IAT, mais sur les fréquences de bonnes et de mauvaises réponses associées. Globalement, le Quad Model a été développé comme une extension du PDP. Là où le PDP incluait uniquement deux paramètres, automaticité et contrôle, le Quad Model propose que quatre processus et donc quatre paramètres principaux contribuent aux réponses du participant au sein de l'IAT (Conrey et al., 2005 ; Sherman, 2006 ; Sherman et al., 2010, 2014). Le premier paramètre est le paramètre d'Activation (AC ; « ACtivation »), qui représente la probabilité qu'une association, évaluation ou impulsion comportementale soit activée à la vision de l'objet attitudinale (par souci de simplicité nous regrouperons ces termes sous le chapeau de « contenu mental »). Le deuxième paramètre est le paramètre de Détection (D ; « Detection »), qui représente la probabilité de discriminer la bonne réponse au sein d'un essai. Le troisième paramètre est le paramètre de dépassement du biais (OB ; « Overcoming Bias ») qui représente la probabilité que le contenu mental réactivé soit inhibé et remplacé par une réponse plus appropriée. Enfin, le dernier paramètre est le paramètre de biais de réponse (G ; « Guessing »), qui correspond à la tendance à fournir une réponse plutôt qu'une autre quand aucun des autres paramètres n'est à l'œuvre.

La relation conditionnelle entre ces quatre paramètres peut être, une fois encore, résumée sous forme d'arbre prédisant le nombre de réponses correctes et de réponses incorrectes dans les essais compatibles (e.g., catégorisation des visages noirs et des mots négatifs par une même clé de réponse) et les essais incompatibles (e.g., catégorisation des visages noirs et des mots positifs par une même clé de réponse ; Figure 2). Comme illustration, considérons un essai où un visage noir apparait à l'écran, et observons pour chaque type d'essai (i.e., compatible vs. incompatible), les fois où le modèle prédira une bonne réponse de la part du participant (+).

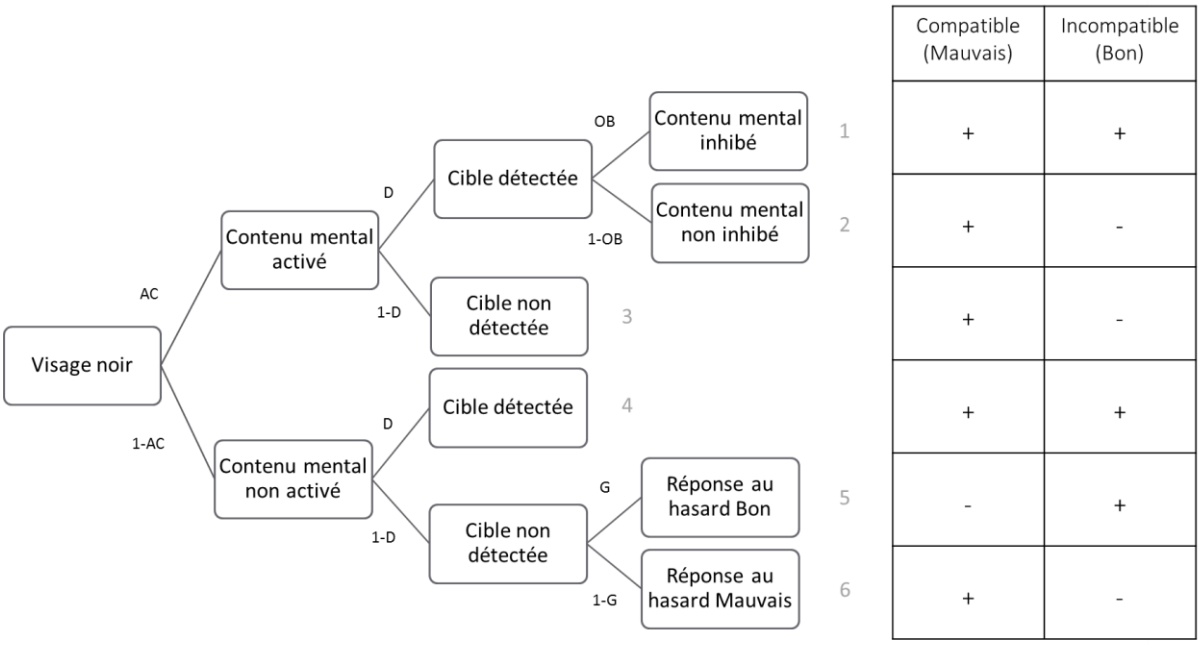


Figure 2. Représentation en arbre du *Quadruple Process Model* (Quad Model). À gauche, le type d'amorce (e.g., visage noir). Au milieu, les processus cognitifs aboutissant à la réponse du participant (i.e., AC = Activation ; OB = Dépassement du biais ; D = Détection ; G = Biais de réponse). À droite, les réponses prédites par le modèle en fonction du type d'essai (i.e., Compatible = Visage noir/Mauvais ; Incompatible = Visage noir/Bon). Les réponses correctes sont représentées par le signe « + » et les réponses incorrectes par le signe « - ». Les numéros (1 à 6) correspondent aux numéros donnés aux branches de l'arbre.

Pour rappel, dans les essais compatibles, les visages noirs et les mots négatifs (i.e., mauvais) partagent une même clé de réponse. À l'apparition d'un visage noir, le participant devra donc répondre avec la clé de réponse associée à « Mauvais ». Si l'on regarde les trois premières branches de l'arbre (les trois branches du haut), nous pouvons voir que, pour chacune de ces branches, le modèle prédira que le participant donnera la bonne réponse (+). En réalité, nous voyons que ces trois branches correspondent au cas où le contenu mental « Noirs/Négatif » va être réactivé (AC). Ainsi, dans un bloc compatible, à partir du moment où le participant a activé son contenu mental « Noir/Négatif », il va donner la bonne réponse. En effet, dans les essais compatibles, le contenu mental réactivé (AC) et la détection de la bonne réponse (D), tendent vers la bonne réponse (« Mauvais »), l'inhibition du contenu mental (OB) n'est donc pas nécessaire. Nous verrons plus précisément le rôle du paramètre OB dans un essai incompatible. Mais avant cela, intéressons-nous, toujours dans les essais compatibles, au

cas où le contenu mental « Noir/Négatif » n'est pas réactivé (1-AC). Nous nous situons ici au niveau des trois branches les plus basses de l'arbre. Nous voyons que si le participant n'a pas activé son contenu mental (1-AC), mais s'il a détecté la bonne réponse (D), il va donner la bonne réponse (quatrième branche de l'arbre). Enfin, si le participant n'a pas activé son contenu mental (1-AC) et s'il n'a pas détecté la bonne réponse (1-D), le modèle prédira qu'il pourra donner la bonne réponse s'il répond au hasard avec la clé de réponse associée à « Mauvais » (1-G ; dernière branche de l'arbre).

Dans les essais incompatibles, les visages noirs et les mots positifs partagent une même clé de réponse. L'apparition d'un visage noir demande donc au participant de répondre avec la clé de réponse associée à « Bon ». Au niveau des processus, l'activation du contenu réactivé (AC) tend pourtant vers la mauvaise réponse « Mauvais », là où la détection de la bonne réponse tend vers la réponse « Bon ». D'abord, plaçons-nous au niveau des branches hautes de l'arbre. Nous pouvons voir que, dans les essais incompatibles, si le participant a activé son contenu mental (AC), il ne pourra donner la bonne réponse que si, d'une part, il a détecté la bonne réponse (D), et d'autre part, il inhibe son contenu mental réactivé (OB) au profit de cette bonne réponse (première branche de l'arbre). Ensuite, plaçons-nous au niveau des branches basses de l'arbre, c'est-à-dire si le participant n'a pas activé son contenu mental (1-AC). Le modèle va prédire que le participant donnera la bonne réponse dans deux cas. Tout d'abord, s'il a détecté que la bonne réponse était associée à la clé « Bon » (D ; quatrième branche de l'arbre). Enfin, s'il n'a pas détecté la bonne réponse (1-D), mais s'il répond au hasard avec la clé de réponse associée à « Bon » (G ; cinquième branche de l'arbre).

Nous l'avons vu avec le PDP, nous pouvons représenter les prédictions du modèle sous forme d'arbre, comme nous venons de le faire, mais aussi sous forme d'équation. Pour connaitre la probabilité que le participant donne la bonne réponse dans un essai compatible, il suffit là encore d'additionner les probabilités conditionnelles de chaque branche prédisant une bonne réponse. Ici, P(Mauvais/compatible) = AC + (1-AC)*D + (1-AC)*(1-D)*(1-G). De la même manière, dans un essai incompatible, la probabilité que le participant donne une bonne réponse sera la somme des branches aboutissant à une bonne réponse. Ici, P(Bon/incompatible) = AC*D*OB + (1-AC)*D + (1-AC)*(1-D)*G. Grâce à la totalité de ces équations, le modèle va pouvoir prédire l'ensemble des réponses des participants pour chaque condition, c'est-à-dire en fonction des différents blocs (compatible, incompatible) et des différents stimuli (visages noirs ou blancs, mots positifs ou négatifs).

Cependant, et à l'inverse du PDP, l'ensemble de ces équations ne peut pas permettre une résolution algébrique de la valeur de chaque paramètre (Batchelder & Riefer, 1999 ; Klauer et al., 2011). Pour estimer la valeur de chaque paramètre, le modèle va se baser sur les réponses effectives des participants, c'est-à-dire les fréquences de bonnes et de mauvaises réponses dans chaque condition. Plus précisément, par itérations successives, le modèle va faire varier la valeur de chaque paramètre du modèle (entre 0 et 1) et estimer statistiquement (e.g., maximum de vraisemblance, méthode Markov Chain Monte Carlo) la combinaison de valeurs de paramètres réduisant au maximum l'écart entre les fréquences de bonnes et mauvaises réponses prédites par le modèle et les fréquences de bonnes et de mauvaises réponses effectives (Batchelder & Riefer, 1999). Si le modèle est identifiable, c'est-à-dire si une seule et unique combinaison de valeurs de paramètres permet de prédire aux mieux les réponses des participants, ces valeurs vont être utilisées comme prédiction de la contribution de chaque processus à l'œuvre.

Comme nous venons de l'évoquer, le modèle va faire varier par itérations successives la valeur de ses paramètres afin que ses prédictions se rapprochent au maximum des données effectives des participants. Néanmoins, la meilleure prédiction du modèle et donc la meilleure combinaison de valeurs de paramètres peut rester trop éloignée des réponses effectives des participants. Le second avantage des modèles multinomiaux est la possibilité d'avoir un indice statistique permettant de quantifier la qualité d'ajustement du modèle (e.g., G^2 ; Hu & Batchelder, 1994), c'est-à-dire le degré de correspondance entre les prédictions du modèle basées sur la meilleure estimation de paramètres et les réponses effectives des participants. Si l'indice de qualité d'ajustement du modèle est faible et non significatif (i.e., $p > .05$), le modèle est décrit comme donnant une description correcte des données et la valeur des paramètres peut être interprétée. À l'inverse, un indice élevé et significatif sera signe que le modèle n'est pas des plus adéquats dans l'explication des processus à l'œuvre au sein de la tâche. Habituellement, l'indice d'ajustement du modèle aux données reporté est l'indice G^2, se basant sur la statistique de Pearson χ^2. Cependant, celui-ci ne sera pas l'indice que nous reporterons au cours de nos études. Il nous semble important de prendre quelques instants pour aborder, dès maintenant, l'une des raisons à cela.

Traditionnellement, l'indice d'ajustement du modèle est calculé sur des données agrégées entre participants, c'est-à-dire sur la somme totale de réponses correctes ou incorrectes de l'ensemble des participants pour chaque catégorie (e.g., Moshagen, 2010 ; Singmann & Kellen, 2013). Analyser les données de manière agrégée présente l'avantage

d'augmenter le nombre d'observations (i.e., le nombre de réponses) par conditions, augmentant ainsi la puissance statistique dans l'estimation des paramètres (Hütter & Klauer, 2016). Cependant, utiliser des données agrégées présuppose que l'estimation des paramètres est homogène entre participants (ou items ; Heck et al., 2018 ; Klauer, 2010). Une violation de l'homogénéité entre participants et/ou items peut entrainer une estimation biaisée des différents paramètres ainsi que des intervalles de confiances associés (Heck et al., 2018 ; Klauer, 2010 ; Smith & Batchelder, 2010).

Des méthodes ont été proposées pour analyser les données en cas d'hétérogénéité d'un participant à l'autre (Klauer, 2010 ; Smith & Batchelder, 2010). La méthode *latent-trait* (Klauer, 2010) propose un modèle hiérarchique bayésien où les participants sont considérés comme effet aléatoire et la somme des réponses des participants comme effet fixe (suivant en cela la structure de modèles mixtes classiques ; e.g., Judd et al., 2012). Cette méthode permet de calculer, en se basant là encore sur la formule de la statistique de Pearson χ2, l'indice d'ajustement T_1 permettant de quantifier l'écart entre les fréquences moyennes prédites et effectives. En outre, que ce soit l'indice classique G^2 ou l'indice T_1 prenant en compte l'hétérogénéité entre participants, ceux-ci se basent (au moins en partie) sur la somme des réponses des participants dans chaque condition. Ainsi, un nombre de participants élevé au sein d'une étude augmentera statistiquement la probabilité de détecter même de petites différences statistiques. La taille d'effet *w* (Cohen, 1992) nous permettra d'obtenir une taille d'effet relative à la déviation d'ajustement (ou d'adéquation) du modèle aux données, tout en contrôlant la taille de l'échantillon. Une taille d'effet inférieur à .10 (effet de petite taille) sera qualifiée d'acceptable pour conclure à une adéquation du modèle aux données.

Construire un modèle multinomial présentant une bonne qualité d'ajustement ne nous dit rien sur la nature intrinsèque des paramètres constituant ce modèle. Effectivement, à ce stade les paramètres ne sont que des « entités mathématiques » (Hütter & Klauer, 2016) contribuant à prédire les réponses des participants. Afin de faire le pont entre représentation théorique des différents processus et leurs opérationnalisations mathématiques, une phase de validation empirique des paramètres du modèle est nécessaire. La validité des différents paramètres comprend trois composantes : la validité convergente (i.e., est-ce que ma manipulation expérimentale influence bien le paramètre concerné), la validité divergente (i.e., est-ce que ma manipulation expérimentale n'influence pas les autres paramètres du modèle) et enfin la validité prédictive (i.e., est-ce que mon paramètre peut être corrélé à d'autres mesures reflétant le construit d'intérêt). Reprenons l'exemple du Quad Model afin d'illustrer

succinctement ce point. De nombreuses études ont pu mettre en évidence la validité de ce modèle (e.g., Conrey et al., 2005 ; Sherman, 2006 ; Sherman et al., 2008). Par exemple, dans une de leur étude, Conrey et collaborateurs (2005) ont montré qu'imposer au participant une fenêtre de réponse impactait les paramètres de détection (D) et d'inhibition du biais (OB), mais pas les paramètres d'activation du contenu mental (AC), ni de biais de réponse (G), fournissant ainsi une contribution à la fois à la validité convergente, mais aussi divergente des paramètres (notamment ici sur la relative automaticité ou non des différents processus). Nous reviendrons plus en détail sur des exemples de validation des différents paramètres quand nous parlerons de la tâche principale utilisée au sein de cette thèse.

Nous avons vu qu'un modèle multinomial se construisait avant tout grâce à une spécification théorique des différents processus à l'œuvre ainsi que de leur relation conditionnelle. Toutefois, il est possible théoriquement que d'autres combinaisons entre processus, et donc d'autres architectures de modèles puissent être envisagées. Dans le PDP par exemple, nous avons évoqué le modèle C-dominant, où les processus automatiques étaient conditionnés par l'absence de processus contrôlés, et le modèle A-dominant, faisant des prédictions inverses. De la même manière, nous pouvons imaginer la comparaison entre deux modèles de nature différente (e.g., PDP vs. Quad Model) dans la description d'une même tâche (e.g., IAT). Enfin, nous pouvons imaginer deux modèles partageant une même architecture (e.g., Quad Model), mais où le premier modèle ferait des prédictions identiques en fonction des conditions (e.g., valeur du paramètre OB identique pour les visages blancs et les visages noirs), là où le second ferait des prédictions différentes en fonction des conditions (e.g., OB supérieur pour les visages noirs par rapport aux visages blancs). En somme, il est parfois possible d'avoir le choix entre plusieurs modèles pour rendre compte des processus à l'œuvre. Encore une fois, quel modèle choisir ? Pour certains auteurs, le plus important est d'utiliser le modèle le plus adéquat théoriquement pour répondre à la question de recherche initiale (e.g., Calanchini, 2020; Sherman et al., 2008). Par exemple, si des chercheurs s'intéressent au mécanisme d'inhibition, le Quad Model peut être le plus pertinent, celui-ci comportant un paramètre d'inhibition (OB). En revanche, si ces chercheurs s'intéressent plus généralement à la comparaison entre processus automatiques et processus contrôlés, le PDP peut suffire. Il arrive néanmoins que la question de recherche concerne spécifiquement la quête du meilleur modèle pour rendre compte des données. Pour cela, des solutions existent.

Pour comparer deux modèles entre eux, nous pourrions imaginer l'utilisation des indices statistiques basés sur la statistique de Pearson χ^2, le meilleur modèle étant le modèle

ayant l'indice le plus bas. Toutefois, plus un modèle aura de paramètres, plus il sera flexible et donc plus il s'ajustera aux données (Hütter & Klauer, 2016). Les indices classiques G^2 ou T_1 ne prennent pas en compte le nombre de paramètres des différents modèles et ne sont donc pas toujours appropriés. Les chercheurs ont pendant longtemps préféré les indicateurs AIC (*Akaike Information Criterion* ; Akaike, 1974) et BIC (*Bayesian Information Criterion* ; Schwarz, 1978), imposant une pénalité dans l'estimation de l'ajustement des modèles aux données en fonction du nombre de paramètres qu'ils comportent. Il est aujourd'hui recommandé d'utiliser des indices prenant également en compte la complexité des différents modèles (e.g., le nombre d'occurrences de chacun des paramètres du modèle), comme des indices basés sur le *Minimum Descriptive Length* (MDL ; e.g., Calanchini, 2020). La méthode *latent-trait* que nous souhaitons appliquer employant des statistiques bayésiennes, nous utiliserons le WAIC (*Widely Applicable Information Criterion* ; Watanabe, 2010) qui se base sur l'indice AIC décrit précédemment et constitue l'indice de comparaison de modèles le plus adéquat en cas d'utilisation de la méthode *latent-trait* (Heck et al., 2018 ; Klauer, 2010). Ainsi, plus un modèle aura un indice WAIC faible et meilleur sera son ajustement aux données. Dans le cas d'une comparaison statistique de plusieurs modèles, sélectionner le modèle le plus adéquat reviendra à sélectionner le modèle ayant l'indice WAIC le plus bas.

Les modèles multinomiaux, grâce à leur formalisation théorique et à leur robustesse méthodologique, ont vu leur popularité grimper au cours des années 2000. Leurs spécificités ont été un atout dans l'étude des processus à l'œuvre en cognition sociale. Par le biais de quelques études utilisant le Quad Model, nous allons montrer en quoi les modèles multinomiaux peuvent être utiles dans le champ des attitudes raciales.

1.3.3 Modèles multinomiaux dans le champ des attitudes raciales

Les modèles multinomiaux appliqués aux mesures indirectes existantes ont permis une meilleure compréhension des mécanismes à l'origine de résultats classiques (Calanchini, 2020). Dans le champ des attitudes raciales, des études ont investigué si le biais racial classiquement observé dans l'IAT pouvait s'expliquer différemment en fonction du type de participant réalisant la tâche. Par exemple, Gonsalkorale, Sherman et collaborateurs (2009) ont voulu mettre à l'épreuve la croyance selon laquelle les individus plus âgés, parce qu'ayant grandi dans une société structurellement discriminante, auraient un préjugé anti-Noirs plus

important que les individus plus jeunes, grandissants dans une société plus égalitaire. Les auteurs ont appliqué un Quad Model à des données IAT récoltées auprès de participants de différents âges. Conformément à la croyance populaire, le biais IAT anti-Noirs augmentait globalement en fonction de l'âge. Toutefois, allant à l'encontre de la croyance habituelle, les individus plus âgés n'avaient pas un paramètre AC (activation du contenu mental « Noirs = Négatif ») supérieur aux individus plus jeunes, celui-ci diminuant même légèrement avec l'âge. À l'inverse, le paramètre d'inhibition du biais (OB) diminuait significativement chez les personnes plus âgées. Ainsi, le biais anti-Noirs classiquement supérieur chez les individus plus âgés serait en réalité le reflet d'une difficulté plus importante à inhiber ce biais, plutôt que d'un biais lui-même plus important (Gonsalkorale, Sherman, et al., 2009 ; Stewart et al., 2009).

Comme nous le verrons au Chapitre 2, des études ont montré que le contexte pouvait diminuer le biais anti-Noirs communément observé au sein des mesures indirectes (e.g., Barden et al., 2004 ; Wittenbrink et al., 2001). Habituellement, cette diminution est décrite comme reflétant une activation différente d'associations stéréotypiques selon le contexte. Par exemple, pour Wittenbrink et collaborateurs (2001), présenter un individu noir dans une église réactiverait en mémoire des aspects positifs du stéréotype des Noirs (e.g., religieux). À l'inverse, présenter le même individu noir dans un contexte de ghetto réactiverait en mémoire des aspects négatifs du stéréotype des Noirs (e.g., pauvre). Pour Allen et collaborateurs (2010), cette diminution de la négativité du biais en fonction du contexte n'est pas la résultante d'une activation stéréotypique différente, mais de l'influence du contexte sur le processus d'inhibition. Utilisant une tâche d'EPT (Fazio et al., 1986, 1995), ces auteurs ont tout d'abord répliqué les résultats classiques d'un biais anti-Noirs moins important quand les visages noirs et blancs étaient présentés en contexte positif (i.e., église) par rapport à un contexte négatif (i.e., prison). De manière plus intéressante encore, en appliquant un Quad Model à leurs données, ces auteurs ont montré que cet effet était associé à un paramètre d'inhibition (OB) plus important en contexte positif par rapport au contexte négatif, le paramètre d'activation (AC) n'étant pour sa part pas influencé par le contexte. Une fois encore, cela semble suggérer une plus forte influence du contexte sur la capacité d'inhibition du biais réactivé plutôt que sur la réactivation du biais en lui-même.

Dans l'étude présentée ci-dessus, Allen et collaborateurs ont également montré que l'influence du contexte sur le paramètre OB était plus importante chez les participants qui étaient motivés à réduire l'expression de leur préjugé par rapport à ceux qui ne l'étaient pas. D'une manière générale, et comme nous le verrons au Chapitre 3, plusieurs auteurs se sont

servis des modèles multinomiaux pour investiguer l'impact de variables individuelles (e.g., motivation à réduire l'expression de son préjugé) sur les processus à l'origine des réponses des participants (e.g., Gonsalkorale et al., 2011). Gonsalkorale et collaborateurs (2011) ont appliqué un Quad Model à des données de participants ayant passé une WIT ainsi qu'à des participants ayant passé un IAT. Ces auteurs s'intéressaient au rôle de la motivation à ne pas paraitre comme ayant des préjugés sur les processus à l'œuvre dans ces tâches. Sans décrire ici les échelles de motivations utilisées (nous reviendrons en détail sur ces échelles dans le Chapitre 3), ces auteurs ont notamment montré que des participants motivés de façon interne, mais pas de façon externe, à répondre sans préjugés allaient davantage détecter la bonne réponse au sein de ces tâches (D plus important). Nous développerons ces implications dans le Chapitre 3.

Cette partie s'est majoritairement centrée sur le Quad Model, modèle le plus souvent pris en exemple pour décrire l'utilité des modèles multinomiaux dans le champ des attitudes raciales (e.g., Sherman et al., 2010, 2014). Cependant, limiter l'utilisation des modèles multinomiaux à l'IAT et au Quad Model serait réducteur. Pour Calanchini (2020), d'autres mesures indirectes, utilisant notamment le paradigme d'amorçage séquentiel, peuvent se révéler utiles dans l'étude de ces processus. Nous avons vu que le Quad Model pouvait être appliqué à certaines de ces tâches (e.g., WIT ; Gonsalkorale et al., 2011). Plus globalement, et comme nous venons de l'illustrer, les modèles multinomiaux ont pendant longtemps permis de répondre à des questions relatives aux tâches existantes, le Quad Model ayant par exemple été créé dans le but de démêler l'impact des différents processus au sein de l'IAT. Aujourd'hui, nous pouvons trouver une dynamique inverse, c'est-à-dire la création de nouvelles tâches dans le but de tester directement la relation conditionnelle entre processus (Calanchini et al., 2018). C'est le cas de la *Stereotype Misperception Task* (SMT ; Krieglmeyer & Sherman, 2012), créée spécifiquement pour étudier la relation entre l'activation et l'application du stéréotype à l'origine du jugement individuel.

1.4 La Stereotype Misperception Task (SMT)

Résumons rapidement ce que nous avons vu jusqu'à présent. Nous savons que pour mesurer les stéréotypes et préjugés individuels, utiliser des mesures directes peut s'avérer problématique. Effectivement, nous avons vu qu'au sein de ces mesures les participants

pouvaient inhiber leurs stéréotypes et préjugés réactivés. Pour capter plus distinctement les stéréotypes et préjugés individuels, les chercheurs se sont inspirés de la psychologie cognitive pour créer des mesures indirectes. Toutefois, nous avons mis en évidence le fait que ces mesures pouvaient elles aussi être contaminées par des processus non automatiques. Enfin, nous avons évoqué la possibilité d'utiliser les modèles multinomiaux pour démêler, au sein d'une mesure indirecte, les différents processus pouvant aboutir à la réponse du participant. La fin de ce chapitre sera consacrée à l'illustration de la mesure que nous avons retenue afin d'étudier les stéréotypes pouvant être associés aux populations maghrébines.

1.4.1 Les processus à l'origine de la stéréotypisation

Comme nous l'avons évoqué, la littérature en psychologie sociale distingue le concept de stéréotype (i.e., ensemble de connaissances abstraites à propos d'un groupe) de celui de stéréotypisation (i.e., utilisation de ces connaissances face à un individu membre du groupe ; e.g., Leyens et al., 1996). En réalité, le mécanisme de stéréotypisation peut être divisé en deux processus distincts : l'activation du stéréotype et l'application du stéréotype (e.g., Devine & Monteith, 1999 ; Gilbert & Hixon, 1991 ; Kunda & Spencer, 2003). L'activation du stéréotype est définie comme l'augmentation de l'accessibilité mentale des connaissances relatives au groupe (e.g., Gilbert & Hixon, 1991). Par exemple, rencontrer un individu maghrébin réactivera en mémoire mes connaissances sur les Maghrébins (e.g., « les Maghrébins sont menaçants »). L'application du stéréotype est définie comme l'utilisation de ces connaissances réactivées dans le jugement (e.g., Gilbert & Hixon, 1991). Par exemple, la réactivation des connaissances associées aux Maghrébins pourra colorer négativement mon impression de l'individu rencontré. L'application du stéréotype dépendra de l'activation ou non du stéréotype (Krieglmeyer & Sherman, 2012), l'activation du stéréotype étant pour sa part indépendante de l'application ou non du stéréotype.

Les modèles duaux ont permis une théorisation des conditions opérantes associées à l'activation et l'application du stéréotype (Devine, 1989 ; Fazio, 1990). Nous l'avons vu, pour Devine (1989) comme pour Fazio (1990) l'activation du stéréotype s'effectuerait de manière non intentionnelle. L'application ou non du stéréotype dans le jugement dépendrait de l'adéquation entre le contenu réactivé et les valeurs personnelles de l'individu (Devine, 1989) ou bien de facteurs motivationnels relatifs à la situation (Fazio, 1990). D'une manière

générale, il semble y avoir un consensus théorique par rapport à ces concepts. L'activation du stéréotype est définie théoriquement comme étant un processus relativement automatique (notamment non intentionnel), là où l'application du stéréotype est définie théoriquement comme étant un processus relativement non automatique (notamment contrôlable ; Rivers et al., 2020).

Néanmoins, en ce qui concerne l'alignement complet, d'une part, des critères d'automaticité de l'activation du stéréotype et, d'autre part, des critères de non automaticité de l'application du stéréotype, les données empiriques sont mitigées (e.g., Schneider, 2004). L'activation du stéréotype peut par exemple être décrite comme étant un processus non conscient (e.g., Wittenbrink et al., 1997) et non intentionnel (e.g., Blair et al., 2002), mais pouvant demander des ressources (e.g., Gilbert & Hixon, 1991 ; mais Sherman et al., 2000). De la même manière, l'application du stéréotype peut être décrite comme étant un processus contrôlable (Kunda & Spencer, 2003), mais relativement efficient (Moskowitz et al., 1999 ; mais Gilbert & Hixon, 1991). Pour certains auteurs, la compréhension des processus sous-tendant le mécanisme de stéréotypisation souffre là encore de limites méthodologiques (Krieglmeyer & Sherman, 2012 ; Rivers et al., 2020). En effet, les conditions opérantes théoriques quant à l'activation et l'application du stéréotype ont guidé le type de mesures employées dans la littérature, les mesures indirectes étant employées pour mesurer l'activation du stéréotype, les mesures directes mesurant l'application du stéréotype (Krieglmeyer & Sherman, 2012). Or, comme nous le savons maintenant, ceci peut être problématique, car aucune mesure n'est *process pure* (e.g., Hütter & Klauer, 2016). Il est de ce fait difficile de savoir si ces mesures reflètent uniquement l'activation du stéréotype, l'application du stéréotype ou bien une combinaison des deux (Rivers et al., 2020). Conscients de ces limites méthodologiques, Krieglmeyer et Sherman (2012) ont créé la SMT, tâche qui va nous permettre de démêler la part d'activation et d'application du stéréotype aboutissant au jugement individuel.

1.4.2 SMT : structure de la tâche et modèles associés

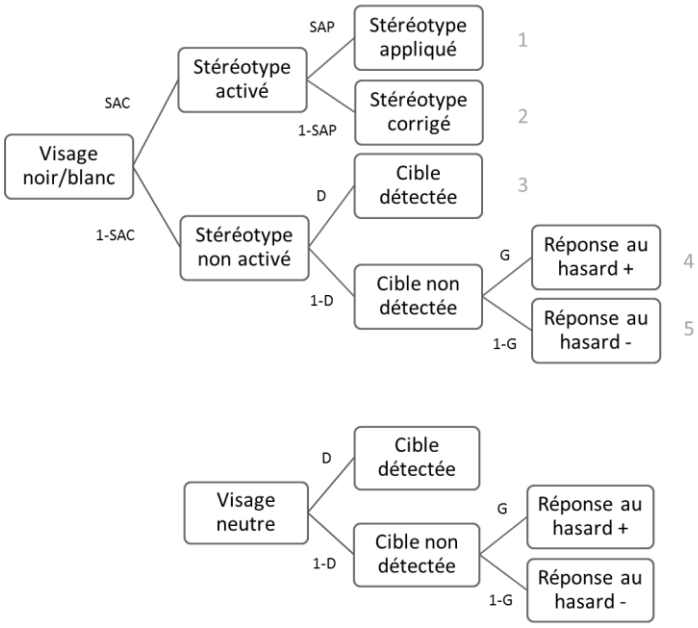
Comme certaines des mesures indirectes décrites précédemment (e.g., EPT, Fazio et al., 1995 ; WIT, Payne, 2001 ; AMP, Payne et al., 2005), la SMT utilise un paradigme d'amorçage séquentiel (Wentura & Degner, 2010). Dans la SMT, les participants doivent se

former une impression concernant des dessins de visages (cibles) présentés très rapidement (i.e., 100ms). Dans la procédure classique, les participants doivent juger si le dessin de visage présenté est plus menaçant ou moins menaçant que la moyenne des dessins de visages que les participants ont vus au cours des différents essais. En tout, la moitié des dessins présentés est relativement menaçante, l'autre moitié étant relativement non menaçante. Néanmoins, l'ensemble des visages sont ambigus ce qui, couplé à la présentation rapide de ces visages cibles, rend délicate la détection de la menace de ces dessins de visages. De plus, avant la présentation de chaque dessin cible, une photo d'un véritable visage (amorce) est présentée très rapidement (i.e., 150ms). Dans le paradigme classique, les amorces sont constituées de photos de visages noirs, blancs ou de visages neutres (i.e., contours grisés de visages). Une description plus complète de la tâche sera présentée au cours de l'Étude 1.

L'hypothèse des auteurs, confirmée par les résultats de plusieurs études (Krieglmeyer & Sherman, 2012 ; Rees, Ma, et al., 2019 ; Rees, Rivers, et al., 2019; Rivers et al., 2020) est que les dessins de visages cibles devraient être jugés comme plus menaçants quand ils sont précédés d'amorces de visages noirs par rapport à des amorces de visages blancs (ou neutres). Dans la suite du manuscrit, nous nommerons « effet SMT » la différence de jugement de menace entre les essais comportant une amorce Maghrébin et les essais comportant une amorce Blanc. Tout l'intérêt de cette tâche est de théoriser les processus à l'œuvre pouvant aboutir à ce type d'effet. Pour Krieglmeyer et Sherman (2012), quatre processus principaux sont à l'œuvre. Tout d'abord, le processus d'activation du stéréotype (SAC ; « Stereotype ACtivation ») relatif aux visages présentés (e.g., les individus noirs sont menaçants). Ensuite, le processus d'application du stéréotype (SAP ; « Stereotype APplication »), c'est-à-dire l'utilisation ou non du stéréotype réactivé dans le jugement de menace des dessins cibles. Ces dessins cibles étant ambigus, mais relativement menaçants (ou non menaçants), il reste possible de détecter la menace réelle de ces visages (D « Detection »). Enfin, un biais de réponse (G « Guessing ») orientant vers l'une ou l'autre des réponses peut survenir si les autres processus ne sont pas à l'œuvre.

L'ensemble de ces processus peut être rassemblé au sein du modèle multinomial traditionnellement associé à la SMT, le modèle SAC-dominant (Figure 3). Lorsque nous avons évoqué les arbres associés au PDP et au Quad Model, nous avons représenté les prédictions des modèles en termes de bonnes et de mauvaises réponses en fonction des essais compatibles et incompatibles. À l'inverse, pour l'arbre représentant le modèle SAC-dominant, nous représentons les prédictions du modèle en termes de réponse « Plus menaçant » (+) et

de réponse « Moins menaçant » en fonction du type d'amorces (Noir et Blanc) et du type de cible (Bas et Haut et sur la dimension de menace).



Noir		Blanc	
Bas	Haut	Bas	Haut
+	+	-	-
-	-	+	+
-	+	-	+
+	+	+	+
-	-	-	-

Bas	Haut
-	+
+	+
-	-

Figure 3. Représentation en arbre du modèle SAC-dominant. En haut, représentation pour un type d'amorce Noir/Blanc. En bas, représentation pour un type d'amorce Neutre. Les processus cognitifs représentés sont les suivants : SAC = Activation du stéréotype, SAP = Application du stéréotype, D = Détection, G = Biais de réponse. À droite, les réponses prédites par le modèle en fonction du type d'essai (i.e., type d'amorce : Noir vs. Blanc ; type de cible : Haut vs. Bas sur la dimension de menace). Les réponses « Plus menaçant » sont représentées par le signe « + » et les réponses « Moins menaçant » par le signe « - ». Les numéros (1 à 5) correspondent aux numéros donnés aux branches de l'arbre représentant le type d'amorce Noir/Blanc.

Pour illustrer le fonctionnement du modèle SAC-dominant, considérons un essai où un visage noir apparait à l'écran, et observons pour chaque type de cibles (i.e., Bas vs. Haut sur la dimension de menace), les fois où le modèle prédira une réponse « Plus menaçant » (+). Commençons par les essais où l'amorce est un visage noir et la cible est menaçante. Si le participant a activé son stéréotype (SAC) « Noirs = Menace » et que le participant applique son stéréotype (SAP), celui-ci va donner la réponse « Plus Menaçant » (première branche de

l'arbre). De plus, si le participant n'a pas activé son stéréotype (1-SAC), mais qu'il a détecté la menace réelle du visage cible (D), le participant va également donner la réponse « Plus Menaçant » (troisième branche de l'arbre). Enfin, si le participant n'a pas activé son stéréotype (1-SAC), qu'il n'a pas détecté la menace de la cible (1-D), il pourra répondre au hasard « Plus Menaçant » (G ; quatrième branche de l'arbre). Dans les essais où l'amorce est un visage noir et la cible est non menaçante, le participant pourra répondre « Plus menaçant » s'il a activé son stéréotype (SAC) « Noirs = Menace » et s'il l'applique pour donner sa réponse (SAP ; première branche de l'arbre). De plus, l'unique autre prédiction du modèle quant à cette réponse correspond au cas où le participant n'a pas activé son stéréotype (1-SAC), n'a pas détecté la menace réelle de la cible (1-D), mais répond au hasard « Plus menaçant (G ; quatrième branche de l'arbre).

Nous savons maintenant qu'un modèle doit être validé empiriquement pour pouvoir être utilisé (e.g., Hütter & Klauer, 2016). C'est chose faite avec les différents paramètres du modèle de la SMT (Krieglmeyer & Sherman, 2012 ; Reichardt et al., 2019). Dans chacune de leurs études, Krieglmeyer et Sherman (2012) ont manipulé le design de la tâche afin d'influencer l'un des paramètres du modèle sans influencer les autres paramètres. Tout d'abord, pour tester le paramètre d'activation du stéréotype (SAC), les auteurs ont manipulé le caractère prototypique des amorces de visages noirs. Le paramètre d'activation était effectivement plus élevé quand les amorces étaient prototypiques de visages noirs par rapport à des visages moins prototypiques, les autres paramètres ne variant pas. Ensuite, pour tester le paramètre d'application (SAP), Krieglmeyer et Sherman ont demandé aux participants d'appliquer ou au contraire corriger leur stéréotype s'ils avaient l'impression que celui-ci était activé. Effectivement, demander directement au participant d'inhiber ou non son stéréotype a influencé le paramètre d'application du stéréotype, celui-ci étant plus important dans la première condition en comparaison à la seconde[2]. Toujours à propos du paramètre SAP, Rivers et collaborateurs (2020) ont montré qu'augmenter l'intervalle de temps entre la présentation de l'amorce et de la cible avait pour conséquence une diminution de l'effet SMT. En regardant l'impact de cette manipulation expérimentale sur les paramètres du modèle, et comme on pouvait s'y attendre, la diminution de l'effet SMT coïncidait avec une diminution du paramètre SAP. En d'autres termes, augmenter l'intervalle de temps entre l'amorce et la cible

[2] Il est à signaler, pour cette étude, que cette manipulation a également influencé le paramètre SAC, n'apportant de ce fait pas d'élément quant à la validité divergente du paramètre SAP. Cependant, d'autres études (e.g., Rees, Ma, et al., 2019) ont montré que le paramètre SAP pouvait être indépendant du paramètre SAC.

(et donc les ressources temporelles) augmentait la capacité du participant à contrôler son stéréotype, reflétant ainsi la non automaticité du paramètre SAP en termes d'efficience. Pour revenir aux étapes de validation de Krieglmeyer et Sherman (2012), ces auteurs ont également mis en évidence le fait que manipuler le caractère discernable des visages cibles (i.e., inclure des visages cibles neutres au niveau de la menace) influençait le paramètre de détection (D) sans influencer les autres paramètres. Enfin, demander aux participants de répondre « Plus menaçant » ou « Moins menaçant » quand ils n'étaient pas certains de leur réponse influençait le biais de réponse (G) sans influencer les autres paramètres.

À la découverte de l'architecture du modèle SAC-dominant, le lecteur s'est peut-être questionné par rapport aux choix théoriques de Krieglmeyer et Sherman (2012) quant à la contribution spécifique de chaque paramètre du modèle. Comme nous le verrons au Chapitre 3, ceci est également notre cas. Effectivement, d'autres modèles théoriques pourraient être aussi appropriés. La prédiction principale du modèle SAC-dominant est la suivante : si le stéréotype est activé, il guidera la réponse du participant que celui-ci ait détecté ou non le niveau de menace de la cible. Reprenant en cela la comparaison entre modèles A-dominant et C-dominant relatifs au PDP, nous pouvons imaginer un modèle inverse où la détection de la cible guidera la réponse du participant que son stéréotype ait été activé ou non. Krieglmeyer et Sherman (2012) ont nommé ce modèle « D-dominant ». Nous reviendrons en détail sur ce modèle au Chapitre 3. Dans le modèle SAC-dominant, si le participant a activé son stéréotype, mais décide de corriger sa réponse, il va donner la réponse inverse à son stéréotype réactivé (e.g., répondre « Moins menaçant » quand l'amorce est un visage noir). Théoriquement, d'autres possibilités peuvent être envisagées. Par exemple, nous pourrions tout aussi bien imaginer un modèle où un participant n'appliquant pas son stéréotype cherche à se former une impression plus précise de la cible (e.g., par un phénomène d'individuation ; Fiske & Neuberg, 1990). Ainsi, Krieglmeyer et Sherman ont imaginé un modèle structurellement similaire au Quad Model où la détection de la cible influence le participant quand celui-ci décide de corriger son stéréotype (modèle « D après 1-SAP »). Enfin, Krieglmeyer et Sherman ont aussi imaginé la possibilité d'avoir un modèle plus simple, où le paramètre SAC et le paramètre SAP sont regroupés autour d'un seul paramètre « M » représentant le processus de fausse attribution décrit précédemment quand nous avons évoqué l'AMP (« modèle AMP » ; voir Payne et al., 2010). Ces quatre modèles pouvant tout autant être envisageables au niveau théorique, Krieglmeyer et Sherman (2012) ont cherché à savoir, dans l'ensemble de leurs études, quel était le modèle s'ajustant le mieux aux données effectives. Comme ils en faisaient

l'hypothèse, le modèle SAC-dominant semblait être le modèle décrivant le mieux les processus à l'œuvre au sein de la SMT et fut donc le modèle retenu par les auteurs.

La SMT sera la tâche que nous utiliserons tout au long de ce manuscrit. Développée dans le but de mieux comprendre les processus à l'origine du jugement stéréotypé, nous pensons que cette tâche peut correspondre à la mesure que nous recherchions. Celle-ci nous permettra ainsi d'étudier au plus près le stéréotype liant les populations maghrébines à une menace potentielle. En effet, grâce à son modèle multinomial associé, nous allons pouvoir démêler l'impact du processus d'activation du stéréotype de celui du processus d'application du stéréotype et être en mesure d'observer l'influence du contexte actuel sur ces processus. Toutefois, avant de décrire l'ensemble des études que nous avons effectuées au sein de ce travail de thèse, intéressons-nous à la littérature traitant de la perception des individus maghrébins.

1.4.3 **Stéréotype de menace : le cas des Maghrébins**

Tout au long de ce premier chapitre, lorsque nous avons illustré les stéréotypes et préjugés raciaux pouvant exister dans nos sociétés, nous avons quasi systématiquement pris l'exemple des individus de couleur noire (e.g., Crosby et al., 1980 ; Fazio et al., 1995 ; Greenwald et al., 1998 ; Payne et al., 2005 ; Sigall & Page, 1971). Cela n'est pas surprenant, la littérature en psychologie sociale étant majoritairement guidée par la littérature nord-américaine (Cheon et al., 2020), les études décrites ne sont que le reflet des problématiques raciales majeures existantes aux États-Unis. Cependant, la littérature nord-américaine ne traite pas uniquement des attitudes raciales envers les populations noires. Ainsi, de nombreuses études ont pu mettre en évidence une vision négative des populations arabo-musulmanes (e.g., Ahluwalia & Pellettiere, 2010 ; Dasgupta et al., 2009 ; Lyons et al., 2010 ; Mange et al., 2012 ; Park et al., 2007). Par exemple, des études ont montré que les populations arabo-musulmanes pouvaient induire chez les répondants des émotions de colère (Dasgupta et al., 2009) ou d'anxiété (Lyons et al., 2010), ces populations pouvant même être associées à des réponses agressives (Mange et al., 2012). Les médias peuvent entretenir cette perception négative des individus arabo-musulmans, ceux-ci étant par exemple dépeints de manière négative au cinéma (Shaheen, 2003). Globalement, il semblerait que ce groupe soit décrit comme étant l'exogroupe pouvant le plus nuire aux États-Unis (Lyons et al., 2010).

En Europe, il semblerait aujourd'hui que les individus arabo-musulmans constituent l'un des groupes si ce n'est le groupe le plus négativement perçu dans nos sociétés. Là encore, les populations européennes ont un préjugé négatif important envers ces populations, celui-ci ayant même augmenté après les attentats du 11 septembre 2001 (nous reviendrons sur cet élément ; Fetzer & Soper, 2003). Comme aux États-Unis, il semblerait que le préjugé négatif envers les populations arabo-musulmanes soit plus important que pour d'autres populations immigrées (Spruyt & van der Noll, 2017). De plus, les individus ayant un fort préjugé anti-Maghrébins auraient une vision prototypique des Maghrébins pouvant être perçus comme criminels et très peu dignes de confiance (Dotsch et al., 2008). Globalement, ces populations sont vues comme étant agressives (Fischer et al., 2007). En conséquence, plusieurs études utilisant le paradigme du tireur, tâche où il faut « tirer » le plus rapidement possible sur un individu s'il est armé (Correll et al., 2002), ont montré que les individus arabo-musulmans, parce que perçus comme plus menaçants, entrainaient plus de réponses agressives de la part des participants (e.g., Essien et al., 2017 ; Mange et al., 2012, 2016 ; Unkelbach et al., 2008, 2009).

Enfin, l'attitude raciale négative envers ces populations ne s'arrête pas à la frontière française. Des études réalisées auprès des forces de police ont par exemple montré un préjugé négatif envers les Maghrébins en France, ces individus ayant un risque presque huit fois plus important de se faire contrôler que des individus de couleur blanche (Gauthier, 2015 ; Jobard et al., 2012). De même, plusieurs auteurs utilisant des mesures indirectes ont pu mettre en évidence un biais racial anti-Maghrébins au sein par exemple d'une mesure d'approche-évitement (Rougier et al., 2020) ou d'un IAT (Essien et al., 2020 ; Popa-Roch & Delmas, 2011). Mange et collaborateurs (2016) ont voulu tester si le biais agressif traditionnellement observé dans le paradigme du tireur était la conséquence d'une simple différence endogroupe/exogroupe ou bien de la perception de menace associée à un exogroupe particulier. Comme on pouvait s'y attendre, les auteurs ont conclu de cette étude que l'exogroupe Maghrébins était davantage soumis à une réponse agressive de la part des participants, cet exogroupe était plus associé à un sentiment de menace et d'anxiété que les autres exogroupes étudiés (e.g., les Espagnoles).

En résumé, les études empiriques montrant une vision négative des populations maghrébines ou arabo-musulmanes ne manquent pas. Ces populations ont été décrites comme menaçantes que ce soit aux États-Unis, en Europe et même en France. Si ce stéréotype est effectivement présent dans nos sociétés, nous ne devrions pas avoir de mal à

reproduire ce type de résultats dans un paradigme SMT. Cela constitue le but de notre première étude : répliquer les résultats classiques du paradigme SMT avec des visages maghrébins.

1.5 Étude 1 (préliminaire) : réplication des résultats originaux

Cette première étude avait en réalité deux objectifs : le premier était de répliquer les résultats traditionnellement obtenus au sein de la SMT (i.e., effet SMT) en changeant de groupe cible. Suivant en cela les résultats classiques associés à la perception des Maghrébins en France, nous faisions l'hypothèse d'un jugement de menace plus important pour les dessins cibles précédés d'amorces de visages maghrébins par rapport à des amorces de visages blancs (ou neutres). Toutefois, l'intérêt de cette tâche se situe moins dans l'analyse des effets « bruts » (i.e., jugements de menace) que dans l'étude du modèle multinomial associé à cette tâche. Par conséquent, le second objectif de cette étude était de s'assurer que le modèle multinomial classiquement associé à la SMT (i.e., modèle SAC-dominant) s'ajuste de manière satisfaisante aux données d'une SMT utilisant des visages maghrébins. Si ces deux conditions sont réunies, la SMT constituera effectivement une tâche appropriée pour répondre aux multiples problématiques de cette thèse.

1.5.1 Méthode

Participants et plan expérimental. Cinquante-six participants ont pris part à cette étude. Parmi eux, trois participants ont été écartés, car s'estimant de culture principale autre que française (i.e, culture maghrébine). De plus, deux autres participants ont été écartés, car donnant systématiquement la même réponse à l'ensemble des essais (répondant plus de 98% du temps « Plus menaçant » ou « Moins menaçant »). Ainsi, 51 participants ont été gardés pour l'analyse finale des données (37 femmes, 14 hommes, $M_{\text{âge}}$ = 19.84, $ET_{\text{âge}}$ = 1.52). Les participants étaient des étudiants de l'Université Grenoble Alpes en L1 Psychologie ou hors cursus de Psychologie et recevaient des crédits de cours ou 10 euros en compensation. Cette étude répliquant au plus près l'étude de Krieglmeyer et Sherman (2012), nous avons reproduit

le plan expérimental original de l'étude, à savoir un plan de type 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant), entièrement intra-sujet.

Matériel et procédure. Les participants venaient au laboratoire pour une étude de 20 minutes sur la « Formation d'impression suboptimale ». À leur arrivée au laboratoire, un expérimentateur les installait dans des box individuels. Les participants signaient un formulaire de consentement et l'étude débutait. Ils étaient informés qu'ils allaient voir des dessins de visage apparaître très rapidement à l'écran et que leur tâche serait de se former une impression concernant ces visages sur la dimension de menace. En outre, ils étaient avertis du fait que de véritables photos de visages apparaitraient à l'écran juste avant la présentation des dessins cibles. Il leur était demandé de ne pas se concentrer sur ces photos de visages, mais de toutefois leur prêter attention, car des questions leur seraient posées à la fin de l'étude. Immédiatement après l'énumération de ces consignes, la tâche SMT débutait.

Les passations s'effectuaient sur un écran de 24 pouces, avec le logiciel E-prime 2.0®. Au début de chaque essai (Figure 4), une croix de fixation apparaissait à l'écran (500ms) et était remplacée par une photo d'un visage blanc, d'un visage maghrébin ou d'un visage entièrement grisé (i.e., visage neutre ; 150ms). Après un bref écran blanc (50ms), un dessin flouté de visage apparaissait (100ms). Enfin, un masque comportant l'inscription « Moins menaçant Appuyez sur D » et « Plus menaçant Appuyez sur K » remplaçait le dessin de visage, et restait à l'écran jusqu'à ce que le participant donne sa réponse. Les amorces étaient des photos de 24 visages maghrébins, 24 visages blancs ou 24 visages neutres. Les visages maghrébins et blancs étaient issus de la base de données CaNAFF (Courset et al., 2018). La CaNAFF est une base de données de visages émotionnellement neutres évalués par des observateurs externes notamment au niveau de leur caractère prototypique. Afin de maximiser nos chances de trouver les effets attendus, nous avons sélectionné les visages les plus prototypiques des catégories correspondantes. Les 24 visages neutres provenaient de l'étude originale de Krieglmeyer et Sherman (2012) et consistaient en 24 contours grisés de visages. Les 48 dessins de visages cibles (i.e., 24 dessins de visages non menaçants et 24 dessins de visages menaçants) provenaient eux aussi de l'étude originale de Krieglmeyer et Sherman (2012). Ces dessins de visages avaient été créés à partir du générateur de visages d'Oosterhof et Todorov (2008). Ces visages étaient émotionnellement neutres et différaient par rapport à certains traits du visage (e.g., forme des yeux, sourcils, nez). Ils avaient été sélectionnés pour être deux écarts types au-dessus (visages menaçants) ou deux écarts types

en dessous (visages non menaçants) de la moyenne des visages créés pour varier sur la dimension de menace.

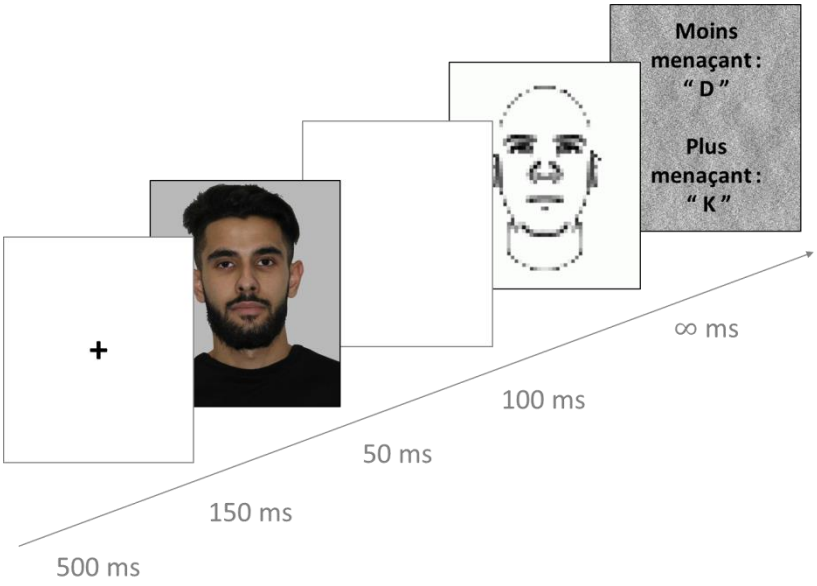


Figure 4. Représentation visuelle d’un essai SMT.

La tâche commençait par deux phases d’entrainements de six essais chacune. Lors de la première phase d’entrainement, seuls les dessins cibles étaient présentés. La seconde phase d’entrainement avait une structure similaire à la phase test et comportait à la fois les dessins cibles et les amorces de visages. Après ces deux phases d’entrainement, la phase test débutait. Cette phase test comprenait 144 essais. L’ensemble des essais étaient aléatorisés pour qu’il y ait exactement le même nombre de cibles menaçantes (72) et de cibles non menaçantes (72) appareillées avec les amorces Maghrébin (24), Blanc (24) et Neutre (24). La phase test comprenait deux blocs de 72 essais séparés par une courte pause. La proportion d’amorces et de cibles de chaque catégorie pouvait être différente entre les deux blocs, la proportion finale de la phase test correspondant toutefois à la description ci-dessus. À la fin de la tâche, les participants répondaient à trois questions relatives aux visages présentés (non gardées pour les analyses), ainsi qu’à des questions démographiques (i.e., âge, genre). Pour finir, les participants étaient débriefés, gratifiés puis remerciés.

1.5.2 Résultats et discussion

Jugement de menace. L'ensemble des analyses présentées dans ce manuscrit ont été réalisées avec le logiciel de statistique R (version 3.6.0, 2019 ; R Core Team, 2013). Pour analyser les réponses des participants, nous avons effectué une ANOVA à mesures répétées 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) avec la somme des réponses « Plus menaçant » en variable dépendante. Dans cette étude et dans la suite du manuscrit, nous ne présenterons pas l'effet omnibus associé à la variable « amorce », mais directement la décomposition en contrastes des modalités de cette variable (e.g., différence entre les visages maghrébins et les visages blancs), ceci afin de correspondre au mieux à nos hypothèses.

Tout d'abord nous avons obtenu un effet principal du type de cible, les participants jugeant les cibles menaçantes comme étant effectivement plus menaçantes (M = 34.75, ET = 8.64) que les cibles non menaçantes (M = 24.22, ET = 9.90), $t(50) = 7.33$, $p < .001$, $\eta^2_p = .52$. De plus, comme nous en faisions l'hypothèse, les participants ont jugé les dessins cibles comme plus menaçants quand ils étaient précédés de visages maghrébins (M = 23.51, ET = 7.93) plutôt que de visages blancs (M = 19.43, ET = 8.31), $t(50) = 2.35$, $p = .02$, $\eta^2_p = .10$ (i.e., effet SMT) ou de visages neutres (M = 16.02, ET = 11.32), $t(50) = 3.81$, $p < .001$, $\eta^2_p = .23$ (Figure 5). À l'inverse, les participants n'ont pas jugé les dessins cibles comme significativement plus menaçants quand ils étaient précédés de visages blancs plutôt que de visages neutres, $t(50) = 1.77$, $p = .08$, $\eta^2_p = .06$.

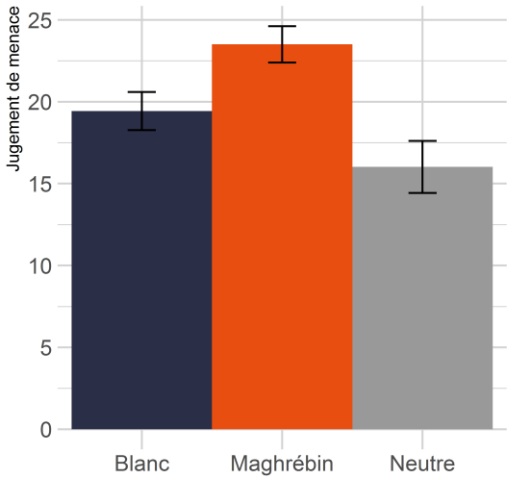


Figure 5. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin, Neutre). Les barres d'erreurs représentent les erreurs standards.

Modèle Multinomial. Là encore, l'ensemble des analyses présentées dans ce manuscrit ont été réalisées avec le logiciel de statistique R (version 3.6.0, 2019 ; R Core Team, 2013) ainsi que le package TreeBUGS (version 1.4.5 ; Heck et al., 2018) permettant l'utilisation de la méthode *latent-trait* (Klauer, 2010). Pour rappel, cette méthode emploie une approche hiérarchique prenant en compte une possible hétérogénéité des paramètres d'un participant à l'autre. Pour tester cette hétérogénéité, Heck et collaborateurs (2018) proposent l'utilisation d'un test χ^2 permettant de quantifier la proportion d'hétérogénéité entre participants, une forte valeur de χ^2 étant diagnostic d'une hétérogénéité importante. Nos données étant effectivement hétérogènes, $\chi^2(300) = 1433.41$, $p < .001$, l'utilisation de la méthode *latent-trait* s'impose[3].

Nous avons voulu estimer l'adéquation aux données du modèle SAC-dominant. De manière surprenante, la différence entre les valeurs T_1 prédites ($M_{pred} = 0.99$) et observées ($M_{obs} = 0.36$) était significative, $p = .005$. Cependant, comme nous l'avons spécifié plus tôt, la valeur de statistique de test T_1 se base sur la statistique de Pearson χ^2. Par conséquent, la

[3] Afin de simplifier la lecture du manuscrit, nous avons uniquement reporté ce test d'hétérogénéité pour l'Étude 1. Il est à noter, cependant, que cette hétérogénéité était présente dans l'ensemble de nos études, justifiant ainsi l'utilisation de la méthode *latent-trait* pour l'ensemble du manuscrit.

faible adéquation du modèle aux données peut potentiellement s'expliquer par le nombre élevé d'observations (n_{obs} = 7344) pouvant maximiser la puissance statistique de détecter même de faibles différences statistiques. C'est pourquoi nous avons calculé la taille d'effet *w* (Cohen, 1992) qui, associée aux valeurs observées T_1, nous a permis d'estimer l'adéquation du modèle aux données tout en contrôlant la taille d'échantillon. La taille d'effet obtenue, *w* = .05, était inférieur à la valeur seuil classiquement admise correspondant à un effet de petite taille, *w* = .10. L'adéquation de notre modèle SAC-dominant aux données était donc acceptable.[4]

À titre indicatif, nous pouvons interpréter les valeurs des paramètres du modèle, nous familiarisant ainsi avec la signification des différentes valeurs de ces paramètres. Pour rappel, la valeur de chaque paramètre est estimée sur une échelle de probabilité (de 0 à 1). Si l'on observe le modèle SAC-dominant, une valeur SAC = 1 signifie que le stéréotype a été systématiquement activé au cours des essais. À l'inverse, une valeur de SAC = 0 (i.e., 1-SAC = 1) signifie que le stéréotype n'a jamais été activé au cours des essais. Ainsi, et comme on pouvait s'y attendre compte tenu de l'effet SMT obtenu, les participants ont la plupart du temps activé leur stéréotype, SAC = .72, $IC_{95\%}$[.45 ; .91]. La signification des valeurs du paramètre SAP est différente. En effet, une valeur SAP = 1 signifie que le stéréotype a systématiquement été appliqué. Cependant, à l'inverse du paramètre SAC, une valeur SAP = 0 (i.e., 1-SAP = 1) signifie que le stéréotype a systématiquement été corrigé. En conséquence, une valeur SAP = .50 signifie que, sur la totalité des essais, le stéréotype a autant été appliqué que corrigé. Ainsi, bien qu'il faille rester prudent quant à ces conclusions (l'intervalle de crédibilité à 95% de SAP comprenant la valeur .50), les participants ont globalement plus appliqué leur stéréotype qu'ils ne l'ont corrigé, SAP = .57, $IC_{95\%}$[.49 ; .66]. Les valeurs du paramètre D s'interprètent comme celles du paramètre SAC, c'est-à-dire que D = 1 signifie que le participant a systématiquement détecté la menace de la cible au cours des essais, alors que D = 0 (i.e., 1-D = 1) signifie que le participant n'a jamais détecté la cible. Comme la structure de la tâche pouvait le laisser supposer et étant donné la nature des cibles utilisées, les participants n'ont que très peu détecté la menace réelle des visages, D = .10, $IC_{95\%}$[.04 ; .16]. Enfin, le paramètre G s'interprète comme le paramètre SAP, c'est-à-dire que G = 1

[4] Il est à noter que la méthode d'analyse *latent-trait* peut également nous permettre de calculer un indice T_2 correspondant à la différence entre la covariance prédite par le modèle et la covariance effective. Toutefois, nous avons préféré reporter uniquement l'indice T_1, car celui-ci correspond plus directement aux indices d'adéquation classiques reportés dans la littérature (e.g., G^2). De plus, à l'inverse de l'indice T_1, il n'existe à ce jour pas de méthode permettant d'estimer l'adéquation T_2 tout en contrôlant la taille d'échantillon.

signifie que le participant a systématiquement répondu « Plus menaçant » quand les autres paramètres n'étaient pas à l'œuvre et G = 0 (i.e., 1-G = 1) signifie que le participant a systématiquement répondu « Moins menaçant » que les autres paramètres n'étaient pas à l'œuvre. Une valeur G = .50 signifie donc que le participant a répondu entièrement au hasard lorsque les autres paramètres n'étaient pas à l'œuvre. Ainsi, lorsque les participants n'ont pas activé leur stéréotype et n'ont pas détecté la menace réelle des visages, ceux-ci ont majoritairement répondu « Moins menaçant », G = .25, $IC_{95\%}$[.16 ; .36]. Ce biais vers la réponse « Moins menaçant » est conforme à ce que l'on peut voir dans d'autres études et correspond à une tendance générale à percevoir l'ensemble des cibles comme étant peu menaçantes (e.g., Krieglmeyer & Sherman, 2012 ; Rees, Ma, et al., 2019 ; Rees, Rivers, et al., 2019)

Pour finir, il nous semblait important d'observer si le modèle SAC-dominant, bien que s'ajustant correctement aux données, était effectivement le meilleur modèle disponible pour expliquer nos données. Nous mentionnions en effet l'existence de différents modèles, testés par Krieglmeyer et Sherman (2012), pouvant théoriquement être envisagés. Parmi ces modèles, le modèle D-dominant proposait que la détection de la cible guide la réponse du participant, et ce, même si son stéréotype avait été activé. Afin de comparer l'adéquation de ces différents modèles aux données, nous avons utilisé le critère WAIC, le modèle ayant la valeur WAIC la plus faible étant considéré comme le modèle s'ajustant le mieux aux données. De façon surprenante, nous avons obtenu une meilleure adéquation aux données de la part du modèle D-dominant (WAIC = 1514.55) que du modèle SAC-dominant (WAIC = 1547.27). Ainsi, il semble que nos données, bien que pouvant être décrites de manière adéquate par le modèle classique associé à la SMT (i.e., modèle SAC-dominant), le soient encore davantage par cet autre modèle. Nous investiguerons en détail au Chapitre 3 de possibles explications à ce résultat.

1.6 Une thèse en trois parties autour de la mesure du stéréotype

Rappelons notre questionnement initial. Nous étions à la recherche d'une mesure adéquate nous permettant d'étudier les processus cognitifs menant au phénomène de stéréotypisation. Nous nous intéressions au stéréotype de menace et à son association aux populations maghrébines, notamment en contexte d'attentats. Notre raisonnement théorique nous a naturellement conduits vers l'utilisation de mesures indirectes, car moins susceptible

de refléter un biais externe. Toutefois, les critiques de ces mesures nous ont poussés à être prudents et à considérer le fait que ces mesures n'étaient pas *process pure* et notamment soumises à une possibilité de contrôle de la part du participant. De ce fait, nous avons choisi de nous tourner vers l'utilisation de modèles multinomiaux et avons fini par sélectionner la SMT pour étudier la perception négative des individus maghrébins. La réplication de la SMT classique avec des visages maghrébins et des visages blancs nous a confortés dans ce choix. Les résultats obtenus lors de cette réplication ont en réalité dirigé ce travail de thèse vers trois grands axes, représentés par les trois parties empiriques de cette thèse.

La première partie empirique (Chapitre 2) était consacrée à l'étude d'un possible amalgame Maghrébin-terroriste, c'est-à-dire à la perception négative des individus d'origine maghrébine en contexte d'attentats. Dans la continuité d'un projet de recherche commencé en amont de ce travail de thèse, nous avons voulu étudier l'impact des attentats par le biais de mesures indirectes. Au cours de trois études ayant chacune une opérationnalisation différente du contexte d'attentats, nous avons voulu mettre en évidence les précautions à prendre dans l'interprétation de l'impact (ou de l'absence d'impact) de ce contexte sur les jugements individuels. Plus précisément, nous avons testé la possibilité que l'absence d'impact des attentats sur ces jugements puisse refléter une dynamique contraire des processus sous-tendant la réponse du participant. Ce chapitre était pour nous l'occasion, au travers d'une problématique d'actualité, de se questionner théoriquement de l'impact de nos manipulations expérimentales sur les processus à l'origine de la réponse du participant. Plus globalement, nous voulions montrer qu'avec l'aide de mesures indirectes et de leur modèle multinomial associé, nous pouvions être en mesure de tester directement l'influence de nos manipulations sur les processus cognitifs à l'origine de la réponse individuelle.

La deuxième partie empirique de cette thèse (Chapitre 3) se voulait dans la continuité de ce questionnement. En effet, là où le Chapitre 2 avait pour objectif la prise en compte des processus sous-tendant la réponse du participant, le Chapitre 3 avait pour but la prise en compte de la relation conditionnelle entre ces processus. Plus particulièrement, nous avons voulu mettre en évidence le fait que le choix des stimuli composant notre tâche (i.e., amorce, cible) pouvait influencer non seulement la contribution relative des principaux processus cognitifs, mais aussi l'organisation même de ces processus. En d'autres termes, nous voulions mettre en évidence la possibilité qu'en fonction des stimuli choisis, les participants prenant part à notre étude utilisent en priorité l'information de l'amorce ou au contraire de la cible pour formuler leur réponse. En outre, nous avons investigué la possibilité que des différences

interindividuelles quant à l'attitude raciale auto-rapportée puissent être liées à cette relation conditionnelle entre processus. Par le biais de ce chapitre, nous avons voulu aller plus loin dans la mise en évidence des précautions à prendre dans l'utilisation de ce type de mesures. Nous avons montré, pour la première fois, l'importance de la prise en compte de facteurs liés à la tâche et surtout de facteurs liés aux participants dans l'étude de la relation conditionnelle entre processus.

Enfin, la dernière partie empirique de cette thèse (Chapitre 4) a été pour nous l'occasion, dans un premier temps, de laisser de côté les processus cognitifs à l'origine de la réponse du participant pour nous intéresser plus directement à cette réponse en elle-même. Plus particulièrement, nous nous sommes questionnés sur le caractère généralisable de l'effet SMT et sur les explications théoriques quant à l'origine de cet effet. D'abord, nous avons testé la possibilité que notre effet ne provienne pas d'une réactivation du stéréotype de menace lié à la catégorie des Maghrébins, mais plus simplement à des exemplaires de cette catégorie (i.e., des visages maghrébins) pouvant être intrinsèquement menaçants. Ensuite, nous avons testé la possibilité que notre effet reproduit systématiquement au sein de notre université puisse ne pas être reproduit dans un contexte culturel différent (i.e., États-Unis). Enfin, nous avons testé la possibilité que cet effet puisse ne pas être reproduit auprès d'un échantillon reflétant la population associée au stéréotype négatif (i.e., individus d'origine maghrébine). Ces deux dernières études avaient ainsi pour but de questionner à la fois la composante culturelle du stéréotype, mais aussi son caractère personnel ou au contraire extra-personnel. Plus globalement, nous voulions montrer une fois encore l'intérêt d'une prise en compte du caractère multi-processuel des réponses des participants au sein de la littérature utilisant des mesures indirectes.

Chapitre 2 – Impact des attentats sur la perception des Maghrébins

Le 07 janvier 2015, l'attentat de Charlie Hebdo sonne le départ d'une longue liste d'attaques sur notre sol. Entre janvier 2015 et décembre 2018, plus d'une dizaine d'attentats terroristes revendiqués par l'État Islamique frappe la France. Parmi les plus marquants, Charlie Hebdo donc, mais aussi l'attaque du Bataclan et les attentats de Paris le 13 novembre 2015, ainsi que l'attentat du 14 juillet 2016, sur la promenade des Anglais à Nice. La multiplication des attaques sur notre sol ne serait pas le fruit du hasard. Pour le politologue français Gilles Kepel (2016), l'État Islamique, en répétant ses attaques, poursuit un but précis. Inspirée par Abou Moussab al-Souri, la stratégie de l'État Islamique consisterait à perpétrer de multiples attentats afin de monter les habitants des pays occidentaux les uns contre les autres. Cette stratégie impliquerait que, suite à ces attaques, les habitants originaires des pays visés cèdent à un amalgame entre les terroristes et la communauté d'individus d'origine maghrébine, pouvant mener à une augmentation de la discrimination envers ces derniers et, à terme, une augmentation de la radicalisation de ces populations ostracisées. Existe-t-il un amalgame Maghrébin-terroriste ? Autrement dit, colorons-nous la perception que nous avons des populations maghrébines avec la menace caractéristique des terroristes ? Plus précisément, qu'en est-il de l'attitude captée par des mesures indirectes, censée être plus difficilement contrôlable, mais pouvant être le reflet de cet amalgame ? Ce chapitre aura pour objectif de mettre en lumière l'impact d'un contexte d'attentats sur les processus à l'origine de l'attitude raciale anti-Maghrébins captée par une mesure indirecte.

2.1 L'impact des attentats, ce que nous savons

Nous avons vu au Chapitre 1 que la littérature ne manquait pas d'études traitants des attitudes raciales envers les individus d'origine arabo-musulmane (e.g., Dasgupta et al., 2009 ; Essien et al., 2017 ; Lyons et al., 2010 ; Mange et al., 2012 ; Unkelbach et al., 2008). Nous avons commencé à voir que les attaques terroristes passées (e.g., 11 septembre 2001) pouvaient jouer un rôle dans la perception négative des populations arabo-musulmanes (Fetzer & Soper, 2003). Globalement, les conséquences des attentats sur ces populations sont bien documentées (e.g., Arciszewski et al., 2009). Ainsi, après avoir été frappés par des

attaques terroristes, les individus de pays touchés afficheraient plus d'attachement et d'identification à l'endogroupe (Moskalenko et al., 2006 ; Orehek et al., 2010). De plus, ces individus auraient un besoin de clôture plus prégnant (Orehek et al., 2010), ainsi qu'une adhérence plus forte à la théorie de justification du système, marquée ici par une volonté d'avoir une société plus ordonnée (Ullrich & Cohrs, 2007). D'un point de vue émotionnel, enfin, ces attaques peuvent entretenir un sentiment de stress (Schuster et al., 2001), mais surtout de peur quant à leurs conséquences ou la possibilité qu'elle se produisent de nouveau (Dumont et al., 2003).

Ces évènements et le ressenti qu'ils peuvent susciter ne sont pas sans conséquence sur l'attitude raciale à l'égard des populations arabo-musulmanes. Par exemple, Oswald (2005) a montré que la perception de peur liée à la menace terroriste, couplée à l'orientation à la dominance sociale ainsi qu'au patriotisme étaient prédicteurs d'attitudes négatives envers ces populations que ce soit au niveau du stéréotype, du préjugé, mais aussi de la discrimination auto-rapportée des participants. Globalement, il semble y avoir un grand nombre d'études soulignant l'impact des attentats sur l'augmentation de l'attitude raciale négative envers les Maghrébins (e.g., Bar-Tal & Labin, 2001 ; Das et al., 2009 ; Echebarria-Echabe & Fernández-Guede, 2006 ; Persson & Musher-Eizenman, 2005 ; Van de Vyver et al., 2016). Ainsi, Persson et Musher-Eizenman (2005) ont montré qu'après les attentats de New York en 2001, les participants rapportaient une attitude plus négative envers les individus arabo-musulmans par rapport aux individus noirs, pourtant classiquement discriminés. De même, Echebarria-Echabe et Fernández-Guede (2006), en contactant leurs participants avant et après les attentats de Madrid en 2004, ont rapporté une augmentation du préjugé négatif envers les populations maghrébines à la suite de ces attaques. Enfin, Van de Vyver et collaborateurs (2016) en contactant leurs participants six semaines avant et un mois après les attentats de Londres en 2005, ont là encore mis en évidence un préjugé plus important envers les populations musulmanes après les attentats.

En outre, plusieurs études laissent entrevoir la possibilité d'un amalgame Maghrébin-terroriste, c'est-à-dire l'augmentation de la menace que peuvent représenter les Maghrébins par le biais de l'association traditionnellement faite avec les terroristes, ces derniers étant majoritairement d'origine maghrébine et/ou de religion musulmane. Tout d'abord, des études ont effectivement pu mettre en évidence une perception de menace associée aux attaques terroristes (e.g., Doosje et al., 2009 ; Kastenmüller et al., 2011) et l'association de cette menace avec les populations arabo-musulmanes (Doosje et al., 2009). De plus, les Musulmans

sont décrits comme supportant davantage les attentats terroristes que d'autres groupes (i.e., les Chrétiens ; Doosje et al., 2009 ; Fischer et al., 2007). Horry et Wright (2009), en utilisant une mesure indirecte (i.e., *dot probe task*), ont montré que des participants anxieux préalablement amorcés (en subliminal) par des mots liés au terrorisme avaient le regard plus attiré vers des visages arabo-musulmans que vers des visages blancs. La littérature ayant préalablement montré que des visages perçus comme menaçants pouvaient, dans ce type de tâche, capter le regard (Richeson & Trawalter, 2008; Trawalter et al., 2008), les auteurs de cette étude ont conclu que l'amorçage du concept de terrorisme avait coloré négativement la perception des Maghrébins, ceux-ci devenant plus menaçants.

En France, des études ont été entreprises après les premières attaques (i.e., Charlie Hebdo, le 07 janvier 2015). Globalement, les résultats sont similaires à ceux rapportés précédemment. Cohu et collaborateurs (2016) ont mis en évidence une augmentation du préjugé anti-Maghrébins juste après les attaques (pendant deux semaines) par rapport à une mesure de préjugé réalisée avant l'attentat. Dans la même veine, Pelletier et Drozda-Senkowska (2016) ont mis en lumière une perception de menace terroriste à un niveau individuel et collectif se maintenant dans le temps. Enfin, Nugier et collaborateurs (2016) ont montré que le préjugé anti-Maghrébins et la perception de menace associée augmentaient immédiatement après l'attaque terroriste. En somme, l'ensemble des études présentées converge vers un impact des attentats sur la perception négative des individus d'origine maghrébine.

2.2 Utilisation de mesures indirectes, une absence de résultats ?

En réalité, nous avons commencé à étudier cette problématique avant la réalisation de ce travail de thèse. Le but de ce projet, que nous nommerons « Projet Amalgame », était de s'intéresser non pas au jugement auto-rapporté individuel (voir Nugier & Guimond, 2016), mais au jugement plus spontané, capté par des tâches indirectes, mais pouvant également influencer nos comportements (e.g., Gawronski & De Houwer, 2014). Plus particulièrement, et suivant en cela certains auteurs (e.g., Horry & Wright, 2009), nous voulions investiguer si des processus automatiques (i.e., principalement non intentionnels) pouvaient être à l'origine, en contexte d'attentat, d'une augmentation de la perception de menace associée aux Maghrébins et plus généralement d'une attitude raciale négative envers ces populations.

Nous nous sommes basés sur la littérature traitant de l'importance du contexte dans la réactivation du contenu mental lié aux attitudes raciales (e.g., Barden et al., 2004 ; Maddux et al., 2005 ; Wittenbrink et al., 2001). Comme nous avons pu le voir au Chapitre 1, des études ont montré que, selon le contexte dans lequel les différents stimuli étaient perçus (e.g., visage noir devant une image d'église ou de prison ; Maddux et al., 2005), le contenu mental réactivé pouvait varier. Le modèle APE propose que la réactivation du contenu mental lié à l'objet attitudinal rencontré (e.g., visage noir) ne se ferait pas selon un processus en tout ou rien (Gawronski & Bodenhausen, 2011). Au contraire, en se basant sur la littérature traitant des modèles connexionnistes (e.g., Smith, 1996), le modèle APE propose qu'un nombre limité d'associations peuvent être réactivées en fonction du contexte dans lequel celles-ci ont été créées. Nous pensions ainsi que, si en contexte d'attentats une association Maghrébin-terroriste avait effectivement été créée, présenter par la suite un stimulus maghrébin dans un contexte rappelant les attentats pourrait réactiver cette association.

Par le biais de deux mesures indirectes (i.e., tâche de mémoire des visages, Ackerman et al., 2006 ; tâche d'approche-évitement ; Rougier et al., 2020) nous avons voulu observer si, comme nous en faisions l'hypothèse, les participants amorcés par un rappel d'attentats (en comparaison à une condition contrôle) avaient une attitude raciale plus négative envers des individus maghrébins. Contrairement à ce que nous avions prédit, nous n'avons pas pu conclure à l'impact d'un rappel d'attentat sur la perception plus négative des individus maghrébins. Naturellement, cela ne signifie pas pour autant qu'il n'y a pas eu d'impact des attentats sur la perception des Maghrébins. Comme nous allons tenter de le montrer maintenant, l'absence de différences entre nos conditions ne signifie en aucun cas l'absence d'impact de notre manipulation sur les processus sous-tendant ces mesures. En développant notre réflexion, nous aurons ainsi pour objectif, au sein de ce chapitre, de montrer l'un des intérêts de l'utilisation de modèles multinomiaux dans l'étude des attitudes raciales.

En accord avec les modèles duaux classiques des attitudes raciales (e.g., Fazio & Olson, 2014) nous pensions, au moment de réaliser ces études, qu'utiliser des mesures indirectes pourrait nous permettre de contourner les biais liés à un possible contrôle de la part du participant. Or, comme nous l'avons exposé grandement au Chapitre 1, les mesures indirectes peuvent être elles aussi, bien que dans une moindre mesure, le reflet de ce type de biais (e.g., Fiedler & Bluemke, 2005 ; Hütter & Klauer, 2016). Serait-il alors possible que les participants aient effectivement une attitude anti-Maghrébins réactivée involontairement en contexte d'attentats, mais que ceux-ci puissent contrôler leur réponse pour inhiber l'expression de

cette attitude ? Si nous regardons les études réalisées après les attentats de Charlie Hebdo, nous pouvons trouver des indices allant dans ce sens. Effectivement, dans l'étude de Nugier et collaborateurs (2016), des participants exposés à la norme culturelle de l'égalité citoyenne se sentaient moins menacés et ressentaient moins d'hostilité par rapport aux populations maghrébines. De plus, il faut nous souvenir de la « marche républicaine » ayant eu lieu en réponse aux attentats de Charlie. Bien que certains auteurs y voient le reflet des préjugés de la société envers les populations maghrébines (i.e., Emmanuel Todd ; voir Nugier & Guimond, 2016), des études ont mis en évidence, au contraire, l'aspect égalitaire et antidiscriminatoire de ces manifestations (Mayer & Tiberj, 2016 ; Zerhouni et al., 2016). Il a ainsi été montré que, parmi ces manifestants, nous trouvions une surreprésentation de populations jeunes, urbaines, instruites, de gauches et tolérantes (Mayer & Tiberj, 2016). De plus, contrairement à ce qui a été évoqué précédemment, les villes où les préjugés « implicites » (c'est-à-dire pour les auteurs, tels que mesurés par l'IAT) étaient les plus importants étaient les villes où le taux de manifestation a été le plus faible (Zerhouni et al., 2016).

En somme, nous pouvons penser que le sentiment de menace associé aux Maghrébins peut être accompagné d'un autre type de sentiment, plus égalitaire et antidiscriminatoire. Comme le soulignent Nugier et collaborateurs (2016), il semblerait que les individus ne tombent pas forcément dans le « piège » du terrorisme. Il est certes possible qu'en contexte d'attentats, un amalgame Maghrébin-terroriste soit présent, c'est-à-dire que les individus réactivent de manière non intentionnelle leur crainte des Maghrébins. Toutefois, comme le montrent les études que nous venons d'évoquer, ces individus sont conscients de ce type de risque. Pour être convaincu, il suffit d'observer le nombre d'occurrences au cours du temps du mot « amalgame » parmi les recherches Google (Figure 6). Immédiatement après les principaux attentats, le nombre de recherches s'envole. Il semble ainsi que les principaux attentats s'accompagnent systématiquement de l'évocation du concept d'amalgame. L'illustration la plus récente fait suite au meurtre d'un enseignant de collège assassiné pour avoir montré des caricatures de Charlie Hebdo à ses élèves. Le lendemain de l'attaque, le « hashtag » *#pasdamalgame* figurait en haut des hasthag les plus utilisés sur Twitter, signe d'un important relai de ce message dans les principaux médias.

Figure 6. Évolution de l'intérêt de recherche Google pour le mot « Amalgame » au cours du temps (entre janvier 2004 et septembre 2020). Les résultats reflètent la proportion de recherches de ce mot par rapport à la proportion la plus élevée (valeur de 100). Ainsi, une valeur de 50 signifie que le mot « Amalgame » a été moitié moins recherché que la proportion la plus élevée.

Pour résumer, en contexte d'attentats deux types de processus contraires peuvent être à l'œuvre. D'un côté, nous pouvons penser que la multiplication des attentats mène à un amalgame Maghrébin-terroriste. En contexte d'attentats, nous pourrions associer involontairement la menace classiquement liée aux terroristes aux individus maghrébins en général. À l'inverse, nous pourrions dans le même temps avoir conscience de cet amalgame et de ce fait inhiber celui-ci. Ainsi, dans nos études préliminaires, il est possible que la réactivation du contenu mental négatif associé aux Maghrébins en condition Attentat ait été volontairement inhibée par nos participants, aboutissant ainsi à une absence de différence significative entre notre condition Attentat et Contrôle. Utiliser un paradigme SMT couplé à son modèle multinomial nous a permis d'investiguer directement ce type de questionnement.

2.3 Étude 2 : utilisation de la SMT suite à un rappel d'attentats

L'Étude 2 avait pour but de tester la possibilité, qu'en contexte d'attentats, deux processus contraires puissent être à l'œuvre. Utiliser la SMT nous semblait idéal pour répondre à cette question. En effet, grâce au modèle multinomial associé nous pouvions observer l'impact d'un rappel d'attentats sur les processus sous-tendant la réponse du participant. Tout d'abord, à cause de la présence d'un possible amalgame Maghrébin-

terroriste, nous faisions l'hypothèse d'une activation plus importante du stéréotype « Maghrébins = Menace » (SAC) pour nos participants amorcés par un rappel d'attentats par rapport à nos participants en condition contrôle. À l'inverse, en conséquence de la volonté affichée de ne pas faire d'amalgame, nous faisions l'hypothèse d'une application moins importante de ce stéréotype (SAP) pour nos participants amorcés par un rappel d'attentats par rapport à nos participants en condition contrôle. En somme, nous pensions qu'en contexte d'attentats, l'application moindre du stéréotype allait compenser une plus forte activation du stéréotype. Nous insistons sur le fait qu'en conséquence, nous n'avions pas d'hypothèse sur les réponses individuelles (i.e., jugement de menace), l'effet SMT pouvant être équivalent entre les conditions et ainsi masquer une éventuelle contribution différente des processus sous-jacents.

2.3.1 **Méthode**

Participant et plan expérimental. Dans le cadre des modèles multinomiaux utilisant la méthode *latent-trait*, nous ne pouvons pas utiliser la taille d'effet *w* pour calculer la puissance nécessaire afin que notre modèle s'adapte correctement aux données (l'indice T_1 ne suivant pas une distribution χ^2). Nous ne pouvons également pas estimer la puissance nécessaire pour observer l'effet de notre manipulation expérimentale sur la valeur des paramètres. Par conséquent, avec l'objectif d'obtenir une puissance statistique suffisante afin de minimiser la probabilité d'obtenir des erreurs de Type II, nous avons fixé notre taille d'échantillon minimale à 50 participants par modalité pour notre variable inter-sujets « contexte » à deux modalités (i.e., Attentat vs. Contrôle). Cent quarante-cinq participants ont pris part à cette étude. Aucun des participants recrutés n'avait pris part à l'Étude 1. Parmi ces 145 participants, 12 ont été écartés, car ils avaient reporté le but réel de l'étude en remplissant le questionnaire de suspicion. Douze autres participants ont été écartés, car ils s'estimaient de culture principale autre que française. Deux participants ont été écartés, car ils avaient admis ne pas avoir lu correctement le texte constituant l'amorce. Enfin, deux participants ont été écartés, car ils avaient donné systématiquement la même réponse à l'ensemble des essais (répondant plus de 98% du temps « Plus menaçant » ou « Moins menaçant »). Au final, 116 participants ont été gardés pour les analyses (85 femmes, 31 hommes, $M_{\text{âge}}$ = 20.02, $ET_{\text{âge}}$ = 1.91). Les participants étaient des étudiants de l'Université Grenoble Alpes en L1 Psychologie ou hors cursus de Psychologie et recevaient des crédits de cours ou 10 euros en compensation. Le plan

expérimental de cette étude était de type 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (contexte : Attentat vs. Contrôle), le dernier facteur étant en inter-sujets. Le pré-enregistrement de cette étude est disponible sur la plateforme OSF (https://osf.io/49wz6).

Matériel et procédure. Les participants venaient au laboratoire pour une double étude de 30 minutes sur « les déterminants de la mémoire et la formation d'impression ». À leur arrivée, un expérimentateur les installait dans des box individuels. Les participants signaient un formulaire de consentement et l'étude débutait. Ils étaient informés qu'ils allaient participer à deux études distinctes. La première étude était présentée comme une tâche de mémoire. Il leur était demandé de lire un texte retraçant un évènement marquant, leur tâche étant de mémoriser le maximum d'informations possibles afin de répondre à quelques questions à la fin de la passation. Entre la phase de mémorisation et la phase de rappel, ils devaient réaliser la seconde étude. Celle-ci consistait en une tâche de formation d'impression (en réalité la tâche SMT) présentée comme une tâche distractive. Les consignes pour cette tâche étaient identiques à celles décrites Étude 1. Immédiatement après l'énumération de ces consignes, la tâche de mémorisation débutait.

Les passations s'effectuaient sur un écran de 24 pouces, avec le logiciel E-prime 2.0®. Les participants avaient cinq minutes pour retenir un maximum d'informations par rapport à un texte affiché à l'écran et dont la nature différait entre les deux conditions. Dans la condition Attentat, les participants devaient lire un rappel chronologique précis des évènements survenus lors de l'attaque du Bataclan (Paris ; 13 novembre 2013 ; Annexe 1). Afin de contrôler la saillance de la mortalité, celle-ci pouvant être liée à une augmentation des préjugés envers les minorités (e.g., Das et al., 2009), notre condition Contrôle impliquait également une tuerie de masse. Ainsi, dans la condition Contrôle les participants devaient lire un rappel chronologique précis de l'attaque du lycée de Columbine, aux États-Unis (20 avril 1999 ; Annexe 1). À la fin des cinq minutes, la tâche SMT débutait. La procédure était en tout point identique à l'Étude 1. À la fin de la passation, les participants répondaient à trois questions concernant le texte qu'ils avaient lu au début de l'étude, correspondant ainsi à la phase de rappel (ces données n'ont pas été conservées pour les analyses). Ensuite, ils devaient donner le degré d'attention qu'ils avaient apporté à la lecture du texte, celui-ci allant de 1 « Pas du tout attentivement » à 5 « Très attentivement ». Cette question nous permettait de nous assurer que les participants avaient été exposés de façon adéquate à la procédure

d'amorçage. Nous avons exclu les participants répondant à moins de 3 à cette échelle (i.e., deux participants). Les participants répondaient ensuite à trois questions relatives aux visages présentés ainsi qu'à des questions démographiques (i.e., âge, genre). À la fin de l'étude, un questionnaire de suspicion était distribué. Celui-ci avait pour but de vérifier que les participants étaient aveugles aux hypothèses réelles de l'étude, particulièrement sur le lien pouvant exister entre la lecture du texte à mémoriser et la tâche de distraction (i.e., tâche SMT ; 12 participants ont été exclus des analyses). Pour finir, les participants étaient débriefés, gratifiés et remerciés.

2.3.2 **Résultats**

Jugement de menace. Pour analyser les réponses des participants, nous avons effectué une ANOVA à plan mixte 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (contexte : Attentat vs. Contrôle) avec la somme des réponses « Plus menaçant » en variable dépendante. Tout d'abord nous avons obtenu un effet significatif du type de cible, les participants jugeant les cibles menaçantes comme étant effectivement plus menaçantes (M = 34.87, ET = 8.45) que les cibles non menaçantes (M = 24.84, ET = 8.75), $t(114) = 10.33$, $p < .001$, $\eta^2_p = .48$. De plus, comme dans notre première étude, les participants ont jugé les dessins cibles comme plus menaçants quand ils étaient précédés de visages maghrébins (M = 25.23, ET = 8.83) plutôt que de visages blancs (M = 18.88, ET = 7.20), $t(114) = 5.98$, $p < .001$, $\eta^2_p = .24$ (i.e., effet SMT) ou de visages neutres (M = 16.02, ET = 11.32), $t(114) = 7.59$, $p < .001$, $\eta^2_p = .34$ (Figure 7). Nous n'avons pas obtenu d'effet d'interaction significatif entre le contexte et le contraste opposant les amorces Maghrébin et les amorces Blanc, $t(114) = 0.34$, $p = .73$, $\eta^2_p = .001$, et entre le contexte et le contraste opposant les amorces Maghrébin et les amorces Neutre, $t(114) = 0.00$, $p = .99$, $\eta^2_p < .001$. Enfin, et de façon surprenante, les participants ont jugé les dessins cibles comme plus menaçants quand ceux-ci étaient précédés de visages blancs plutôt que de visages neutres, $t(114) = 3.37$, $p = .001$, $\eta^2_p = .09$, cet effet n'étant pas modéré de manière significative par le contexte, $t(114) = 0.38$, $p = .70$, $\eta^2_p = .001$.

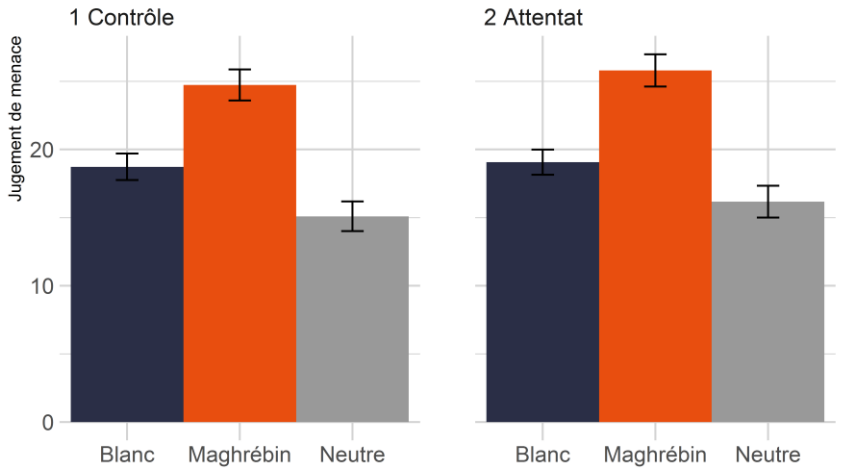


Figure 7. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin, Neutre) et de la condition expérimentale (Contrôle vs. Attentat). Les barres d'erreurs représentent les erreurs standards.

Modèle multinomial. En préambule de la description de nos analyses, nous souhaitons revenir sur le meilleur moyen de tester nos hypothèses. Initialement, et comme décrit dans notre pré-enregistrement, nous avons entrepris de tester nos hypothèses sur le modèle SAC-dominant classique, c'est-à-dire un modèle à quatre paramètres : SAC, SAP, D et G. Cela signifie que le paramètre SAC correspondait à la fois à l'activation du stéréotype « Maghrébins = Menace », mais aussi du stéréotype « Blancs = Non menace ». De la même manière, le paramètre SAP correspondait à la fois à l'application du stéréotype « Maghrébins = Menace », mais aussi à celle du stéréotype « Blancs = Non menace ». Or, pour rappel, nous faisions l'hypothèse qu'en contexte d'attentats les participants devaient plus activer leur stéréotype « Maghrébins = Menace », mais moins appliquer ce stéréotype. En d'autres termes, nous n'avions aucune hypothèse à propos des visages blancs. Un modèle pouvant estimer des paramètres SAC et SAP différents pour les visages maghrébins et les visages blancs semble, a posteriori, plus approprié[5]. C'est la raison pour laquelle, dans la description des résultats de

[5] En réalité, l'une des raisons pour laquelle nous n'avons pas utilisé ce modèle en premier lieu est liée au type d'analyse que nous réalisions à l'origine. Au début de ce travail de thèse, nous réalisions nos analyses sur des données agrégées et utilisions l'indice d'ajustement G^2. Pour être utilisé, ce type d'indice nécessite un nombre de degrés de liberté non nul, c'est-à-dire un nombre de conditions expérimentales (ici six) supérieur au nombre de paramètres (ici quatre). Or, l'indice d'ajustement T_1 peut

cette étude et de l'étude suivante, nous présenterons les résultats à la fois avec le modèle classique SAC-dominant à quatre paramètres (i.e., SAC, SAP, D, G), mais aussi avec le modèle SAC-dominant à six paramètres (SAC_{Mg}, SAC_{Bl}, SAP_{Mg}, SAP_{Bl}, D, G).

Commençons avec le modèle classique. Nous avons tout d'abord estimé la qualité d'ajustement T_1 du modèle aux données de nos participants en condition Attentat et en condition Contrôle. Dans la condition Attentat, la différence entre les valeurs T_1 prédites (M_{pred} = 0.10) et observées (M_{obs} = 0.22) était certes tendancielle, p = .08, mais la taille d'effet associée était acceptable, w = .04. Dans la condition Contrôle, la différence entre les valeurs T_1 prédites (M_{pred} = 0.09) et observées (M_{obs} = 0.17) était non significative, p = .14, la taille d'effet associée étant là aussi acceptable, w = .03. Notre modèle s'ajustait donc bien aux données.

Nous avons ensuite observé si nos paramètres différaient en condition Attentat par rapport à la condition Contrôle (Figure 8). De manière descriptive, le paramètre SAC semblait plus élevé en condition Attentat (M_{SAC} = .63, $IC_{95\%}$[.22 ; .93]) qu'en condition Contrôle (M_{SAC} = .58, $IC_{95\%}$[.22 ; .88]). Toutefois, contrairement à nos hypothèses, cette différence n'était pas significative, ΔSAC = .042, $IC_{95\%}$[-.48 ; .51], p_B = .42[6]. De la même manière, le paramètre SAP semblait descriptivement légèrement plus élevé en condition Attentat (M_{SAP} = .66, $IC_{95\%}$[.57 ; .77]) qu'en condition Contrôle (M_{SAP} = .62, $IC_{95\%}$[.53 ; .72]). Une fois encore, cette différence était toutefois non significative, ΔSAP = .047, $IC_{95\%}$[-.09 ; .18], p_B = .24. Enfin, bien que nous n'ayons pas d'hypothèse sur les paramètres D et G, il est à noter que nous n'avons pas trouvé de différences pour ces deux paramètres entre les conditions Attentat (M_D = .15, $IC_{95\%}$[.09 ; .20] ; M_G = .28, $IC_{95\%}$[.22 ; .35]) et Contrôle (M_D = .12, $IC_{95\%}$[.07 ; .17] ; M_G = .26, $IC_{95\%}$[.19 ; .32]), les différences entre ces deux conditions étant non

être calculé avec un nombre de conditions expérimentales égales au nombre de paramètres. En conséquence, nous n'avons pris conscience de la possibilité d'utiliser un modèle à six paramètres que plus tardivement.

[6] Ce test correspond au test de la différence entre le paramètre SAC en condition Attentat et le paramètre SAC en condition Contrôle. Si l'intervalle de crédibilité contient la valeur zéro, il n'y a pas de différence « significative » entre nos deux conditions. La p-valeur associée est de type bayésien, et correspond à la proportion de cas pour lesquels le paramètre SAC en condition Attentat est inférieur au paramètre SAC en condition Contrôle. Nous pouvons rapprocher ce test d'un test statistique unilatéral, dans le sens où si p_B < .05, SAC(Attentat) est significativement supérieur à SAC(Contrôle). Au contraire, si p_B > .95, SAC(Attentat) est significativement inférieur à SAC(Contrôle). Enfin, nous avons conscience qu'utiliser le terme « significatif » dans un cadre bayésien est un abus de langage. Toutefois, nous utiliserons tout de même ce terme afin de simplifier au maximum la lecture des analyses.

significatives pour le paramètre D, ΔD = .027, $IC_{95\%}$[-.05 ; .10], p_B = .24, comme pour le paramètre G, ΔG = .022, $IC_{95\%}$[-.07 ; .12], p_B = .31.

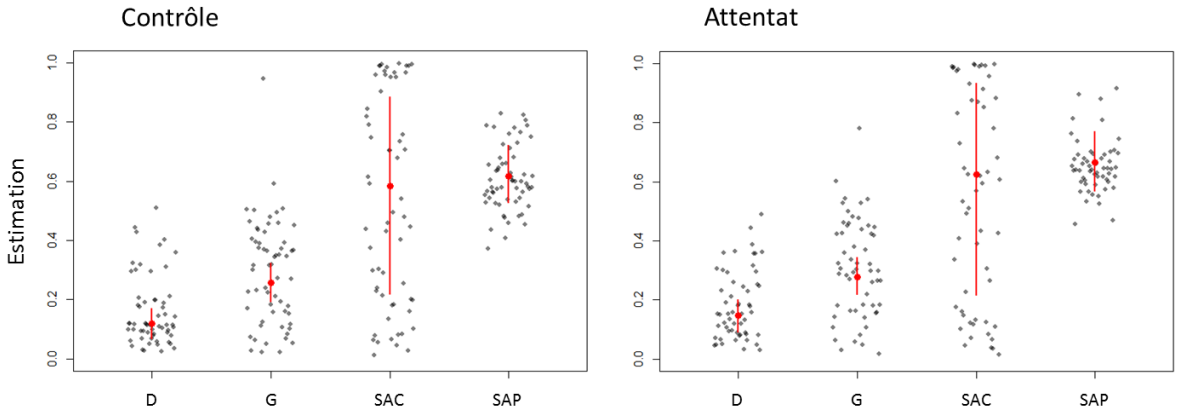


Figure 8. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un modèle SAC-dominant à quatre paramètres.

Comme nous l'avons exposé précédemment, nos hypothèses se situaient plus précisément au niveau de l'impact de nos conditions sur la perception des visages maghrébins. Utiliser un modèle multinomial séparant les paramètres SAC et SAP en fonction du type d'amorce pouvait ainsi s'avérer plus pertinent. Concernant l'ajustement du modèle aux données de la condition Attentat, la différence entre les valeurs T_1 prédites (M_{pred} = 0.10) et observées (M_{obs} = 0.10) était non significative, p = .46, la taille d'effet associée étant également acceptable, w = .03. Il en était de même pour la condition Contrôle, où la différence entre les valeurs T_1 prédites (M_{pred} = 0.09) et observées (M_{obs} = 0.08) était là encore non significative, p = .59, et la taille effet associée acceptable, w = .02.

Commençons par nos paramètres d'intérêt, à savoir les paramètres liés aux visages maghrébins. Une nouvelle fois, les résultats n'ont pas confirmé nos hypothèses (Figure 9). En effet, le paramètre SAC_{Mg} ne différait pas en condition Attentat (M_{SAC} = .41, $IC_{95\%}$[.25 ; .62]) par rapport à la condition contrôle (M_{SAC} = .41, $IC_{95\%}$[.22 ; .65]), ΔSAC_{Mg} = .00, $IC_{95\%}$[-.27 ; .28], p_B = .50. Il en est de même pour le paramètre SAP_{Mg}, celui-ci n'étant pas différent entre les conditions Attentat (M_{SAP} = .82, $IC_{95\%}$[.54 ; .99]) et Contrôle (M_{SAP} = .80, $IC_{95\%}$[.58 ; .99]), ΔSAP_{Mg} = .02, $IC_{95\%}$[-.34 ; .33], p_B = .44. Il est toutefois envisageable que notre manipulation

ait impacté les paramètres liés aux visages blancs. Le paramètre SAC_{Bl} était descriptivement plus important en condition Attentat (M_{SAC} = .48, $IC_{95\%}$[.14 ; .89]) qu'en condition contrôle (M_{SAC} = .39, $IC_{95\%}$[.18 ; .78]), mais cette différence était toutefois non significative, ΔSAC_{Bl} = .09, $IC_{95\%}$[-.39 ; .58], p_B = .36. De même, le paramètre SAP_{Bl} était descriptivement plus important en condition Attentat (M_{SAP} = .49, $IC_{95\%}$[.35 ; .61]) qu'en condition contrôle (M_{SAP} = .38, $IC_{95\%}$[.07 ; .61]), mais là encore, la différence entre les deux conditions était non significative, ΔSAP_{Bl} = .10, $IC_{95\%}$[-.17 ; .44], p_B = .26.

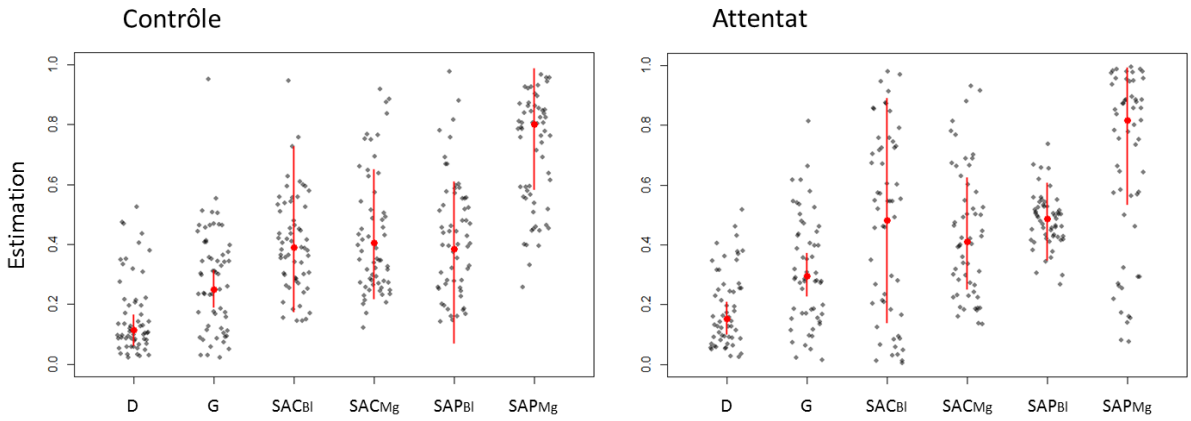


Figure 9. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un modèle SAC-dominant à six paramètres. SAC_{Bl} et SAP_{Bl} = paramètres pour les visages blancs, SAC_{Mg} et SAP_{Mg} = paramètres pour les visages maghrébins.

2.3.3 **Discussion**

Dans cette deuxième étude, nous avions pour objectif de tester l'impact d'un rappel d'attentats sur les principaux processus à l'origine du jugement de menace associé aux individus maghrébins. Contrairement à nos hypothèses, nous n'avons pas pu mettre en évidence une activation du stéréotype (SAC) plus importante suite à un rappel d'attentats. En décomposant le paramètre SAC en fonction des visages maghrébins et blancs, nous n'avons observé descriptivement qu'une augmentation du paramètre SAC associé aux visages blancs. De la même manière, nous n'avons pas pu mettre en évidence une application du stéréotype

(SAP) moins importante suite à un rappel d'attentats. Là encore, en décomposant le paramètre SAP en fonction des visages maghrébins et blancs, nous n'avons observé, de façon descriptive, qu'une augmentation du paramètre SAP associé aux visages blancs.

Ainsi, si l'effet que nous attendions existe réellement, il semblerait que notre manipulation du contexte d'attentats n'ait pas permis sa mise en évidence. Une première explication peut venir de la présence d'un potentiel « effet plafond ». Il est probable qu'au moment de la réalisation de cette étude (i.e., automne 2017), et suite à la multiplication des attentats entre 2015 et 2017, les participants aient constamment à l'esprit le contexte des attentats. Notre manipulation expérimentale n'a peut-être pas été assez forte pour permettre une activation du contenu mental associé aux attentats plus importante que celle que nous avons au quotidien. En d'autres termes, nos participants en condition Attentat comme nos participants en condition Contrôle ont peut-être été tout autant influencés par le contexte latent relatif aux attentats.

Nous pouvons également envisager le rôle du type d'induction que nous avons choisi dans notre condition Contrôle. En effet, comme dit précédemment, nous voulions éviter tout effet confondu de la saillance de la mortalité (e.g., Das et al., 2009). C'est la raison pour laquelle nous avons choisis une condition Contrôle se rapprochant au plus proche de notre condition Attentat (i.e., tuerie de masse) et s'éloignant du prototype de l'attaque terroriste classique (i.e., adolescents américains). Il est cependant envisageable que, dans le contexte de la multiplication des attentats entre 2015 et 2017, nos participants aient perçu la tuerie de Columbine comme étant similaire aux attaques revendiquées par l'État Islamique, activant ainsi également le contenu mental lié aux attentats.

Enfin, il semble descriptivement que l'impact du contexte dans la condition Attentat ait été plus important (bien que non significatif) sur les processus liés aux visages blancs (SAC_{Bl}, SAP_{Bl}) par rapport aux processus liés aux visages maghrébins (SAC_{Mg}, SAP_{Mg}). Cette tendance surprenante peut éventuellement s'expliquer par le type d'induction relatif cette fois-ci à notre amorçage des attentats. Pour rappel, notre condition Attentat consistait en un rappel chronologique détaillé de l'attaque du Bataclan. Or, ce déroulé chronologique de l'attaque, bien que se concentrant majoritairement sur le parcours des terroristes, insistait également sur le sort réservé aux victimes (e.g., « dans la fosse, les spectateurs pris au piège se couchent au sol, feignant d'être morts »). En conséquence, l'amorçage des attentats n'a peut-être pas renforcé le stéréotype « Maghrébins = Menace » comme nous en faisions l'hypothèse, mais davantage le stéréotype « Blancs = Non menace ».

En résumé, l'ensemble de ces critiques converge vers la nécessité d'améliorer notre manipulation du contexte. Par conséquent, nous avons fait le choix de répliquer cette deuxième étude avec une manipulation différente du contexte. En changeant cette manipulation, nous avions pour objectif de maximiser la probabilité d'observer des différences entre notre condition Attentat et notre condition Contrôle.

2.4 Étude 3 : utilisation de la SMT dans un contexte d'attentats

Le but principal de cette troisième étude était de répliquer notre étude précédente en changeant notre manipulation du contexte. Plus particulièrement, le but de cette réplication était d'augmenter la saillance du contexte d'attentats afin de maximiser nos chances de trouver un effet si celui-ci existe. Suivants en cela certaines des études ayant montré un effet du contexte sur les attitudes raciales (e.g., Barden et al., 2004 ; Wittenbrink et al., 2001), nous n'avons pas manipulé le contexte avant la présentation des stimuli, mais en parallèle de leur présentation. En d'autres termes, nous avons choisi d'inclure, en arrière-plan de la tâche SMT, une photo rappelant les attaques du Bataclan (condition Attentat) ou au contraire une photo d'un bâtiment neutre (condition Contrôle). Nous faisions une fois encore l'hypothèse que nos participants en condition Attentat auraient un SAC plus important, mais un SAP moins important que nos participants en condition Contrôle. Enfin, et pour la première fois dans nos différentes tâches SMT, nous avons introduit de manière exploratoire une échelle d'attitude raciale négative envers les Maghrébins, l'Échelle de Préjugés Généralisés (Dambrun & Guimond, 2001)[7]. Dans d'autres travaux (Rougier et al., 2020), nous avons mis en évidence une corrélation entre cette échelle et une mesure indirecte d'attitudes raciales. C'est pourquoi, nous voulions tester si cette mesure directe pouvait être corrélée, d'une part, à l'effet SMT classique, et d'autre part aux différents paramètres du modèle et notamment au paramètre d'application du stéréotype (SAP).

[7] L'Échelle de Préjugés Généralisés correspond davantage à la définition que nous avons donnée de l'attitude raciale plutôt que du préjugé (composante de l'attitude). Toutefois, dans un but de simplicité et pour suivre la terminologie des créateurs de cette échelle, nous parlerons d'« échelle de préjugé » et de « score de préjugé ».

2.4.1 Méthode

Participants et plan expérimental. Une fois encore dans le but de minimiser le risque d'obtenir des erreurs de Type II, nous avons fixé notre taille d'échantillon à 50 participants minimum par modalités de notre variable inter-sujets « contexte » à deux modalités. Cent-treize participants ont pris part à cette étude. Aucun des participants recrutés n'avait pris part à l'une de nos études précédentes. Parmi les participants recrutés, 12 ont été exclus des analyses, car s'estimant de culture principale autre que française. Cent-un participants ont été gardé pour les analyses (89 femmes, 12 hommes, $M_{\text{âge}}$ = 19.24, $ET_{\text{âge}}$ = 2.00). Les participants étaient des étudiants de Licence (L1, L2) en Sciences humaines à l'Université Grenoble Alpes et recevaient des crédits de cours en compensation. Le plan expérimental de cette étude était identique à l'étude précédente, excepté l'ajout de la mesure directe d'attitude raciale, et donc de type 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (contexte : Attentat vs. Contrôle) x score de préjugé, les deux dernières variables étant en inter-sujet. Le pré-enregistrement de cette étude est disponible sur la plateforme OSF (https://osf.io/ajhzn).

Matériel et procédure. Les participants venaient au laboratoire pour une étude de 20 minutes sur la « Formation d'impression suboptimale ». À leur arrivée, un expérimentateur les installait dans des box individuels. Les participants signaient un formulaire de consentement et l'étude débutait. Les consignes présentées à l'écran étaient similaires à celles de l'Étude 1. Pour rappel, les participants étaient informés qu'ils allaient voir des dessins de visage apparaître très rapidement à l'écran et que leur tâche serait de se former une impression par rapport à ces visages sur la dimension de menace. Ils étaient avertis du fait que de véritables photos de visages apparaitraient à l'écran juste avant la présentation des dessins cibles. Il leur était demandé de ne pas se concentrer sur ces photos de visages, mais de toutefois leur prêter attention, car des questions leur seraient posées à la fin de l'étude. Immédiatement après l'énumération de ces consignes, la tâche SMT débutait.

Les passations s'effectuaient sur un écran de 24 pouces, avec le logiciel E-prime 3.0®. À une exception près, la tâche SMT était identique à celle présentée dans nos deux premières études. L'unique changement se situait au niveau de notre manipulation expérimentale du contexte. Ainsi, les participants passaient l'intégralité de la tâche avec un arrière-plan bien précis. Dans la condition Attentat, les participants passaient la SMT avec en arrière-plan une

image de la devanture du Bataclan après les attaques (Figure 10, droite). À l'opposé, dans la condition Contrôle, les participants passaient la SMT avec en arrière-plan une image de la devanture d'un Musée (Figure 10, gauche). Comme précédemment, à la fin de la SMT, les participants répondaient à trois questions relatives aux visages présentés ainsi qu'à des questions démographiques (i.e., âge, genre). Enfin, les participants devaient remplir une version papier de l'Échelle de Préjugés Généralisés (Annexe 4). Cette échelle de 15 items (α = .87) était présentée comme une « liste de propositions portant sur la diversité et l'immigration » (e.g., « Il est insensé d'attribuer aux Algériens et aux Marocains les problèmes économiques de la France »). Les participants devaient donner leur degré d'accord sur une échelle de Likert allant de 1 (« Pas du tout d'accord ») à 7 (« Tout à fait d'accord »). Pour insister sur l'anonymat du questionnaire, les participants devaient déposer leur questionnaire complété dans une urne à la toute fin de l'étude. Ils étaient enfin débriefés, gratifiés et remerciés.

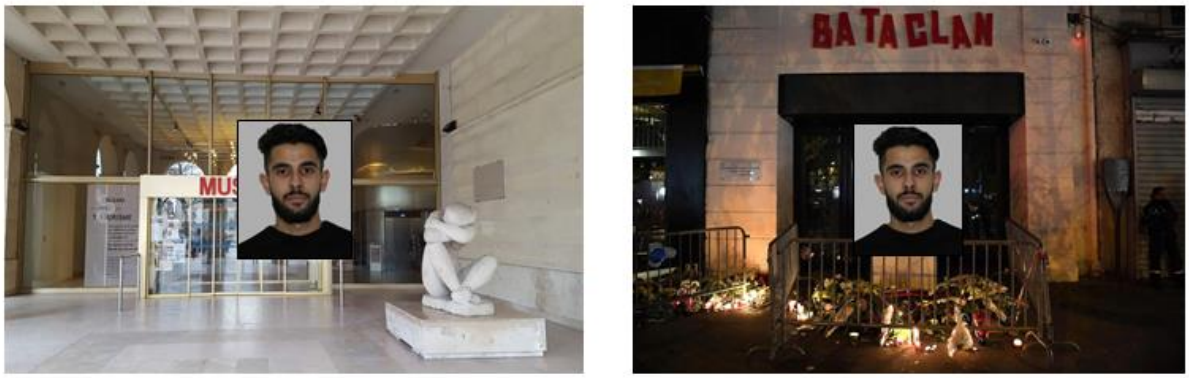


Figure 10. Manipulation expérimentale de l'arrière-plan. À gauche, devanture d'un musée (condition Contrôle). À droite, devanture du Bataclan (condition Attentat).

2.4.2 **Résultats**

Jugement de menace. Pour analyser les réponses des participants, nous avons effectué une ANOVA à plan mixte 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 3 (contexte : Attentat vs. Contrôle) avec la somme des réponses « Plus menaçant » en variable dépendante. Nous avons tout d'abord répliqué l'effet de la cible, les participants jugeant les cibles menaçantes comme étant plus menaçantes

(*M* = 32.00, *ET* = 11.14) que les cibles non menaçantes (*M* = 21.23, *ET* = 8.73), *t*(99) = 10.14, *p* < .001, η^2_p = .51. De plus, les participants ont une fois encore jugé les cibles comme étant plus menaçantes quand elles étaient précédées de visages maghrébins (*M* = 20.37, *ET* = 8.19) plutôt que de visages blancs (*M* = 18.09, *ET* = 8.19), *t*(99) = 2.46, *p* = .02, η^2_p = .06 (i.e., effet SMT), ou de visages neutres (*M* = 14.78, *ET* = 9.42), *t*(99) = 4.55, *p* < .001, η^2_p = .17. Nous n'avons pas obtenu d'effet d'interaction significatif entre le contexte et le contraste opposant les amorces Maghrébin et Blanc, *t*(99) = 0.62, *p* = 0.54, η^2_p = .004, et entre le contexte et le contraste opposant les amorces Maghrébin et Neutre, *t*(114) = 0.65, *p* = .52, η^2_p = .004 (Figure 11). Pour finir, les participants ont une fois encore jugé les dessins cibles comme étant plus menaçants quand ceux-ci étaient précédés de visages blancs plutôt que de visages neutres, *t*(99) = 3.59, *p* < .001, η^2_p = .12, cet effet n'étant pas modéré de manière significative par le contexte, *t*(99) = 0.24, *p* = .81, η^2_p = .001.

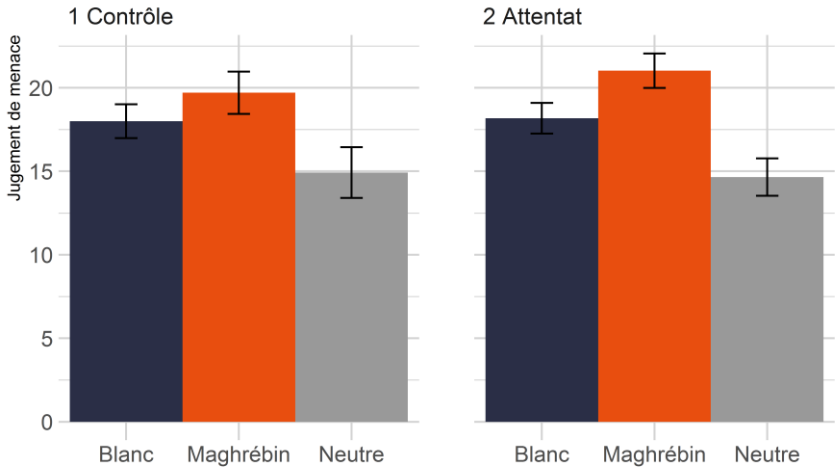


Figure 11. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin, Neutre) et de la condition expérimentale (Contrôle vs. Attentat). Les barres d'erreurs représentent les erreurs standards.

Pour analyser de manière exploratoire le lien entre le jugement de menace et la mesure directe d'attitude raciale anti-Maghrébins (i.e., échelle de préjugé), nous avons effectué une régression linéaire de type 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (contexte : Attentat vs. Contrôle) x score de préjugé,

avec la somme des réponses « Plus menaçant » en variable dépendante. L'effet de la cible n'était pas modéré de manière significative par le niveau de préjugé individuel, b = 1.04, $t(97)$ = 0.81, p = .42 η^2_p = .01. À l'inverse, nous avons obtenu un effet d'interaction entre le score de préjugé et le contraste opposant les amorces Maghrébin et Blanc. Ainsi, plus les participants avaient un score de préjugé élevé et plus la différence de jugement entre les amorces Maghrébin et Blanc était importante, b = 3.20, $t(97)$ = 3.02, p = .003, η^2_p = .09. La différence de jugement entre les amorces Maghrébin et Neutre n'était pas significativement modérée par le score de préjugé individuel, b = 0.82, $t(97)$ = 0.56, p = .60, η^2_p = .003. De manière surprenante, nous avons obtenu un effet d'interaction significatif entre le score de préjugé et le contraste opposant les amorces Blanc et Neutre. Ainsi, plus les participants avaient un score de préjugé élevé et moins la différence de jugement entre les amorces Blanc et Neutre était importante, b = -2.37, $t(97)$ = -2.18, p = .03, η^2_p = .05. Il est à noter, enfin, que nous n'avons obtenu aucun impact significatif du contexte sur ces différents effets (ts < 1).

Modèle multinomial. Une fois encore, nous avons analysé les données à la fois avec le modèle SAC-dominant classique à quatre paramètres (SAC, SAP, D, G) et avec le modèle SAC-dominant à six paramètres (SAC_{Mg}, SAC_{Bl}, SAP_{Mg}, SAP_{Bl}, D, G). Commençons avec le modèle classique. Concernant l'ajustement du modèle en condition Attentat, la différence entre les valeurs T_1 prédites (M_{pred} = 0.11) et observées (M_{obs} = 0.13) était non significative, p = .41, la taille d'effet associée étant acceptable, w = .03. De même, pour la condition Contrôle, la différence entre les valeurs T_1 prédites (M_{pred} = 0.10) et observées (M_{obs} = 0.13) était là encore non significative, p = .39, la taille d'effet associée étant également acceptable, w = .03.

Contrairement à nos hypothèses, mais descriptivement uniquement, le paramètre SAC semblait moins élevé en condition Attentat (M_{SAC} = .48, $IC_{95\%}$[.20 ; .74]) qu'en condition Contrôle (M_{SAC} = .58, $IC_{95\%}$[.32 ; .81]), la différence entre ces conditions étant toutefois non significative, ΔSAC = -.10, $IC_{95\%}$[-.47 ; .30], p_B = .71 (Figure 12). De plus, le paramètre SAP ne différait significativement pas entre nos participants en condition Attentat (M_{SAP} = .51, $IC_{95\%}$[.42 ; .60]) et en condition Contrôle (M_{SAP} = .51, $IC_{95\%}$[.44 ; .57]), ΔSAP = .01, $IC_{95\%}$[-.10 ; .12], p_B = .46. Enfin, nous n'avons pas observé de différences significatives pour les paramètres D et G entre les conditions Attentat (M_D = .16, $IC_{95\%}$[.10 ; .22] ; M_G = .23, $IC_{95\%}$[.18 ; .29]) et Contrôle (M_D = .14, $IC_{95\%}$[.08 ; .20] ; M_G = .21, $IC_{95\%}$[.12 ; .30]), les différences entre ces deux conditions étant non significatives pour le paramètre D, ΔD = .022, $IC_{95\%}$[-.06 ; .11], p_B = .30, comme pour le paramètre G, ΔG = .026, $IC_{95\%}$[-.09 ; .13], p_B = .30.

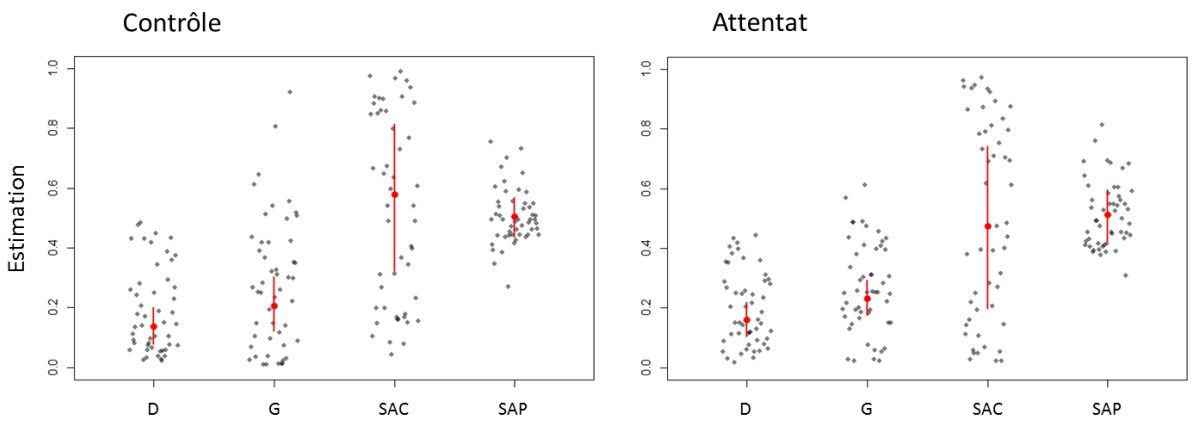


Figure 12. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un modèle SAC-dominant à quatre paramètres.

Afin d'observer le lien potentiel entre la mesure directe d'attitude raciale et les paramètres du modèle, nous avons extrait les valeurs individuelles de chaque paramètre et avons réalisé des régressions linéaires selon un plan expérimental de type 2 (contexte : Attentat vs. Contrôle) x score de préjugé, avec la moyenne de chaque paramètre en variable dépendante. Conformément à ce que l'on pouvait imaginer, nous avons observé un lien positif significatif entre le score de préjugé et le paramètre SAP, b = 0.03, $t(97)$ = 2.56, p = .01, η^2_p = .06. Ainsi, plus les participants avaient une attitude raciale négative envers les Maghrébins et plus ils appliquaient leur stéréotype « Maghrébins = Menace » / « Blancs = Non menace ». Aucun autre paramètre n'était corrélé à l'échelle de préjugé et nous n'avons pas observé d'impact significatif du contexte (ts <1). Nous discuterons de l'utilisation de cette mesure directe au Chapitre 3.

Passons au modèle SAC-dominant à six paramètres. L'ajustement du modèle aux données était bon. Effectivement, dans la condition Attentat, la différence entre les valeurs T_1 prédites (M_{pred} = 0.10) et observées (M_{obs} = 0.10) était non significative, p = .53, la taille d'effet associée étant également acceptable, w = .03. De même pour la condition Contrôle, la différence entre les valeurs T_1 prédites (M_{pred} = 0.10) et observées (M_{obs} = 0.08) était là encore non significative, p = .63, et la taille effet associée acceptable, w = .02.

Commençons par les paramètres relatifs aux visages maghrébins. Le paramètre SAC_{Mg} allait dans le sens attendu, celui-ci étant descriptivement plus élevé en condition Attentat (M_{SAC} = .65, $IC_{95\%}$[.17 ; .98]) qu'en condition Contrôle (M_{SAC} = .41, $IC_{95\%}$[.22 ; .65], Figure 13).

Toutefois, cette différence n'était pas significative, ΔSAC_{Mg} = .24, $IC_{95\%}$[-.30 ; .66], p_B = .18. De même, le paramètre SAP_{Mg} était descriptivement plus faible en condition Attentat (M_{SAP} = .53, $IC_{95\%}$[.39 ; .68]) qu'en condition contrôle (M_{SAP} = .68, $IC_{95\%}$[.44 ; .93]). Cependant, cette différence était là encore non significative, ΔSAP_{Mg} = -.15, $IC_{95\%}$[-.45 ; .14], p_B = .84. Une fois encore, il reste envisageable que notre manipulation ait impacté les paramètres associés aux visages blancs. Concernant le paramètre SAC_{Bl}, celui-ci était descriptivement moins important en condition Attentat (M_{SAC} = .28, $IC_{95\%}$[.14 ; .69]) qu'en condition Contrôle (M_{SAC} = .40, $IC_{95\%}$[.26 ; .59]). Mais une nouvelle fois cette différence n'était pas significative, ΔSAC_{Bl} = -.12, $IC_{95\%}$[-.38 ; .30], p_B = .82. Enfin, le paramètre SAP ne différait pas de manière significative en condition Attentat (M_{SAP} = .36, $IC_{95\%}$[.03 ; .67]) et en condition Contrôle (M_{SAP} = .33, $IC_{95\%}$[.12 ; .53]), ΔSAP_{Bl} = .03, $IC_{95\%}$[-.37 ; .41], p_B = .44.

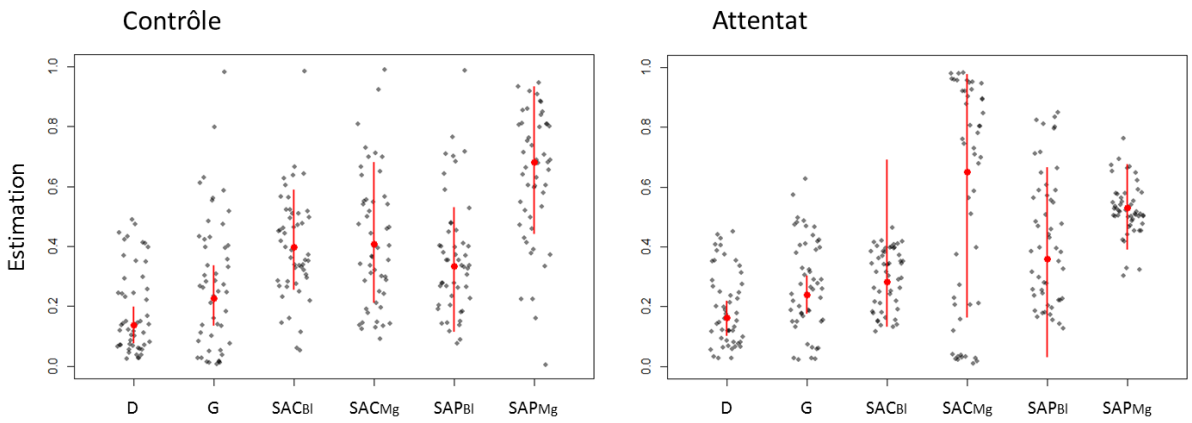


Figure 13. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un modèle SAC-dominant à six paramètres. SAC_{Bl} et SAP_{Bl} = paramètres pour les visages blancs, SAC_{Mg} et SAP_{Mg} = paramètres pour les visages maghrébins.

Pour finir et de manière surprenante, nous avons observé un lien entre le score de préjugé et le paramètre SAP_{Bl}, b = 0.05, $t(97)$ = 2.19, p = .03, η^2_p = .05, mais pas entre le score de préjugé et le paramètre SAP_{Mg}, b = 0.02, $t(97)$ = 1.22, p = .23, η^2_p = .02. Ainsi, plus les participants avaient une attitude raciale négative envers les Maghrébins et plus ils appliquaient leur stéréotype « Blancs = Non menace », mais pas leur stéréotype « Maghrébins = Menace ». Aucun autre paramètre n'était corrélé à l'échelle de préjugé (ts <1).

2.4.3 **Discussion**

Par le biais de cette troisième étude, nous voulions une nouvelle fois observer si un contexte d'attentats pouvait impacter les processus à l'origine de l'attitude raciale anti-Maghrébins. Contrairement à nos hypothèses, nous n'avons pas pu mettre en évidence une activation du stéréotype (SAC) plus importante et une application de ce stéréotype (SAP) moins importante en contexte d'attentats. Bien que non significatifs, nous avons toutefois pu observer descriptivement une activation du stéréotype « Maghrébins = Menace » plus élevée (i.e., SAC_{Mg} plus élevé) et une application de ce stéréotype moins élevée (i.e., SAP_{Mg} moins élevé) en condition Attentat par rapport à la condition Contrôle.

Ainsi, utiliser une induction expérimentale nous semblant a priori plus forte ne nous a pas permis de mettre en évidence l'impact des attentats sur les processus cognitifs sous-tendant le jugement anti-Maghrébins. Suivant le raisonnement qui nous a poussés à répliquer notre étude, il est possible que notre induction d'un contexte d'attentats ne soit une nouvelle fois pas assez forte. Dans la discussion de l'étude précédente, nous avions mentionné la possibilité d'un « effet plafond » résultant d'un contexte latent lié aux attentats, les attaques ayant été particulièrement nombreuses à la période de réalisation de ces études. Mais de manière plus globale, nous pouvons nous questionner quant à la faisabilité d'étudier cet objet de recherche de manière expérimentale. En effet, certains auteurs ont remis en cause l'étude des attentats en contexte de laboratoire (Dumont et al., 2003 ; Nugier & Guimond, 2016). Pour Dumont et collaborateurs (2003), certaines émotions comme la peur suscitée par les attentats sont difficilement réactivable en contexte expérimental. De la même manière, pour Nugier et Guimond (2016), il manquerait une « composante vitale » aux études en laboratoire. Pour ces auteurs, ces études ne permettent pas d'évoquer la menace réelle que peuvent ressentir les êtres humains confrontés à un évènement de forte ampleur comme une attaque terroriste. Même avec une induction expérimentale optimale, les individus ont conscience de ne pas être dans un contexte réel d'attaque terroriste. Afin de contourner ces possibles limites, des auteurs ont utilisé un contexte naturel d'attentats, c'est-à-dire qu'ils ont comparé les réponses de participants collectées avant et après une attaque terroriste (e.g., Das et al., 2009 ; Echebarria-Echabe & Fernández-Guede, 2006). C'est ce que nous avons également fait dans la quatrième et dernière étude de ce chapitre.

2.5 Étude 4 : application d'un Quad Model à des données récoltées au moment des attentats

La quatrième étude de ce chapitre avait pour objectif d'observer, non plus de manière provoquée (i.e., manipulation d'un contexte d'attentats), mais invoquée (i.e., utilisation de données collectées au moment d'attentats), le lien entre un contexte d'attentats et les processus sous-tendant l'attitude raciale anti-Maghrébins. Pour cela, nous avons utilisé des données IAT collectées en ligne sur la plateforme Project implicit (Nosek et al., 2002). Ce site internet permet la collecte en ligne de données relatives à un ensemble d'IAT (voir Nosek et al., 2002 ; Ratliff et al., 2020, pour une revue des différents paradigmes) ainsi que de mesures directes associées. Ces données étant accessibles librement aux chercheurs en faisant la demande, nous avons recueilli l'ensemble des données de participants français ayant passé un IAT-Maghrébins. Cependant, nous n'avons gardé que les données de participants ayant passé un IAT avant ou après les principales attaques terroristes en France depuis 2015. Nous avons appliqué un Quad Model à ces données afin de comparer la valeur des paramètres du modèle avant et après ces attaques. Conformément aux hypothèses développées dans les deux études précédentes, nous faisions l'hypothèse d'une activation du contenu mental négatif lié aux Maghrébins (i.e., paramètre AC) plus importante après les attentats qu'avant. De la même manière, nous faisions l'hypothèse d'une inhibition de ce contenu mental (OB) plus importante après les attentats qu'avant. Une fois encore, nous n'avions pas d'hypothèse sur le score IAT, la plus forte inhibition du contenu mental après les attaques pouvant masquer sa plus forte activation.

2.5.1 Méthode

Participants et plan expérimental. Comme nous venons de le décrire, les données collectées sont issues de l'IAT-Maghrébins, accessible sur le site internet Project Implicit (https://implicit.harvard.edu/ ; Nosek et al., 2002). Nous avons concentré nos analyses sur trois grandes attaques terroristes : l'attentat de Charlie Hebdo (7 janvier 2015 ; « IAT-Charlie »), l'attentat du Bataclan (13 novembre 2015 ; « IAT-Bataclan ») et l'attentat de Nice (14 juillet 2016 ; « IAT-Nice »). Pour chaque évènement, nous avons recueilli les données de participants français ayant passé l'IAT durant les trois mois précédents l'attaque (condition

« Avant ») ou durant les trois mois suivant l'attaque (condition « Après »)[8]. Ainsi, pour l'IAT-Charlie, 391 participants français ont passé un IAT-Maghrébins dans la période qui nous intéresse. Parmi eux, 298 étaient de couleur blanche[9] et ont de ce fait été retenu pour les analyses (175 femmes, 120 hommes, 3 non spécifiés, $M_{\text{âge}}$ = 29.92, $ET_{\text{âge}}$ = 11.39 ; n_{Avant} = 123, $n_{\text{Après}}$ = 175). Pour l'IAT-Bataclan, 296 participants français ont passé un IAT-Maghrébins dans la période correspondante. Deux-cent-vingt-quatre participants étaient de couleur blanche et donc retenus pour les analyses (132 femmes, 91 homme, 1 non spécifié, $M_{\text{âge}}$ = 27.52, $ET_{\text{âge}}$ = 10.40 ; n_{Avant} = 105, $n_{\text{Après}}$ = 119). Enfin, pour l'IAT-Nice, 220 participants français ont passé un IAT-Maghrébins dans la période de six mois qui nous intéresse. Parmi eux, 160 étaient de couleur blanche (89 femmes, 70 hommes, 1 non spécifié, $M_{\text{âge}}$ = 28.64, $ET_{\text{âge}}$ = 10.58 ; n_{Avant} = 89, $n_{\text{Après}}$ = 71). Pour chaque IAT, le plan expérimental était de type 2 x (type de prénom : Maghrébin vs. Français) x 2 (type de mot : Bon vs. Mauvais) x 2 (bloc : Compatible vs. Incompatible) x (contexte : Avant vs. Après), avec le dernier facteur en inter-sujet. Le pré-enregistrement de cette quatrième étude est disponible sur la plateforme OSF (https://osf.io/whd4g).

Matériel et procédure. Les participants se rendant sur le site internet Project Implicit avaient la possibilité de passer plusieurs « Tests d'associations implicites ». L'IAT-Maghrébins était décrit comme un IAT nécessitant la capacité de distinguer des noms d'origine française et maghrébine. Les participants commençaient la tâche en remplissant des informations démographiques (e.g., âge, genre). Les participants passaient ensuite une version IAT composée de 7 blocs (Greenwald et al., 1998). Par le biais de deux clés de réponse (i.e., « e » et « i »), les participants avaient pour consigne de catégoriser d'une part des prénoms comme étant maghrébins (6 prénoms ; e.g., « Djamel ») ou français (6 prénoms ; e.g., « Nicolas »), et d'autre part des mots comme étant bons (8 mots ; e.g., « joie ») ou mauvais (8 mots ; e.g., « affreux »). Il leur était demandé de catégoriser les différents stimuli aussi rapidement que

[8] Le choix de périodes de trois mois nous semblait être un bon compromis entre une période assez courte pour voir directement l'impact des attentats sans que celui-ci ne s'estompe, tout en ayant assez de participants pour réaliser nos analyses. Nous espérions ainsi avoir un minimum de 100 participants par modalités (Avant vs. Après) des différents IAT (il est cependant à noter que pour l'IAT-Nice, le nombre de participants retenu, 160, est inférieur au nombre espéré, 200).

[9] Nous avons gardé uniquement les participants de couleur blanche afin de correspondre au mieux à notre échantillon des études précédentes. Nous espérions ainsi nous centrer sur des participants de culture principale française. Cependant, nous insistons sur le fait que nous faisons ici un raccourci et que la correspondance n'est certainement pas parfaite.

possible en essayant de ne pas faire d'erreur. À chaque erreur, une croix rouge apparaissait à l'écran et les participants devaient à nouveau fournir la bonne réponse. Tout au long de la tâche, les labels associés aux clés de réponses étaient affichés en haut de l'écran, du côté correspond à la clé de réponse associée. Les prénoms et leurs labels associés étaient de couleur blanche, les mots et leurs labels associés étant de couleur verte.

Les participants devaient tout d'abord compléter deux blocs d'entrainement de 20 essais chacun. Le premier bloc consistait à catégoriser les différents mots comme étant bons ou mauvais et le second à catégoriser les différents prénoms comme étant maghrébins ou français. Les Blocs 3 et 4, respectivement de 20 et 40 essais, étaient des blocs critiques. Il était demandé au participant de catégoriser à la fois les types de prénoms et les types de mots. Les blocs pouvaient être compatibles (e.g., prénoms maghrébins et mots négatifs partageant une même clé de réponse) ou incompatibles (e.g., prénoms maghrébins et mots positifs partageant une même clé de réponse). Le Bloc 5 était à nouveau un bloc d'entrainement où les clés de réponses utilisées pour catégoriser les prénoms maghrébins et français étaient inversées. Enfin, les Blocs 6 et 7, respectivement de 20 et 40 essais, étaient à nouveau des blocs critiques reflétant l'inversion des clés de réponse du Bloc 5. Si les Blocs 3 et 4 étaient compatibles (incompatibles), les Blocs 6 et 7 étaient incompatibles (compatibles). L'ordre d'apparition des blocs critiques compatibles et incompatibles était contrebalancé d'un participant à l'autre.

2.5.2 **Résultats**

Scores IAT. Pour calculer les scores IAT, et à la différence des analyses relatives au Quad Model, nous ne nous sommes pas basés sur les fréquences de réponses correctes et incorrectes, mais sur les temps de réponse aux différents essais. Les scores IAT étaient calculés à partir de l'algorithme de Greenwald et collaborateurs (2003). Cet algorithme propose notamment que les données issues de l'ensemble des blocs critiques (i.e., Bloc 3 et Bloc 4 vs. Bloc 6 et Bloc 7) soient prises en compte dans les analyses. De plus, les essais pour lesquels le temps de réponse était supérieur à 10 000ms ainsi que les participants pour qui 10% des essais avaient un temps de réponse inférieur à 300ms ont été supprimés. La différence entre la moyenne de temps de réponse des essais compatibles et incompatibles divisé par l'écart type de l'ensemble des essais composait notre score IAT. Ainsi, plus le score

IAT était positif et élevé, et plus le biais anti-Maghrébins (et/ou pro-Français) était important. À l'inverse, plus le score IAT était négatif et faible, et plus le biais anti-Français (et/ou pro-Maghrébins) était important.

Nous avons tout d'abord effectué une analyse par type d'attentat, c'est-à-dire une analyse différente pour les participants passant l'IAT-Charlie, l'IAT-Bataclan et l'IAT-Nice. Ainsi, nous avons effectué trois ANOVA à 1 facteur (contexte : Avant vs. Après) avec le score IAT moyen en variable dépendante. Nous avons obtenu des résultats similaires pour nos trois IAT. Les participants passant l'IAT-Charlie avaient un score IAT moyen positif (M = 0.38 ; ET = 0.41) reflétant un biais anti-Maghrébins/pro-Français, $t(296) = 15.94$, $p < .001$, $\eta^2_p = .46$. Le lien entre le contexte (i.e., Avant vs. Après) et le score IAT était non significatif, $t(296) = 0.50$, $p = .61$, $\eta^2_p = .001$, le score IAT moyen n'étant pas différent avant (M = 0.37 ; ET = 0.39) et après l'attentat (M = 0.39 ; ET = 0.42). Les participants passant l'IAT-Bataclan avaient eux aussi un score IAT moyen positif (M = 0.45 ; ET = 0.39) reflétant là encore un biais anti-Maghrébins/pro-Français, $t(222) = 17.16$, $p < .001$, $\eta^2_p = .57$. Le lien entre le contexte et le score IAT était une fois encore non significatif, $t(222) = 1.41$, $p = .16$, $\eta^2_p = .01$, le score IAT moyen étant descriptivement plus faible avant les attentats (M = 0.41 ; ET = 0.40) qu'après les attentats (M = 0.48 ; ET = 0.38). Enfin, les participants passant l'IAT-Nice avaient également un score IAT moyen positif (M = 0.42 ; ET = 0.39) reflétant un biais anti-Maghrébins/pro-Français, $t(158) = 13.78$, $p < .001$, $\eta^2_p = .55$. Une nouvelle fois, le lien entre le contexte et le score IAT était non significatif, $t(158) = 0.73$, $p = .47$, $\eta^2_p = .003$, le score IAT moyen étant descriptivement plus faible avant les attentats (M = 0.40 ; ET = 0.36) qu'après les attentats (M = 0.45 ; ET = 0.41).

Afin d'augmenter la puissance statistique de nos analyses, nous avons réalisé une analyse de données intégratives (Curran & Hussong, 2009). Nous avons ainsi regroupé en un jeu de données unique l'ensemble des participants ayant pris part à ces trois IAT. Nous avons traité nos données à l'aide d'une ANOVA à plan inter-sujets de type 2 (contexte : Avant vs. Après) x 3 (type IAT : IAT-Charlie vs. IAT-Bataclan vs. IAT-Nice) avec le score IAT moyen en variable dépendante. Globalement, les participants avaient un score IAT moyen positif (M = 0.41 ; ET = 0.40), reflétant un biais anti-Maghrébins/pro-Français, $t(676) = 26.44$, $p < .001$, $\eta^2_p = .51$. Comme précédemment, le lien entre le contexte (Avant vs. Après) et le score IAT était non significatif, $t(676) = 1.50$, $p = .13$, $\eta^2_p = .003$, le score IAT moyen étant descriptivement plus faible avant les attentats (M = 0.39 ; ET = 0.38) qu'après les attentats (M = 0.43 ; ET = 0.41).

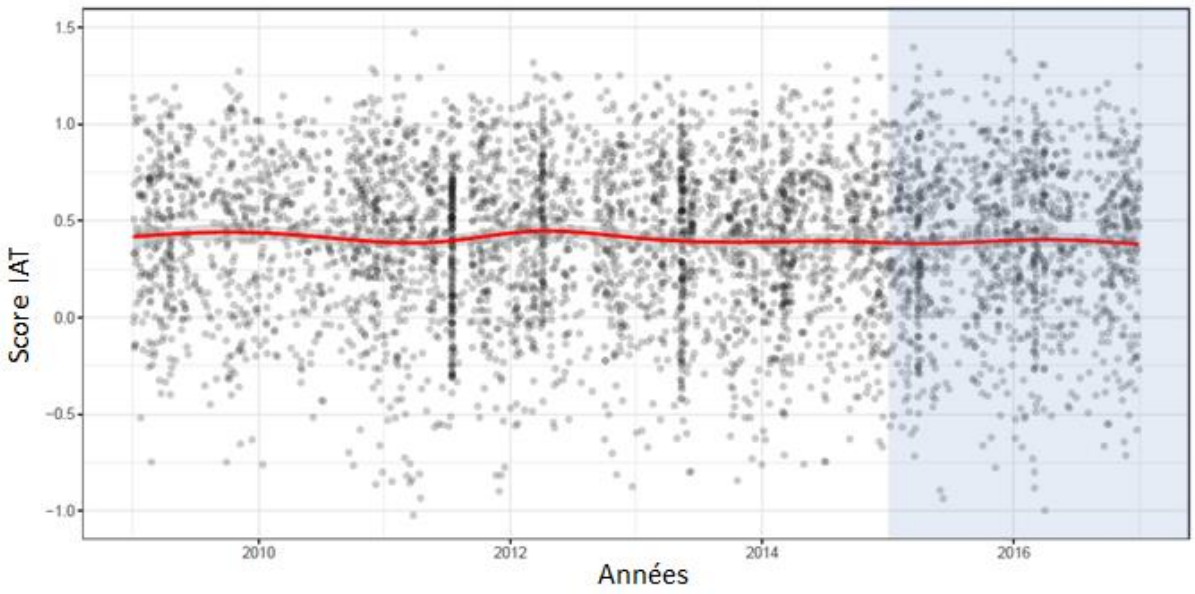


Figure 14. Évolution de la moyenne individuelle (gris) et groupale (rouge) des Scores IAT au cours du temps. La zone bleue correspond à la zone post-attentat de Charlie Hebdo (janvier 2015).

Modèle multinomial. Nous avons analysé les données de ces trois IAT selon un Quad Model à six paramètres, c'est-à-dire un modèle séparant les paramètres d'activation (AC) et d'inhibition (OB) en fonction des stimuli français et maghrébins. [10] Comme pour les analyses relatives au score IAT, nous avons commencé par analyser les données séparément en fonction du type d'attentat (i.e., IAT-Charlie, IAT-Bataclan, IAT-Nice), puis avons regroupé l'ensemble des données en un jeu de données unique. Commençons par les analyses séparées en fonction du type d'attentat.

[10] Les paramètres du Quad Model s'interprètent de la même manière que les paramètres du modèle SAC-dominant. Ainsi, une valeur AC = 1 signifie que le contenu mental a systématiquement été activé au cours des essais. À l'inverse, une valeur AC = 0 (i.e., 1-AC = 1) signifie que le contenu mental n'a jamais été activé au cours des essais. Le paramètre OB s'interprète à l'inverse du paramètre SAP. En effet, une valeur OB = 1 signifie que le contenu mental réactivé a systématiquement été inhibé et une valeur OB = 0 (i.e., 1-OB = 1) signifie que le contenu mental réactivé a systématiquement été appliqué. Une valeur OB = .50 signifie donc que, sur la totalité des essais, le contenu mental a autant été appliqué qu'inhibé. Les valeurs du paramètre D s'interprètent comme pour le modèle SAC-dominant, D = 1 signifie que le participant a systématiquement détecté la bonne réponse au cours des essais, alors que D = 0 (i.e., 1-D = 1) signifie que le participant n'a jamais détecté la bonne réponse. Enfin, le paramètre G s'interprète de manière similaire au paramètre G du modèle SAC-dominant, c'est-à-dire que G = 1 signifie que le participant a systématiquement répondu avec la clé de réponse associée à « Bon » quand les autres paramètres n'étaient pas à l'œuvre et G = 0 (1-G) signifie que le participant a systématiquement répondu avec la clé de réponse associée à « Mauvais » quand les autres paramètres n'étaient pas à l'œuvre. Une valeur G = .50 signifie donc que le participant a répondu totalement au hasard lorsque les autres paramètres n'étaient pas à l'œuvre.

Pour l'IAT-Charlie, l'ajustement (T_1) du modèle aux données avant (M_{pred} = 0.06, M_{obs} = 0.21, p =.002, w = 0.04) et après l'attaque (M_{pred} = 0.04, M_{obs} = 0.30, p < .001, w = 0.04) était acceptable. Contrairement à notre hypothèse principale et concernant les stimuli maghrébins, nous n'avons pas observé de différence significative du paramètre AC_{Mg} avant (M_{AC} = .02, $IC_{95\%}$[.00 ; .04]) et après l'attaque (M_{AC} = .00, $IC_{95\%}$[.00 ; .014]) , ΔAC_{Mg} = -.01, $IC_{95\%}$[-.03 ; .01], p_B = .87, celui-ci étant même descriptivement plus faible après l'attaque (Figure 15). De la même manière, nous n'avons pas observé de différence significative du paramètre OB_{Mg} avant (M_{OB} = .84, $IC_{95\%}$[.22 ; .99]) et après l'attaque (M_{OB} = .70, $IC_{95\%}$[.09 ; .99]), ΔOB_{Mg} = -.14, $IC_{95\%}$[-.84 ; .60], p_B = .68, celui-ci étant descriptivement plus faible après l'attaque. Concernant les stimuli français, nous n'avons pas observé de différence significative du paramètre AC_{Fr} avant (M_{AC} = .09, $IC_{95\%}$[.07 ; .11]) et après l'attaque (M_{AC} = .07, $IC_{95\%}$[.05 ; .09]), ΔAC_{Fr} = -.02, $IC_{95\%}$[-.05 ; .01], p_B = .91, ce paramètre étant lui aussi descriptivement plus faible après l'attaque. De même, nous n'avons pas observé de différence significative du paramètre OB_{Fr} avant l'attaque (M_{OB} = .81, $IC_{95\%}$[.32 ; .99]) et après l'attaque (M_{OB} = .92, $IC_{95\%}$[.65 ; .99]), ΔOB_{Fr} = .11, $IC_{95\%}$[-.24 ; .61], p_B = .27, ce paramètre étant descriptivement plus élevé après l'attaque. Enfin, nous n'avons observé aucune différence significative au niveau des paramètres D et G entre avant (M_D = .88, $IC_{95\%}$[.86 ; .90] ; M_G = .58, $IC_{95\%}$[.53 ; .63]) et après l'attaque (M_D = .90, $IC_{95\%}$[.88 ; .91] ; M_G = .60, $IC_{95\%}$[.56 ; .65]), ΔD = .02, $IC_{95\%}$[-.01 ; .04], p_B = .07 et ΔG = .03, $IC_{95\%}$[-.04 ; .09], p_B = .21.

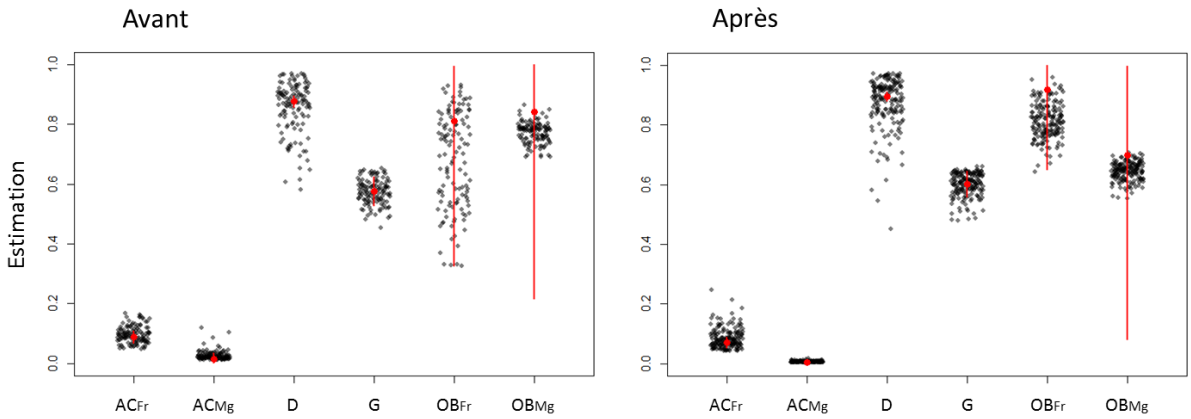


Figure 15. Prédictions pour l'IAT-Charlie. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un Quad Model. AC_{Fr} et OB_{Fr} = paramètres pour les prénoms français, AC_{Mg} et OB_{Mg} = paramètres pour les prénoms maghrébins.

Pour l'IAT-Bataclan, l'ajustement (T_1) du modèle aux données avant (M_{pred} = 0.07, M_{obs} = 0.17, p =.003, w = 0.04) et après l'attaque (M_{pred} = 0.07, M_{obs} = 0.29, p < .001, w = 0.05) était là encore acceptable. Contrairement à notre hypothèse principale et concernant là encore nos stimuli maghrébins, nous n'avons pas observé de différence significative du paramètre AC_{Mg} avant (M_{AC} = .01, $IC_{95\%}$[.00 ; .02]) et après l'attaque (M_{AC} = .01, $IC_{95\%}$[.00 ; .03]), ΔAC_{Mg} = -.00, $IC_{95\%}$[-.02 ; .02], p_B = .43 (Figure 16). De la même manière, nous n'avons pas observé de différence significative du paramètre OB_{Mg} avant (M_{OB} = .63, $IC_{95\%}$[.06 ; .99]) et après l'attaque (M_{OB} = .76, $IC_{95\%}$[.12 ; .99]), ΔOB_{Mg} = .13, $IC_{95\%}$[-.66 ; -.84], p_B = .35, celui-ci étant descriptivement plus fort après l'attaque. En ce qui concerne les stimuli français, nous n'avons pas observé de différence significative du paramètre AC_{Fr} avant (M_{AC} = .09, $IC_{95\%}$[.06 ; .11]) et après l'attaque (M_{AC} = .08, $IC_{95\%}$[.06 ; .11]), ΔAC_{Fr} = -.00, $IC_{95\%}$[-.04 ; .03], p_B = .56. De même, le paramètre OB_{Fr} ne différait significativement pas avant (M_{OB} = .67, $IC_{95\%}$[.12 ; .99]) et après l'attaque (M_{OB} = .74, $IC_{95\%}$[.18 ; .99]), ΔOB_{Fr} = .07, $IC_{95\%}$[-60 ; .72], p_B = .42. Enfin, il n'y avait également aucune différence significative au niveau des paramètres D et G entre avant (M_D = .88, $IC_{95\%}$[.87 ; .90] ; M_G = .58, $IC_{95\%}$[.53 ; .63]) et après l'attaque (M_D = .88, $IC_{95\%}$[.86 ; .90] ; M_G = .61, $IC_{95\%}$[.57 ; .66]), ΔD = -.00, $IC_{95\%}$[-.03 ; .02], p_B = .56 et ΔG = -.01, $IC_{95\%}$[-.05 ; .07], p_B = .39.

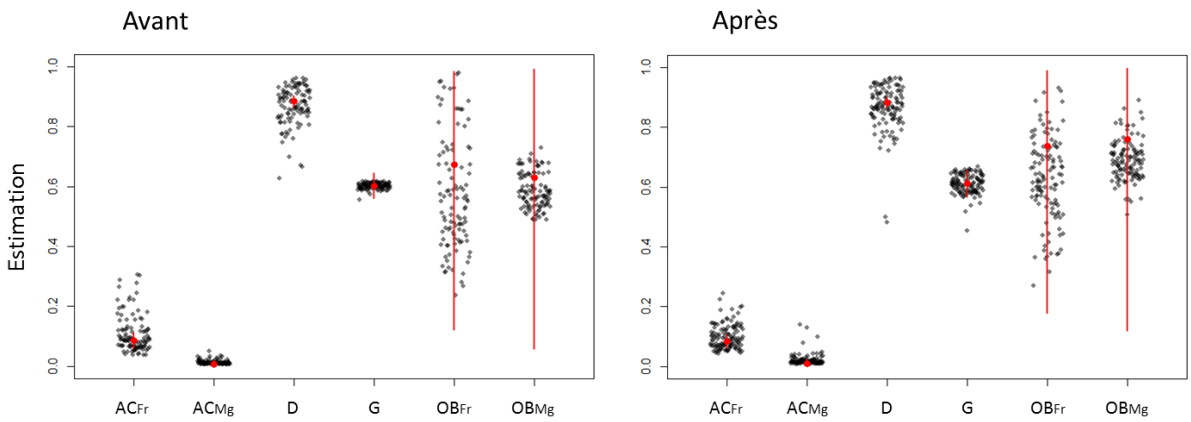


Figure 16. Prédictions pour l'IAT-Bataclan. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un Quad Model. AC_{Fr} et OB_{Fr} = paramètres pour les prénoms français, AC_{Mg} et OB_{Mg} = paramètres pour les prénoms maghrébins.

Concernant l'IAT-Nice, l'ajustement (T_1) du modèle aux données avant (M_{pred} = 0.09, M_{obs} = 0.32, p < .001, w = 0.06), et après l'attaque (M_{pred} = 0.11, M_{obs} = 0.29, p = .001, w = 0.06) était une nouvelle fois acceptable. Contrairement à notre hypothèse principale et concernant les stimuli maghrébins, nous n'avons pas observé de différence significative du paramètre AC_{Mg} avant (M_{AC} = .01, $IC_{95\%}$[.00 ; .03]) et après l'attaque (M_{AC} = .01, $IC_{95\%}$[.00 ; .02]), ΔAC_{Mg} = -.01, $IC_{95\%}$[-.03 ; .01], p_B = .73 (Figure 17). De la même manière, nous n'avons pas observé de différence significative du paramètre OB_{Mg} avant (M_{OB} = .69, $IC_{95\%}$[.09 ; .99]) et après l'attaque (M_{OB} = .66, $IC_{95\%}$[.08 ; .99]), ΔOB_{Mg} = -.04, $IC_{95\%}$[-.78 ; .75], p_B = .55. En ce qui concerne les stimuli français, nous n'avons pas observé de différence significative du paramètre AC_{Fr} avant (M_{AC} = .09, $IC_{95\%}$[.06 ; .12]) et après l'attaque (M_{AC} = .06, $IC_{95\%}$[.03 ; .09]), ΔAC_{Fr} = -.03, $IC_{95\%}$[-.07 ; .01], p_B = .06, bien que celui-ci soit descriptivement plus faible après l'attaque. De même, nous n'avons pas observé de différence significative du paramètre OB_{Fr} avant (M_{OB} = .58, $IC_{95\%}$[.07 ; .97]) et après l'attaque (M_{OB} = .57, $IC_{95\%}$[.06 ; .98]), ΔOB_{Fr} = -.01, $IC_{95\%}$[-.72 ; .70], p_B = .52. Il n'y avait également aucune différence significative au niveau du paramètre D avant l'attaque (M_D = .87, $IC_{95\%}$[.85 ; .89]) et après l'attaque (M_D = .87, $IC_{95\%}$[.85 ; .90]), ΔD = .00, $IC_{95\%}$[-.03 ; .04], p_B = .46. Enfin, le paramètre G était significativement plus important avant l'attaque (M_G = .60, $IC_{95\%}$[.56 ; .66]), qu'après (M_G = .53, $IC_{95\%}$[.47 ; .58]), ΔG = -.08, $IC_{95\%}$[-.15 ; .00], p_B = .02.

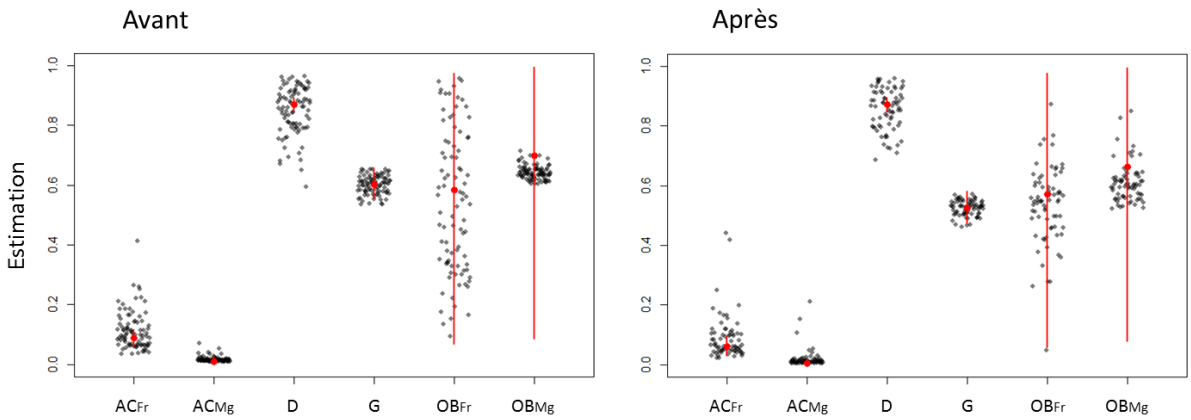


Figure 17. Prédictions pour l'IAT-Nice. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un Quad Model. AC_{Fr} et OB_{Fr} = paramètres pour les prénoms français, AC_{Mg} et OB_{Mg} = paramètres pour les prénoms maghrébins.

Enfin, intéressons-nous à un modèle regroupant les données de l'ensemble des participants (i.e., les participants de l'IAT-Charlie, l'IAT-Bataclan et l'IAT-Nice). Pour ce dernier modèle, l'ajustement (T_1) du modèle aux données avant (M_{pred} = 0.02, M_{obs} = 0.18, p < .001, w = 0.04), et après l'attaque (M_{pred} = 0.02, M_{obs} = 0.26, p < .001, w = 0.05) était une nouvelle fois acceptable. Contrairement à notre hypothèse principale et concernant les stimuli maghrébins, nous n'avons pas observé de différence significative du paramètre AC_{Mg} avant (M_{AC} = .01, $IC_{95\%}$[.00 ; .02]) et après l'attaque (M_{AC} = .00, $IC_{95\%}$[.00 ; .01]), ΔAC_{Mg} = -.01, $IC_{95\%}$[-.02 ; .01], p_B = .85, celui-ci étant descriptivement plus faible après l'attaque (Figure 18). De la même manière, nous n'avons observé aucune différence significative du paramètre OB_{Mg} avant (M_{OB} = .82, $IC_{95\%}$[.19 ; 1]) et après l'attaque (M_{OB} = .77, $IC_{95\%}$[.15 ; 1]), ΔOB_{Mg} = -.05, $IC_{95\%}$[-.78 ; .70], p_B = .57, celui-ci étant descriptivement plus faible après l'attaque. En ce qui concerne les stimuli français, nous n'avons pas observé de différence significative du paramètre AC_{Fr} avant (M_{AC} = .09, $IC_{95\%}$[.08 ; .10]) et après l'attaque (M_{AC} = .07, $IC_{95\%}$[.06 ; .09]), ΔAC_{Fr} = -.02, $IC_{95\%}$[-.04 ; .00], p_B = .949, celui-ci étant descriptivement (au niveau du seuil de significativité) moins élevé après l'attaque. Nous n'avons pas observé de différence significative du paramètre OB_{Fr} avant (M_{OB} = .71, $IC_{95\%}$[.18 ; .99]) et après l'attaque (M_{OB} = .83, $IC_{95\%}$[.33 ; 1]), ΔOB_{Fr} = .12, $IC_{95\%}$[-.46 ; .72], p_B = .33, bien que descriptivement plus élevé après l'attaque. Il n'y avait également aucune différence significative au niveau du paramètre D avant (M_D = .88, $IC_{95\%}$[.87 ; .89]) et après l'attaque (M_D = .89, $IC_{95\%}$[.88 ; .90]), ΔD = .01, $IC_{95\%}$[-.01 ; .03], p_B = .12, ce paramètre étant descriptivement plus élevé après l'attaque. Enfin, nous n'avons pas observé de différence significative du paramètre G avant (M_G = .59, $IC_{95\%}$[.57 ; .62]), et après l'attaque (M_G = .59, $IC_{95\%}$[.57 ; .62]), ΔG = -.00, $IC_{95\%}$[-.04 ; .04], p_B = .49.

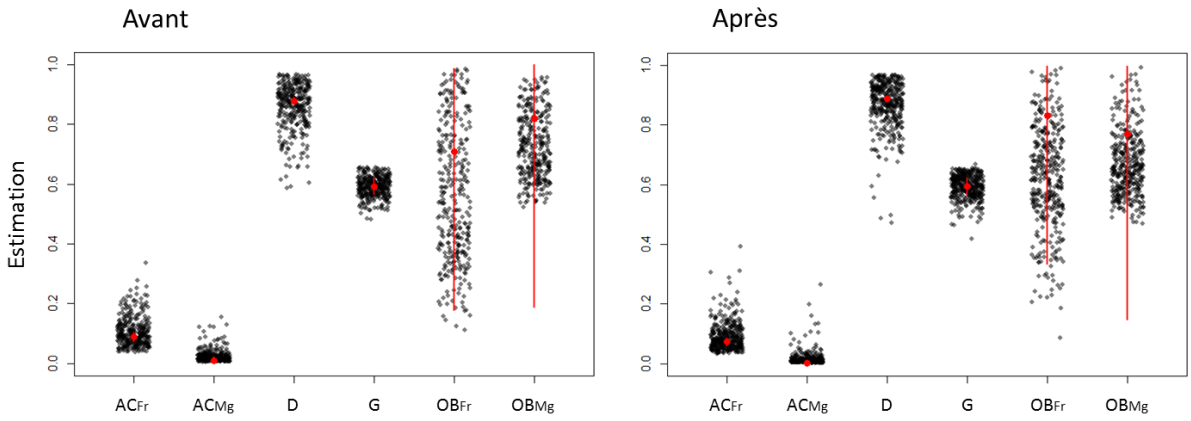


Figure 18. Prédictions pour l'ensemble des participants. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un Quad Model. AC_{Fr} et OB_{Fr} = paramètres pour les prénoms français, AC_{Mg} et OB_{Mg} = paramètres pour les prénoms maghrébins.

2.5.3 **Discussion**

Par le biais de cette quatrième et dernière étude nous voulions observer, de manière invoquée, le possible lien entre des attaques terroristes et les processus sous-tendant les l'attitude raciale anti-Maghrébins. En appliquant un Quad Model à des données IAT collectées au moment de trois attentats différents (i.e., Charlie Hebdo, Bataclan, Nice), nous n'avons une nouvelle fois pas observé de différences significatives au niveau des valeurs des paramètres du modèle avant et après les différentes attaques. À la vue de l'ensemble de ces résultats, nous ne pouvons pas écarter la possibilité qu'il n'y ait réellement pas eu d'impact des attentats sur le score IAT anti-Maghrébins et sur les processus à l'origine de ces performances individuelles. Cependant, nous voyons d'autres explications possibles à l'absence de résultats significatifs.

La principale remarque quant à ces résultats concerne leur hétérogénéité, plus importante encore que celle observée dans la réalisation de nos SMT. Ainsi, certains paramètres, comme le paramètre OB, ont plusieurs fois eu un intervalle de crédibilité recouvrant l'ensemble des valeurs possibles (e.g., pour IAT-Nice et concernant les stimuli français, M_{OB} = .58, $IC_{95\%}$[.07 ; .97]). À la différence de la SMT, l'IAT n'est pas une mesure

s'intéressant aux réponses catégorielles, mais aux différents temps de réponse. Or, comme nous l'avons mentionné précédemment, pour réaliser un Quad Model nous devons utiliser les fréquences de bonnes et mauvaises réponses. Dans la SMT, cette fréquence de bonnes ou de mauvaises réponses est assez bien répartie (e.g., pour l'Étude 1 : 42.69% d'erreurs). À l'inverse, dans l'IAT il n'y a au final que très peu de mauvaises réponses (i.e., 8.04 % pour IAT-Charlie ; 8.33% pour l'IAT-Bataclan ; 9.12% pour l'IAT-Nice). Un modèle comme le Quad Model estimant la valeur de ses paramètres en fonction de cette répartition incomplète des données sera donc limité en termes de fiabilité (e.g., Calanchini et al., 2018).

Plusieurs solutions existent pour pallier à ce problème. L'une des solutions consiste à augmenter artificiellement le taux d'erreurs au sein de la tâche en imposant par exemple aux participants une fenêtre de réponse (Meissner & Rothermund, 2013). Cependant, cela a comme inconvénient d'influencer involontairement les différents paramètres du modèle et notamment le paramètre d'inhibition (OB ; Calanchini et al., 2018). Une autre solution consiste à augmenter le nombre d'essais et par conséquent le nombre d'erreurs au sein de la tâche. Nous implémenterons nous-même ce changement au Chapitre 3. Enfin, la solution qui est peut-être le plus souvent retenue est de maximiser le nombre de participants réalisant la tâche. L'utilisation des données du Project Implicit est particulièrement propice à cela et, de ce fait, grandement utilisée dans les recherches utilisant le Quad Model (e.g., Calanchini et al., 2017 ; Gonsalkorale et al., 2014).

Pour notre présente étude, nous avons nous-même collecté nos données sur la plateforme Project Implicit. Cependant, notre population d'intérêt était particulièrement restreinte (i.e., individus français de couleur blanche). Dès lors, bien que nous ayons choisi une période de collecte de données relativement large (période de trois mois avant et après les attaques), nous n'avons eu que relativement peu de participants (e.g., pour Nice, 160 participants). En conséquence, il est probable que nous n'ayons pas eu la puissance nécessaire pour observer les effets escomptés. Il est à noter que le choix d'une période de collecte de données de trois mois avant et trois mois après ces attaques nous a peut-être été préjudiciable. En effet, certains auteurs ont montré que l'augmentation du préjugé anti-Maghrébins suite aux attentats ne se maintenait pas dans le temps (i.e., deux semaines ; Cohu et al., 2016). Par conséquent, il est probable que nos périodes d'intérêts aient été trop étendues, l'impact éventuel des attentats se diluant au cours du temps.

Afin de maximiser la puissance statistique de nos analyses, nous avons regroupé les données de l'ensemble de nos participants en un jeu de données unique et avons comparé les

paramètres des participants ayant passé un IAT avant ces attaques avec ceux des participants ayant passé un IAT après ces attaques. Là encore, l'hétérogénéité du paramètre OB était importante. Il est à noter que, de manière descriptive, certains paramètres allaient dans le sens inverse de nos hypothèses. Par exemple, l'inhibition du contenu mental négatif lié aux Maghrébins (OB_{Mg}) était plus faible après l'attaque. De même, l'activation du contenu mental négatif envers les Maghrébins (AC_{Mg}) était moins élevée après les attentats qu'avant. De façon surprenante, celle-ci était même significativement nulle après les attaques (intervalle de crédibilité comprenant la valeur zéro). De manière générale, il semble ainsi que les scores IAT aient été davantage le reflet d'un biais positif envers les Français que d'un biais négatif envers les Maghrébins, le paramètre AC_{Fr} étant systématiquement plus élevé que le paramètre AC_{Mg}. Ces résultats sont cohérents avec certains travaux récents ayant montré que le biais IAT anti-Noirs classique semblait davantage être le reflet d'une association entre les stimuli blancs et les mots positifs que d'une association entre les stimuli noirs et les mots négatifs (Calanchini, en préparation).

À la suite de cette étude utilisant le Quad Model, nous voyons deux grandes pistes s'offrir à nous. La première consiste à élargir notre population d'intérêt. En effet, bien que notre objet d'étude soit l'impact des attentats sur la population française (hors participants d'une origine ethnique différente), il pourrait être intéressant d'analyser nos données IAT sur l'ensemble des participants (i.e., au niveau mondial) ayant passé l'IAT-Maghrébins à ces périodes, afin de maximiser la puissance de notre modèle et ainsi augmenter la fiabilité et la précision de ses estimations. Cela pourrait être d'autant plus intéressant, car d'autres recherches effectuées en France ont pu observer un lien entre les attentats et le score IAT au niveau mondial juste après l'attaque du Bataclan (Arnoult et al., 2019).

La seconde piste concerne le type d'analyse des données que nous pouvons entreprendre. Comme nous l'avons décrit précédemment, il existe à ce jour de nouveaux modèles multinomiaux permettant d'analyser non plus uniquement les fréquences de bonnes ou de mauvaises réponses, mais aussi les temps de réponse des participants (Hartmann et al., 2020 ; Heck & Erdfelder, 2016 ; Klauer & Kellen, 2018). Ces modèles, bien que n'ayant pas encore été appliqués aux mesures indirectes en cognition sociale, sont prometteurs, car ils peuvent fournir une description plus précise des processus à l'œuvre et apportent notamment une information au niveau temporel (Calanchini, 2020). Il pourrait ainsi être intéressant de réanalyser nos données par le biais de ces modèles afin d'observer si ceux-ci nous fournissent une information plus précise nous permettant de répondre à nos hypothèses.

2.6 Discussion du Chapitre 2 : L'amalgame Maghrébin-terroriste existe-t-il ?

Reprenons notre raisonnement initial. L'État islamique, en multipliant les attaques sur notre sol, veut à terme créer dans l'esprit des individus un amalgame Maghrébin-terroriste (Kepel, 2016). En d'autres termes, nous pourrions associer, de façon non intentionnelle, la menace que nous avons des terroristes aux individus maghrébins en général. Au sein de notre laboratoire, le Projet Amalgame a été mis en place. Par le biais de ce projet, nous voulions étudier s'il pouvait exister, au sein de la population française, ce risque d'amalgame. Avec l'utilisation de plusieurs mesures indirectes (i.e., tâche d'approche-évitement, Rougier et al., 2018 ; tâche de mémoire des visages, Ackerman et al., 2006 ; SMT, Krieglmeyer & Sherman, 2012 ; IAT, Conrey et al., 2005), nous voulions mettre en évidence la réactivation d'une attitude raciale anti-Maghrébins plus forte dans un contexte d'attentats. Plus de trois ans après le début de ce projet et plusieurs études effectuées au sein de notre laboratoire, nous ne pouvons toujours pas conclure à la présence de ce type d'amalgame. Essayons d'analyser pourquoi.

Notre première remarque porte tout d'abord sur nos différentes tentatives d'opérationnalisation d'un contexte d'attentats en laboratoire. Suivant en cela certaines études manipulant expérimentalement le contexte (e.g., Barden et al., 2004 ; Maddux et al., 2005 ; Wittenbrink et al., 2001), nous pensions qu'induire en laboratoire un contexte d'attentats serait suffisant pour mettre en évidence la réactivation d'une attitude raciale plus négative envers les populations maghrébines. En effet, plusieurs études utilisant ce type de manipulation ont par le passé montré un impact des attentats sur les attitudes raciales (e.g., Das et al., 2009 ; Horry & Wright, 2009). Cependant, que ce soit dans nos études antérieures à ce travail de thèse ou dans les études décrites en détail dans ce manuscrit, l'impact d'un contexte d'attentats sur les performances de nos participants s'est avéré non significatif. Nous l'avons souligné, induire expérimentalement ce type de contexte peut s'avérer délicat (Dumont et al., 2003 ; Nugier & Guimond, 2016). Toutefois, même lorsque nous avons utilisé des données collectées au moment des attentats nous n'avons pas observé de lien entre ces attaques et les performances individuelles. Nous pourrions focaliser nos ressources sur la recherche d'une meilleure manipulation d'un contexte d'attentats, mais, pour nous, cela négligerait d'autres aspects primordiaux à prendre en compte dans ce type d'études.

L'un des intérêts historiques justifiant l'utilisation de mesures indirectes est la possibilité de contourner l'influence de processus davantage contrôlés et notamment ici,

comme nous l'imaginions tout du moins, de la volonté affichée de ne pas faire d'amalgame. Cependant, comme nous l'avons répété de nombreuses fois au Chapitre 1, nous savons maintenant que ce type de mesures n'est pas uniquement le reflet de processus de Type 1, mais également en partie de processus de Type 2 (et notamment de processus contrôlés ; Fiedler & Bluemke, 2005 ; Hütter & Klauer, 2016). C'est la raison pour laquelle nous avons voulu tester si nos manipulations expérimentales pouvaient influencer différemment les processus à l'œuvre au sein de ces mesures. Toutefois, les trois études de ce chapitre ne nous ont pas permis de mettre en évidence l'impact des attentats à la fois sur les processus reflétant une activation non intentionnelle du contenu mental (i.e., SAC, AC), mais aussi sur les processus reflétant une potentielle inhibition de cette activation (i.e., SAP, OB). Nous avons souligné l'hétérogénéité des données de l'Étude 4 (i.e., données IAT). Comme évoqué précédemment, il serait pour nous pertinent de répliquer cette étude en augmentant le nombre de participants ou en utilisant un autre type d'analyse relatif aux temps de réponse (Hartmann et al., 2020 ; Heck & Erdfelder, 2016 ; Klauer & Kellen, 2018). Néanmoins, ce problème de puissance ne semble pas spécifique à notre Étude 4. En effet, il est probable que même nos études utilisant le paradigme SMT aient été en sous-puissance. Parmi ces études, l'Étude 3 a été celle qui nous a permis d'observer les résultats les plus prometteurs. Répliquer cette étude avec une puissance plus importante, c'est-à-dire avec plus d'essais et plus de participants au sein de la tâche, pourrait peut-être nous permettre d'observer un impact des attentats sur les processus cognitifs à l'origine du jugement.

En résumé, nous pouvons effectivement mettre en évidence certaines critiques quant à nos études réalisées au sein de ce Projet Amalgame (e.g., manipulation du contexte, problème de sous-puissance). Néanmoins, au-delà de ces questionnements méthodologiques, l'ensemble des résultats du Projet Amalgame nous pousse aujourd'hui à poser la question suivante : l'amalgame Maghrébin-terroriste existe-t-il ? Récemment, un journaliste de Charlie Hebdo nous a contactés afin de connaitre l'avancée des recherches en psychologie sociale quant au préjugé « implicite » anti-Maghrébins suite aux attentats. Lui faisant part de l'absence de résultats probants au sein du Projet Amalgame, ce journaliste a titré son article : « le racisme ne fleurit pas sur le fumier des attentats » (Fischetti, 2020). Prenant moins de retenue que nous ne l'aurions espéré, le cœur de cet article conclut à une absence de conséquences des attentats sur le préjugé « implicite ». L'amalgame Maghrébin-terroriste n'existe donc-t-il pas ? Les terroristes ont-ils échoué à instaurer dans l'esprit des individus une perception plus négative des Maghrébins ? Nous ne pouvons pas tirer cette conclusion. Cependant, le Projet

Amalgame nous a permis de mettre en lumière une chose : avec les mesures dont nous disposons actuellement, nous avons des difficultés à mettre en évidence cet amalgame.

Très récemment, et dans le contexte de pandémie mondiale liée au Covid-19, IJzerman et collaborateurs (2020) ont alerté quant au risque d'utiliser les travaux de sciences sociales au sein des politiques publiques. Les arguments avancés par ces auteurs étaient que les résultats de ces études étaient souvent trop fragiles pour pouvoir être utilisés (e.g., problème de reproductibilité, problème de généralisabilité, problème de validité des mesures). Sans partager entièrement ce point de vue, nous pensons effectivement qu'en l'état, utiliser les mesures indirectes d'attitudes raciales peut s'avérer problématique. Tout au long du Chapitre 1, nous nous sommes intéressés à l'évolution des mesures des attitudes raciales et notamment à l'importance des développements théoriques relatifs aux processus sous-tendant ces attitudes. Avec ce Chapitre 2 nous avons essayé de mettre en évidence le fait qu'il nous fallait être prudent non seulement quant au choix de la mesure adéquate pour répondre à notre question de recherche, mais aussi et surtout à l'interprétation que nous faisions des résultats apportés par ces mesures. Nous maintenons qu'utiliser une mesure indirecte classique (e.g., ; IAT, Greenwald et al., 1998 ; biais du tireur, Correll et al., 2002 ; EPT, Fazio et al., 1995) sans une réflexion approfondie quant aux processus cognitifs à l'œuvre peut s'avérer problématique. Globalement, nous nous inscrivons ainsi parmi les nombreuses critiques relatives à l'utilisation des mesures indirectes et à l'interprétation de leurs résultats (e.g., Arkes & Tetlock, 2004 ; Corneille & Hütter, 2020 ; Hütter & Klauer, 2016 ; Mitchell & Tetlock, 2017). Les modèles multinomiaux, bien que comportant également des points négatifs (e.g., incertitude quant à la valeur des paramètres, hétérogénéité des données), peuvent aujourd'hui être une solution contribuant à cette réflexion théorique.

Naturellement, l'utilisation de modèles multinomiaux n'est qu'un outil parmi d'autres. Pour investiguer les processus à l'œuvre au sein d'une même tâche, des alternatives existent (e.g., *dot probe task*, Richeson & Trawalter, 2008 ; *mouse tracking*, Freeman & Ambady, 2011 ; théorie de la détection du signal, Correll et al., 2002 ; modèles de diffusion, Klauer et al., 2011 ; voir Calanchini et al., 2018, pour une revue de la littérature). Par exemple, un type de procédé particulier appelé *mouse tracking* peut permettre, au sein de l'IAT, d'explorer les dimensions sensorimotrices de la tâche (e.g., Freeman, Dale, et al., 2011 ; Freeman, Penner, et al., 2011 ; Wojnowicz et al., 2009). Dans les tâches de *mouse tracking*, les participants doivent réaliser un IAT, non pas à l'aide du clavier, mais de la souris. Pour chaque essai, les participants doivent aller cliquer le plus rapidement possible sur le label correspondant à l'item présenté.

L'intérêt d'utiliser la souris est de pouvoir enregistrer, en plus de la réponse finale, la trajectoire aboutissant à cette réponse. Grâce à divers outils (e.g., aire sous la courbe, maximum de déviation, profil de vitesse), les auteurs utilisant ce type de mesure peuvent calculer différents profils de réponse pour les essais compatibles et incompatibles, et notamment « l'attraction » du participant vers le label incorrect au sein des essais incompatibles, ainsi que l'inhibition de cette réponse pour donner la réponse correcte. Les modèles théoriques associés au *mouse tracking* ont ainsi permis la prise en compte d'une influence multimodale au sein de l'IAT, et notamment de processus à la fois *top-down* et *bottom-up* contribuant à la réponse finale du participant (Freeman & Ambady, 2011). À notre sens, des outils comme les modèles multinomiaux, le *mouse tracking* ou plus généralement des outils permettant une prise en compte des processus à l'œuvre au sein d'une mesure indirecte peuvent participer à l'amélioration des mesures indirectes actuelles.

En conclusion, nous ne pouvons pas certifier aujourd'hui de la présence d'un amalgame Maghrébin-terroriste réactivé involontairement en contexte d'attentats. Il est possible que des problèmes méthodologiques (e.g., sous-puissance statistique) nous aient empêchés d'observer ce type d'amalgame. Il est également possible que nos mesures ou manipulations expérimentales n'aient pas été propices à l'obtention des résultats attendus (e.g., manipulation du contexte). Toutefois, par le biais de cette problématique, le but de ce chapitre était avant tout de mettre en évidence les précautions à prendre lors de l'interprétation des résultats issus des mesures indirectes. Nous avons ainsi suggéré théoriquement qu'un même jugement négatif anti-Maghrébins pouvait être influencé par une contribution différente des processus cognitifs sous-jacents. À nos yeux, les modèles multinomiaux peuvent être pertinents pour étudier ce type de questions. Comme d'autres types d'outils (e.g., *mouse tracking*), nous pensons qu'utiliser ces modèles peut nous permettre d'enrichir notre réflexion quant à l'utilisation des mesures indirectes et ainsi participer à l'amélioration de ce type de mesures. Les chercheurs apportant leur expertise à des problématiques de société aussi importantes que la discrimination envers les populations maghrébines, nous devons nous assurer d'avoir les meilleurs outils possibles afin d'apporter des réponses valides à ces questions.

Chapitre 3 – Rôle des facteurs liés à la tâche et des facteurs liés aux participants dans la relation conditionnelle entre processus

Eva et Clara sont à l'opposé du spectre concernant l'attitude raciale à l'égard des individus maghrébins, Clara étant celle ayant l'attitude négative. Aujourd'hui, Eva et Clara viennent au laboratoire pour passer une tâche sur ordinateur où, pour chaque essai, elles voient une photo d'un visage maghrébin ou d'un visage blanc, suivi d'un dessin de visage plus menaçant ou au contraire moins menaçant que la moyenne de ces dessins. Nous le savons, cette tâche est la *Stereotype Misperception Task* (SMT ; Krieglmeyer & Sherman, 2012) et illustre d'abord ce qui équivaut à un effet d'amorçage : les dessins sont jugés plus menaçants lorsqu'ils sont précédés d'un visage maghrébin plutôt que d'un visage blanc. Plus intéressant, utiliser les modèles multinomiaux associés permet d'étudier les différents processus impliqués dans la réponse finale d'Eva et de Clara, par exemple l'activation du stéréotype et l'application du stéréotype lié aux amorces, mais aussi la détection de la menace de la cible (i.e., si elles ont pu détecter le véritable niveau de menace des dessins). En fait, de précédents travaux ont notamment montré qu'Eva ayant une attitude raciale auto-rapportée moins négative que Clara, son niveau d'application du stéréotype sera plus faible, ce qui signifie qu'elle a moins de chance (qu'Eva) d'utiliser son stéréotype une fois celui-ci activé (e.g., Krieglmeyer & Sherman, 2012 ; Étude 3 de ce manuscrit). L'objectif de ce chapitre est cependant très différent. Au lieu de se demander dans quelle mesure chaque processus va entrer en jeu en fonction de l'attitude raciale auto-rapportée, ce travail vise à tester, pour la première fois, si l'organisation même de ces processus peut être différente en fonction de différents facteurs, et notamment de l'attitude raciale anti-Maghrébins. Ainsi, le modèle multinomial expliquant au mieux les données d'Eva sera différent de celui de Clara. Plus précisément, cela signifie que lorsqu'Eva détecte le niveau de menace de la cible, peu importe si son stéréotype est activé, elle répondra ce qu'elle a vu, et ce n'est que lorsqu'elle n'a pas détecté le niveau de menace de la cible que son stéréotype entre en jeu (i.e., modèle D-dominant). Pour Clara, en revanche, les choses pourraient être très différentes, car maintenant, ce qui compte le plus, c'est que son stéréotype soit activé, et ce n'est que lorsqu'il ne l'est pas que la détection du niveau de menace de la cible entre en jeu (i.e., modèle SAC-dominant). Nous montrerons ainsi au cours de ce chapitre qu'une mesure directe d'attitude raciale anti-Maghrébins peut être liée à la relation conditionnelle entre les processus à l'origine du jugement de menace.

3.1 Comparaison de modèles multinomiaux dans la littérature

Au Chapitre 1, nous avons abordé le fait qu'un modèle multinomial se construisait avant tout par le biais d'une spécification théorique des différents processus à l'œuvre, mais aussi de la relation conditionnelle entre ces processus. Cela signifie que, pour une tâche donnée, le modèle multinomial associé à cette tâche rassemblera l'ensemble des principaux processus pouvant théoriquement jouer un rôle dans la réponse du participant et l'organisation la plus probable de ces processus. Néanmoins, il est possible que d'autres processus ou différentes organisations de ces processus et donc différents modèles multinomiaux puissent être tout autant envisageables d'un point de vue théorique. C'est pourquoi certaines études ont testé empiriquement si la meilleure version d'un modèle multinomial (i.e., le modèle classiquement associé à une tâche) pouvait être remise en cause par un modèle théoriquement plus pertinent et présentant une meilleure qualité d'ajustement aux données (e.g., Bishara & Payne, 2009 ; Nadarevic & Erdfelder, 2011 ; Payne & Bishara, 2009 ; Rees, Rivers, et al., 2019 ; Sherman et al., 2008). Décrivons rapidement certaines de ces études.

Bishara et Payne (2009) se sont intéressés au modèle multinomial pouvant expliquer au mieux les données de participants passant la WIT (*Weapon Identification Task* ; Payne, 2001). Nous l'avons vu au Chapitre 1, ces auteurs ont porté un intérêt aux rôles des processus automatiques et contrôlés dans cette mesure indirecte et ont appliqué plusieurs versions du modèle PDP à ces données (Payne, 2001). Bishara et Payne (2009) ont envisagé théoriquement trois relations conditionnelles possibles entre les processus : tout d'abord, une relation à dominante contrôlée, c'est-à-dire qu'un outil sera identifié comme une arme uniquement lorsque les processus contrôlés échouent et que le stéréotype Noirs/armes est activé (PDP-C). Ensuite une relation à dominante automatique, c'est-à-dire qu'un outil sera identifié comme une arme dès lors que le stéréotype Noirs/armes est activé, et ce, quel que soit le rôle des processus contrôlés (PDP-A). Enfin, une relation probabiliste, c'est-à-dire que si les processus contrôlés et les processus automatiques sont à l'œuvre sur un même essai, le potentiel conflit sera résolu de manière probabiliste, donnant plus ou moins de poids à chacun de ces processus (Quad Model). En analysant les données de quatre études à la fois au niveau agrégé et au niveau individuel, Bishara et Payne ont observé que le modèle supportant le mieux les données de la WIT était le PDP-C, c'est-à-dire un modèle à dominante contrôlée, confirmant en cela les données collectées par ces auteurs (Payne, 2001).

Dans la même veine, Sherman et collaborateurs (2008) ont voulu comparer l'ajustement du Quad Model, développé par ses équipes (Conrey et al., 2005), par rapport à l'ajustement de la version C-dominant du modèle PDP (PDP-C ; Payne, 2001). Pour Sherman et collaborateurs, et comme nous l'avons décrit au Chapitre 1, le choix du modèle doit tout d'abord tenir compte du lien entre la question de recherche et les paramètres du modèle. Par exemple, si des chercheurs sont particulièrement intéressés par l'influence du processus d'inhibition du contenu mental réactivé (comme ce fut notre cas au Chapitre 2), le Quad Model pourra être pertinent, celui-ci comprenant, à l'inverse d'autres modèles (e.g., PDP), un paramètre d'inhibition (Calanchini, 2020 ; Calanchini et al., 2018 ; Sherman et al., 2008). Pour Sherman et collaborateurs, cette considération théorique ne s'arrête pas uniquement à la question du paramètre d'intérêt, mais également à la validité prédictive des différents paramètres du modèle. Or, le Quad Model a montré une bonne validité prédictive (Conrey et al., 2005 ; Sherman, 2006 ; Sherman et al., 2008), soulignant ainsi, pour ces auteurs, l'intérêt de l'utilisation de ce modèle. Afin de confronter de manière empirique le Quad Model au modèle PDP-C classique, Sherman et collaborateurs ont comparé statistiquement l'ajustement de ces deux modèles sur un nombre important de jeux de données. De manière similaire aux conclusions de Bishara et Payne (2009), ces auteurs ont observé que le PDP-C s'ajustait légèrement mieux que le Quad Model pour les études utilisant un paradigme d'amorçage (e.g., WIT). Au contraire, et conformément à leurs attentes, le Quad Model avait une qualité d'ajustement supérieure pour les données issues de paradigmes IAT. Enfin, il est à signaler que le Quad Model s'ajustait légèrement mieux que le modèle PDP-C pour des données issues d'un paradigme GNAT (voir Chapitre 1 pour une description de ce paradigme), confirmant la capacité du Quad Model à s'adapter à des variantes de la procédure IAT classique (e.g., Gonsalkorale, von Hippel, et al., 2009).

Nadarevic et Erdfelder (2011) ont cependant remis en question la capacité du Quad Model à décrire de manière optimale les processus cognitifs à l'œuvre au sein d'une GNAT. Dans cette tâche, et à l'inverse d'un IAT classique, une fenêtre de réponse est imposée au participant, poussant ce dernier à répondre très rapidement. Or, pour Nadarevic et Erdfelder, la présence de cette fenêtre de réponse empêcherait une éventuelle correction (i.e., inhibition) de la réponse du participant. En conséquence, pour ces auteurs le paramètre d'inhibition (OB) n'influencerait pas la réponse du participant au sein de la GNAT et serait donc superflu. À l'inverse, ces auteurs ont fait l'hypothèse d'une influence supérieure du biais de réponse (G) en comparaison au biais de réponse classique mesuré au sein de l'IAT. Ainsi, Nadarevic et Erdfelder (2011) ont construit un nouveau modèle, le Trip Model, n'incluant pas

de paramètre d'inhibition de la réponse (OB). Par le biais de deux études, ces auteurs ont comparé la qualité d'ajustement du Quad Model et du Trip Model face à des données issues d'un paradigme IAT et d'un paradigme GNAT. De manière conforme à leurs attentes, le Trip model s'ajustait mieux aux données de participants passant un paradigme GNAT. À l'inverse, le Quad Model s'ajustait mieux aux données de participants passant un paradigme IAT. En comparant ainsi ces deux modèles, les auteurs ont pu conclure qu'à l'inverse de l'IAT, le processus d'inhibition du contenu mental réactivé n'était pas à l'œuvre dans un paradigme GNAT.

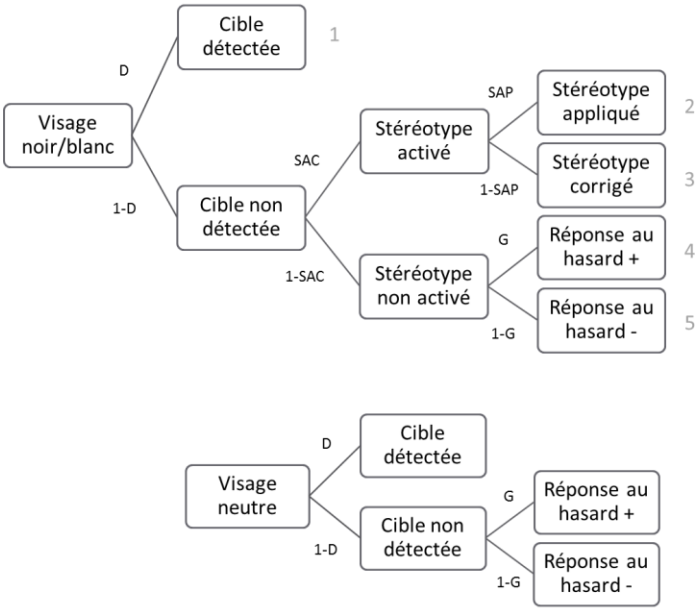
Pour résumer, nous venons de voir que, globalement, un type de modèle (e.g., Quad Model) s'adaptait mieux à un type de tâche (e.g., IAT). Nous venons aussi de voir que, selon les variations de structure de la tâche (e.g., GNAT), d'autres modèles pouvaient être préférés théoriquement et/ou empiriquement (e.g., Trip Model). Mais qu'en est-il de la SMT, notre mesure d'intérêt ? Au Chapitre 1, nous avons brièvement décrit la possibilité que les données SMT puissent être expliquées par d'autres modèles que le modèle SAC-dominant. Dans leurs étapes de validation, Krieglmeyer et Sherman (2012) ont systématiquement testé l'adéquation de quatre modèles différents (i.e., SAC-dominant, D-dominant, Quad Model, AMP). Cinq des six études effectuées ont montré, de manière conforme aux attentes de Krieglmeyer et Sherman, un meilleur ajustement aux données de la part du modèle SAC-dominant. En complément de cette première étude, Rees, Rivers et collaborateurs (2019) ont comparé l'ajustement aux données de sept modèles différents : le modèle SAC-dominant, le modèle D-dominant, le modère AMP, et les deux versions du modèle PDP (PDP-C, PDP-A) avec ou sans l'ajout d'un paramètre de biais de réponse (G). Conformément à l'étude de Krieglmeyer et Sherman (2012), Rees, Rivers et collaborateurs (2019) ont mis en évidence la supériorité du modèle SAC-dominant dans leurs quatre études. En résumé, il semble que la robustesse du modèle SAC-dominant dans la description des processus sous-tendant la SMT soit relativement bien établie. Mais cela se confirme-t-il dans nos données ?

3.2 SAC-dominant vs. D-dominant

Notre étude préliminaire (Étude 1) nous a permis de répliquer les résultats classiques SMT avec des visages maghrébins et des visages blancs. Néanmoins, comme nous l'avons rapidement spécifié alors et bien que le modèle SAC-dominant s'ajuste de façon adéquate aux

données, nous avons obtenu un meilleur ajustement aux données de la part du modèle D-dominant ($WAIC_D$ = 1514.55, $WAIC_{SAC}$ = 1547.27). De plus, nous ne l'avons pas mentionné à ce moment, mais nous avons reproduit ce résultat au cours de notre Étude 2, le modèle D-dominant s'ajustant une fois encore mieux aux données pour la condition Contrôle ($WAIC_D$ = 1784.43, $WAIC_{SAC}$ = 1803.67) mais aussi pour la condition Attentat ($WAIC_D$ = 1771.41, $WAIC_{SAC}$ = 1819.30).[11] Ainsi, comment expliquer que nous obtenions, à la fois pour notre première étude, mais aussi pour notre deuxième étude, une préférence pour le modèle D-dominant, ceci allant à l'encontre de la préférence habituelle pour le modèle SAC-dominant ? Abordons les spécificités du modèle D-dominant et son opposition avec le modèle SAC-dominant.

[11] Il est à noter cependant que dans notre Étude 3, le modèle SAC-dominant s'ajustait mieux aux données pour la condition Contrôle ($WAIC_D$ = 1735.63, $WAIC_{SAC}$ = 1576.44), mais aussi pour la condition Attentat ($WAIC_D$ = 1466.00, $WAIC_{SAC}$ = 1450.97). Ayant entrepris cet axe de recherches après l'Étude 2, nous n'avions alors pas ces informations à disposition.



Noir		Blanc	
Bas	Haut	Bas	Haut
-	+	-	+
+	+	-	-
-	-	+	+
+	+	+	+
-	-	-	-

Bas	Haut
-	+
+	+
-	-

Figure 19. Représentation en arbre du modèle D-dominant. En haut, représentation pour un type d'amorce Noir/Blanc. En bas, représentation pour un type d'amorce Neutre. Les processus cognitifs représentés sont les suivants : SAC = Activation du stéréotype, SAP = Application du stéréotype, D = Détection, G = Biais de réponse. À droite, les réponses prédites par le modèle en fonction du type d'essai (i.e., type d'amorce : Noir vs. Blanc ; type de cible : Haut vs. Bas sur la dimension de menace). Les réponses « Plus menaçant » sont représentées par le signe « + » et les réponses « Moins menaçant » par le signe « - ». Les numéros (1 à 5) correspondent aux numéros donnés aux branches de l'arbre représentant le type d'amorce Noir/Blanc.

Pour illustrer le fonctionnement du modèle D-dominant, reprenons l'exemple d'une SMT composée d'amorces de visages noirs et blancs ainsi que l'arbre multinomial associé au modèle D-dominant (Figure 19). Considérons un essai où un visage noir apparait à l'écran et observons pour chaque type de cibles (i.e., Bas et Haut sur la dimension de menace), les cas où le modèle prédira une réponse « Plus menaçant » de la part du participant (+). Commençons par les essais où l'amorce est un visage noir et la cible est menaçante. Si le participant a détecté la menace réelle de la cible (D ; première branche de l'arbre), celui-ci va donner la réponse « Plus menaçant ». Si le participant n'a pas détecté la menace réelle de la cible (1-D), mais s'il a activé son stéréotype (SAC ; « Noirs = Menace »), il pourra appliquer son

stéréotype (SAP) et répondre « Plus menaçant » (deuxième branche de l'arbre). Enfin, si le participant n'a pas détecté la menace de la cible (1-D) et s'il n'a pas activé son stéréotype (1-D), il pourra répondre au hasard « Plus menaçant » (G ; quatrième branche de l'arbre). Dans les essais où l'amorce est un visage noir et la cible est non menaçante, le participant pourra répondre « Plus menaçant » uniquement s'il n'a pas détecté la menace de la cible (1-D). Dès lors, il pourra répondre « Plus menaçant » s'il a activé son stéréotype (SAC) et s'il applique ce stéréotype (SAP ; deuxième branche de l'arbre), ou s'il n'a pas activé son stéréotype, mais qu'il a répondu au hasard « Plus menaçant » (G ; quatrième branche de l'arbre).

Selon nous, la principale différence entre le modèle SAC-dominant et le modèle D-dominant rejoint la différence que nous pouvions observer, au sein du PDP, entre les modèles A-dominant et C-dominant. Dans le modèle A-dominant, les processus contrôlés ne vont influencer la réponse du participant que si les processus automatiques ne sont pas à l'œuvre. Appliqué à la WIT et pour reprendre l'exemple de Bishara et Payne (2009), un outil sera identifié comme une arme dès lors que le biais est activé, et ce, même si le participant a identifié correctement l'outil. Le modèle A-dominant va donc prédire qu'à la fin d'un essai, s'il y a un conflit entre les informations provenant de l'amorce (i.e., association stéréotypique Noirs/armes) et les informations provenant de la cible (i.e., identification correcte de l'outil), le participant va utiliser en priorité les informations provenant de l'amorce. À l'inverse, dans le modèle C-dominant, les processus automatiques n'influenceront la réponse du participant que si les processus contrôlés ne sont pas à l'œuvre. Pour reprendre l'exemple de Bishara et Payne (2009), un outil sera identifié comme une arme que si le participant n'a pas détecté l'outil et que le stéréotype est activé. Le modèle C-dominant va donc prédire qu'à la fin d'un essai, s'il y a un conflit entre les informations provenant de l'amorce (i.e., association stéréotypique Noirs/armes) et les informations provenant de la cible (i.e., identification correcte de l'outil), le participant va utiliser en priorité les informations provenant de la cible.

L'opposition du modèle SAC-dominant et du modèle D-dominant va suivre la même logique. Comme nous l'avons vu au Chapitre 1, le modèle SAC-dominant va prédire qu'à la fin d'un essai, s'il y a un conflit entre les informations provenant de l'amorce (e.g., stéréotype « Maghrébins = Menace ») et les informations provenant de la cible (e.g., détection d'une cible non menaçante), le participant va utiliser en priorité les informations provenant de l'amorce (i.e., répondre « Plus menaçant ».) En d'autres termes, le paramètre de détection de la cible (D) est conditionnel au paramètre d'activation du stéréotype (SAC). À l'opposé, le modèle D-dominant va prédire qu'à la fin d'un essai, s'il y a le même conflit entre les informations

provenant de l'amorce (e.g., stéréotype « Maghrébins = Menace ») et les informations provenant de la cible (e.g., détection d'une cible non menaçante), le participant va utiliser en priorité les informations provenant de la cible (i.e., répondre « Moins menaçant »). En d'autres termes, le paramètre d'activation du stéréotype (SAC) est conditionnel au paramètre de détection de la cible (D).

Ainsi, comment expliquer que nos participants aient davantage suivi le modèle D-dominant (ont été influencés en priorité par la cible) là où, habituellement (e.g., Krieglmeyer & Sherman, 2012 ; Rivers et al., 2020), les participants semblent être influencés en priorité par l'amorce ? La littérature nous a montré que changer la structure de la tâche pouvait changer l'influence des paramètres à l'origine de la réponse du participant (Nadarevic & Erdfelder, 2011). Comme nous l'avons expliqué en détail, l'unique différence entre nos deux premières études et les études SMT classiques était le type d'amorce utilisé (si l'on excepte l'origine géographique des participants). Dans la tâche classique (Krieglmeyer & Sherman, 2012 ; Rees, Ma, et al., 2019 ; Rees, Rivers, et al., 2019 ; Rivers et al., 2020), les participants étaient exposés à des amorces de visages noirs et blancs. Dans nos études, les participants étaient exposés à des amorces de visages maghrébins et blancs. Serait-il possible que changer simplement le type d'amorce influence la relation conditionnelle entre processus ? C'est ce que nous avons testé dans la première étude de ce chapitre.

3.3 Étude 5 : SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc

Le but de l'Étude 5 était de reproduire le paradigme SMT avec deux conditions différentes : la première condition (SMT-Maghrébin/Blanc) était une réplication directe de nos paradigmes SMT utilisant des visages maghrébins et blancs (CaNAFF ; Courset et al., 2018). La deuxième condition (SMT-Noir/Blanc) était une réplication directe de la SMT originale de Krieglmeyer et Sherman (2012) utilisant des visages noirs et blancs (matériel expérimental identique à l'étude originale ; Phills et al., 2011). Ainsi, hormis les amorces utilisées dans nos deux conditions, tous les autres éléments étaient identiques. Pour la condition SMT-Maghrébin/Blanc, et suivant en cela les résultats de nos deux premières études, nous faisions l'hypothèse d'une meilleure adéquation aux données pour le modèle D-dominant que pour le modèle SAC-dominant. À l'inverse, pour la condition SMT-Noir/Blanc, et suivant en cela les

résultats classiques de la littérature, nous faisions l'hypothèse d'une meilleure adéquation aux données pour le modèle SAC-dominant que pour le modèle D-dominant.

De plus, cette cinquième étude était pour nous l'occasion de comparer l'attitude raciale de nos participants envers deux groupes classiquement discriminés. Comme nous l'avons vu au Chapitre 2, en Europe le groupe des Maghrébins est particulièrement étudié et souvent décrit comme étant l'un des groupes les plus négativement perçus (e.g., Dotsch et al., 2008 ; Essien et al., 2020 ; Mange et al., 2016 ; Spruyt & van der Noll, 2017). C'est la raison pour laquelle nous faisions l'hypothèse d'un effet SMT (i.e., contraste opposant les amorces Maghrébin/Noir et les amorces Blanc) plus important dans la SMT-Maghrébin/Blanc que dans la SMT-Noir/Blanc.

3.3.1 Méthode

Participants et plan expérimental. Une fois encore, nous ne pouvions pas estimer la taille d'échantillon optimale pour tester notre hypothèse principale. Cependant, dans le but de maximiser la puissance nécessaire pour observer l'impact du type d'amorce sur l'effet SMT (hypothèse exploratoire), nous avons effectué une analyse de sensibilité à l'aide du logiciel G*Power (3.1 ; Faul et al., 2007). Ainsi, 200 participants étaient suffisants pour détecter un effet d'une taille faible à moyenne (*ds* = 0.39) avec un seuil à .05 et une puissance de .80. En conséquence, dans le but de minimiser le risque d'obtenir des erreurs de Type II, nous avons fixé notre taille d'échantillon minimale à 100 participants (de culture française) par modalité de notre variable inter-sujets « type de tâche » à deux modalités. Deux-cent-quarante-trois participants ont pris part à cette étude. Parmi eux, 37 participants ont été écartés des analyses, car s'estimant de culture principale autre que française. Un participant a été exclu, car il avait déjà passé l'une de nos précédentes études. Au final, 205 participants ont été retenus pour les analyses (129 femmes, 76 hommes, $M_{\text{âge}}$ = 20.57, $ET_{\text{âge}}$ = 3.35). Les participants étaient des étudiants de l'Université Grenoble Alpes en L1 Psychologie ou hors cursus de Psychologie, et recevaient des crédits de cours ou 10 euros en compensation. Le plan expérimental de cette étude était de type 3 (amorce : Maghrébin/Noir vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (type de tâche : SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc), le dernier facteur étant en inter-sujet. Le pré-enregistrement de cette étude est disponible sur la plateforme OSF (https://osf.io/2pved).

Matériel et procédure. Les participants venaient au laboratoire pour une étude de 30 minutes sur la « Formation d'impression suboptimale ». À la différence de nos études précédentes, les passations s'effectuaient dans une salle de passations commune. À leur arrivée au laboratoire, les participants étaient installés face à un ordinateur, séparés des autres ordinateurs par des panneaux (pour un maximum de huit participants en simultanée). Les participants signaient un formulaire de consentement et l'étude débutait. La procédure était identique à notre Étude 1, les participants étant affectés aléatoirement à l'une ou l'autre des conditions dès le lancement de l'étude.

Les passations s'effectuaient sur un écran de 24 pouces, avec le logiciel E-prime 3.0®. Dans la condition « SMT-Maghrébin/Blanc », les participants passaient une SMT identique à la tâche décrite dans l'Étude 1 et l'Étude 2. Ainsi, les participants voyaient des amorces de visages maghrébins et blancs provenant de la base de données développée au sein de notre laboratoire (CaNAFF ; Courset et al., 2018). Les amorces de visages neutres provenaient de l'étude originale de Krieglmeyer et Sherman (2012) tout comme les cibles menaçantes et non menaçantes (voir l'Étude 1 pour plus de détails). Dans la condition « SMT-Noir/Blanc », les participants passaient une SMT identique à la tâche originale de Krieglmeyer et Sherman (2012). Ainsi, les participants voyaient des amorces de visages noirs et blancs provenant de l'étude de Phills et collaborateurs (2011) et réutilisées dans la SMT originale. Par conséquent, les visages blancs présentés dans cette condition étaient différents des visages blancs présentés dans la condition « SMT-Maghrébin/Blanc ». Excepté les amorces de visages, tous les autres aspects des deux SMT étaient identiques.

À la fin de la passation, de manière similaire à nos études précédentes, les participants répondaient à trois questions relatives aux visages présentés, ainsi qu'à des questions démographiques (i.e., âge, genre). Enfin, les participants étaient débriefés, gratifiés et remerciés.[12]

[12] Le lecteur peut se questionner quant à la non-utilisation de l'échelle de préjugés pour cette étude, celle-ci ayant pourtant été utilisée Étude 3. Comme nous l'avons spécifié précédemment, l'Étude 3 a été effectuée après l'Étude 5. Au moment de la réalisation de cette étude, nous n'avions donc pas encore eu la volonté d'inclure une mesure directe d'attitudes raciales.

3.3.2 Résultats

Jugement de menace. Pour analyser nos différents effets, nous avons effectué une ANOVA à plan mixte 3 (amorce : Maghrébin/Noir vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (type de tâche : SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc) avec la somme des réponses « Plus menaçant » en variable dépendante. Tout d'abord nous avons obtenu un effet significatif du type de cible, les participants jugeant les cibles menaçantes comme étant effectivement plus menaçantes (M = 34.92, ET = 10.17) que les cibles non menaçantes (M = 23.34, ET = 9.88), $t(203) = 14.57$, $p < .001$, $\eta^2_p = .51$. Cet effet n'était pas modéré par le type de tâche, $t(203) = 0.11$, $p = .91$, $\eta^2_p < .001$. Comme nous en faisions l'hypothèse, nous avons obtenu un impact significatif du type de tâche sur l'effet SMT, $t(202) = 3.97$, $p < .001$, $\eta^2_p = .07$. Ainsi, la différence de jugement de menace était effectivement plus importante entre les amorces Maghrébin (M = 23.78, ET = 8.18) et Blanc (M = 19.71, ET = 7.01), $t(202) = 4.20$, $p < .001$, $\eta^2_p = .08$, qu'entre les amorces Noir (M = 18.94, ET = 7.19) et Blanc (M = 20.19, ET = 7.00), $t(202) = -1.42$, $p = .16$, $\eta^2_p = .01$[13] (Figure 20). De la même manière, nous avons obtenu un effet d'interaction significatif entre le type de tâche et le contraste opposant les amorces Maghrébin/Noir et les amorces Neutre, $t(202) = 6.67$, $p < .001$, $\eta^2_p = .18$[14]. Ainsi, la différence de jugement de menace était une fois encore plus importante entre les amorces Maghrébin (M = 23.78, ET = 8.18) et Neutre (M = 16.07, ET = 8.22), $t(202) = 8.32$, $p < .001$, $\eta^2_p = .26$, qu'entre les amorces Noir (M = 18.94, ET = 7.19) et Neutre (M = 17.79, ET = 9.68), $t(202) = 1.16$, $p = .25$, $\eta^2_p = .01$. Enfin, le jugement de menace était plus important pour les amorces Blanc que pour les amorces Neutre, $t(202) = 4.87$, $p < .001$, $\eta^2_p = .11$, cet effet n'étant pas modéré par le type de tâche, $t(202) = 1.25$, $p = .21$, $\eta^2_p = .01$.

[13] Un participant a été retiré des analyses, car il avait un RSS > 4.
[14] Un participant a été retiré des analyses, car il avait un RSS > 4.

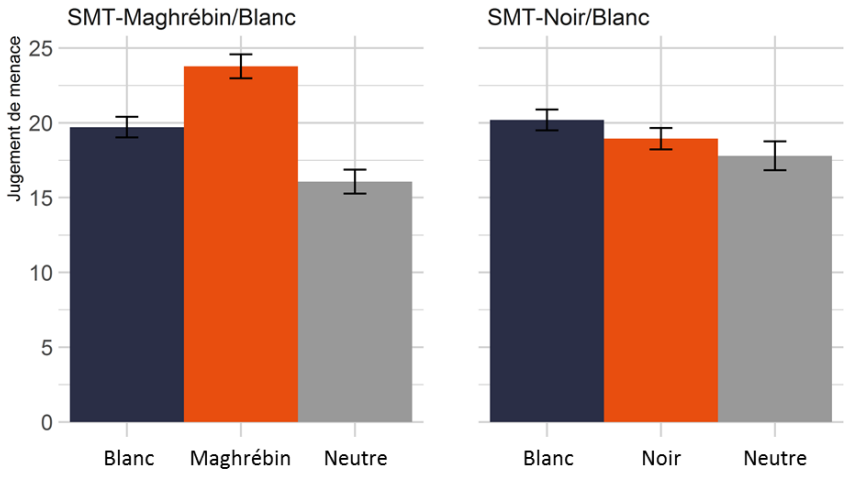


Figure 20. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin/Noir, Neutre) et de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc). Les barres d'erreurs représentent les erreurs standards.

Modèle multinomial. Pour le modèle SAC-dominant, l'ajustement (T_1) du modèle aux données pour la SMT-Noir/Blanc (M_{pred} = 0.05, M_{obs} = 0.13, p = .08, w = 0.03), et la SMT-Maghrébin/Blanc (M_{pred} = 0.05, M_{obs} = 0.20, p = .008, w = 0.04) était acceptable. De la même manière, l'ajustement (T_1) aux données du modèle D-dominant quant à la SMT-Noir/Blanc (M_{pred} = 0.05, M_{obs} = 0.16, p =.02, w = 0.03) et la SMT-Maghrébin/Blanc (M_{pred} = 0.05, M_{obs} = 0.09, p = .12, w = 0.03) était là encore bon.

Pour comparer la meilleure adéquation de nos modèles aux données et comme nous l'avons décrit au Chapitre 1, nous avons utilisé l'indice WAIC, le modèle ayant l'indice le plus bas étant décrit comme s'ajustant le mieux aux données (Watanabe, 2010). Pour rappel, nous faisions l'hypothèse d'un meilleur ajustement du modèle D-dominant pour nos participants ayant passé la SMT-Maghrébin/Blanc. Effectivement, nous avons observé un meilleur ajustement aux données du modèle D-dominant ($WAIC_D$ = 3193.32) que du modèle SAC-dominant ($WAIC_{SAC}$ = 3257.89). À l'inverse, nous faisions l'hypothèse d'un meilleur ajustement du modèle SAC-dominant pour nos participants ayant passé la SMT-Noir/Blanc. Effectivement, nous avons observé un meilleur ajustement aux données du modèle SAC-dominant ($WAIC_{SAC}$ = 3000.30) que du modèle D-dominant ($WAIC_D$ = 3019.49).

Analyses complémentaires. Comme nous en faisions l'hypothèse, nous avons pu mettre en évidence un effet SMT plus important dans la SMT-Maghrébin/Blanc que dans la SMT-Noir/Blanc. À visée exploratoire, nous avons voulu regarder la valeur des différents paramètres pouvant expliquer ces effets. Pour simplifier la lecture des analyses, nous nous sommes concentrés sur le modèle SAC-dominant à la fois pour la SMT-Noir/Blanc, mais aussi pour la SMT-Maghrébin/Blanc (bien que le modèle D-dominant s'ajuste mieux aux données dans cette condition). Les analyses relatives au modèle D-dominant sont disponibles en annexe, les résultats étant relativement similaires (Annexe 2).

Le paramètre SAC était descriptivement plus important dans la SMT-Maghrébin/Blanc (M_{SAC} = .62, $IC_{95\%}$[.43 ; .80]) que dans la SMT-Noir/Blanc (M_{SAC} = .45, $IC_{95\%}$[.27 ; .64]), mais la différence n'était pas significative, ΔSAC = .17, $IC_{95\%}$[-.10 ; .43], p_B = .10[15] (Figure 21). Par contre, le paramètre SAP était significativement plus important dans la SMT-Maghrébin/Blanc (M_{SAP} = .56, $IC_{95\%}$[.50 ; .62]) que dans la SMT-Noir/Blanc (M_{SAP} = .43, $IC_{95\%}$[.37 ; .49]), ΔSAP = .13, $IC_{95\%}$[.05 ; .22], p_B = .001. Ainsi, les participants ont plus appliqué leur stéréotype dans la SMT-Maghrébin/Blanc que dans la SMT-Noir/Blanc. Enfin, nous n'avons pas observé de différences pour les paramètres D et G entre la SMT Maghrébin/Blanc (M_D = .15, $IC_{95\%}$[.10 ; .20] ; M_G = .27, $IC_{95\%}$[.23 ; .32]) et la SMT-Noir/Blanc (M_D = .15, $IC_{95\%}$[.11 ; .19] ; M_G = .29, $IC_{95\%}$[.34 ; .35]), ces différences étant non significatives pour le paramètre D, ΔD = .00, $IC_{95\%}$[-.06 ; .06], p_B = .51, comme pour le paramètre G ΔG = -.02, $IC_{95\%}$[-.09 ; .05], p_B = .70.

[15] Il est à noter que pour le modèle D-dominant cette différence était significative (Annexe 2)

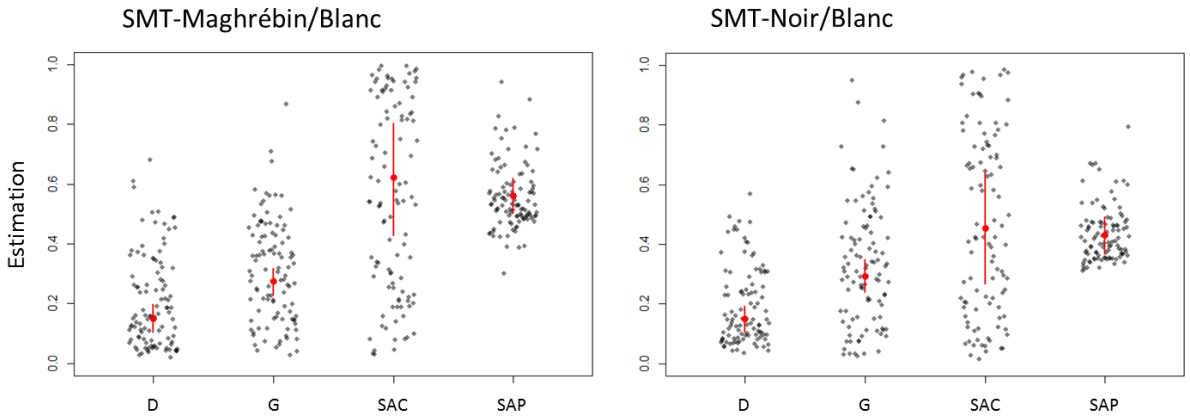


Figure 21. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc) et pour un modèle SAC-dominant à quatre paramètres.

De manière similaire à nos études du Chapitre 2, nous pouvons aller encore plus en détail et analyser ces résultats avec un modèle SAC-dominant à six paramètres. L'ajustement (T_1) du modèle aux données pour la SMT-Noir/Blanc (M_{pred} = 0.05, M_{obs} = 0.06, p = .40, w = 0.02) et la SMT-Maghrébin/Blanc (M_{pred} = 0.05, M_{obs} = 0.09, p = .11, w = 0.03) était acceptable. Intéressons-nous à la différence entre les paramètres SAC et SAP associés aux visages maghrébins ou aux visages noirs. Tout d'abord, nous n'avons pas obtenu de différence significative quant au paramètre SAC entre les visages maghrébins (M_{SAC} = .36, $IC_{95\%}$[.25 ; .50]) et les visages noirs (M_{SAC} = .30, $IC_{95\%}$[.11 ; .71]), ΔSAC = .06, $IC_{95\%}$[-.36 ; .30], p_B = .27 (Figure 22). En revanche, nous avons obtenu un SAP lié aux visages maghrébins (M_{SAP} = .91, $IC_{95\%}$[.73 ; .99]) significativement plus élevé que le SAP lié aux visages noirs (M_{SAP} = .44, $IC_{95\%}$[.11 ; .79]), ΔSAP = .46, $IC_{95\%}$[.09 ; .82], p_B = .01. En d'autres termes, les participants passant la SMT-Maghrébin/Blanc ont plus appliqué leur stéréotype « Maghrébins = Menace » que les participants passant la SMT-Noir/Blanc n'ont appliqué leur stéréotype « Noirs = Menace ». Comme on pouvait s'y attendre, nous n'avons pas obtenu de différence quant au paramètre SAC entre les visages blancs de la SMT-Maghrébin/Blanc (M_{SAC} = .35, $IC_{95\%}$[.25 ; .50]) et les visages blancs de la SMT-Noir/Blanc (M_{SAC} = .42, $IC_{95\%}$[.22 ; .68]), ΔSAC = -.06, $IC_{95\%}$[-.35 ; .20], p_B = .70, bien que ce paramètre ait été descriptivement plus important dans cette dernière condition. De la même manière, nous n'avons pas obtenu de différence

significative entre le paramètre SAP lié aux visages blancs de la SMT-Maghrébin/Blanc (M_{SAP} = .25, $IC_{95\%}$[.05 ; .43]) et celui lié aux visages blancs de la SMT-Noir/Blanc (M_{SAP} = .37, $IC_{95\%}$[.09 ; .52]), ΔSAP = -.13, $IC_{95\%}$[-.39 ; .21], p_B = .81, bien que ce paramètre ait été là encore descriptivement plus important dans cette dernière condition.

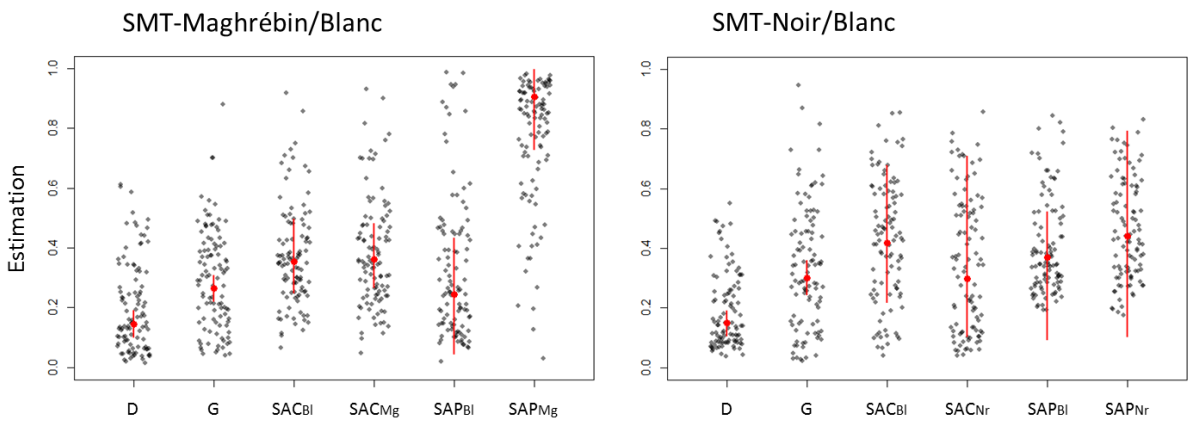


Figure 22. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc) et pour un modèle SAC-dominant à six paramètres. SAC_{Bl} et SAP_{Bl} = paramètres pour les visages blancs, SAC_{Mg} et SAP_{Mg} = paramètres pour les visages maghrébins, SAC_{Nr} et SAP_{Nr} = paramètres pour les visages noirs.

3.3.3 **Discussion**

La première étude de ce chapitre avait pour but principal de tester si la différence d'ajustement des modèles SAC-dominant et D-dominant pouvait être le reflet des différentes amorces utilisées au sein de la SMT. Conformément à nos hypothèses et en accord avec nos études précédentes, nous avons observé un meilleur ajustement du modèle D-dominant chez nos participants passant une tâche SMT constituée d'amorces Maghrébin et Blanc. De plus, là encore de manière conforme à nos hypothèses et en accord avec les résultats classiques de la littérature (e.g., Krieglmeyer & Sherman, 2012 ; Rees, Ma, et al., 2019 ; Rees, Rivers, et al., 2019 ; Rivers et al., 2020), nous avons observé un meilleur ajustement du modèle SAC-dominant chez nos participants passant une tâche SMT constituée d'amorces Noir et Blanc.

Ainsi, comme nous l'avons développé précédemment (e.g., Nadarevic & Erdfelder, 2011), il semble que changer certains aspects de la tâche puisse influencer la contribution spécifique des différents processus sous-tendant la réponse du participant. Pour reprendre les spécificités de chaque modèle, il semble ici que nos participants ayant passé une SMT-Maghrébin/Blanc aient davantage privilégié l'information provenant de la cible que celle provenant de l'amorce. À l'inverse, nos participants ayant passé une SMT-Noir/Blanc semblent avoir majoritairement privilégié l'information provenant de l'amorce que celle provenant de la cible.

Suivant ces résultats, et étant donné que l'unique différence entre nos conditions était le type d'amorces utilisées, nous nous sommes demandés si cette différence pouvait être le reflet de différences perceptives au niveau de l'amorce et/ou de la cible. Nous pouvons en effet envisager que le contraste important entre les visages noirs et blancs (qui pourraient ainsi être plus faciles à catégoriser) puisse davantage attirer l'attention des participants vers les amorces. Rees, Ma et collaborateurs (2019), en utilisant un paradigme SMT, ont montré qu'augmenter la saillance catégorielle des amorces pouvait augmenter la propension des participants à appliquer leur stéréotype (SAP plus important). Nous pouvons aller plus loin et imaginer que l'augmentation de la saillance catégorielle des amorces n'augmente pas seulement l'application du stéréotype, mais aussi la tendance des participants à utiliser en priorité les informations provenant de l'amorce. En d'autres termes, augmenter la saillance catégorielle des amorces pourrait augmenter la tendance des participants à suivre le modèle SAC-dominant. Suivant le même raisonnement, le contraste moins important entre les visages maghrébins et les visages blancs (qui pourraient ainsi être plus difficile à catégoriser) pourrait davantage attirer l'attention des participants vers les cibles, aboutissant à une utilisation prioritaire des informations issues des cibles (i.e., détection de la menace) et donc, à une meilleure adéquation aux données du modèle D-dominant. Notre étude suivante avait pour but principal de tester cette hypothèse.

Conformément à notre hypothèse exploratoire, nous avons obtenu une différence de jugement de menace plus importante entre les amorces Maghrébin et Blanc qu'entre les amorces Noir et Blanc. Autrement dit, l'effet SMT était plus important dans la SMT-Maghrébin/Blanc que dans la SMT-Noir/Blanc. Cette différence semble être le reflet d'une application et, au moins descriptivement, d'une activation plus importante du stéréotype « Maghrébins = Menace » dans la SMT-Maghrébin/Blanc que du stéréotype « Noirs = Menace » dans la SMT-Noir/Blanc. Conformément à certains travaux et notamment des

travaux effectués en France (e.g., Essien et al., 2020 ; Mange et al., 2016 ; Popa-Roch & Delmas, 2011 ; Rougier et al., 2020), nous avons pu mettre en évidence une attitude raciale négative envers les individus Maghrébins, et même plus négative que d'autres groupes pourtant classiquement discriminés (i.e., les individus de couleur noire). Se pourrait-il qu'effectivement, en France, les individus maghrébins soient perçus plus négativement que les individus noirs ? De ce fait, à quels résultats peut-on s'attendre dans un contexte géographique différent où les populations noires sont classiquement associées au stéréotype de menace ? Nous aborderons ces points dans le Chapitre 4 de ce manuscrit.

3.4 Étude 6 : manipulation de la perception de l'amorce et de la cible

L'objectif de cette sixième étude était de manipuler directement la tendance des participants à utiliser en priorité l'amorce (i.e., modèle SAC-dominant) ou au contraire la cible (i.e., modèle D-dominant) pour formuler leur réponse. Nous pensions que le contraste important entre les visages noirs et blancs pouvait faciliter la catégorisation de ces visages en comparaison à des visages maghrébins et blancs. Plusieurs études ont montré que le caractère prototypique d'un visage pouvait influencer le phénomène de catégorisation ethnique et le jugement stéréotypique lié à cette catégorisation (Blair et al., 2004 ; Blair et al., 2002 ; Freeman, Penner, et al., 2011). Nous pensions ainsi que manipuler le caractère prototypique des amorces pouvait jouer sur le phénomène de catégorisation des visages et ainsi sur la tendance à utiliser les informations provenant de l'amorce. Krieglmeyer et Sherman (2012, Étude 4) ont manipulé le caractère prototypique de l'amorce et ont montré que celle-ci influençait le paramètre SAC. Dans notre présente étude, nous sommes allés plus loin et faisions l'hypothèse que des amorces plus prototypiques de la catégorie, plutôt que des amorces moins prototypiques, augmenteraient la propension de nos participants à suivre le modèle SAC-dominant. À l'opposée, Krieglmeyer et Sherman (2012, Étude 2) ont montré qu'utiliser des cibles plus ou moins discernables sur la dimension de menace influençait le paramètre D. là encore, nous sommes allés plus loin et faisions l'hypothèse que des dessins cibles plus faciles à discerner (sur la dimension de menace), plutôt que des dessins cibles plus difficiles à discerner, augmenteraient la propension de nos participants à suivre le modèle D-dominant.

De plus, indépendamment de la manipulation du type d'amorce et du type de cible, nous avons introduit une nouvelle fois la mesure directe d'attitude raciale anti-Maghrébins utilisée lors de l'Étude 3, à savoir l'Échelle de Préjugés Généralisés (Dambrun & Guimond, 2001). Là encore, nous voulions observer si le score à cette échelle pouvait être lié aux jugements de menace, aux différents paramètres du modèle, mais aussi à la relation conditionnelle entre processus (i.e., modèle SAC-dominant vs. modèle D-dominant).

3.4.1 **Méthode**

Participants et plan expérimental. À partir d'une analyse de sensibilité (G*power 3.1 ; Faul et al., 2007) et des corrélations attendues entre nos variables d'intérêt (i.e., r = .70 ; Krieglmeyer & Sherman, 2012, Étude 4 ; r = .17; Krieglmeyer & Sherman, 2012, Étude 2), 100 participants devaient être suffisants pour détecter une taille d'effet petite à moyenne (f = .20), avec une puissance de .80 et un calcul prenant en compte la multiplicité des tests (Perugini et al., 2018). En conséquence, nous voulions recruter au moins 100 participants s'estimant de culture française. Cent-trente-trois participants ont pris part à cette étude. Parmi eux, 115 se s'estimaient de culture principale française. Deux participants ont été écartés des analyses, car donnant systématiquement la même réponse à l'ensemble des essais (répondant plus de 98% du temps « Plus menaçant » ou « Moins menaçant »). Au final, 113 participants ont été gardés pour les analyses (90 femmes, 23 hommes, $M_{\text{âge}}$ = 20.90, $ET_{\text{âge}}$ = 2.50). Aucun des participants recrutés n'avait pris part à l'une de nos études précédentes. Les participants étaient des étudiants de l'Université Grenoble Alpes et recevaient des crédits de cours ou 10 euros en compensation. Le plan expérimental de l'Étude 6 était de type 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (caractère prototypique de l'amorce : Haut vs. Bas) x 2 (caractère discernable de la cible : Haut vs. Bas) x score de préjugé, les quatre premiers facteurs étant en intra-sujets. Le pré-enregistrement de cette étude est disponible sur la plateforme OSF (https://osf.io/ud7ah).

Matériel et procédure. Les participants venaient au laboratoire pour une étude de 20 minutes sur la « Formation d'impression suboptimale ». À leur arrivée, un expérimentateur les installait dans des box individuels. Les participants signaient un formulaire de consentement

et l'étude débutait. Les consignes présentées à l'écran étaient similaires à celles de l'Étude 1. Immédiatement après l'énumération de ces consignes, la tâche SMT débutait.

Les passations s'effectuaient sur un écran de 24 pouces, avec le logiciel E-prime 3.0®. Les amorces étaient des photos de 24 visages maghrébins, 24 visages blancs ou 24 visages neutres. Comme dans nos études précédentes, les visages maghrébins et blancs étaient issus de la base de données CaNAFF (Courset et al., 2018). Douze visages maghrébins et 12 visages blancs ont été sélectionnés à partir des 24 visages maghrébins et des 24 visages blancs utilisés dans nos précédentes SMT. Ces visages constituaient ainsi les amorces prototypiques Maghrébin et Blanc. De plus, 12 visages maghrébins et 12 visages blancs ont été sélectionnés dans la base de données CaNAFF parmi les visages les moins prototypiques de ces catégories. Ainsi, ces visages constituaient les amorces non prototypiques Maghrébin et Blanc. De plus, les 48 dessins de visages cibles (i.e., 24 dessins de visages non menaçants et 24 dessins de visages menaçants) ont été sélectionnés à partir d'une base de données OSF (Rivers & Sherman, 2017 ; https://osf.io/pqbhf/) et créés à partir du générateur de visages de Todorov et collaborateurs (Todorov et al., 2013 ; Todorov & Oosterhof, 2011). Parmi ces visages, la moitié était facilement discernable (-3 ou +3 écart type de la moyenne de menace), l'autre moitié étant difficilement discernable (-1 ou +1 écart type de la moyenne de menace).

Pour s'assurer d'un design factoriel complet (i.e., chaque type d'amorce étant associé de manière égale à chaque type de cible), le nombre d'essais a été augmenté à 288. Une fois encore les participants commençaient par deux phases d'entrainements de six essais chacune. Dans la première phase d'entrainement, les participants ne voyaient que des dessins cibles. Ces dessins cibles étaient choisis aléatoirement parmi l'ensemble des dessins cibles utilisés dans nos précédentes études. Dans la seconde phase d'entrainement, les participants voyaient ces dessins cibles précédés d'amorces de visages. Ces visages étaient choisis parmi les douze visages prototypiques maghrébins et blancs issus de nos précédentes SMT et n'ayant pas été sélectionnés pour être inclus dans la phase test de cette étude. La phase test était constituée de deux blocs de 144 essais, séparés par une courte pause. Comme précédemment, à la fin de la SMT, les participants répondaient à trois questions relatives aux visages présentés ainsi qu'à des questions démographiques (i.e., âge, genre). Enfin, les participants devaient remplir une version papier de l'Échelle de Préjugés Généralisés (α = .88). Comme précédemment, pour insister sur l'anonymat de leurs réponses, les participants devaient déposer leur questionnaire complété dans une urne à la toute fin de l'étude. Pour terminer, ils étaient débriefés, gratifiés et remerciés.

3.4.2 Résultats

Jugement de menace. Pour nos analyses, nous avons effectué une ANOVA à plan mixte 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (caractère prototypique de l'amorce : Haut vs. Bas) x 2 (caractère discernable de la cible : Haut vs. Bas) avec la somme des réponses « Plus menaçant » en variable dépendante. Tout d'abord, les participants ont jugé les cibles menaçantes comme effectivement plus menaçantes (M = 66.46, ET = 24.09) que les cibles non menaçantes (M = 47.91, ET = 21.04), $t(112) = 8.93$, $p < .001$, $\eta^2_p = .41$. Cet effet était plus important pour les cibles fortement discernables que pour les cibles faiblement discernables, $t(112) = 8.18$, $p < .001$, $\eta^2_p = .37$. De manière équivalente à nos études précédentes, nous avons répliqué l'effet SMT. Ainsi, les participants ont jugé les cibles comme plus menaçantes quand celles-ci étaient précédées de visages maghrébins (M = 47.14, ET = 16.51) plutôt que de visages blancs (M = 38.44, ET = 15.20), $t(112) = 5.91$, $p < .001$, $\eta^2_p = .24$. De plus, cet effet était supérieur pour les visages fortement prototypiques que pour les visages peu prototypiques, $t(112) = 7.78$, $p < .001$, $\eta^2_p = .35$ (Figure 23). Le contraste opposant les amorces Maghrébin et Neutre était lui aussi significatif. Ainsi, les participants ont jugé les cibles comme plus menaçantes quand elles étaient précédées de visages maghrébins (M = 47.14, ET = 16.51) plutôt que de visages neutres (M = 30.78, ET = 21.59), $t(112) = 6.72$, $p < .001$, $\eta^2_p = .29$. De la même manière, le contraste opposant les amorces Blanc et Neutre était significatif. En d'autres termes, les participants ont jugé les cibles comme plus menaçantes quand elles étaient précédées de visages blancs (M = 38.44, ET = 15.20), plutôt que de visages neutres (M = 30.78, ET = 21.59), $t(112) = 3.27$, $p = .001$, $\eta^2_p = .09$.

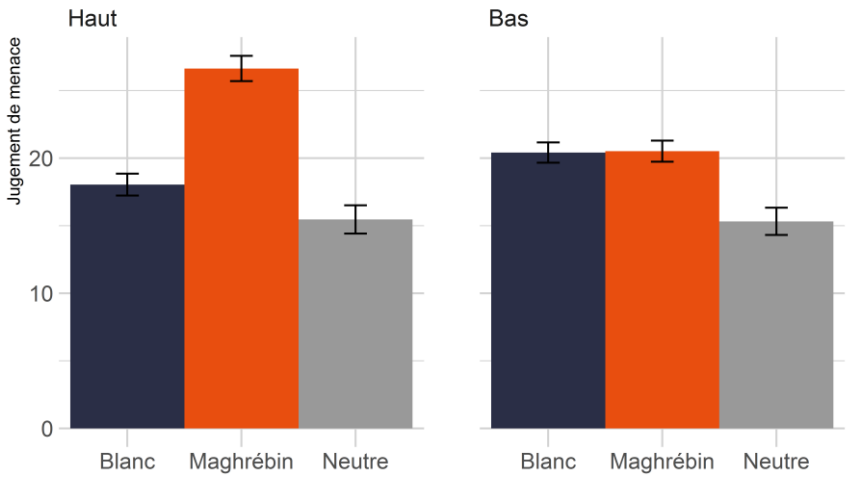


Figure 23. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin, Neutre) et du caractère prototypique de l'amorce (Haut vs. Bas). Les barres d'erreurs représentent les erreurs standards.

Pour analyser de manière exploratoire le lien entre le jugement de menace et la mesure directe d'attitude raciale anti-Maghrébins (i.e., Échelle de Préjugés Généralisés), nous avons effectué des régressions linéaires entre ces résultats et le score de préjugé selon un modèle de type 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (caractère prototypique de l'amorce : Haut vs. Bas) x 2 (caractère discernable de la cible : Haut vs. Bas) x score de préjugé, avec la somme des réponses « Plus menaçant » en variable dépendante. Tout d'abord, nous avons observé un lien significatif entre le score de préjugé et le jugement de la cible. Ainsi, plus les participants avaient un score de préjugé élevé et moins ils jugeaient les cibles comme menaçantes quand celles-ci étaient effectivement menaçantes que lorsqu'elles ne l'étaient pas, b = -4.84, $t(111)$ = -2.30, p = .02, η^2_p = .05. Cet effet était plus important chez les cibles hautement discernables par rapport aux cibles faiblement discernables, b = -2.62, $t(111)$ = -2.47, p = .02, η^2_p = .05. Plus intéressant pour notre propos, nous avons également observé un lien significatif entre le score de préjugé et l'effet SMT. Plus les participants avaient un préjugé élevé et plus l'effet SMT, c'est-à-dire la différence de jugement entre les amorces Maghrébin et Blanc, était important, b = 4.88, $t(111)$ = 3.43, p < .001, η^2_p = .10. Ce lien positif entre le score de préjugé et l'effet SMT était plus important pour les amorces prototypiques que pour les amorces peu prototypiques, b = 3.93, $t(111)$ = 3.76, p < .001, η^2_p = .11. Enfin, la relation entre le score de préjugé et le

contraste opposant les amorces Maghrébin et Neutre n'était pas significative, b = 1.48, $t(111)$ = 0.60, p = .55, η^2_p = .003. De même, la relation entre le score de préjugé et le contraste opposant les amorces Blanc et Neutre n'était une fois encore pas significative, b = -3.40, $t(111)$ = -1.44, p = .15, η^2_p = .02.

Modèle multinomial. Dans notre pré-enregistrement nous avions initialement prévu, comme pour nos études précédentes, de nous référer à l'indice WAIC pour comparer l'ajustement aux données des modèles SAC-Dominant et D-dominant pour chaque combinaison d'amorce (i.e., caractère prototypique de l'amorce) et de cible (i.e., caractère discernable de la cible). Nous faisions ainsi l'hypothèse que pour les essais comportant une amorce peu prototypique et une cible hautement discernable (i.e., amorce_bas /cible_haut), le modèle D-dominant devait avoir un WAIC plus faible (i.e., une meilleure adéquation aux données) que le modèle SAC-dominant. Contrairement à nos hypothèses, nous avons obtenu une meilleure adéquation aux données du modèle SAC-dominant (WAIC = 2743.90) que du modèle D-dominant (WAIC = 2751.10). Par ailleurs, nous faisions à l'inverse l'hypothèse que pour les essais comportant une amorce très prototypique et une cible faiblement discernable (i.e., amorce_haut/cible_bas), le modèle SAC-dominant devait avoir un WAIC plus faible (i.e., une meilleure adéquation aux données) que le modèle D-dominant. Effectivement, pour cette condition, nous avons obtenu une meilleure adéquation aux données de la part du modèle SAC-dominant (WAIC = 2752.11) que du modèle D-dominant (WAIC = 2764.93). Enfin, nous n'avions pas d'hypothèses spécifiques pour les deux conditions intermédiaires, à savoir les essais comportant une amorce peu prototypique et une cible faiblement discernable (i.e., amorce_bas/cible_bas) et les essais comportant une amorce très prototypique et une cible fortement discernable (i.e., amorce_haut/cible_haut). Pour la condition amorce_bas/cible_bas, le modèle D-dominant s'est mieux ajusté aux données (WAIC = 2792.64) que le modèle SAC-dominant (WAIC = 2806.76). À l'inverse, pour la condition amorce_haut/cible_haut, le modèle SAC-dominant s'est mieux ajusté aux données (WAIC = 2711.03) que le modèle D-dominant (WAIC = 2779.64).

L'inconvénient de l'approche développée ci-dessus est que, dans le cas d'un plan expérimental intra-sujets, chaque modèle n'est appliqué que sur une seule partie des données (i.e., les essais correspondant aux combinaisons spécifiques entre amorces et cibles). Ainsi, dans le but de maximiser la puissance de nos analyses, nous avons construit un nouveau modèle multinomial permettant de tester nos hypothèses en prenant en considération

l'ensemble des essais. Suivant en cela certains travaux précoces s'intéressant aux méthodes de sélections de modèles multinomiaux (Atkinson, 1970), nous avons créé un nouveau modèle multinomial dans lequel étaient « nichés » à la fois le modèle SAC-dominant et le modèle D-dominant (Figure 24). Plus précisément, nous avons créé un nouveau paramètre « λ » ayant pour but d'intégrer (i.e., de connecter entre eux) les modèles SAC-dominant et D-dominant. En conséquence, une valeur λ élevée (λ > .5) avait pour signification une meilleure adéquation aux données du modèle SAC-dominant. À l'inverse, une valeur faible de λ (λ <.5) avait pour signification une meilleure adéquation aux données du modèle D-dominant. Nous avons ainsi créé un modèle global dans lequel nous avions un paramètre D variant en fonction du caractère discernable de la cible (D1, D2) et des paramètres SAC/SAP variant en fonction du caractère prototypique de l'amorce (SAC1, SAC2 et SAP1, SAP2)[16]. De plus le paramètre λ variait en fonction du caractère discernable de la cible ET du caractère prototypique de l'amorce (λ1, λ2, λ3 et λ4)[17].

[16] Pour le paramètre D, D1 était associé aux cibles peu discernables et D2 aux cibles fortement discernables. Pour les paramètres SAC et SAP, SAC1 et SAP1 étaient associés aux amorces faiblement prototypiques, SAC2 et SAP2 étant associés aux amorces fortement prototypiques.

[17] Spécifiquement chaque paramètre λ était associé aux croisements (2x2) suivants : λ1 = « amorce_bas/cible_bas », λ2 = « amorce_bas/cible_haut », λ3 = « amorce_haut/cible_bas », λ4 = « amorce_haut/cible_haut »

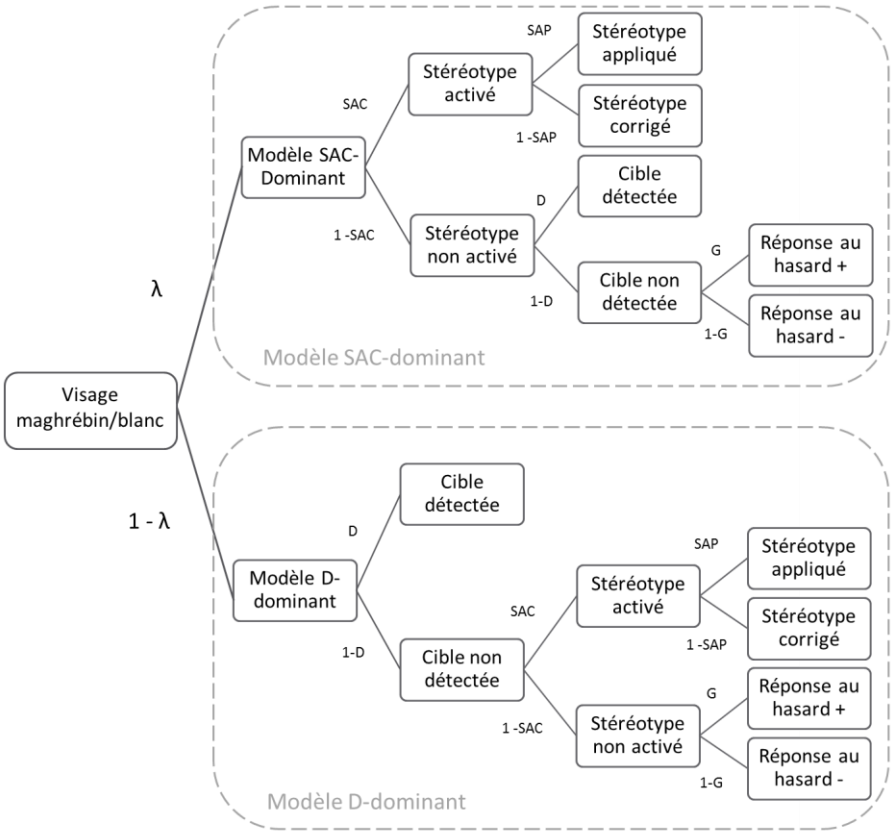


Figure 24. Représentation en arbre du modèle englobant à la fois le modèle SAC-dominant (en haut) et le modèle D-dominant (en bas). Le paramètre λ connecte ces deux modèles : λ = tendance à suivre le modèle SAC-dominant, 1-λ = tendance à suivre le modèle D-dominant. La signification des autres processus cognitifs est identique aux modèles SAC-dominant et D-dominant classiques : SAC = Activation du stéréotype, SAP = Application du stéréotype, D = Détection, G = Biais de réponse.

L'ajustement (T_1) aux données de ce nouveau modèle était acceptable (M_{pred} = 0.18, M_{obs} = 0.37, p =.01, w = 0.04). Tout d'abord, d'un point de vue descriptif, nous pouvons voir que le paramètre SAC était plus important pour les visages prototypiques (M_{SAC2} = .87, $IC_{95\%}$ [.69 ; .97]) que pour les visages peu prototypiques (M_{SAC1} = .79, $IC_{95\%}$ [.52 ; .94]), la différence étant toutefois non significative, ΔSAC = .08, $IC_{95\%}$[-.06 ; .25], p_B = .13 (Figure 25). Par contre, le paramètre SAP était plus important pour les visages prototypiques (M_{SAP2} = .67, $IC_{95\%}$ [.62 ; .73]) que pour les visages peu prototypiques (M_{SAP1} = .52, $IC_{95\%}$ [.50 ; .54]), la différence étant cette fois significative, ΔSAP = .16, $IC_{95\%}$[.11 ; .21], p_B < .001. Ainsi, les participants ont plus appliqué leur stéréotype quand les amorces étaient prototypiques que lorsqu'elles étaient peu prototypiques. Concernant les cibles, le paramètre D était

effectivement plus important quand les cibles étaient hautement discernables (M_{D2} = .12, $IC_{95\%}$ [.07 ; .18]) que quand elles étaient faiblement discernables (M_{D1} = .04, $IC_{95\%}$ [.02 ; .06]), cette différence étant significative, ΔD = .08, $IC_{95\%}$[.04 ; .14], p_B < .001. En d'autres termes, les participants ont davantage détecté le trait menaçant de la cible quand celles-ci étaient hautement discernables que lorsqu'elles étaient faiblement discernables. Pour finir, et pour en venir à nos hypothèses principales, intéressons-nous aux paramètres λ. Pour rappel, nous faisions l'hypothèse d'une meilleure adéquation du modèle SAC-dominant pour les essais « amorce_haut/cible_bas » par rapport aux essais « amorce_bas/cible_haut », où le modèle D-dominant devait prédominer. En d'autres termes, nous devions obtenir un λ plus important dans la première condition (et supérieur à .5) par rapport à la seconde (et inférieur à .5). Contrairement à nos hypothèses, nous n'avons obtenu aucune différence du paramètre λ entre la première condition ($M_{\lambda 3}$ = .44, $IC_{95\%}$ [.02 ; .95]) et la seconde ($M_{\lambda 2}$ = .44, $IC_{95\%}$ [.02 ; .94]), $\Delta\lambda$ = -.00, $IC_{95\%}$[-.71 ; .73], p_B = .51.

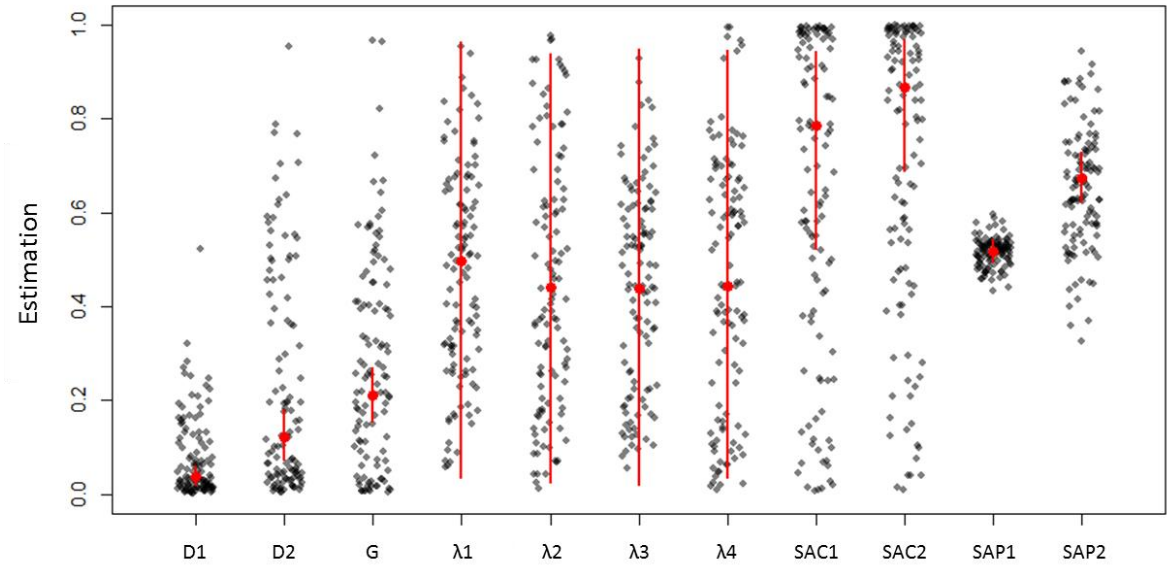


Figure 25. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction du caractère prototypique de l'amorce (Haut vs. Bas) et du caractère discernable de la cible (Haut vs. Bas). SAC1 et SAP1 = amorces peu prototypiques (Bas), SAC2 et SAP2 = amorces fortement prototypiques (Haut). D1 = cibles peu discernables (Bas), D2 = cibles fortement discernables (Haut). λ1 = amorces Bas / cible Bas, λ2 = amorces Bas / cible Haut, λ3 = amorces Haut / cible Bas, λ4 = amorces Haut / cible Haut.

Si nous analysons la Figure 25, nous pouvons voir, au niveau des différents λ, que l'intervalle de crédibilité associé à l'estimation groupale moyenne comprend, à quelques exceptions près, l'ensemble des valeurs possibles (e.g., $M_{\lambda 1}$ = .50, $IC_{95\%}$ [.04 ; .97]). De plus, les estimations individuelles des différents λ sont elles-mêmes particulièrement étendues. En d'autres termes, il semble y avoir une forte hétérogénéité chez nos participants quant au meilleur modèle pouvant sous-tendre leurs réponses. Certains participants semblent suivre en grande majorité le modèle SAC-dominant (λ proche de 1), d'autres semblent suivre uniquement le modèle D-dominant (λ proche de 0) alors que d'autres semblent alterner entre les deux modèles en fonction des essais (λ proche de .5). Comment expliquer cette forte variabilité ? Des analyses complémentaires nous ont permis d'explorer la relation possible entre cette hétérogénéité interindividuelle et l'attitude raciale auto-rapportée. Afin d'observer le lien potentiel entre la mesure directe d'attitude raciale et les différents paramètres du modèle, nous avons extrait les valeurs individuelles de chaque paramètre et avons regroupé, pour chaque participant, la valeur des différents paramètres (e.g., λ1, λ2, λ3 et λ4) en un score moyen (λ_{moy}). Nous avons ensuite effectué des régressions linéaires entre le score de préjugé individuel et le score moyen de chaque paramètre (i.e., λ_{moy}, SAC_{moy}, SAP_{moy}, D_{moy} et G).

Tout d'abord, la relation entre le score de préjugé et le paramètre λ était significative. Ainsi, plus les participants avaient un score de préjugé anti-Maghrébins élevé, plus ils avaient une valeur λ élevée, $b = 0.06$, $t(111) = 2.65$, $p = .009$, $\eta^2_p = .059$ (Figure 26). En d'autres termes, plus les participants avaient un préjugé auto-rapporté anti-Maghrébins, plus ils avaient tendance à suivre le modèle SAC-dominant. De plus, nous avons observé un lien positif et significatif entre le score de préjugé et le paramètre SAP, $b = 0.03$, $t(111) = 4.25$, $p < .001$, $\eta^2_p = .14$. Ainsi, plus les participants avaient un préjugé auto-rapporté anti-Maghrébins, plus ils avaient tendance à appliquer leur stéréotype. De même, nous avons observé un lien négatif et significatif entre le score de préjugé et le paramètre D, $b = -0.04$, $t(111) = -2.69$, $p = .008$, $\eta^2_p = .06$. Ainsi, plus les participants avaient un préjugé auto-rapporté anti-Maghrébins, moins ils avaient tendance à détecter la menace de la cible. Enfin, nous n'avons trouvé de lien significatif ni entre le score de préjugé et le paramètre SAC, $b = -0.01$, $t(111) = -0.30$, $p = .77$, $\eta^2_p = .001$, ni entre le score de préjugé et le paramètre G, $b = 0.03$, $t(111) = 1.24$, $p = .22$, $\eta^2_p = .01$. Un tableau de corrélation récapitulatif est disponible en annexes (Annexe 3).

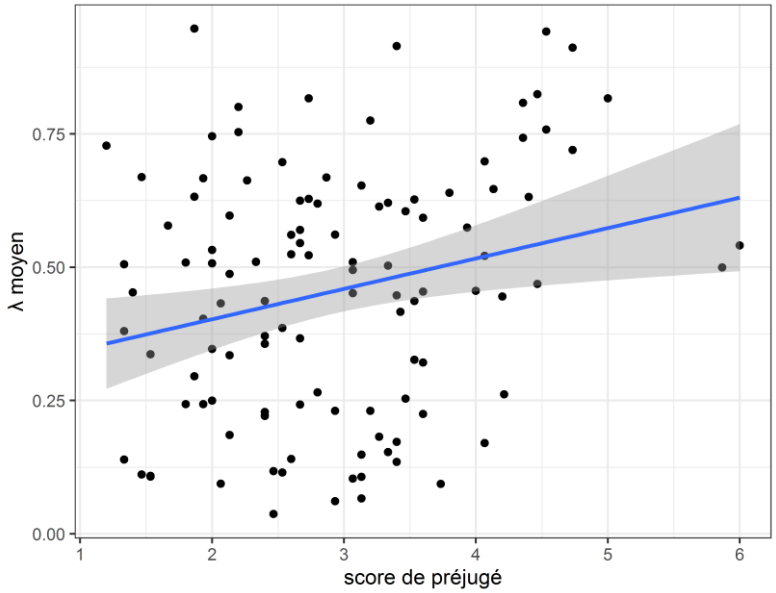


Figure 26. λ moyen en fonction du score de préjugé. Les zones grisées représentent les intervalles de confiance à 95%.

3.4.3 **Discussion**

Cette sixième étude avait pour but principal d'observer si des différences perceptives au sein des stimuli composant la SMT pouvaient être à l'origine d'une différence d'adéquation aux données des modèles SAC-dominant et D-dominant. Contrairement à ce que nous envisagions, manipuler le caractère prototypique de l'amorce n'a pas favorisé l'adéquation du modèle SAC-dominant par rapport au modèle D-dominant. De même, manipuler le caractère distinctif de la cible n'a pas favorisé l'adéquation du modèle D-dominant par rapport au modèle SAC-dominant.

Pour répondre au mieux à nos hypothèses et pour la première fois au sein de ce manuscrit, nous avons créé un nouveau modèle, regroupant au sein d'un même modèle multinomial les modèles SAC-dominant et D-dominant. Comme on pouvait s'y attendre au regard de précédentes études (Krieglmeyer & Sherman, 2012 ; Rees, Ma et al., 2019), manipuler le caractère prototypique de l'amorce et le caractère discernable de la cible a impacté les paramètres liés à ces stimuli. Tout d'abord, manipuler le caractère distinctif de la cible a effectivement impacté le paramètre D. Répliquant en cela les travaux de Krieglmeyer et Sherman (2012, Étude 2) les participants ont davantage détecté la menace de la cible

quand celle-ci était hautement discernable que quand elle était faiblement discernable. Ensuite, et cela peut paraître surprenant, manipuler le caractère prototypique de l'amorce n'a pas influencé le paramètre SAC, mais le paramètre SAP. Pour rappel, Krieglmeyer et Sherman (2012, Étude 4) ont montré que manipuler le caractère prototypique de visages noirs et blancs influençait le paramètre SAC du modèle sans influencer les autres paramètres. Si nous observons la Figure 25, nous pouvons voir que, descriptivement, le paramètre SAC est effectivement plus élevé dans les essais où l'amorce est prototypique que dans les essais où elle l'est moins. La valeur du paramètre SAC étant toutefois élevée dans les deux conditions, nous pouvons penser qu'un effet plafond puisse être à l'origine de l'absence de différence significative entre les conditions. Toutefois, comment expliquer que le paramètre SAP ait été influencé par cette manipulation ? En accord avec certains travaux de la littérature (e.g., Blair et al., 2004 ; Blair et al., 2002 ; Freeman, Penner, et al., 2011), nous avons manipulé le caractère prototypique de nos amorces pour augmenter ou au contraire diminuer leur saillance catégorielle. Or, comme nous l'avons expliqué, Rees, Ma et collaborateurs (2019) ont montré qu'augmenter la saillance catégorielle pouvait affecter le paramètre SAP. Suivant les travaux de ces auteurs, nous pouvons donc penser qu'en manipulant le caractère prototypique de nos amorces, nous avons pu jouer sur la saillance catégorielle de ces dernières et en conséquence sur le processus d'application du stéréotype.

Bien que nos manipulations expérimentales aient effectivement influencé les processus liés aux amorces (i.e., SAP) et les processus liés aux cibles (i.e., D), nous n'avons pas observé, comme nous en faisions l'hypothèse, une influence de nos manipulations sur la différence d'adéquation entre les modèles SAC-dominant et D-dominant. Effectivement, pour l'ensemble de nos quatre paramètres λ correspondant au croisement entre nos conditions 2 (caractère prototypique de l'amorce : Haut vs. Bas) x 2 (caractère discernable de la cible : Haut vs. Bas), nous avons obtenu des intervalles de crédibilités associés englobants des estimations de paramètres proches de 0 (i.e., suivre le modèle D-dominant) et de 1 (i.e., suivre le modèle SAC-dominant). Cette hétérogénéité semblait être en partie la résultante de valeurs individuelles λ elles aussi hétérogènes. Indépendamment des conditions expérimentales, certains participants avaient une valeur λ élevée (proche de 1) et donc suivaient le modèle SAC-dominant sur la totalité des essais. D'autres avaient des valeurs λ faibles (proche de 0) et suivaient donc le modèle D-dominant sur la totalité des essais. Par le biais d'analyses complémentaires, nous avons de ce fait voulu observer si cette variabilité interindividuelle pouvait être reliée à notre mesure directe d'attitude raciale.

Effectivement, l'estimation individuelle du paramètre λ était reliée au score de préjugé auto-rapporté. Plus précisément, plus les participants avaient une attitude raciale auto-rapportée négative envers les Maghrébins, plus ils suivaient le modèle SAC-dominant. Autrement dit, plus les participants avaient une attitude raciale anti-Maghrébins plus ils utilisaient en priorité l'information provenant de l'amorce pour formuler leur réponse. Comment expliquer ce résultat ? Nous pouvons envisager que les participants ayant une attitude raciale anti-Maghrébins plus importante portent davantage leur attention sur les amorces de visages maghrébins et blancs, peut-être parce qu'ils auraient une volonté plus faible de ne pas se faire influencer par ces visages. En accord avec cette idée, nous avons observé un lien positif et significatif entre le score de préjugé individuel et le paramètre SAP, signe que ces participants appliquent davantage leur stéréotype. Dans le même ordre d'idée, nous pouvons envisager que les participants ayant une attitude raciale anti-Maghrébins plus faible portent moins leur attention sur les amorces de visages maghrébins et blancs, peut-être parce qu'ils auraient une volonté plus importante de ne pas se faire influencer par ces visages. En accord avec cette idée, nous avons observé un lien négatif et significatif entre le score de préjugé auto-rapporté et le paramètre D, signe que ces participants détectent davantage le niveau de menace de la cible.

Par le passé, des études ont mis en évidence un lien entre une mesure directe individuelle et les processus à l'origine de la réponse du participant (e.g., Gonsalkorale et al., 2011 ; Krieglmeyer & Sherman, 2012). Toutefois, à notre connaissance, c'est la première fois qu'une recherche met en lumière un lien entre une mesure directe individuelle et la relation conditionnelle entre ces différents processus. Plusieurs bémols peuvent cependant être mentionnés. Tout d'abord, le but premier de cette étude n'était pas d'observer le lien entre une mesure directe et l'adéquation des différents modèles (SAC-dominant vs. D-dominant). Comme nous l'avons décrit ici et comme cela est spécifié dans notre pré-enregistrement, les effets liés à cette mesure directe n'étaient qu'exploratoires et demandent donc à être confirmés. De plus, comme nous l'avons mentionné plus haut, l'intervalle de crédibilité des paramètres λ était important, reflétant une forte hétérogénéité au niveau de l'estimation des valeurs de paramètres groupales, mais aussi au niveau des paramètres individuels. La méthode *latent-trait* donne, pour chaque paramètre individuel, son estimation moyenne ainsi que l'incertitude associée à cette estimation (i.e, intervalle de crédibilité). Cette incertitude est prise en compte dans l'incertitude liée à l'estimation groupale moyenne. En extrayant les moyennes des paramètres individuels, nous avons supprimé artificiellement cette incertitude. Or, il reste envisageable que les différents liens obtenus entre la mesure directe de préjugé et

les valeurs de paramètres ne soient pas reproductibles. Dans le but de confirmer d'une part, ces résultats exploratoires, et d'autre part, que la suppression artificielle de cette incertitude n'ait pas créée des erreurs de Type I (faux positifs) entre nos variables, nous avons entrepris de répliquer cette étude.

3.5 Étude 7 : mesure directe et relation conditionnelle entre processus

La dernière étude de ce chapitre avait pour objectif de répliquer et donc confirmer les résultats exploratoires de notre étude précédente. Conformément aux résultats de notre précédente étude, nous faisions l'hypothèse d'un lien positif entre la mesure directe d'attitude raciale anti-Maghrébins et le paramètre λ. Autrement dit, plus les individus auraient une attitude raciale auto-rapportée négative envers les Maghrébins et plus ils utiliseraient en priorité les informations provenant de l'amorce plutôt que de la cible (i.e., modèle SAC-dominant). Suivant là encore les résultats de notre étude précédente, nous faisions également l'hypothèse d'un lien positif entre le score de préjugé et le paramètre SAP. Ainsi, plus les participants auraient un score élevé sur l'Échelle de Préjugés Généralisés (Dambrun & Guimond, 2001) et plus ils appliqueraient leur stéréotype. De même, nous faisions l'hypothèse d'un lien négatif entre le score de préjugé et le paramètre D. En d'autres termes, plus les participants auraient un score faible sur l'Échelle de Préjugés Généralisés et plus ils détecteraient la menace de la cible.

Dans notre étude précédente, nous avons montré que la tendance à suivre un modèle plutôt qu'un autre était liée à l'attitude raciale auto-rapportée. Il est probable qu'un tel lien puisse être le reflet d'aspects motivationnels et notamment de la motivation à réduire l'expression de son préjugé. Comme nous l'avons vu au Chapitre 1, des études ont montré un lien entre ces aspects motivationnels et les processus cognitifs sous-tendant le paradigme IAT (Gonsalkorale et al., 2011). C'est la raison pour laquelle nous avons inclus deux échelles motivationnelles : l'échelle de motivation interne et externe à répondre sans préjugés (MRWP ; Plant & Devine, 1998) et l'échelle de motivation à contrôler ses réactions teintées de préjugé (MCPR ; Dunton & Fazio, 1997). Nous faisions l'hypothèse que les scores globaux à chacune de ces deux échelles seraient corrélés négativement au paramètre λ. Autrement dit, plus les individus auraient une forte motivation à réduire leur préjugé, et plus ils devraient suivre le modèle D-dominant (i.e., utiliseraient en priorité l'information provenant de la cible).

De plus, nous pensions que les scores à ces deux échelles seraient corrélés négativement au paramètre SAP et corrélés positivement au paramètre D. En d'autres termes, plus les individus auraient une motivation importante à réduire l'expression de leur préjugé et moins ils appliqueraient leur stéréotype. À l'inverse, plus ils auraient une motivation importante à réduire l'expression de leur préjugé et plus ils détecteraient la menace de la cible. Enfin, nous voulions observer le lien entre les sous-échelles de chacune de ces échelles et les paramètres du modèle.

3.5.1 **Méthode**

Le but principal de cette étude était de répliquer le lien obtenu entre le score de préjugé individuel et les trois paramètres du modèle : λ, SAP et D. Dans notre précédente étude, nous avons obtenu des tailles d'effets faibles à modérées (i.e., f^2 = 0.060 pour le lien entre le score de préjugé et λ ou D, f^2 = 0.15 pour le lien entre le score de préjugé et SAP). Ainsi, en se basant sur un calcul de puissance (G*power 3.1; Faul et al., 2007) avec une estimation (pessimiste) d'une taille d'effet 5% en dessous de notre plus faible taille d'effet obtenu (i.e., f^2 = 0.057), une puissance de .80 et en prenant en compte la multiplicité des tests (Perugini et al., 2018), 230 participants seraient suffisants pour détecter un effet de cette taille. Trois-cent-dix-huit participants ont pris part à cette étude. Parmi eux, 276 se s'estimaient de culture principale française. Cinq participants ont été écartés, car donnant systématiquement la même réponse à l'ensemble des essais (répondant plus de 98% du temps « Plus menaçant » ou « Moins menaçant »). Ainsi, 271 participants ont été retenus pour les analyses (224 femmes, 44 hommes, 3 autres, $M_{\text{âge}}$ = 21.21, $ET_{\text{âge}}$ = 3.91). Les participants étaient des étudiants de l'Université Grenoble Alpes et recevaient des crédits de cours ou 10 euros en compensation. Le but de cette étude était de répliquer au plus proche les résultats de notre étude précédente tout en simplifiant le plan expérimental là où il pouvait être simplifié. Des analyses complémentaires de l'Étude 6 ont montré que le paramètre λ pouvait être influencé par le caractère prototypique de l'amorce, mais ne semblait pas influencé par le caractère discernable de la cible. Par conséquent, nous avons décidé de ne pas manipuler cette dernière variable. Ainsi, le plan expérimental de cette étude était de type 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (caractère prototypique de l'amorce : Haut vs. Bas) x score de préjugé x score MRWP x score MCPR, avec les trois dernières variables en inter-sujets. Le pré-enregistrement de cette étude est

disponible sur la plateforme OSF
(https://osf.io/2sgzf/?view_only=55e5e08d8f6a4d0bb8347b78241d5438).

Matériel et procédure. Les participants venaient au laboratoire pour une étude de 30 minutes sur la « Formation d'impression suboptimale ». À leur arrivée, un expérimentateur les installait dans des box individuels. Les participants signaient un formulaire de consentement et l'étude débutait. Les consignes présentées à l'écran étaient similaires à celles de l'Étude 1 et de notre étude précédente. Immédiatement après l'énumération de ces consignes, la tâche SMT débutait.

Les passations s'effectuaient sur un écran de 24 pouces, avec le logiciel E-prime 3.0®. Les amorces utilisées étaient identiques aux amorces utilisées dans notre précédente étude. Toutefois, comme nous n'avons pas manipulé le caractère distinctif de la cible, nous avons utilisé les 24 cibles « classiques » que nous avons utilisées lors de nos Études 1 à 5 et issues de l'étude originale de Krieglmeyer et Sherman (2012). La procédure était identique à l'étude précédente. Cependant, après avoir complété l'Échelle de Préjugés Généralisés (α = .89), les participants ont complété une version papier de l'échelle MCPR (α = .75 ; Annexe 5). Cette échelle en 17 items incluait deux dimensions, « soucieux d'agir sans préjugé » (CAP, 11 items, α = .76) et « retenue pour éviter les différents » (RAD, 6 items, α = .38). Pour finir, les participants complétaient l'échelle MRWP (α = .69 ; Annexe 5). Cette échelle en 10 items incluait deux dimensions, la motivation externe à répondre sans préjugés (EMS; 5 items, α = .75) et la motivation interne à répondre sans préjugés (IMS; 5 items, α = .74). Comme dans notre étude précédente, pour insister sur l'anonymat des réponses, les participants devaient déposer leur questionnaire complété dans une urne à la toute fin de l'étude. Ils étaient enfin débriefés, gratifiés et remerciés.

3.5.2 Résultats

Jugement de menace. Nous avons tout d'abord réalisé une ANOVA à mesures répétées 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (caractère prototypique de l'amorce : Haut vs. Bas) avec la somme des réponses « Plus menaçant » en variable dépendante. Nous avons obtenu un effet significatif de la cible. Les participants ont jugé les cibles menaçantes comme effectivement plus menaçantes

(M = 65.41, ET = 20.42) que les cibles non menaçantes (M = 48.37, ET = 18.26), $t(270) = 13.88$, $p < .001$, $\eta^2_p = .42$. Concernant les effets liés aux amorces, nous avons une fois encore répliqué l'effet SMT. Ainsi, les participants ont jugé les dessins de visages comme plus menaçants quand ils étaient précédés de visages maghrébins (M = 46.98, ET = 15.85) plutôt que de visages blancs (M = 38.28, ET = 12.77), $t(270) = 9.33$, $p < .001$, $\eta^2_p = .24$. Cet effet était plus important lorsque les amorces étaient fortement prototypiques que lorsqu'elles étaient faiblement prototypiques, $t(270) = 10.86$, $p < .001$, $\eta^2_p = .30$ (Figure 27). Enfin, les participants ont jugé les dessins de visages comme plus menaçants quand ils étaient précédés de visages maghrébins (M = 46.98, ET = 15.85) plutôt que de visages neutres (M = 28.52, ET = 18.51), $t(270) = 12.94$, $p < .001$, $\eta^2_p = .38$. De même, ils ont jugé les dessins de visages comme plus menaçants quand ils étaient précédés de visages blancs (M = 38.28, ET = 12.77), plutôt que de visages neutres (M = 28.52, ET = 18.51), $t(270) = 8.13$, $p < .001$, $\eta^2_p = .20$.

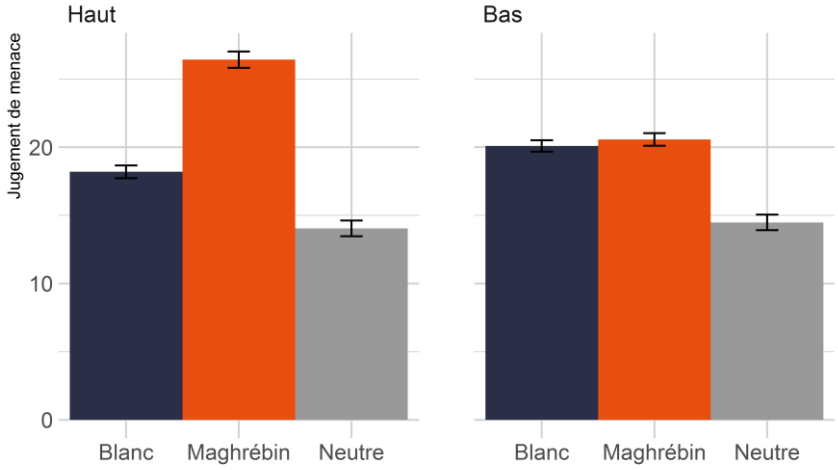


Figure 27. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin, Neutre) et du caractère prototypique de l'amorce (Haut vs. Bas). Les barres d'erreurs représentent les erreurs standards.

Pour analyser le lien entre le jugement de menace et la mesure directe d'attitude raciale (i.e., Échelle de Préjugés Généralisés), nous avons effectué des régressions linéaires entre ces résultats et le score de préjugé selon un modèle de type 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (caractère prototypique de

l'amorce : Haut vs. Bas) x score de préjugé, avec la somme des réponses « Plus menaçant » en variable dépendante. Comme dans notre étude précédente, plus les participants avaient un score de préjugé élevé et plus l'effet SMT était important, b = 3.60, $t(269)$ = 3.74, p < .001, η^2_p = .05. De plus, le lien entre le score de préjugé et l'effet SMT était plus important pour les amorces prototypiques que pour les amorces non prototypiques, b = 1.76, $t(269)$ = 2.36, p = .02, η^2_p = .02. Enfin, nous n'avons pas observé de lien significatif entre le score de préjugé et le contraste opposant les amorces Maghrébin et Neutre, b = 2.31, $t(269)$ = 1.53, p = .13, η^2_p = .01. De même, nous n'avons pas observé de lien significatif entre le score de préjugé et le contraste opposant les amorces Blanc et Neutre, b = -1.30, $t(269)$ = -1.02, p = .31, η^2_p = .004.

Dans un deuxième temps, et de manière exploratoire, nous avons effectué des régressions linéaires afin observer le lien entre le jugement de menace et nos deux autres mesures directes (i.e., MRWP et MCPR)[18], en remplaçant dans nos analyses le score de préjugé par le score à l'une ou l'autre de ces échelles ou sous-échelles. Notons tout d'abord que nous n'avons observé aucune influence de notre échelle MCPR, ou de l'une de ses sous-échelles CAP et RAD (*ps* >.1). Concernant l'échelle MRWP, nous ne pouvons pas conclure à un lien significatif entre le score global à cette échelle et l'effet SMT, b = 1.74, $t(269)$ = 1.93, p = .054, η^2_p = .01. Cependant, lorsque nous prenons en compte les deux sous-échelles (EMS et IMS) nous pouvons voir un lien positif et significatif entre la sous-échelle EMS et l'effet SMT, b = 2.66, $t(267)$ = 4.43, p < .001, η^2_p = .07, et un lien négatif et significatif entre la sous-échelle IMS et l'effet SMT, b = -1.58, $t(267)$ = -2.19, p = .03, η^2_p = .02. Ainsi, plus les participants étaient motivés de manière externe à répondre sans préjugé et plus la différence de jugement entre les amorces Maghrébin et Blanc était importante. À l'inverse, plus les participants étaient motivés de manière interne à répondre sans préjugé et plus la différence de jugement entre les amorces Maghrébin et Blanc était faible. Comme pour l'échelle de préjugé, le lien positif entre le score global MRWP et l'effet SMT était plus élevé pour les amorces prototypiques que pour les amorces non prototypiques, b = 1.42, $t(269)$ = 2.06, p = .04, η^2_p = .02. Là encore, lorsque nous prenons en compte les sous-échelles EMS et IMS,

[18] Suite à une erreur de notre part, pour nos 28 premiers participants, le dernier item de l'échelle MRWP « En raison de la norme du "politiquement correct", j'essaie d'apparaître comme n'ayant pas de préjugés à l'égard des personnes d'origine maghrébine. » a été remplacé par l'item « En raison de mes valeurs personnelles, je pense qu'il n'est pas acceptable d'utiliser les stéréotypes envers les personnes d'origine maghrébine », cet item ayant été donc vu deux fois. En conséquence, pour ces 28 participants, nous avons remplacé les réponses à cet item par des valeurs manquantes. Il est à signaler qu'enlever ces participants des analyses n'impactait pas nos résultats.

nous voyons que cet effet est principalement porté par l'échelle EMS, b = 1.74, $t(267)$ = 3.75, p < .001, η^2_p = .05, mais pas par l'échelle IMS, b = -0.74, $t(267)$ = -1.31, p = .19, η^2_p = .01. Ainsi, le lien entre le score EMS et l'effet SMT était plus important pour les amorces prototypiques que pour les amorces peu prototypiques. Les liens entre le score global MRWP ou ses sous-échelles et le contraste opposant les amorces Maghrébin et Neutre ainsi que celui opposant les amorces Blanc et Neutre étaient non significatifs (*ps* > .1). Enfin, nous avons trouvé un lien positif entre le score de préjugé et la sous-échelle EMS, b = 0.16, $t(267)$ = 5.13, p < .001, η^2_p = .09, et un lien négatif entre le score de préjugé et la sous-échelle IMS, b = -0.43, $t(267)$ = -11.71, p < .001, η^2_p = .34. En d'autres termes, plus les participants avaient une attitude raciale auto-rapportée négative envers les Maghrébins et plus ils étaient motivés de façon externe à répondre sans préjugé, mais moins ils étaient motivés de façon interne à répondre sans préjugé.

Modèle multinomial. Comme pour notre étude précédente, nous avons créé un modèle global ou les paramètres SAC et SAP pouvaient varier en fonction du caractère prototypique de l'amorce (SAC1, SAC2 et SAP1, SAP2). De même, le paramètre λ variait en fonction du caractère prototypique de l'amorce (λ1, λ2). Pour rappel, dans cette étude nous ne manipulions pas le caractère discernable de la cible. En conséquence, nous n'avions qu'un seul paramètre D (et un seul paramètre G). Concernant l'ajustement du modèle aux données, l'écart entre les valeurs T_1 prédites (M_{pred} = 0.04) et observées (M_{obs} = 0.21) était significatif, p = .01, mais la taille d'effet associée était inférieure à un effet de petite taille, w = .03, reflétant ainsi une adéquation acceptable de notre modèle aux données.

Afin d'observer le lien potentiel entre la mesure directe d'attitude raciale et les différents paramètres du modèle, nous avons extrait les valeurs individuelles de chaque paramètre et avons regroupé, pour chaque participant, la valeur des différents paramètres (e.g., λ1, λ2) en un score moyen (λ_{moy}). Nous avons ensuite effectué des régressions linéaires entre le score de préjugé individuel et le score moyen pour chaque paramètre. Tout d'abord, le préjugé individuel était lié au paramètre λ. Ainsi, comme le prévoyait notre hypothèse principale, plus les individus avaient un score de préjugé élevé et plus ils avaient une valeur λ élevée, b = 0.06, $t(269)$ = 3.91, p < .001, η^2_p = .06 (Figure 28). En d'autres termes, plus les participants avaient une attitude raciale auto-rapportée négative envers les Maghrébins et plus ils avaient tendance à suivre le modèle SAC-dominant. De plus, comme le prévoyait encore notre hypothèse, nous avons trouvé un lien positif entre le score de préjugé et le

paramètre SAP, b = 0.02, $t(269)$ = 3.47, p < .001, η^2_p = .05. Ainsi, plus les participants avaient une attitude raciale auto-rapportée négative envers les Maghrébins et plus ils appliquaient leur stéréotype. Cependant, nous n'avons pas répliqué le lien significatif entre le score de préjugé et le paramètre D, b = 0.01, $t(269)$ = 1.61, p = .11, η^2_p = .01. Enfin, nous n'avons pas observé de lien significatif (mais un lien tendanciel) entre le préjugé individuel et le paramètre SAC, b = 0.04, $t(269)$ = 1.89, p = .06, η^2_p = .01, et un lien significatif entre le préjugé individuel et le paramètre G, b = 0.03, $t(269)$ = 2.13, p = .03, η^2_p = .03. Pour ce dernier lien, plus les participants avaient un score de préjugé élevé et plus ils avaient tendance à répondre au hasard « Plus menaçant » quand aucun autre processus n'était à l'œuvre.

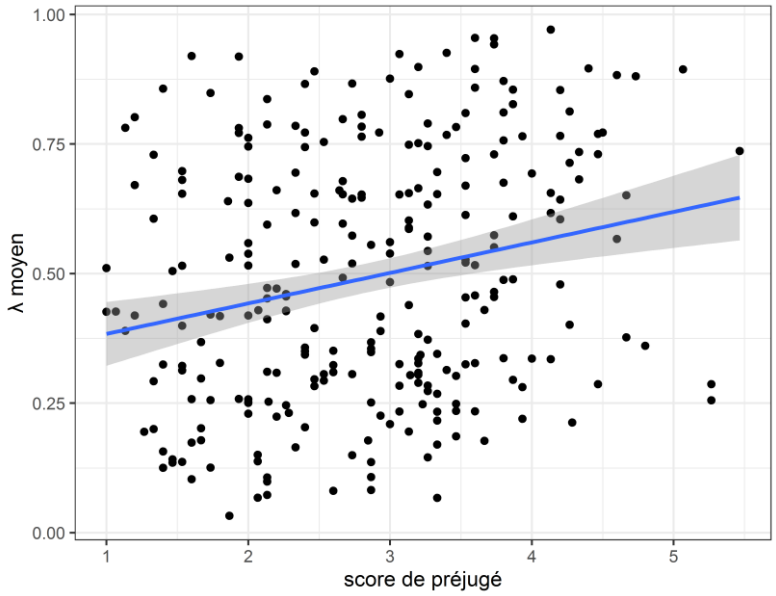


Figure 28. λ moyen en fonction du score de préjugé. Les zones grisées représentent les intervalles de confiance à 95%.

Concernant nos deux échelles exploratoires, nous n'avons tout d'abord pas observé de lien entre l'échelle MCPR (et ses deux sous-échelles) et les différents paramètres du modèle (*ps* >.1). Toutefois, nous avons observé des liens entre l'échelle MRWP (et/ou ses sous-échelles) et les différents paramètres du modèle. Tout d'abord, bien que nous n'ayons pas observé de lien significatif entre le score global MRWP et le paramètre λ, b = 0.02, $t(269)$ = 1.58, p = .12, η^2_p = .01, nous avons observé un lien positif et significatif entre la sous-échelle EMS et le paramètre λ, b = 0.03, $t(267)$ = 3.59, p < .001, η^2_p = .05, et un lien négatif

non significatif (mais tendanciel) entre l'échelle IMS et le paramètre λ, *b* = -0.02, t(267) = 1.86, *p* = .06, η^2_p = .01. En d'autres termes, plus les participants étaient motivés de façon externe à répondre sans préjugé et plus ils avaient tendance à suivre le modèle SAC-dominant. De plus, de façon descriptive (tendancielle), plus les participants étaient motivés de façon interne à répondre sans préjugé et plus ils avaient tendance à suivre le modèle D-dominant. Nous n'avons pas observé de lien significatif (mais tendanciel) entre le score global MRWP et le paramètre SAP, *b* = 0.01, *t*(269) = 1.88, *p* = .062, η^2_p = .01, mais toutefois un lien positif et significatif entre la sous-échelle EMS et le paramètre SAP, *b* = 0.01, *t*(267) = 4.26, *p* < .001, η^2_p = .06 et un lien négatif et significatif entre la sous-échelle IMS et le paramètre SAP, *b* = -0.01, *t*(267) = -2.21, *p* = .03, η^2_p = .02. Ainsi, plus les participants étaient motivés de façon externe à répondre sans préjugé et plus ils appliquaient leur stéréotype. À l'inverse, plus les participants étaient motivés de façon interne à répondre sans préjugé et moins ils appliquaient leur stéréotype. Bien que nous n'ayons pas observé de lien significatif (mais un lien tendanciel) entre le score global MRWP et le paramètre SAC, *b* = 0.03, *t*(269) = 1.95, *p* = .052, η^2_p = .01, nous avons pu observer un lien positif et significatif entre la sous-échelle EMS et le paramètre SAC, *b* = 0.03, t(267) = 2.71, *p* = .007, η^2_p = .03, le lien entre la sous-échelle IMS et le paramètre SAC n'étant pas significatif, *b* = -0.00, *t*(267) = -0.40, *p* = .72, η^2_p = .001. En d'autres termes, plus les participants étaient motivés de façon externe à répondre sans préjugé et plus ils activaient leur stéréotype « Maghrébins = Menace » et/ou « Blancs = Non menace ». Les liens entre ces échelles (ou sous-échelles) et les paramètres D et G étaient non significatifs (*p*s > .1). Il est à noter, enfin, que nous n'avons obtenu aucun effet d'interaction significatif entre l'échelle de préjugé et les sous-échelles IMS et/ou EMS sur les différents paramètres (*p*s > .1). De même, nous n'avons obtenu aucun effet d'interaction significatif entre les deux sous-échelles IMS et EMS sur les différents paramètres (*p*s > .1). Un tableau de corrélation récapitulatif est disponible en annexes (Annexe 3).

3.5.3 **Discussion**

La dernière étude de ce chapitre avait pour but principal de répliquer nos résultats exploratoires observés dans l'Étude 6. Conformément à nos hypothèses, nous avons reproduit le lien positif entre le score de préjugé individuel et le paramètre λ. Ainsi, plus les participants avaient une attitude raciale auto-rapportée négative envers les Maghrébins et plus ils suivaient le modèle SAC-dominant. Autrement dit, plus les participants avaient une attitude

raciale négative et plus ils utilisaient en priorité les informations provenant de l'amorce plutôt que de la cible. De plus, et de manière similaire à notre étude précédente, nous avons également répliqué le lien positif entre le score de préjugé et le paramètre SAP. Plus les participants avaient une attitude raciale auto-rapportée négative envers les Maghrébins et plus ils appliquaient leur stéréotype dans la SMT. Cependant, à l'inverse de notre étude précédente, nous n'avons pas répliqué le lien négatif entre le score de préjugé et le paramètre D. Afin d'apporter des éléments de compréhension supplémentaires à l'ensemble de ces résultats, abordons nos autres mesures directes.

De manière exploratoire, nous avons voulu tester si la motivation à répondre sans préjugés pouvait influencer la tendance des participants à suivre le modèle SAC-dominant ou au contraire le modèle D-dominant. Les résultats les plus probants ont été obtenus avec l'échelle de motivation interne et externe à répondre sans préjugés (MRWP ; Plant & Devine, 1998). Commençons avec la sous-échelle IMS, c'est-à-dire la motivation interne à répondre sans préjugés. D'abord, nous avons pu voir que cette échelle était liée négativement à la fois au score de préjugé auto-rapporté et à l'effet SMT anti-Maghrébins. En d'autres termes, plus les individus étaient motivés de manière interne à ne pas exprimer leur préjugé et moins ils avaient une attitude raciale auto-rapportée négative envers les Maghrébins (i.e., Échelle de Préjugés Généralisés). De même, plus les participants étaient motivés de manière interne à répondre sans préjugé et moins leur effet SMT était important. Nous pouvons mettre en parallèle ce dernier résultat avec le lien négatif entre le score IMS et le paramètre SAP : globalement, plus les participants étaient motivés de manière interne à ne pas exprimer leur préjugé et moins ils appliquaient leur stéréotype au sein de la SMT. Enfin, bien que cet effet n'ait pas été significatif (mais proche du seuil), nous avons pu voir de manière descriptive que plus les participants étaient motivés de manière interne à répondre sans préjugés et plus ils suivaient le modèle D-dominant. En somme, et reprenant en cela nos hypothèses avancées dans la discussion de l'Étude 6, les participants motivés de manière interne à ne pas exprimer leur préjugé n'ont peut-être pas voulu utiliser les informations provenant des amorces (i.e., leur stéréotype réactivé) et donc davantage les informations provenant de la cible. Toutefois, nous devons être prudents quant à ces derniers résultats, le lien mentionné ci-dessus n'étant que tendanciel et, surtout, le lien entre la sous-échelle IMS et le paramètre D étant non significatif.

Concernant la sous-échelle EMS, nous pouvons observer des résultats inverses. Tout d'abord, il est intéressant de voir que cette sous-échelle était reliée positivement au score de

préjugé individuel. En d'autres termes, plus les individus étaient motivés de manière externe à répondre sans préjugé et plus ils avaient une attitude raciale auto-rapportée négative envers les Maghrébins. Plant et Devine (1998), dans leurs étapes de validation des échelles, ont observé de la même manière une corrélation positive entre la sous-échelle EMS et une mesure directe d'attitude raciale (i.e., *Modern Racism Scale*). Comme son nom l'indique, l'EMS est censée refléter la motivation des individus à ne pas exprimer leur préjugé en présence d'autrui. C'est pourquoi, nous pouvons penser que le lien positif entre l'EMS et le score de préjugé puisse s'expliquer par le contexte privé des passations (i.e., box expérimentaux) celui-ci ayant facilité l'expression du préjugé pour les participants ayant un score EMS élevé. De plus, l'expression plus forte de ce préjugé (i.e., de cette attitude raciale négative) se retrouvait également au sein de la SMT. Effectivement, nous avons observé un lien positif entre le score EMS et l'effet SMT. Nous pouvons mettre en parallèle ce dernier résultat à la fois avec le lien positif entre EMS et SAP, donc d'une application plus importante du stéréotype, mais aussi du lien positif entre EMS et SAC. Cette dernière relation nous paraît intéressante, car elle semble signifier que la relation observée entre la sous-échelle EMS et l'échelle de préjugé ne reflète pas seulement une plus grande expression de l'attitude anti-Maghrébins en contexte privé, mais aussi une réactivation de cette attitude elle-même plus importante. Enfin, et là encore de manière intéressante, nous avons obtenu un lien positif entre la sous-échelle EMS et le paramètre λ. En d'autres termes, plus les participants avaient une motivation externe à répondre sans préjugés et plus ils suivaient le modèle SAC-dominant. Si nous résumons ces derniers résultats, nous pouvons penser que les participants ayant une forte motivation externe à répondre sans préjugés ont davantage utilisé les informations provenant des amorces, peut-être parce que ces individus se souciaient moins de paraitre comme ayant des préjugés (dans ce contexte privé) et ont donc plus appliqué leur stéréotype (SAP), mais aussi, car ces individus avaient une réactivation du stéréotype également plus importante (SAC).

Pour finir, intéressons-nous rapidement à l'autre échelle de motivation à réduire l'expression de ses préjugés, à savoir l'échelle de motivation à contrôler ses réactions teintées de préjugé (MCPR ; Dunton & Fazio, 1997). Comme nous l'avons mentionné plus haut, nous n'avons pas observé de lien entre cette échelle (ou ses sous-échelles) et les différents paramètres du modèle. Nous pouvons envisager des raisons à cela. Tout d'abord, la première dimension de cette échelle (CAP) est décrite comme comportant à la fois des items mesurant la motivation interne à contrôler ses préjugés, mais aussi des items mesurant la motivation externe à contrôler ses préjugés (Dunton & Fazio, 1997 ; Plant & Devine, 1998). Or, comme nous l'avons vu ci-dessus, ces deux types d'items sont liés de manière opposée aux différents

paramètres du modèle, raison pour laquelle il est envisageable que cette sous-échelle n'ait pas donné de résultats probants. De plus, la deuxième sous-échelle de la MCPR, la RAD, avait une consistance interne particulièrement faible (α = .38), masquant de ce fait d'éventuels liens avec les paramètres du modèle. En somme, nous pensons donc que les mauvaises propriétés psychométriques de cette échelle ne nous ont pas permis d'observer de lien entre cette échelle (ou ses sous-échelles) et les différents paramètres du modèle.

3.6 Discussion du Chapitre 3 : aller plus loin dans nos précautions

À l'origine de ce chapitre, nous sommes partis du questionnement suivant : comment expliquer que le modèle multinomial décrivant au mieux les données de nos premières études (i.e., modèle D-dominant) puisse être différent du modèle multinomial habituellement associé au paradigme SMT (i.e., modèle SAC-dominant) ? Autrement dit, comment expliquer que nos participants utilisent en priorité l'information provenant de la cible, là où les participants passant la SMT classique utilisaient en priorité l'information provenant de l'amorce ? Les trois études constituant ce chapitre avaient pour objectif de déterminer les facteurs pouvant être liés à l'adéquation supérieure d'un modèle multinomial (e.g., modèle D-dominant) par rapport à un autre (e.g., modèle SAC-dominant).

La première étude de ce chapitre (i.e., Étude 5) nous a permis d'observer l'impact du type d'amorce sur la relation conditionnelle entre les processus sous-tendant l'effet SMT. De manière conforme à nos deux premières études, les participants passant notre paradigme SMT (i.e., SMT-Maghrébin/Blanc) ont utilisé en priorité l'information provenant de la cible (i.e., modèle D-dominant). De plus, de manière conforme aux résultats classiques de la littérature, les participants passant le paradigme SMT traditionnel (i.e., SMT-Noir/Blanc) ont utilisé en priorité l'information provenant de l'amorce (i.e., modèle SAC-dominant). Ainsi, comme nous en faisions l'hypothèse, changer le type d'amorce impactait effectivement la relation conditionnelle entre processus. Cependant, l'Étude 6 ne nous a pas permis de confirmer la possibilité que des facteurs liés à la tâche (i.e., caractère prototypique de l'amorce, caractère discernable de la cible) puissent être à l'origine de ces effets.

À l'inverse et de manière tout aussi intéressante, l'Étude 6 nous a permis de mettre en lumière la possibilité que des facteurs liés aux participants (i.e., attitude raciale auto-rapportée) puissent être liés à la relation conditionnelle entre les paramètres du modèle. Il

semblerait ainsi que plus les participants ont une attitude raciale auto-rapportée négative envers les Maghrébins et plus ils suivent le modèle SAC-dominant. En d'autres termes, ces participants utiliseraient en priorité l'information provenant de l'amorce plutôt que l'information provenant de la cible. L'Étude 7 n'a pas simplement confirmé ces résultats, mais elle nous a aussi permis d'aller plus loin dans l'interprétation de ceux-ci. En effet, nous avons pu observer que des participants ayant une plus forte motivation interne à ne pas utiliser leur préjugé allaient non seulement moins appliquer leur stéréotype dans la SMT, mais aussi (de manière descriptive) utiliser davantage l'information provenant de la cible plutôt que de l'amorce. À l'inverse, nous avons pu observer que des participants ayant une plus forte motivation externe à ne pas utiliser leur préjugé allaient non seulement avoir une plus forte activation du stéréotype et appliquer ce stéréotype plus souvent dans la SMT, mais aussi utiliser davantage l'information provenant de l'amorce plutôt que de la cible.

Nous allons revenir plus longuement dans la partie Discussion générale sur l'apport que peuvent avoir nos études dans la littérature traitant des attitudes raciales. Toutefois, pour évoquer succinctement la contribution de ce chapitre, il nous faut revenir sur les éléments qui nous ont amenés jusqu'ici. Au Chapitre 1, nous nous sommes longuement arrêtés sur les évolutions autant théoriques que méthodologiques des mesures des attitudes raciales. Un des points si ce n'est le point le plus important dans l'évolution de ces mesures a été le basculement de mesures centrées sur le contenu vers des mesures centrées sur les processus (Dovidio et al., 2010 ; Schneider, 2004). Dans le champ des attitudes raciales, des constats ont été faits : les réponses issues de mesures directes peuvent être le reflet d'une inhibition du contenu mental réactivé (e.g., Crosby et al., 1980 ; Sigall & Page, 1971). En ce sens, des précautions ont été prises, aboutissant à l'utilisation de paradigmes différents, censés contourner ce type de biais (e.g., Fazio et al., 1995 ; Greenwald et al., 1998). Les mesures indirectes se sont inscrites dans cette mouvance et ont permis de capter des attitudes raciales reflétant plus directement des processus de Type 1. Mais là encore, des constats ont été faits : les réponses des participants au sein de ces mesures sont certes le reflet de processus de Type 1, mais aussi au moins partiellement de processus de Type 2 (e.g., Fiedler & Bluemke, 2005). D'autres précautions ont été prises et l'utilisation de modèles multinomiaux s'est avérée être une solution crédible (e.g., Hütter & Klauer, 2016). Nous l'avons vu au Chapitre 2, ces modèles nous poussent à être plus prudents quant au niveau de spécification théorique que l'on doit avoir et notamment par rapport à l'impact de nos manipulations sur les processus à l'œuvre. Bien que nous n'ayons pas confirmé cela empiriquement, nous avons évoqué la possibilité que des résultats à première vue similaires (e.g., pas de différence significative au

niveau de l'effet SMT entre les conditions Attentat et Contrôle) puissent être en réalité le fruit de variations dans la contribution spécifique des processus cognitifs sous-jacents. C'est la raison pour laquelle nous avons conclu le Chapitre 2 en soulignant les précautions que nous devions prendre dans l'interprétation des résultats relatifs à nos manipulations.

Pour nous, le Chapitre 3 s'inscrit dans la continuité de ces précautions. Des auteurs ont par le passé montré qu'un type de modèle multinomial (e.g., Quad Model) s'adaptait mieux à un type de tâche (e.g., IAT ; Bishara & Payne, 2009 ; Conrey et al., 2005 ; Nadarevic & Erdfelder, 2011 ; Sherman et al., 2008). Autrement dit, des auteurs ont montré qu'une certaine architecture de modèle et donc une contribution spécifique de processus pouvait expliquer de manière adéquate les réponses des participants au sein de ces tâches. Ces mêmes auteurs sont allés plus loin et ont questionné d'une part, si changer des aspects de la tâche pouvait changer la contribution de certains de ces processus (e.g., paramètre OB au sein de la GNAT ; Nadarevic & Erdfelder, 2011) et d'autre part, si d'autres modèles multinomiaux et donc d'autres relations entre processus pouvaient expliquer les réponses des participants de manière plus adéquate (Bishara & Payne, 2009 ; Krieglmeyer & Sherman, 2012 ; Rees, Rivers et al., 2019). En somme, ces auteurs se sont questionnés quant au meilleur modèle multinomial pouvant rendre compte de ces données et si ce modèle pouvait changer, mais presque jamais sur les facteurs pouvant expliquer qu'un modèle soit plus pertinent qu'un autre. À notre connaissance, les études constituant ce chapitre sont les premières à mettre en avant un lien entre des facteurs liés aux participants et la relation conditionnelle entre processus et donc la meilleure adéquation aux données d'un modèle plutôt qu'un autre.

Comme nous l'avons spécifié en introduction de ce chapitre, pour des participantes comme Eva et Clara, passer une tâche SMT n'aura pas le même impact. Clara, à la différence d'Eva, dispose d'une attitude raciale négative envers les Maghrébins. Certes, pour ces deux participantes les consignes sont identiques : elles doivent juger des cibles sur la dimension de menace, tout en prêtant attention aux amorces précédant ces cibles. Jusqu'à présent, si nous avions voulu connaitre la contribution des processus cognitifs à l'origine des réponses d'Eva et de Clara, nous aurions appliqué le modèle multinomial classique à leurs données, c'est-à-dire le modèle SAC-dominant, et interprété les estimations de ce modèle. Toutefois, ce que nous n'avions pas envisagé, c'est que les informations provenant des différents stimuli (e.g., amorce d'un visage maghrébin vs. cible non menaçante) n'ont pas le même poids pour Eva et Clara. Les processus à l'origine de leur réponse peuvent effectivement être les mêmes (i.e., activation du stéréotype, application du stéréotype, détection de la menace de la cible), mais

la relation conditionnelle entre ces processus va être différente. Nous verrons dans la partie Discussion générale que cela peut nous amener à une estimation imprécise de la contribution de ces processus. Mais plus généralement, cet exemple met en évidence le recul théorique que nous devons avoir par rapport à notre perception des mesures indirectes et de la performance de nos participants prenant part à ces mesures. Nos participants ne sont pas égaux face à la tâche, ils n'utilisent pas tous les mêmes informations pour formuler leur réponse et il faut en tenir compte.

Chapitre 4 – Que trouve-t-on derrière l'effet SMT ?

La *Stereotype Misperception Task* et son modèle multinomial nous ont permis, tout au long de ce manuscrit, de mettre en évidence plusieurs utilisations possibles des modèles multinomiaux dans le champ des attitudes raciales. Nous nous sommes centrés sur les processus à l'origine de la réponse individuelle et sur les facteurs pouvant être liés à la contribution de ces processus sans toutefois nous attarder sur la réponse des participants en elle-même. Pourtant, bien que cela n'ait, au final, jamais été notre objectif principal, nous avons systématiquement répliqué l'effet SMT. En définitive, nous ne nous sommes que très peu intéressés aux composantes théoriques à l'origine de cet effet. Est-ce bien la réactivation d'un stéréotype de menace lié aux populations maghrébines qui peut être à l'origine de cet effet ? Ce stéréotype est-il spécifique à notre environnement culturel ou peut-il être retrouvé dans des zones géographiques différentes ? L'effet SMT est-il le fruit d'une simple connaissance du stéréotype ou reflète-t-il une composante plus personnelle de l'attitude raciale ? Ce dernier chapitre empirique aura pour objectif de répondre à ces différentes questions et plus directement à celle relative au caractère généralisable de l'effet SMT. Plus globalement, nous montrerons une fois encore la nécessité pour le chercheur d'être prudent quant aux interprétations des différents résultats issus de mesures indirectes. Nous mettrons ainsi en avant les avantages de la SMT et de son modèle multinomial dans l'investigation en profondeur de ces différentes interprétations.

4.1 L'effet SMT : un effet robuste ?

Depuis bientôt une décennie, la psychologie sociale est touchée par la crise de la réplication (e.g., Nelson et al., 2018 ; Simmons et al., 2011). Dans un désir de mettre en lumière la reproductibilité et donc la robustesse des paradigmes classiques de psychologie sociale, des réplications à grande échelle ont été entreprises (e.g., Ebersole et al., 2016 ; Hagger et al., 2016 ; Klein et al., 2014 ; Open Science Collaboration, 2015). Ainsi, les résultats d'études accréditant des théories solidement ancrées (i.e., *ego-depletion theory* ; Baumeister et al., 1998 ; *terror management theory* ; Greenberg et al., 1997) n'ont pas été répliqués,

remettant en question le fondement empirique de ces théories (Hagger et al., 2016 ; Klein et al., 2019).

Concernant la littérature traitant des mesures indirectes, le constat semble meilleur. Certains paradigmes, comme l'IAT, ont démontré leur robustesse, les tâches IAT classiques (e.g., *race-IAT*) ayant été répliquées de nombreuses fois (voir Hofmann et al., 2005 ; Ratliff et al., 2020). Toutefois, les études utilisant des paradigmes d'amorçages séquentiels ont montré des résultats hétérogènes. En guise d'illustration, prenons la méta-analyse de Kidder et collaborateurs (2018) regroupant un ensemble d'études utilisant un paradigme d'amorçage séquentiel relié au stéréotype. Ces auteurs ont mis en évidence une forte hétérogénéité quant aux tailles d'effets rapportées par ces études, la présence de nombreux modérateurs (e.g., type de stimuli, type de relation amorce-cible et amorce-réponse) ainsi qu'un potentiel biais de publication pour certains types de tâches (i.e., amorçages séquentiel de type amorce-réponse).

Pour Cesario (2014), l'état de confusion ou d'incertitude quant à la reproductibilité ou la précision des effets d'amorçage est en partie la résultante de mauvaises pratiques de recherches : les chercheurs ne répliquent que trop peu leurs propres recherches. Teige-Mocigemba et collaborateurs (2017), pourtant, ont à plusieurs reprises tenté de répliquer les effets classiques obtenus dans un paradigme AMP. De manière surprenante, dans six de leurs sept études pilotes, ces auteurs ont échoué à répliquer l'effet AMP classique, et ce, en utilisant notamment des stimuli arabo-musulmans. Dans une dernière étude à forte puissance statistique et en utilisant des stimuli allemands et turcs, ces auteurs ont répliqué l'effet AMP classique, mais avec une taille d'effet particulièrement faible (i.e., d = 0.17). Nous l'avons vu, l'AMP, bien que s'intéressant à l'aspect évaluatif de l'attitude (et non pas à sa composante sémantique), est structurellement proche de la SMT. Ainsi, cette période post-crise de la réplication nous pousse à nous poser la question suivante : quelle est la reproductibilité de nos effets ?

Au moment de la rédaction de ce manuscrit, la grande majorité des études utilisant le paradigme SMT avait pour objectif la comparaison de la menace associée à des visages de couleur noire et de couleur blanche (Krieglmeyer & Sherman, 2012 ; Rees, Ma, et al., 2019 ; Rees, Rivers, et al., 2019 ; Rivers et al., 2020). Dans chacune de ces études, les auteurs ont répliqué l'effet SMT classique : les participants ont répondu plus souvent « Plus menaçant » dans les essais où la cible était précédée d'un visage noir que dans les essais où la cible était précédée d'un visage blanc. Dans la majorité de nos études, nous avons utilisé la tâche SMT

non pas avec des visages de couleur noire et blanche, mais avec des visages maghrébins et blancs. Suivant les résultats de Krieglmeyer et Sherman, nous nous attendions à répliquer l'effet SMT traditionnel. Effectivement, dans l'ensemble de nos études, les participants ont répondu plus souvent « Plus menaçant » dans les essais où la cible était précédée d'un visage maghrébin plutôt que d'un visage blanc.

Afin d'illustrer la robustesse de cet effet, nous avons effectué, à l'aide du logiciel R (version 3.6.0, 2019 ; R Core Team, 2013) et du package metafor (version 2.4.0 ; Viechtbauer, 2010), une méta-analyse des six premières études de ce manuscrit (nous n'avons pas inclus l'Étude 3, celle-ci utilisant le paradigme IAT). Pour chaque étude, nous avons gardé uniquement les données de participants réalisant notre tâche SMT habituelle, c'est-à-dire la tâche comportant des amorces de visages maghrébins et blancs prototypiques de leur catégorie. Les résultats de cette méta-analyse sont présentés Figure 29. Malgré une forte hétérogénéité de nos données, I^2 = 68.82%, H^2 = 3.21, $Q(5)$ = 16.45, p = .01, pouvant être expliquée en partie par nos différentes manipulations expérimentales (e.g., arrière-plan du bataclan, cibles moins prototypiques), nous avons obtenu un effet SMT global de taille moyenne, dz = 0.48 [0.34, 0.62]. Cette taille d'effet est consistante avec la taille d'effet que l'on peut retrouver dans des tâches d'amorçage sémantique présentant une compatibilité amorce-réponse, comme l'est la tâche SMT (i.e., d = 0.48 ; Kidder et al., 2018).

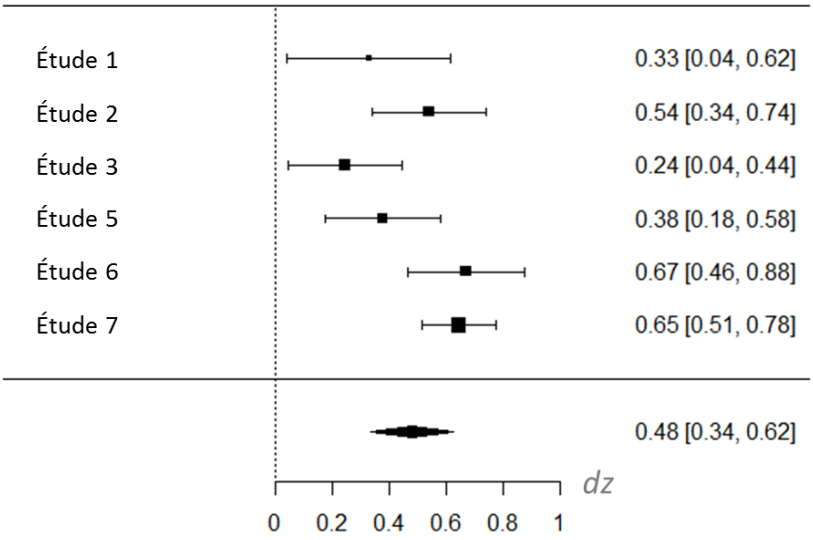


Figure 29. Représentation graphique de la méta-analyse interne à nos six études. Les carrés noirs correspondent aux tailles d'effets (*dz*) associées à chaque étude. Les barres d'erreurs sont les intervalles de confiance à 95%. Le losange représente le résultat de cette méta-analyse.

L'effet SMT semble donc, tout du moins dans notre laboratoire, un effet robuste. Toutefois, nous pouvons pousser plus loin notre réflexion par rapport à cet effet. L'explication théorique de l'origine de cet effet peut-elle être questionnée par une explication alternative ? Cet effet est-il généralisable à d'autres laboratoires ? D'autres échantillons ? Ce chapitre aura pour objectif de répondre à l'ensemble de ces questions. Commençons tout d'abord par l'explication théorique de l'effet SMT.

4.2 Effet SMT : exemplaire versus catégorie

Sur le plan théorique, Krieglmeyer et Sherman (2012) expliquent l'effet SMT de la manière suivante : la présentation d'une amorce (e.g., un visage noir) va réactiver la catégorie sociale reliée à cette amorce (e.g., les Noirs) de même que le stéréotype associé à cette catégorie (e.g., menaçant). En conséquence, Krieglmeyer et Sherman s'inscrivent pleinement dans la conception classique du stéréotype en cognition sociale, conception que nous avons développée au Chapitre 1. Pour rappel, notre système cognitif a une capacité limitée, poussant les individus à simplifier (et structurer) leur perception du monde qui les entoure (e.g., Fiske & Taylor, 2013 ; Moskowitz, 2005 ; Ric & Muller, 2017). C'est pourquoi nous aurions tendance à regrouper les connaissances que nous avons sur un concept (e.g., les Maghrébins) sous forme de représentations abstraites et par un phénomène de catégorisation mentale (Dovidio et al., 1996 ; Macrae & Bodenhausen, 2000, 2001). Les représentations abstraites que nous aurions par rapport à une catégorie d'individus (e.g., les Maghrébins) contiendraient les attributs propres à la catégorie, ainsi que la relation entre ces attributs (Fiske & Taylor, 2013). Ainsi, regrouper ces connaissances sous forme de catégories abstraites nous permettrait de faire face à la multitude d'informations qui nous arrive au quotidien afin de faciliter le traitement de l'information. Comme nous l'avons vu en évoquant les modèles duaux de la formation d'impression (e.g., Brewer, 1988 ; Fiske & Neuberg, 1990), l'impression que l'on se forme sur autrui proviendrait en priorité et de manière non intentionnelle d'informations saillantes propres à la catégorie sociale (e.g., âge, ethnie, genre). En d'autres termes, et pour revenir à l'exemple de la SMT, la perception d'une amorce de visage noir entrainerait, de manière non intentionnelle, la réactivation de la catégorie associée (i.e., les Noirs) et donc du contenu mental associé à cette catégorie, c'est-à-dire ici le stéréotype de menace.

Comme décrit précédemment, nous avons répliqué l'effet SMT avec des visages maghrébins et des visages blancs. Nous pourrions penser, de la même façon, que l'apparition d'une amorce de visage maghrébin ait réactivé en mémoire la catégorie des Maghrébins et la menace associée à cette catégorie. Cela constitue effectivement notre hypothèse principale. Toutefois, nous ne pouvons pour lors pas éliminer une hypothèse alternative : est-ce que nos visages maghrébins ne sont pas intrinsèquement plus menaçants que nos visages blancs ? Plus précisément, indépendamment de la réactivation de leur catégorie associée, nos visages maghrébins n'ont-ils pas des caractéristiques physiques rendant ces visages menaçants ? L'effet SMT ne viendrait alors pas d'une réactivation de la catégorie associée à ces visages, mais juste d'une perception globale de ces caractéristiques visuelles.

Pour développer notre propos, nous pouvons faire le parallèle avec les dessins cibles que nous avons utilisé dans nos SMT. Comme Krieglmeyer et Sherman (2012) l'ont rapporté, ces cibles ont justement été créées pour avoir des caractéristiques visuelles évoquant un visage menaçant. Or, l'une des limites évoquées par ces auteurs concerne la dimension de menace associée aux cibles : « On pourrait avancer qu'un problème potentiel des stimuli cibles utilisés dans le paradigme SMT est qu'ils peuvent différer en termes d'appartenance apparente à un groupe (i.e., l'ethnie). Par exemple, les cibles menaçantes peuvent avoir une apparence plus prototypique des Noirs que les cibles non menaçantes » (p. 16). Bien que Krieglmeyer et Sherman réfutent cette possibilité, les cibles ayant été créées pour varier sur la dimension de menace, mais non d'ethnie (i.e., toutes les cibles étaient créées à partir de visages blancs ; Oosterhof & Todorov, 2008), il reste possible qu'un visage créé pour être menaçant (i.e., forme des yeux, sourcils, nez) se rapproche plus d'un visage noir (et dans notre cas d'un visage maghrébin) que d'un visage blanc. En d'autres termes, le caractère prototypique d'un visage noir (et dans notre cas d'un visage maghrébin) peut varier en même temps que la perception intrinsèque de menace de ce visage.

Un autre exemple, cette fois directement relié aux amorces, a été rapporté par Teige-Mocigemba et collaborateurs (2017) : « Un expert anonyme a souligné que nos efforts pour rendre équivalente l'attractivité entre nos stimuli [visages allemands et turcs] pouvaient avoir abouti à la sélection de Turcs plus attractifs que les Allemands sélectionnés. En effet, les préjugés des participants peuvent avoir biaisé les évaluations de l'attractivité, ce qui conduit à des évaluations généralement plus mauvaises des Turcs sur des aspects positifs tels que l'attractivité. » (p. 11, note 5). Pour sélectionner nos visages, nous avons choisi les visages les plus prototypiques des catégories Maghrébins et Blancs (Courset et al., 2018). À l'inverse de

Teige-Mocigemba et collaborateurs (2017), nous avons uniquement contrôlé la neutralité émotionnelle des visages sélectionnés. Or, comme nous venons de le voir, il est possible que le caractère prototypique d'un visage covarie avec le caractère menaçant du visage. Par conséquent, il reste possible que, dans la phase de validation de cette base de données, les visages maghrébins aient été décrits comme prototypiques Maghrébins en partie parce que ceux-ci étaient intrinsèquement plus menaçants.

Pour résumer, comment savoir si l'effet SMT est bien la résultante de la réactivation de la catégorie associée aux amorces, comme nous en faisons l'hypothèse, et pas uniquement d'une perception des visages maghrébins comme intrinsèquement plus menaçants que des visages blancs ? Pour éliminer cette possibilité, il nous faut un nouveau type d'amorces, plus abstraites (Foroni & Bel-Bahar, 2009) et donc intrinsèquement neutres sur la dimension de menace. Nous verrons que des « prénoms » peuvent remplir cette tâche. Toutefois, évoquons avant cela une deuxième manière de départager ces hypothèses.

Globalement, nous pouvons associer ce questionnement à une problématique plus large de la recherche en psychologie, à savoir le principe de correspondance entre mesures (Ajzen & Fishbein, 1977) et la différence entre catégorie sociale et exemplaire de cette catégorie (Gawronski, 2019 ; Ledgerwood et al., 2018). Nous avons déjà abordé le principe de correspondance au Chapitre 1 comme possible explication de la faible corrélation entre mesures indirectes et directes. Pour rappel, selon ce principe, la corrélation entre mesures indirectes et directes sera plus élevée quand les deux types de mesures correspondent en termes de structure et de contenu. Comme nous l'avons vu, les mesures indirectes et directes n'ont, la plupart du temps, pas la même structure (Hütter & Klauer, 2016 ; Payne et al., 2008). Par exemple, les mesures directes demandent au participant d'interpréter les différents items d'un questionnaire et l'échelle de réponse associée, mais ne demandent pas de répondre dans un temps imparti. À l'inverse, les mesures indirectes peuvent demander au participant de répondre le plus rapidement possible, ainsi que de détecter la réponse correcte (e.g., la cible présentée était-elle une arme ou un outil ?). Mais qu'en est-il de la correspondance quant au contenu ?

Dans le champ des attitudes raciales, les mesures indirectes et directes ne se réfèrent habituellement pas au même niveau de construit pour capter l'attitude. Les mesures directes s'intéressent la plupart du temps à la catégorie dans sa globalité (e.g., les Maghrébins ; Dambrun & Guimond, 2001). Par exemple, dans notre mesure d'attitude raciale anti-Maghrébins, nos participants devaient répondre à des items portant directement sur la

catégorie des Maghrébins ou des catégories associées (e.g., « On peut facilement comprendre la colère que ressentent les Maghrébins en France »). À l'inverse, les mesures indirectes utilisent la plupart du temps des exemplaires de la catégorie étudiée. Par exemple, dans notre paradigme SMT, les individus n'étaient pas exposés aux catégories « Maghrébins » et « Blancs », mais à des visages d'individus membres de ces catégories.

Afin d'illustrer l'importance d'une correspondance entre mesure indirecte et directe liée au contenu, Gawronski (2019) a réanalysé les données d'une de ses précédentes études (Gawronski et al., 2008, Étude 3). Dans cette étude, les participants devaient compléter une AMP comportant des visages noirs et blancs, une échelle *feeling thermometer* relative aux catégories Noirs et Blancs et une échelle d'appréciation de chaque visage noir et blanc perçu lors de l'AMP. Gawronski (2019) a montré que les scores des participants à l'AMP étaient corrélés avec l'échelle d'appréciation des visages noirs et blancs, mais pas avec l'échelle *feeling thermometer* relative aux catégories Noirs et Blancs. En d'autres termes, l'AMP, centrée sur les exemplaires de la catégorie (visages noirs et blancs) était corrélée avec une mesure directe elle-même centrée sur les exemplaires de la catégorie (échelle d'appréciation), mais pas avec une mesure centrée sur la catégorie en elle-même (échelle *feeling thermometer*).

Nous voyons que ces résultats ne suivent pas ce que l'on attendrait de notre mesure SMT, l'AMP n'étant ici pas liée à une mesure directe centrée sur la catégorie. Toutefois, nous pouvons réutiliser ce type d'approche pour l'appliquer à notre propre problématique. Nous avons vu dans certaines de nos études (i.e., Études 3, 6 et 7) que l'effet SMT était corrélé avec une mesure directe d'attitudes raciales (i.e., échelle de Préjugés Généralisés ; Dambrun & Guimond, 2001). Suivant le raisonnement que nous avons eu tout au long de ce manuscrit, cela n'est pas surprenant : les mesures indirectes comme les mesures directes ne sont pas *process-pure* et partagent donc des processus en commun (e.g., processus de Type 2). Par conséquent, si, comme nous le pensons, l'effet SMT est le fruit d'une réactivation de la catégorie associée aux amorces et non pas uniquement d'exemplaires (i.e., amorces de visages) intrinsèquement menaçants, l'effet SMT devrait être corrélé à un jugement direct de la catégorie au-delà de la corrélation avec un jugement direct des différents exemplaires.

En résumé, nous voulons écarter l'hypothèse alternative que l'effet SMT ne soit pas la résultante (au moins en partie) d'une réactivation des catégories Maghrébins et Blancs (et du contenu mental associé), mais d'exemplaires de ces catégories intrinsèquement (i.e., traits du visage) menaçants ou non menaçants. Pour tester cette hypothèse, nous avons deux solutions. Nous pouvons utiliser un autre type d'amorce ayant des caractéristiques visuelles

ne pouvant pas évoquer un aspect menaçant (ou non menaçant). De plus, suivant en cela la littérature sur le principe de correspondance, nous pouvons aussi voir si, comme nous l'envisageons, l'effet SMT peut être corrélé avec une mesure directe centrée sur la catégorie au-delà de la corrélation avec une mesure directe centrée sur les exemplaires. Ces deux solutions ont été testées au sein de notre Étude 8.

4.3 Étude 8 : impact des exemplaires vs. impact de la catégorie

Dans cette étude, nous voulions répliquer l'effet SMT avec un type de stimuli ne pouvant pas être décrit comme intrinsèquement menaçant (ou non menaçant). La littérature, notamment celle traitant des attitudes raciales, utilise parfois des prénoms en guise de stimuli (e.g., IAT ; Nosek et al., 2002). La logique sous-jacente est la même : un prénom (e.g., Mohammed) va réactiver la catégorie associée (e.g., les Maghrébins) et le contenu mental lié à cette catégorie (e.g., menaçant). Les prénoms sont souvent décrits comme des stimuli de plus haut niveau, plus abstraits (e.g., Foroni & Bel-Bahar, 2009). C'est pourquoi, à l'inverse des visages utilisés jusqu'à présent, nous pouvons penser que des prénoms sont des stimuli plus « neutres », dans le sens où leur composition (ensemble de lettres) ne peut pas visuellement évoquer un sentiment de menace. C'est la raison pour laquelle, en plus de notre SMT classique composée de visages maghrébins et blancs (i.e., SMT-visages), les participants ont passé une SMT avec pour amorces des prénoms maghrébins et des prénoms français (SMT-prénoms). De plus, indépendamment du type d'amorce (i.e., visages vs. prénoms) nous pensions que l'effet SMT serait corrélé avec une mesure directe de stéréotype centrée sur la catégorie (i.e., Maghrébins vs. Français) indépendamment, c'est-à-dire en contrôlant la menace perçue par chaque exemplaire.[19]

[19] Il est à noter que notre pré-enregistrement comportait également une hypothèse liée au modèle multinomial : nous faisions l'hypothèse que le paramètre SAP devait être corrélé au jugement de la catégorie indépendamment du jugement des exemplaires. Toutefois, nous n'avions pas anticipé le fait que cette hypothèse ne puisse pas être testée, du fait des analyses en modèles mixtes décrites ci-dessous. En effet, contrairement à la variable dépendante « Jugement de menace » estimée pour chaque essai, nous n'avons qu'une estimation globale du paramètre SAP par participant et pour l'ensemble des essais. Nous avons donc, de manière exploratoire, analysé le lien entre le paramètre SAP et le jugement de la catégorie, sans prendre en compte le jugement des exemplaires.

4.3.1 Méthode

Participants et plan expérimental. Dans cette étude, deux types d'effets étaient attendus. Tout d'abord, l'effet SMT à la fois pour les visages, mais aussi pour les prénoms. Si l'on se réfère aux cinq études réalisées jusqu'alors[20], la taille d'effet attendue était moyenne, *dz* = .44. Ainsi, sur la base d'un calcul de puissance avec le logiciel G*Power (3.1 ; Faul et al., 2007), 43 participants auraient été nécessaires pour détecter un effet de cette taille avec une puissance de .80 et un seuil à .05. De plus, nous attendions une corrélation entre l'effet SMT global et le jugement direct de la catégorie. En nous basant sur nos deux études (i.e., Étude 3 et Étude 6) liant l'effet SMT à une mesure directe d'attitude raciale, la taille d'effet attendue était là encore moyenne (f^2 = 0.10). Sur la base d'un calcul de puissance, 80 participants auraient été suffisants pour détecter un effet de cette taille. Néanmoins, certaines différences entre les études précédentes et cette étude nous ont poussés à être précautionneux à propos des tailles d'effets réellement attendues. Premièrement, étant donné que nous n'avions pas de fortes attentes quant à l'ampleur de l'effet SMT avec des prénoms en guise de stimuli, il était possible que cet effet soit plus faible. Deuxièmement, dans cette étude nous voulions tester la relation entre l'effet SMT et le jugement de la catégorie au-delà du jugement des exemplaires, ce qui pouvait entrainer une diminution de la taille de l'effet attendu par rapport à nos études précédentes. Pour finir, nous avons utilisé des analyses en modèles mixtes et non des ANOVA classiques, ceci pouvant entrainer une perte de puissance statistique (Judd et al., 2012). Sur la base d'une analyse de sensibilité, 150 participants étaient suffisants pour détecter une taille d'effet petite à moyenne (f^2 = 0.05) avec une puissance de .80 et un seuil à .05. Nous avons donc supposé que 150 participants (de culture principale française) pouvaient suffire pour détecter les effets attendus, compte tenu des spécificités de la présente étude.

Cent-soixante-dix-sept participants n'ayant pas pris part à l'une de nos cinq études précédentes ont pris part à cette étude. Parmi eux, 144 s'estimaient de culture principale française et ont été gardés pour les analyses (126 femmes, 18 hommes, $M_{\text{âge}}$ = 20.90, $ET_{\text{âge}}$ = 2.96). Les participants étaient des étudiants de l'Université Grenoble Alpes et recevaient des crédits de cours ou 10 euros en compensation. Le plan expérimental de cette étude était de type 3 (amorce : Maghrébin vs. Blanc/Français vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (type d'amorce : Visage vs. Prénom) x jugement de la catégorie x

[20] Au moment de la réalisation de cette étude, l'Étude 7 n'avait pas été effectuée.

jugement de l'exemplaire, les deux dernières variables étant inter-sujets. Le pré-enregistrement de cette étude est disponible sur la plateforme OSF (https://osf.io/s3nh2).

Matériel et procédure. Les participants venaient au laboratoire pour une étude de 15 minutes sur la « Formation d'impression suboptimale ». À leur arrivée, un expérimentateur les installait dans des box individuels. Les participants signaient un formulaire de consentement et l'étude débutait. Les passations s'effectuaient sur un écran de 24 pouces, avec le logiciel E-prime 3.0®. Les participants passaient tout d'abord les deux types de SMT, celles-ci étant contrebalancées d'un participant à l'autre. La SMT-visages et les consignes associées étaient identiques à celles de notre Étude 1. En d'autres termes, nous avons réutilisé les 24 amorces de visages maghrébins et les 24 amorces de visages blancs prototypiques des catégories associées (Courset et al., 2018) ainsi que les 24 visages neutres (Krieglmeyer & Sherman, 2012). Pour la SMT-prénoms, les consignes étaient identiques à la SMT-visages, si ce n'est la mention des prénoms à la place des visages. Plus précisément, les participants étaient informés qu'ils allaient voir des dessins de visages apparaitre très rapidement à l'écran et que leur tâche serait de se former une impression par rapport à ces visages sur la dimension de menace. Ils étaient avertis du fait que des prénoms apparaitraient à l'écran juste avant la présentation des dessins cibles. Il leur était demandé de ne pas se concentrer sur ces prénoms, mais de toutefois leur prêter attention, car des questions leur seraient posées à la fin de l'étude. Pour cette SMT, nous avons utilisé 72 prénoms en guise d'amorces (i.e., 24 prénoms maghrébins, 24 prénoms français, 24 prénoms neutres). L'ensemble des prénoms provenaient de la base de données Lexique database (New et al., 2004). Ceux-ci ont été sélectionnés parmi les prénoms maghrébins et français les plus fréquents dans la population française (Annexe 6). Les prénoms « neutres » ont été sélectionnés parmi des prénoms n'étant pas maghrébins ou français et ayant une fréquence dans la population qui était nulle (i.e., prénoms non présents en France). Les cibles de la SMT-prénoms et de la SMT-visages correspondaient aux 48 dessins de visages floutés de notre Étude 1 (24 dessins de visage menaçants, 24 dessins de visages non menaçants).

Après avoir répondu aux deux SMT, les participants passaient aux jugements directs de la catégorie et au jugement direct des exemplaires (contrebalancés d'un participant à un autre). Pour les jugements directs de la catégorie, les participants devaient juger les catégories

Maghrébins et Français sur la dimension de menace (« Selon vous, les Maghrébins/Français[21] sont »). Plus particulièrement, pour chaque catégorie (contrebalancée d'un participant à un autre), les participants devaient déplacer un curseur allant de « Peu menaçants » à « Très menaçants » (reproduisant les réponses de la SMT, les réponses sur ce curseur étant recodées pour les analyses en une échelle de menace en 100 points). Pour le jugement direct des exemplaires, les participants devaient, de la même façon, juger les différents visages et prénoms vus lors de la SMT sur la dimension de menace (« Selon vous, cet individu est »). Pour répondre, les participants utilisaient le même curseur décrit plus haut pour le jugement de la catégorie (les réponses sur ce curseur étant là encore recodées pour les analyses en une échelle de menace en 100 points). Pour finir, les participants étaient débriefés, gratifiés et enfin remerciés pour leur participation.

4.3.2 **Résultats**

Jugement de menace. Pour cette huitième étude, et à l'inverse de nos études précédentes, nous n'avons pas réalisé d'ANOVA classique. Pour rappel, nous faisions entre autres l'hypothèse suivante : l'effet SMT devait être corrélé avec le jugement de la catégorie indépendamment du jugement de chaque exemplaire. Par conséquent, et pour répondre au mieux à cette hypothèse, nous avons réalisé des analyses en modèles mixtes (Judd et al., 2012). Suivant en cela la structure de l'effet SMT (i.e, différence de jugement de menace entre amorces Maghrébin et Blanc) nous avons créé un score de jugement de la catégorie en soustrayant, pour chaque participant, la moyenne de jugements directs de la catégorie Maghrébins, par la moyenne de jugements directs de la catégorie Français. Pour chacun des essais composant la SMT, nous avions ainsi le jugement de menace associé à cet essai (i.e., réponse du participant) qui était mis en lien avec le score de jugement de la catégorie, en contrôlant le jugement de menace de l'exemplaire constituant cet essai. Plus précisément, nous avons réalisé une analyse en modèle mixte généralisé (i.e., en prenant pour variable dépendante le jugement de menace) avec pour effets fixes l'amorce (Maghrébin vs. Blanc), le type d'amorce (Visages vs. Prénoms), la cible (Menaçant vs. Non menaçant), leurs interactions et enfin le jugement des exemplaires et le score de jugement de la catégorie. Nous avons

[21] Pour la catégorie Maghrébins, il était spécifié par le biais d'un astérisque « français d'origine maghrébine ». Pour la catégorie Français, il était spécifié « français d'origine européenne ».

estimé les effets aléatoires relatifs aux intercepts et pentes pour les participants, les amorces et les cibles. Suivant les préconisations des certains auteurs (Judd et al., 2012), nous avons entré un modèle global comprenant l'ensemble des effets aléatoires, puis avons simplifié ce modèle en suivant les recommandations de Bates et collaborateurs (2018). Dans les lignes qui vont suivre, nous reporterons uniquement les résultats liés aux effets fixes, les effets aléatoires n'étant pas directement liés à nos différentes hypothèses. De plus, n'existant pas encore de consensus dans la littérature quant aux reports des tailles d'effets dans le cas d'analyses en modèles mixtes, celles-ci ne seront pas reportées.

Tout d'abord, de manière similaire à nos études précédentes, les participants ont jugé les cibles menaçantes comme effectivement plus menaçantes (M = 67.03, ET = 20.49) que les cibles non menaçantes (M = 45.28, ET = 19.68), z = 6.68, p < .001. Conformément à notre hypothèse, nous avons répliqué l'effet SMT, c'est-à-dire que nos participants ont jugé les cibles comme plus menaçantes quand elles été précédées d'amorces Maghrébin (M = 43.22, ET = 17.63) plutôt que d'amorces Blanc/Français (M = 33.37, ET = 13.93), z = 4.30, p < .001. Plus particulièrement, nous avons reproduit l'effet SMT pour la SMT-visages, le jugement de menace étant plus important lorsque l'amorce était composée de visages maghrébins (M = 23.66, ET = 8.19) plutôt que de visages blancs (M = 17.63, ET = 10.04), z = 4.35, p < .001, et pour la SMT-prénoms, le jugement de menace étant plus important lorsque l'amorce était composée de prénoms maghrébins (M = 19.56, ET = 9.84) plutôt que de prénoms français (M = 15.74, ET = 8.31), z = 2.47, p = .01 (Figure 30). L'impact du type d'amorce sur l'effet SMT s'est avéré non significatif, z = 1.51, p = .13. De plus, là encore en accord avec notre hypothèse, nous avons observé une relation significative entre l'effet SMT et le jugement sur la catégorie, en contrôlant le jugement de chaque exemplaire, z = 3.39, p < .001. Autrement dit, plus l'effet SMT était important et plus les participants jugeaient de manière directe les Maghrébins (i.e., la catégorie « Maghrébins ») comme plus menaçants que les Français (i.e., la catégorie « Français »), et ce, indépendamment du jugement de menace associé à chaque exemplaire. Cet effet était observé à la fois pour les amorces de type visage, z = 3.15, p = .002, et de type prénom, z = 3.41, p < .001.

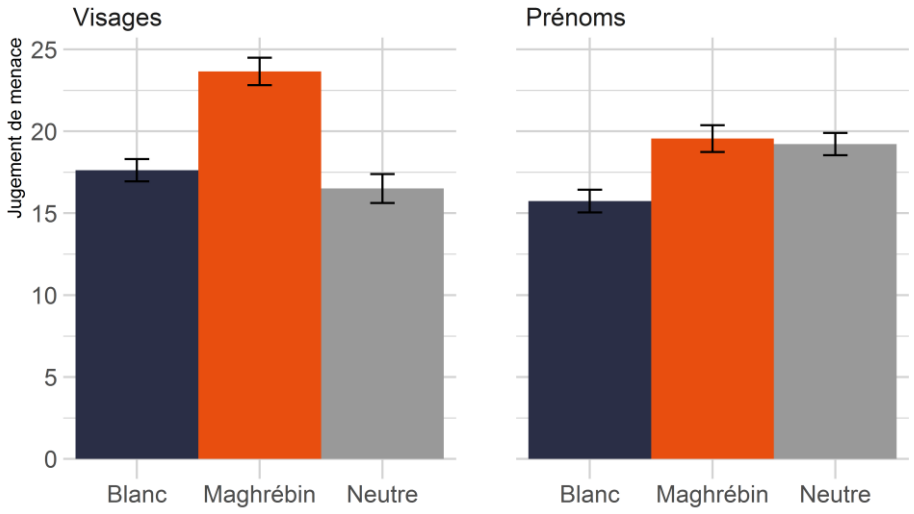


Figure 30. Moyenne des réponses « Plus menaçant » en fonction de l'amorce (Blanc, Maghrébin/Noir, Neutre) et du type d'amorce (visages vs. prénoms). Les barres d'erreurs représentent les erreurs standards.

Modèle Multinomial. Nous avons appliqué un modèle multinomial SAC-dominant à six paramètres, c'est-à-dire avec les paramètres SAC et SAP variant en fonction du type d'amorce (i.e., visages vs. prénoms), mais des paramètres D et G uniques. L'adéquation du modèle aux données (T_1) était acceptable (M_{pred} = 0.07, M_{obs} = 0.54, p < .001, w = .06). Bien que n'ayant pas d'hypothèses précises sur les différents paramètres du modèle, nous avons comparé de manière exploratoire les paramètres d'activation du stéréotype (SAC) et d'application du stéréotype (SAP) en fonction du type d'amorce (i.e., visages vs. prénoms). Le paramètre SAC était plus élevé pour les visages (M_{SAC} = .58, $IC_{95\%}$[.37 ; .78]) que pour les prénoms (M_{SAC} = .02, $IC_{95\%}$[.001 ; .057]), ΔSAC = .56, $IC_{95\%}$[.35 ; .77], p_B < .001 (Figure 31). Autrement dit, les participants ont plus activé leur stéréotype quand les amorces étaient constituées de visages plutôt que de prénoms. À l'inverse, le paramètre SAP était plus important pour les prénoms (M_{SAP} = .87, $IC_{95\%}$[.61 ; .99]) que pour les visages (M_{SAP} = .59, $IC_{95\%}$[.52 ; .67]), ΔSAP = .27, $IC_{95\%}$[.02 ; .42], p_B = .02. Autrement dit, les participants ont plus appliqué leur stéréotype quand les amorces étaient constituées de prénoms plutôt que de visages.

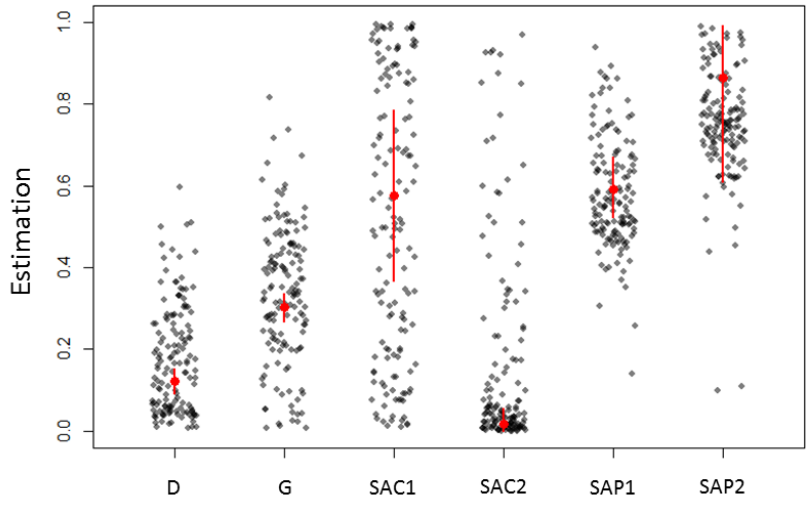


Figure 31. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction du type d'amorce (visages vs. prénoms) et pour un modèle SAC-dominant à quatre paramètres. SAC1 et SAP1 = paramètres pour les visages, SAC2 et SAP2 = paramètres pour les prénoms.

Comme dans nos précédentes études, nous avons ensuite réalisé des régressions linéaires pour observer le lien entre les différents paramètres du modèle et la mesure de jugement de la catégorie. Concernant les paramètres liés aux amorces de visages (SAC1 et SAP1), nous avons observé un lien entre le jugement direct de la catégorie et le paramètre SAP1, b = 0.001, $t(142)$ = 2.60, p = .01, η^2_p = .05, mais pas entre le jugement direct et le paramètre SAC1, b = 0.001, $t(142)$ = 1.00, p = .32, η^2_p = .01. Ainsi, plus les participants appliquaient leur stéréotype dans la SMT-visages et plus le score de jugement de la catégorie était élevé (catégorie Maghrébins plus menaçante que catégorie Français). Pour les paramètres liés aux amorces de prénoms (SAC2 et SAP2), nous avons observé un lien entre le jugement direct de la catégorie et le paramètre SAP2, b = 0.002, $t(142)$ = 3.38, p < .001, η^2_p = .08. Ainsi, plus les participants appliquaient leur stéréotype dans la SMT-prénoms et plus le score de jugement de la catégorie était élevé (catégorie Maghrébins plus menaçante que catégorie Français). De plus, nous avons aussi observé un lien entre le jugement direct de la catégorie et le paramètre SAC2, b = 0.003, $t(142)$ = 2.96, p = .004, η^2_p = .06. Ainsi, plus les participants activaient leur stéréotype dans la SMT-prénoms et plus le score de jugement de la catégorie était élevé (catégorie Maghrébins plus menaçante que catégorie Français). Enfin, nous n'avons pas observé de lien significatif entre le jugement direct de la catégorie et à la

fois le paramètre D, b = -0.00, $t(142)$ = -0.05, p = .96, η^2_p < .001, et le paramètre G, b = 0.001, $t(142)$ = 0.79, p = .43, η^2_p = .004.

4.3.3 **Discussion**

Le but de cette huitième étude était de pouvoir éliminer une hypothèse alternative quant à l'explication de l'effet SMT. Il était en effet envisageable que l'effet SMT ne soit pas la résultante d'une réactivation du contenu mental relié aux catégories Maghrébins et Blancs, mais plus simplement de visages maghrébins et blancs intrinsèquement menaçants et non menaçants. Afin de départager ces possibilités, nous avons entrepris de répliquer notre tâche SMT avec des amorces de visages, mais aussi avec des amorces de prénoms, ces dernières ne pouvant pas être intrinsèquement qualifiées de menaçantes. Conformément à nos hypothèses, nous avons répliqué l'effet SMT avec des visages (effet SMT classique), mais aussi avec des prénoms. Il semblerait ainsi que l'effet SMT ne soit pas (seulement) le reflet de visages maghrébins intrinsèquement plus menaçants que des visages blancs, mais aussi d'une réactivation de la catégorie et du stéréotype associé à ces visages. Pour nous en assurer, nous avons également voulu voir si l'effet SMT pouvait être corrélé à un jugement direct de la catégorie (Maghrébins vs Français) indépendamment du jugement direct de chaque exemplaire, c'est-à-dire de chaque amorce constituant la SMT. Effectivement, en lien avec nos hypothèses, l'effet SMT était corrélé avec la différence de jugement entre les catégories Maghrébins et Français indépendamment du jugement des exemplaires constituant la SMT, et ce, à la fois pour les visages et pour les prénoms. L'ensemble de ces résultats confortent donc notre hypothèse première quant à l'explication de l'effet SMT. Rejoignant Krieglmeyer et Sherman (2012), ces résultats sont cohérents avec une réactivation en mémoire du contenu mental associé aux catégories respectives des amorces Maghrébin et Blanc.

Comme dans nos précédentes études, les paramètres d'application du stéréotype (SAP) étaient corrélés aux jugements de la catégorie que ce soit pour la SMT-visages ou la SMT-prénoms. De plus, et de manière inexpliquée au regard de nos études précédentes, le paramètre SAC était également corrélé au jugement de la catégorie pour la SMT-prénom. Une fois encore de manière surprenante, le paramètre d'activation du stéréotype (SAC) pour les prénoms était relativement faible (proche de 0) et bien plus faible que celui des visages. Ceci est d'autant plus étonnant que, dans la littérature et notamment la littérature traitant de l'IAT,

les effets sont traditionnellement plus importants pour les prénoms que pour les visages (e.g., Foroni & Bel-Bahar, 2009 ; Mitchell et al., 2003 ; Nosek et al., 2002). Pour Foroni et Bel-Bahar (2009), cela peut venir d'un niveau de représentation différent entre visages et prénoms. Pour ces auteurs et comme nous l'avons décrit plus haut, les prénoms sont des stimuli de haut niveau, plus abstraits et donc plus proche de la catégorie en elle-même. À l'inverse, les visages sont des stimuli de plus bas niveau, plus concrets et donc partageant des composantes visuelles autres que celles directement reliées à la catégorie. Or, dans ce cas précis, nous pouvons penser qu'un stimulus de plus bas niveau comme des visages peut avoir une influence supérieure sur l'effet SMT, de par ces composantes autres que celles directement liées à la catégorie (i.e., traits du visage menaçants ou non). Toutefois, le paramètre d'application du stéréotype (SAP) pour les prénoms était supérieur à celui des visages. Une explication à cela peut être que le prénom présenté en amorce vienne fonctionner comme un label pour la cible (e.g., « cette cible s'appelle Mohammed ») et donc renforcer la menace ou non menace potentielle de celle-ci. Bien sûr, ces explications sont réalisées a posteriori et effectuées sur des résultats exploratoires. Mais d'une manière générale, il nous semble important de ne pas sous-estimer la probabilité que des mécanismes différents puissent être à l'origine de la réponse du participant dans la SMT-visages et dans la SMT-prénoms, comme cela a pu être avancé pour l'IAT (e.g., Foroni & Bel-Bahar, 2009).

Pour résumer, nous avons une nouvelle fois répliqué l'effet SMT, cette fois-ci avec à la fois des prénoms et des visages. Pour nous, ceci vient renforcer le caractère généralisable de nos résultats. Toutefois, jusqu'à présent, l'ensemble de nos études a été réalisé en France, à l'Université Grenoble Alpes, sur des participants de culture principale française. L'effet SMT qui nous intéresse, à savoir celui contrastant les visages maghrébins et blancs, est-il généralisable dans un autre contexte ?

4.4 L'effet SMT est-il généralisable à d'autres populations ?

Cesario (2014), dans son article traitant de la reproductibilité des effets d'amorçage, donne au lecteur les recommandations suivantes : « À ce stade du développement des recherches sur les effets d'amorçages, il me semble évident que l'étape la plus importante est que nous établissions d'abord la reproductibilité de base d'un effet d'amorçage donné : un seul laboratoire de recherche peut-il reproduire un effet plusieurs fois ? Ce n'est que lorsque

nous avons une taille d'effet associée à cet effet et que nous savons que l'effet obtenu n'est pas une erreur de Type I, que nous pouvons voir d'autres laboratoires tenter des réplications » (p. 44). Bien que Cesario ne traite ici naturellement pas des effets propres aux attitudes raciales, nous pouvons nous poser la question suivante : notre effet SMT, reproductible dans notre laboratoire auprès de participants de culture principale française, est-il généralisable à d'autres laboratoires ? D'autres populations ? Les deux dernières études de ce chapitre vont tenter de répondre à cette question. Dans notre neuvième étude, nous avons testé le caractère généralisable de notre effet SMT au sein d'un autre environnement culturel (i.e., les États-Unis). Dans notre dixième étude, nous avons testé le caractère généralisable de notre effet SMT au sein du groupe social cible du stéréotype.

4.4.1 Généralisabilité de l'effet SMT au sein d'un autre environnement culturel

Pour savoir si l'effet SMT Maghrébins/Blancs est généralisable dans d'autres laboratoires, il nous faut revenir sur notre raisonnement théorique et, une nouvelle fois, à ce que nous entendons par « stéréotype ». Comme nous l'avons vu au Chapitre 1, les définitions actuelles du stéréotype décrivent celui-ci comme un ensemble de croyances partagées à propos de caractéristiques de certains groupes sociaux (e.g., Hilton & von Hippel, 1996 ; Leyens et al., 1996). Le stéréotype aurait donc un fondement social (e.g., amis, familles) et culturel (e.g., Devine, 1989 ; Hilton & von Hippel, 1996 ; Leyens et al., 1996 ; Payne et al., 2017). Plus particulièrement, l'ensemble des membres d'une zone géographique donnée (e.g., ville, pays) serait exposé aux stéréotypes culturels présents au sein de cette zone (e.g., Arkes & Tetlock, 2004 ; Devine, 1989 ; Payne et al., 2017 ; Rougier et al., 2020). En lien avec la description de Payne et collaborateurs (2017), nous avons par le passé défini le stéréotype culturel (ou « connaissances culturelles ») comme « l'exposition objective à un environnement donné, diffusant certaines idées et valeurs à propos de groupes sociaux » (Rougier et al., 2020, p. 858). En conséquence, un stéréotype (e.g., les Maghrébins sont menaçants) pourrait être spécifique d'un environnement culturel donné (i.e., zone géographique précise) et refléter la situation actuelle au sein de cet environnement (Payne et al., 2017).

Nous avons déjà évoqué ce point aux Chapitres 1 et 3 à propos de la différence de perception des individus de couleur noire et des individus d'origine maghrébine aux États-Unis et en Europe. Aux États-Unis, un grand nombre d'études a mis en évidence un biais

négatif réactivé automatiquement (i.e., de manière non intentionnelle) envers les individus de couleur noire (e.g., Devine, 1989 ; Fazio et al., 1995 ; Greenwald & Banaji, 1995 ; Nosek et al., 2002). Plus intéressant quant à l'importance de l'environnement culturel, pour Payne et collaborateurs (2019) l'attitude négative envers ces populations serait notamment issue d'un héritage culturel marqué par l'esclavagisme et le passé ségrégationniste de l'histoire des États-Unis. Ces auteurs ont ainsi montré que les comtés et états des États-Unis ayant été les plus marqués par l'esclavagisme (e.g., états du Sud) avaient des scores IAT pro-Blancs plus importants chez les populations blanches, suggérant ainsi la réactivation d'un contenu mental négatif anti-Noirs lié à cet héritage. Au niveau de la menace associée à ces populations, nous avons listé bon nombre d'études montrant, aux États-Unis, une association entre les individus de couleur noire et ce stéréotype de menace (e.g., Ackerman et al., 2006 ; Correll et al., 2002 ; Trawalter et al., 2008) et notamment le paradigme SMT classique (Krieglmeyer & Sherman, 2012). À l'opposé, comme nous l'avons aussi montré, le stéréotype de menace en France est particulièrement associé aux populations d'origine maghrébine (e.g., Gauthier, 2015 ; Jobard et al., 2012 ; Mange et al., 2016). Nous avons nous-même pu mettre en évidence, dans les sept premières études constituant cette thèse, la présence d'un stéréotype liant les populations maghrébines à une menace. Par conséquent, nous avons fait l'hypothèse d'un stéréotype ancré différemment d'un point de vue culturel, celui-ci étant davantage porté par les populations noires aux États-Unis et par les populations maghrébines en France.

L'un des objectifs de notre Étude 5 était de comparer, en France, la SMT classique (SMT-Noir/Blanc) avec notre SMT (SMT-Maghrébin/Blanc). Nous pensions que le stéréotype de menace culturellement associé en France aux populations maghrébines pouvait aboutir à un effet SMT plus important dans la SMT-Maghrébin/Blanc que dans la SMT-Noir/Blanc. Effectivement, nous avons obtenu un effet SMT plus fort chez les participants passant la SMT-Maghrébin/Blanc que chez ceux passant la SMT-Noir/Blanc. Cet effet était expliqué par un paramètre SAC (descriptivement) plus élevé dans la SMT-Maghrébin/Blanc et surtout un paramètre SAP significativement plus élevé dans cette même SMT. Suivant en cela les travaux sur la composante culturelle du stéréotype (e.g., Payne et al., 2017), nous avons interprété ces résultats comme pouvant effectivement suggérer une différence culturelle entre France et États-Unis par rapport au groupe social majoritairement associé au stéréotype de menace.

Toutefois, nous devons être prudents quant à ces conclusions. D'abord, pour conclure à un stéréotype reflétant l'environnement culturel français et pas juste « Grenoblois », il nous faudrait répliquer ce résultat ailleurs en France et pas uniquement à l'Université Grenoble

Alpes. Ensuite, si effectivement l'effet SMT peut être le fruit de différences culturelles entre France et États-Unis, répliquer cette étude aux États-Unis serait pour nous l'opportunité d'observer si, comme on peut s'y attendre, l'influence culturelle géographique des États-Unis peut aboutir à un effet SMT différent de ce que nous avons en France. Ceci constituait l'objectif principal de notre Étude 9.

4.4.2 Généralisabilité de l'effet SMT au sein du groupe cible du stéréotype

Les études que nous avons effectuées en France l'ont été sur un échantillon de participants se catégorisant de culture principale française. Dans le champ des attitudes raciales, l'effet SMT peut de ce fait être rapproché d'un biais dit « pro-endogroupe », c'est-à-dire d'un jugement plus positif pour son groupe social (i.e., les Blancs) en comparaison à un groupe social différent (i.e., les Maghrébins). Mais peut-on s'attendre à un même type d'effet de la part d'individus faisant partie du groupe cible du stéréotype négatif (i.e., les individus français d'origine maghrébine) ? En d'autres termes, de la part des individus membres de ce groupe, doit-on s'attendre à un effet SMT pro-exogroupe (i.e., effet SMT classique) ou pro-endogroupe (i.e., effet SMT inversé) ? Dans la littérature, le groupe des Maghrébins (ou plus largement des Arabo-musulmans) est souvent décrit comme un groupe de bas statut social (e.g., Degner et al., 2016 ; Essien et al., 2020). Nous allons voir que, dans notre questionnement, cela peut avoir son importance.

Dans le champ des attitudes raciales, les premières études réalisées à l'aide de l'IAT sur des groupes cibles d'un préjugé négatif ont majoritairement mis en évidence un favoritisme pro-exogroupe (e.g., Ashburn-Nardo et al., 2003 ; Livingston, 2002 ; Nosek et al., 2002 ; Richeson et al., 2005 ; Rudman et al., 2002). Par exemple, Nosek et collaborateurs (2002) ont mis en évidence, chez un nombre important de participants passant un IAT Noirs/Blancs (i.e., données issues du Project Implicit), un biais globalement pro-Blancs de la part des participants Noirs. Pour Rudman et collaborateurs (2002), le biais pro-endogroupe ou pro-exogroupe obtenu au sein de l'IAT dépendrait avant tout du statut social du groupe au sein de la société. Ces auteurs ont réalisé différents IAT sur plusieurs groupes et notamment sur des groupes sociaux de bas statut social (i.e., personnes en surpoids, pauvres). Là où les participants de plus haut statut (i.e., juifs, asiatiques) exprimaient un biais pro-endogroupe, les participants issus de groupes de bas statuts exprimaient davantage un biais pro-exogroupe.

Pour un grand nombre d'auteurs, ces résultats seraient une fois encore la preuve que, d'une part, les mesures indirectes captent effectivement les croyances culturelles dominantes au sein d'une société donnée, et d'autre part, que l'ensemble des individus issus du même environnement sont sensibles à ces croyances (e.g., Arkes & Tetlock, 2004 ; Banaji et al., 2004 ; Jost, 2019 ; Jost et al., 2004 ; Karpinski & Hilton, 2001 ; Payne et al., 2017 ; Rudman et al., 2002). Cependant, la littérature diverge quant à un point précis : les réponses des participants dans ce type de mesure reflètent-elles uniquement des « connaissances extra-personnelles » ou bien reflètent-elles l'attitude raciale personnelle des différents participants (Arkes & Tetlock, 2004 ; Banaji et al., 2004 ; Jost, 2019 ; Nosek & Hansen, 2008) ?

Pour Arkes et Tetlock (2004), le fait que des individus de groupes de bas statuts expriment une préférence pour l'exogroupe serait la preuve que le contenu capté par l'IAT n'est qu'une simple connaissance du stéréotype culturel, sans adhérence à celui-ci. Ainsi, pour Arkes et Tetlock « si je suis au courant du stéréotype culturel, j'ai tout le matériel cognitif nécessaire pour manifester un préjugé dans l'IAT » (p. 262). Ces auteurs prennent l'exemple de Jesse Jackson, pasteur libéral afro-américain ayant passé la majeure partie de sa vie à combattre pour les droits des Afro-américains aux États-Unis. Pour eux, si Jesse Jackson passait une tâche IAT, celui-ci exprimerait un biais pro-exogroupe (i.e., pro-Blancs), quand bien même son attitude raciale individuelle serait profondément pro-Noirs. Payne et collaborateurs (2017) partagent eux aussi ce point de vue : « Toute personne — même un leader noir des droits civiques — imprégnée d'une culture où les biais systémiques sont très répandus a de fortes chances de voir passer dans son esprit des pensées et des sentiments conformes aux stéréotypes. » (p. 240).

À l'inverse, Nosek et Hansen (2008) réfutent la possibilité d'un IAT ne reflétant qu'une simple connaissance du stéréotype culturel. Par le biais de sept études, ces auteurs ont montré que, globalement, les scores IAT étaient plus fortement corrélés avec une mesure directe de préjugé individuel plutôt qu'une simple connaissance du stéréotype culturel. Pour Nosek et Hansen, cela serait la preuve que ces « associations dans notre tête » (p. 553) sont propres à chaque individu. Cela rejoint en partie la vision de Banaji et collaborateurs (2004) pour qui il n'existerait pas de frontière entre attitude personnelle et croyances culturelles, nos croyances personnelles étant façonnées par notre culture. De même, pour d'autres auteurs et suivant en cela les théories de justification du système, les individus issus de groupes sociaux de bas statuts n'auraient pas qu'une simple connaissance de ces croyances culturelles, mais les auraient intériorisées au cours du temps (e.g., Jost, 2019 ; Jost et al., 2004 ; Rudman et al.,

2002). Ces scores IAT ne seraient ainsi que le reflet de la perpétuation du système inégalitaire (Jost et al., 2004). Jost (2019), rejoignant en cela le point de vue d'autres auteurs (Banaji et al., 2004), a une vision de l'individu comme faisant partie de son environnement. L'attitude raciale captée par les mesures indirectes serait le reflet à la fois du contexte culturel, mais aussi d'aspects importants pour l'individu lui-même (e.g., ses idéologies). En résumé, la question de l'origine personnelle ou extra-personnelle du contenu mental réactivé semble ne pas faire consensus : est-ce que ces résultats sont le reflet d'une simple connaissance du stéréotype ou est-ce que ceux-ci sont le reflet d'une intériorisation, voire d'une adhésion à ces connaissances ?

En réalité, et pour revenir un peu en arrière, les études utilisant l'IAT ne sont pas unanimes à propos de cette tendance qu'ont les groupes de bas statut social à exprimer un biais pro-exogroupe. Entre autres, plusieurs études récentes ont montré un favoritisme pro-endogroupe de la part d'individus de couleur noire (Hehman et al., 2019 ; Payne et al., 2019 ; Ratliff et al., 2020). C'est le cas de Payne et collaborateurs (2019) dont nous avons décrit les résultats plus hauts. Ces auteurs ont certes mis en évidences des scores IAT pro-endogroupe plus importants chez des participants blancs issus d'états anciennement esclavagistes, mais, à l'inverse, ils ont observé un biais également pro-endogroupe chez les participants noirs issus de ces mêmes états. Pour ces auteurs, cela s'expliquerait par un impact différent d'un même environnement sur les populations blanches et les populations noires (sans toutefois détailler ces différences). Néanmoins, et nous reviendrons sur ce point, l'impact du type d'état (plus ou moins historiquement esclavagiste) sur le score IAT des participants Noirs disparaissait quand contrôlé par la réponse du participant à une mesure directe d'attitude raciale.

L'ambiguïté de ces résultats ne se limite pas à l'IAT. D'autres types de mesures indirectes ont en effet montré des résultats antagonistes quant à la préférence des groupes sociaux de bas statut, par exemple la littérature traitant des tâches d'approche-évitement (e.g., Degner et al., 2016 ; Bianchi et al., 2018 ; Rougier et al., 2020). Degner et collaborateurs (2016) ont mis en évidence le fait que des élèves faisant partie d'un groupe social minoritaire allaient exprimer un biais pro-endogroupe lorsque ceux-ci se trouvaient dans une école comportant de nombreux enfants issus de l'immigration, mais un biais pro-exogroupe lorsque ceux-ci se trouvaient dans une école comportant une majorité d'élèves blancs, ces auteurs expliquant une nouvelle fois ce résultat en se référant à la théorie de justification du système. Pourtant, ce dernier résultat n'est pas compatible avec l'une de nos précédentes études réalisée en amont de ce travail de thèse (Rougier et al., 2020). En effet, dans notre université,

nous avons montré que des participants maghrébins passant une tâche d'approche-évitement allaient davantage exprimer un biais pro-endogroupe, signe pour nous de la capacité de notre tâche, la *Visual Approach/Avoidance by the Self Task* (i.e., la VAAST), à capter l'attitude personnelle individuelle au-delà de l'impact culturel.

En somme, que ce soit pour l'IAT ou pour des mesures d'approche-évitement, les résultats sont ambivalents. En plus de l'absence de consensus entre la composante extra-personnelle ou personnelle du contenu mental réactivé, il ne semble pas y avoir de consensus non plus sur une éventuelle préférence endogroupe ou exogroupe au sein des mesures indirectes, du moins de la part du groupe de bas statut. À ce débat, nous n'avons pas la prétention d'apporter de réponses permettant d'expliquer cette absence de consensus. Toutefois, et dans le prolongement de ce travail de thèse, nous pensons que les auteurs (nous y compris) interprétant les résultats de ces mesures n'ont pas pris assez en considération un aspect important : ces mesures ne sont pas *process pure*. Pour nous, ceci constitue une limite quant à l'avancée des connaissances dans ce champ de recherche et pourrait même aboutir à une interprétation erronée des résultats.

Deux illustrations peuvent nous montrer l'intérêt de prendre en compte l'ensemble des processus à l'œuvre et notamment l'importance du processus lié à une inhibition du biais réactivé. Tout d'abord, Gonsalkorale et collaborateurs (2014) ont montré que des participants jeunes (groupe de haut statut), comme des participants âgés (groupe de bas statut ; e.g., Jost et al., 2004) exprimaient un biais IAT négatif envers les personnes âgées, constituant donc un biais pro-exogroupe pour le groupe social de plus bas statut (i.e., les personnes âgées) et répliquant les résultats classique de la littérature (voir Jost et al., 2004). Toutefois, en appliquant un Quad Model à ces données, Gonsalkorale et collaborateurs ont montré que ce même biais pro-Jeunes pouvait être expliqué différemment en fonction du groupe social. Effectivement, ces auteurs ont montré que les participants plus âgés avaient une activation du biais « Jeunes = Positif » / « Agés = Négatif » plus faible que les participants plus jeunes, mais aussi une inhibition de ce biais là encore plus faible que ces derniers. Ces résultats apportent ainsi un éclairage nouveau à l'interprétation du biais pro-exogroupe habituellement observé au sein de l'IAT et montrent l'importance du processus de réactivation du contenu mental, mais aussi, et de manière nouvelle, du processus d'inhibition de ce contenu réactivé.

Plus intéressant encore pour notre propos, Jin et collaborateurs (2016) se sont intéressés aux scores IAT pro-endogroupe (i.e., pro-Femmes) de participantes chinoises venant d'un environnement urbain ou rural. Pour ces auteurs, l'environnement urbain chinois

serait plus propice que l'environnement rural à la diffusion de valeurs prônant l'égalité des genres et à la place de la femme dans la société. C'est la raison pour laquelle Jin et collaborateurs faisaient l'hypothèse d'un biais pro-endogroupe plus important pour les femmes issues d'un environnement urbain que celles issues d'un environnement rural, ces dernières vivant dans un environnement culturellement plus subordonné aux hommes. Conformément à leurs attentes, les participantes issues d'un environnement urbain avaient un score IAT pro-endogroupe plus élevé que les participantes issues d'un environnement rural. Ce résultat est déjà intéressant mais Jin et collaborateurs sont allez encore plus loin dans l'explication de ce résultat. En effet, en appliquant un Quad Model à ces données, ces auteurs ont montré que ce résultat n'était pas le fruit d'une activation plus importante du biais pro-endogroupe chez les participantes issues d'un environnement urbain, mais plutôt d'une inhibition plus importante de ce biais chez les participantes issues d'un environnement rural. En somme, il semblerait ici que l'environnement culturel ait davantage influencé le processus d'inhibition, mais pas le contenu mental en lui-même. Une fois encore, cette étude nous montre l'importance du paramètre d'inhibition du contenu mental réactivé.

À la lumière des résultats de ces études, il se pourrait donc que les biais pro-endogroupe ou pro-exogroupe relevés dans la littérature ne soient pas uniquement le reflet d'une réactivation automatique d'un contenu de croyances culturelles. Dès lors, nous pouvons nous questionner sur l'interprétation de certains des résultats décrits plus hauts. Prenons l'étude de Rudman et collaborateurs (2002) montrant que plus les individus faisaient partie d'un groupe de bas statut social et plus ils exprimaient un biais pro-exogroupe au sein de l'IAT. En tenant compte du caractère multi-processuel de l'IAT, nous pouvons nous demander si ces différences de résultats en fonction du statut reflètent une activation plus forte du biais négatif chez les groupes de bas statut, une inhibition moins importante de ce biais chez ces mêmes groupes ou bien la combinaison des deux. De la même manière, Nosek et collaborateurs (2002) ont certes montré un biais pro-exogroupe chez des participants noirs, mais celui-ci était plus faible que celui de participants blancs (i.e., d = 0.16 vs. d = 0.83). Ces résultats reflètent-ils une activation plus faible du contenu mental anti-Noirs chez les participants Noirs ou bien une volonté plus importante d'inhiber cette activation chez ces mêmes participants ? Dans l'étude de Payne et collaborateurs (2019), enfin, nous avons vu que le lien entre le type d'état (plus ou moins historiquement esclavagiste) et le score IAT s'était révélé non significatif une fois contrôlée par l'attitude raciale auto-rapporté. Ce résultat laisse entrevoir la possibilité que le lien entre état et score IAT soit en réalité davantage porté par des processus de Type II. Autrement dit, et comme nous l'avons vu dans l'étude de Jin et

collaborateurs (2016), ne peut-on pas envisager que l'environnement culturel influence le paramètre d'inhibition, mais pas le contenu mental en lui-même ?

Pour reprendre l'exemple de Jesse Jackson, celui-ci exprimerait-il un biais IAT classique pro-exogroupe, comme certains auteurs pouvaient le laisser penser (e.g., Arkes & Tetlock, 2004 ; Payne et al., 2017) ou pourrait-il contrôler ce biais réactivé et ainsi se rapprocher d'autres résultats de la littérature montrant un biais pro-endogroupe (e.g., Jost et al., 2004 ; Payne et al., 2019). Encore une fois, le but de l'Étude 10 n'était pas d'apporter une réponse à ce débat. Toutefois, nous pensions qu'aborder la question de la composante personnelle ou extra-personnelle de l'attitude en prenant en considération le caractère multi-processuel des mesures indirectes pouvait nous permettre d'apporter un éclairage nouveau à celui-ci. C'est pourquoi le but de l'Étude 10 était d'observer si nous retrouvions l'effet SMT habituel au sein du groupe social cible du stéréotype (i.e., les individus maghrébins), groupe social décrit comme étant de plus bas statut social (e.g., Degner et al., 2016 ; Essien et al., 2020)

4.4.3 Vue d'ensemble des Études 9 et 10

Les deux dernières études de ce chapitre avaient pour but de tester la généralisabilité de notre effet SMT relatif aux visages maghrébins et blancs. Avec l'Étude 9, nous voulions tester le caractère généralisable de l'effet SMT dans un autre environnement culturel. Plus particulièrement, en miroir de l'Étude 5, nous voulions observer si, aux États-Unis, le stéréotype de menace était davantage associé aux individus de couleur noire plutôt qu'aux individus d'origine maghrébine. En complément, le but de l'Étude 10 était d'observer, au sein de la population cible du stéréotype négatif (i.e., les Maghrébins), si nous retrouvions ou non l'effet SMT classique, mais également de déterminer l'influence des différents processus (i.e., activation du stéréotype, application du stéréotype) sur le biais pro-endogroupe ou pro-exogroupe observé.

4.5 Étude 9 : SMT dans un autre environnement culturel, réplication aux États-Unis

Jusqu'à présent, nous avons effectué l'intégralité de nos études en France, à l'Université Grenoble Alpes. Au sein de l'environnement culturel français, nous avons répliqué l'effet SMT de nombreuses fois. De la même manière, dans l'Étude 5, nous avons montré que l'effet SMT était plus important pour des participants passant la tâche SMT avec des amorces de visages maghrébins et blancs que pour des participants passant la tâche SMT avec des amorces de visages noirs et blancs. Nous avons envisagé ce résultat par l'influence culturelle française, c'est-à-dire par la présence d'un stéréotype de menace plus associé aux individus maghrébins, en France, qu'aux individus de couleur noire. Le but de cette étude était de répliquer notre Étude 5 dans un contexte culturel différent : aux États-Unis. Plus précisément, nous avons entrepris de répliquer l'effet SMT au sein même du laboratoire dans lequel la tâche a été créée. Nous faisions ainsi l'hypothèse, en accord avec les résultats de Krieglmeyer et Sherman (2012), d'un effet SMT plus important pour les participants passant la SMT-Noir/Blanc que pour les participants passant la SMT-Maghrébin/Blanc. Cette étude était également pour nous la possibilité de répliquer les résultats de notre Étude 5 quant au meilleur modèle multinomial expliquant les données. Ayant interprété ces résultats comme le reflet de différences perceptives relatives aux amorces utilisées (voir Étude 5), nous pensions que ces résultats ne différeraient pas dans un autre contexte géographique. C'est pourquoi, en accord avec les résultats de cette précédente étude, nous faisions l'hypothèse d'un meilleur ajustement aux données du modèle SAC-dominant pour la SMT-Noir/Blanc et d'un meilleur ajustement aux données du modèle D-dominant pour la SMT-Maghrébin/Blanc.

4.5.1 Méthode

Participants et plan expérimental. Suivant en cela notre Étude 5 et compte tenu de nos possibilités en termes de participants, nous avons planifié la collecte d'au moins 200 participants, c'est-à-dire 100 participants dans chaque modalité de la condition expérimentale « type de SMT » à deux modalités (i.e., SMT-Noir/Blanc vs. SMT-Maghrébin/Blanc). En nous basant sur une analyse de sensibilité (G*power 3.1 ; Faul et al., 2007) cela nous permettait de

détecter une taille d'effet de petite à moyenne (*ds* = 0.40) avec un seuil à .05 et une puissance statistique à .80.

Deux-cent-trois participants ont pris part à cette étude. Parmi les participants recrutés, deux ont été écartés des analyses, car ils avaient donné systématiquement la même réponse à l'ensemble des essais (répondant plus de 98% du temps « Plus menaçant » ou « Moins menaçant »). Au final 201 participants ont été gardés pour les analyses (168 femmes, 28 hommes, 5 autres/non spécifiés, $M_{\text{âge}}$ = 19.58, $ET_{\text{âge}}$ = 2.84). Les participants étaient des étudiants de licence de Psychologie (i.e., undergraduated students) de l'Université de Davis, Californie et recevaient des crédits de cours en compensation. Il est à noter, et nous reviendrons sur ce point par la suite, que nos participants n'étaient pas uniquement de couleur blanche (95 Asiatiques, 51 Hispaniques, 39 Blancs, 9 Arabo-musulmans, 3 Noirs, 1 Natif américain et 3 non spécifiés). Le plan expérimental de cette étude était identique à celui de notre Étude 5, soit un plan de type 3 (amorce : Maghrébin/Noir vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (type de tâche : SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc), le dernier facteur étant en inter-sujets. Le pré-enregistrement de cette étude est disponible sur OSF.
(https://osf.io/7afh9/?view_only=36d9bd8cf37a496e9d6b7286638910c1)

Matériel et procédure. À leur arrivée au laboratoire, les participants étaient conduits vers deux salles expérimentales de passations communes. Ainsi, jusqu'à huit participants pouvaient réaliser l'étude en simultanée (i.e., quatre par salle expérimentale). La tâche SMT a été créée avec l'aide du logiciel Medialab (v2018.1.117). Les consignes étaient identiques aux consignes de nos études précédentes, mais traduites en langue anglaise et correspondant ainsi aux consignes de l'étude originale de Krieglmeyer et Sherman (2012). De la même manière, l'ensemble du matériel expérimental était similaire au matériel expérimental de l'Étude 5.

4.5.2 Résultats

Jugement de menace. Comme nous l'avons spécifié ci-dessus, les participants gardés pour les analyses n'étaient pas uniquement de couleur blanche. Ainsi, les analyses reportées ci-dessous ont été réalisées sur l'ensemble des participants ayant pris part à notre tâche et

sont donc le reflet d'origines ethniques différentes. Comme nous le verrons plus tard, cela peut avoir son importance.

Pour analyser nos différents effets, nous avons effectué une ANOVA à plan mixte 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 2 (type de tâche : SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc) avec la somme des réponses « Plus menaçant » en variable dépendante. Tout d'abord nous avons obtenu un effet du type de cible, les participants jugeant les cibles menaçantes comme étant effectivement plus menaçantes (M = 32.98, ET = 11.00) que les cibles non menaçantes (M = 28.26, ET = 10.82), $t(198) = 7.97$, $p < .001$, $\eta^2_p = .24$[22]. Cet effet n'était modéré par le type de tâche, $t(198) = 1.49$, $p = .14$, $\eta^2_p = .01$. Concernant notre hypothèse principale, nous n'avons pas obtenu d'impact du type de tâche sur l'effet SMT, $t(199) = 0.86$, $p = .39$, $\eta^2_p = .004$. Plus précisément, la différence entre les amorces Maghrébins (M = 24.08, ET = 10.64) et Blancs (M = 25.27, ET = 8.75) était non significative, $t(199) = -0.81$, $p = .42$, $\eta^2_p = .003$ (Figure 32). De manière surprenante, la différence entre les amorces Noirs (M = 21.12, ET = 11.14) et Blancs (M = 20.49, ET = 9.79) était là encore non significative, $t(199) = 0.42$, $p = .68$, $\eta^2_p = .001$. Globalement, les cibles étaient perçues comme plus menaçantes quand elles étaient précédées d'amorces Noir/Maghrébin (M = 22.63, ET = 10.95) plutôt que d'amorces Neutre, (M = 15.66, ET = 13.02), $t(199) = 6.14$, $p < .001$, $\eta^2_p = .16$, cet effet n'étant pas modéré par le type de tâche, $t(199) = -1.22$, $p = .22$, $\eta^2_p = .01$. De même, les cibles étaient globalement perçues comme plus menaçantes quand elles étaient précédées d'amorces Blanc (M = 25.27, ET = 8.75) plutôt que d'amorces Neutre, (M = 15.66, ET = 13.02), $t(199) = 6.28$, $p < .001$, $\eta^2_p = .17$, cet effet étant significativement plus fort dans la SMT-Maghrébin/Blanc que dans la SMT-Noir/Blanc, $t(199) = -2.00$, $p = .047$, $\eta^2_p = .02$.

[22] Un participant a été retiré des analyses, car il avait un RSS > 4.

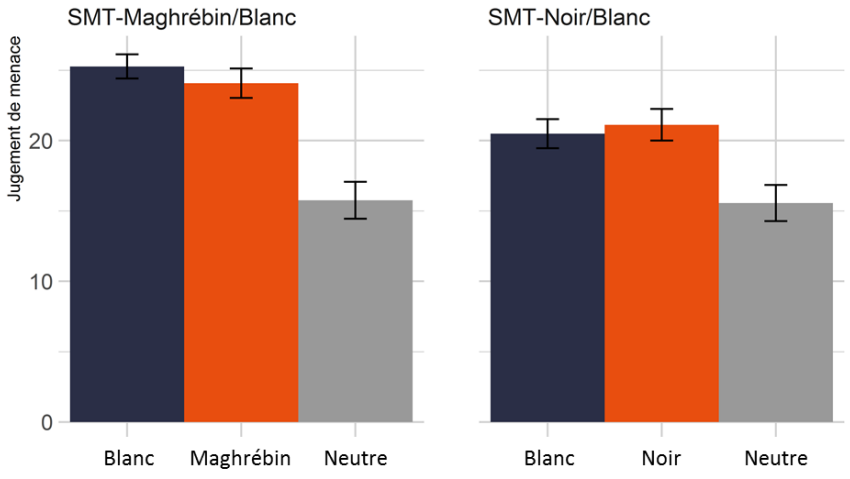


Figure 32. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin/Noir, Neutre) et de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc). Les barres d'erreurs représentent les erreurs standards.

Modèles multinomiaux. Concernant le modèle SAC-dominant, l'ajustement (T_1) du modèle aux données pour la SMT-Noir/Blanc (M_{pred} = 0.05, M_{obs} = 0.12, p =.07, w = 0.03) et la SMT-Maghrébin/Blanc (M_{pred} = 0.05, M_{obs} = 0.26, p < .001, w = 0.04) était acceptable. De la même manière, l'ajustement (T_1) du modèle aux données du modèle D-dominant quant à la SMT-Noir/Blanc (M_{pred} = 0.05, M_{obs} = 0.20, p =.002, w = 0.04), et la SMT-Maghrébin/Blanc (M_{pred} = 0.05, M_{obs} = 0.43, p < .001, w = 0.05) était là encore bon.

Pour rappel, nous faisions une nouvelle fois l'hypothèse d'un meilleur ajustement du modèle D-dominant pour nos participants ayant passé la SMT-Maghrébin/Blanc. De manière inverse à cette hypothèse, nous avons observé un meilleur ajustement aux données du modèle SAC-dominant ($WAIC_{SAC}$ = 3285.94) que du modèle D-dominant ($WAIC_{D}$ = 3383.64). Nous faisions également l'hypothèse d'un meilleur ajustement du modèle SAC-dominant pour nos participants ayant passé la SMT-Noir/Blanc. Effectivement, nous avons observé un meilleur ajustement aux données du modèle SAC-dominant ($WAIC_{SAC}$ = 3203.93) que du modèle D-dominant ($WAIC_{D}$ = 3308.58). En résumé, nous n'avons pas répliqué les résultats de notre Étude 5. Il semble ici que les réponses de nos participants passant la SMT-Noir/Blanc comme ceux passant la SMT-Maghrébin/Blanc soient mieux expliqués par le modèle SAC-

dominant. Par conséquent, pour la suite de nos analyses, nous avons traité nos données à l'aide du modèle SAC-dominant uniquement.

Pour rappel et au niveau du jugement de menace, nous n'avons pas répliqué l'effet SMT pour la SMT-Maghrébin/Blanc, mais aussi pour la SMT-Noir/Blanc. Nous pouvons observer les valeurs des paramètres afin de mieux comprendre ces résultats. De manière surprenante, le paramètre SAC était significativement plus important (et proche de 1) dans la SMT-Maghrébin/Blanc (M_{SAC} = .98, $IC_{95\%}$[.92 ; 1]) que dans la SMT-Noir/Blanc (M_{SAC} = .87, $IC_{95\%}$[.72 ; .97]), ΔSAC = .11, $IC_{95\%}$[0.01 ; .27], p_B = .03 (Figure 33). En d'autres termes, les participants passant la SMT-Maghrébin/Blanc ont plus activé leur stéréotype « Maghrébins = Menace » / « Blancs = Non menace » que les participants passant la SMT-Noir/Blanc n'ont activé leur stéréotype « Noirs = Menace » / « Blancs = Non menace ». Concernant le paramètre SAP, nous n'avons pas observé de différence significative entre la SMT-Maghrébin/Blanc (M_{SAP} = .49, $IC_{95\%}$[.43 ; .55]) et la SMT-Noir/Blanc (M_{SAP} = .51, $IC_{95\%}$[.43 ; .58]), ΔSAP = -.01, $IC_{95\%}$[-.11 ; .08], p_B = .59. Enfin, nous n'avons pas observé de différences pour les paramètres D et G entre la SMT-Maghrébin/Blanc (M_D = .03, $IC_{95\%}$[.01 ; .06] ; M_G = .23, $IC_{95\%}$[.15 ; .32]) et la SMT-Noir/Blanc (M_D = .04, $IC_{95\%}$[.01 ; .07] ; M_G = .23, $IC_{95\%}$[.17 ; .30]), les différences entre ces deux conditions étant non significatives pour le paramètre D, ΔD = -.00, $IC_{95\%}$[-.04 ; .04], p_B = .52, comme pour le paramètre G, ΔG = -.00, $IC_{95\%}$[-.11 ; .11], p_B = .53.

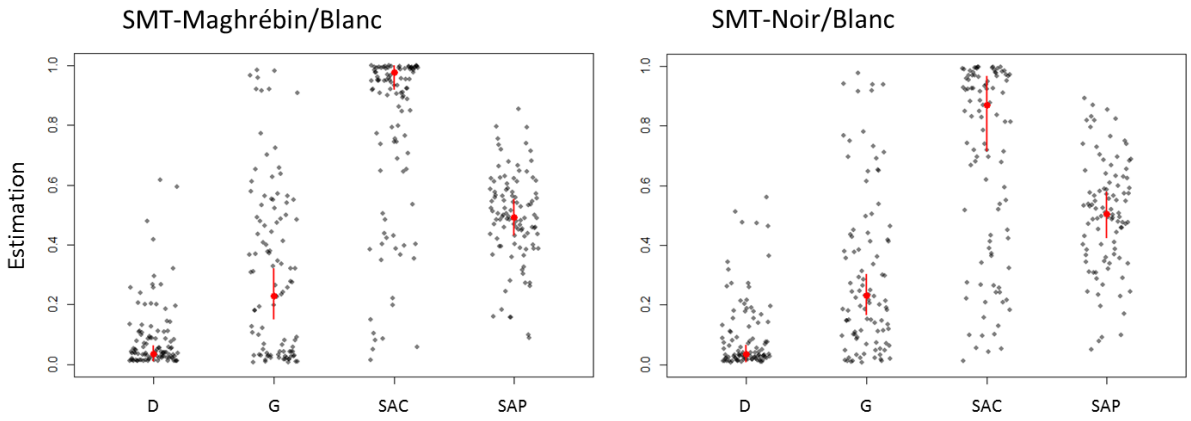


Figure 33. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc) et pour un modèle SAC-dominant à quatre paramètres.

De manière similaire à nos études précédentes, nous pouvons aller encore plus dans le détail et analyser ces résultats avec un modèle SAC-dominant à six paramètres. Notez tout d'abord que l'ajustement (T_1) du modèle aux données pour la SMT-Noir/Blanc (M_{pred} = 0.06, M_{obs} = 0.05, p = .37, w = 0.02) et la SMT-Maghrébin/Blanc (M_{pred} = 0.05, M_{obs} = 0.04, p = .53, w = 0.02) étaient acceptables. Intéressons-nous aux différences entre les paramètres SAC et SAP associés aux visages maghrébins et aux visages noirs. Tout d'abord, et de manière surprenante, nous avons obtenu un paramètre SAC descriptivement plus élevé pour les visages maghrébins (M_{SAC} = .85, $IC_{95\%}$[.66 ; .99]) que pour les visages noirs (M_{SAC} = .68, $IC_{95\%}$[.47 ; 93]), cette différence n'étant toutefois pas significative, ΔSAC = .17, $IC_{95\%}$[-.14 ; .43], p_B = .13 (Figure 34). À l'inverse, nous avons obtenu un paramètre SAP lié aux visages maghrébins (M_{SAP} = .62, $IC_{95\%}$[.49 ; .78]) descriptivement moins élevé que le paramètre SAP lié aux visages noirs (M_{SAP} = .69, $IC_{95\%}$[.49 ; .87]), cette différence n'étant une fois encore pas significative, ΔSAP = -.07, $IC_{95\%}$[-.28 ; .18], p_B = .72. Concernant les visages blancs, nous avons obtenu un paramètre SAC descriptivement moins important pour la SMT-Maghrébin/Blanc (M_{SAC} = .82, $IC_{95\%}$[.56 ; .98]) que pour la SMT-Noir/Blanc (M_{SAC} = .88, $IC_{95\%}$[.65 ; 1]), cette différence n'étant pas significative, ΔSAC = -.06, $IC_{95\%}$[-.37 ; .25], p_B = .66. Enfin, nous avons obtenu un paramètre SAP lié aux visages blancs moins important dans la SMT-Maghrébin/Blanc (M_{SAP} = .42, $IC_{95\%}$[.28 ; .50]) que dans la SMT-Noir/Blanc (M_{SAP} = .54, $IC_{95\%}$[.40 ; .65]), ΔSAP = -.12, $IC_{95\%}$[-.28 ; .05], p_B = .93, cette différence étant une fois encore non significative.

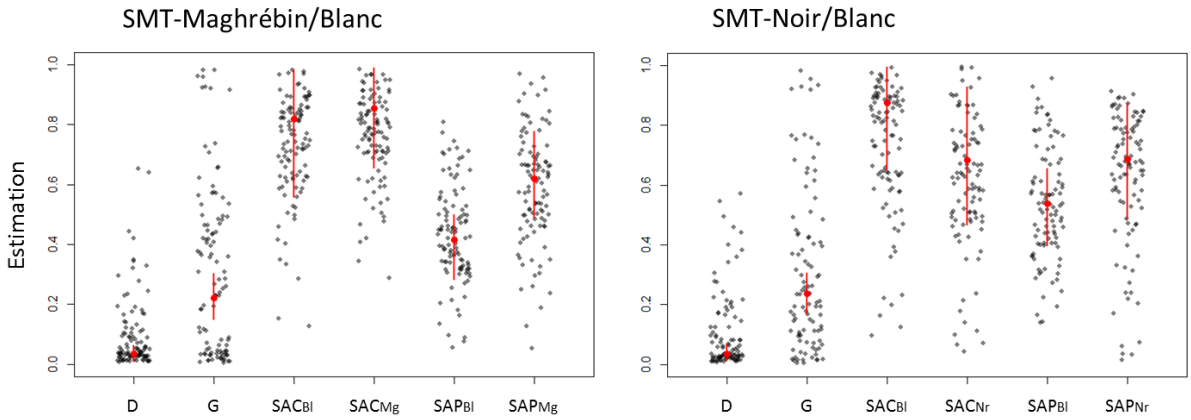


Figure 34. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc) et pour un modèle SAC-dominant à six paramètres. SAC_{Bl} et SAP_{Bl} = paramètres pour les visages blancs, SAC_{Mg} et SAP_{Mg} = paramètres pour les visages maghrébins, SAC_{Nr} et SAP_{Nr} = paramètres pour les visages noirs.

Analyses exploratoires. Nous l'avons spécifié plus haut, pour cette étude nous n'avons pas été en mesure de recruter majoritairement des participants se catégorisant comme d'ethnie blanche/Européo-américaine. Ainsi, les résultats décris ci-dessus ne reflètent pas uniquement les résultats d'individus du groupe de plus haut statut social (i.e., individus de couleur blanche). C'est la raison pour laquelle, de manière exploratoire, nous avons voulu séparer ces résultats en fonction des différentes origines ethniques des participants. Nous mettons en garde le lecteur quant au fait que la puissance statistique de ces résultats est particulièrement faible (e.g., 39 participants blancs pour un plan inter-sujets). Toutefois, il nous semble intéressant de décrire ces résultats, notamment en vue de l'Étude 10. Pour limiter au maximum la sous-puissance de nos estimations, nous ne présenterons que le modèle multinomial à quatre paramètres (i.e., le modèle SAC-dominant classique), afin de maximiser le nombre d'essais par paramètres SAC et SAP et ainsi la précision quant à leur estimation. De plus, afin de simplifier la lecture des résultats, nous ne présenterons que les effets nous semblant pertinents.

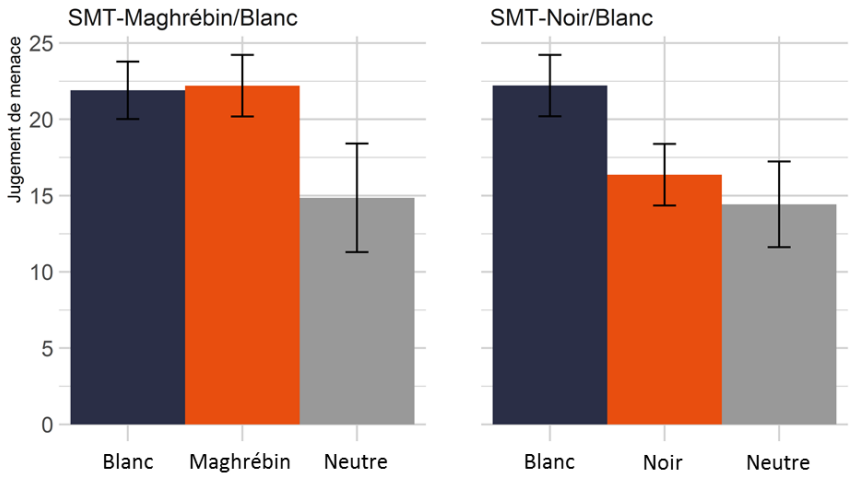


Figure 35. Moyenne des réponses « Plus menaçant » des participants blancs en fonction du type d'amorce (Blanc, Maghrébin/Noir, Neutre) et de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc). Les barres d'erreurs représentent les erreurs standards.

Commençons tout d'abord avec les participants se catégorisant comme d'ethnie blanche (39 participants). Au niveau du jugement de menace, et de manière surprenante, nous pouvons voir que ces participants avaient un effet SMT inversé pour la SMT-Noir/Blanc, $t(37) = -2.09$, $p = .04$, $\eta^2_p = .11$ (Figure 35). Autrement dit, ces participants ont davantage jugé les dessins de visages comme menaçants quand ils étaient précédés de visages blancs ($M = 22.21$, $ET = 16.37$) plutôt que de visages noirs ($M = 16.37$, $ET = 8.81$). L'effet SMT pour la SMT-Maghrébin/Blanc n'était pour sa part pas significatif, $t(37) = 0.11$, $p = .91$, $\eta^2_p < .001$. Au niveau du modèle multinomial associé, le paramètre SAC était descriptivement plus important dans la SMT-Maghrébin/Blanc, ($M_{SAC} = .87$, $IC_{95\%}$[.46 ; 1]) que dans la SMT-Noir/Blanc ($M_{SAC} = .74$, $IC_{95\%}$[.25 ; .98]), cette différence n'étant pas significative, $\Delta SAC = .14$, $IC_{95\%}$[-.33 ; .65], $p_B = .27$. Concernant le paramètre SAP, celui-ci était descriptivement plus important dans la SMT-Maghrébin/Blanc, ($M_{SAP} = .50$, $IC_{95\%}$[.39 ; .60]) que dans la SMT-Noir/Blanc ($M_{SAP} = .36$, $IC_{95\%}$[.22 ; .50]), cette différence n'étant pas significative, $\Delta SAP = .14$, $IC_{95\%}$[-.04 ; .31], $p_B = .06$. Plus intéressant pour nous, le paramètre SAP lié à la SMT-Noir/Blanc était significativement inférieur à .50, $\Delta SAP_{50} = .14$, $IC_{95\%}$[.09 ; .28], $p_B = .03$. En d'autres termes, face à des visages noirs et blancs, les participants de couleur blanche ont davantage corrigé leur stéréotype qu'ils ne l'ont appliqué. Nous pouvons donc penser, au final, que l'effet SMT pro-Noirs

observé chez nos participants blancs peut être le reflet d'une inhibition (correction) du stéréotype réactivé « Noirs = Menace » / « Blancs = Non menace ».

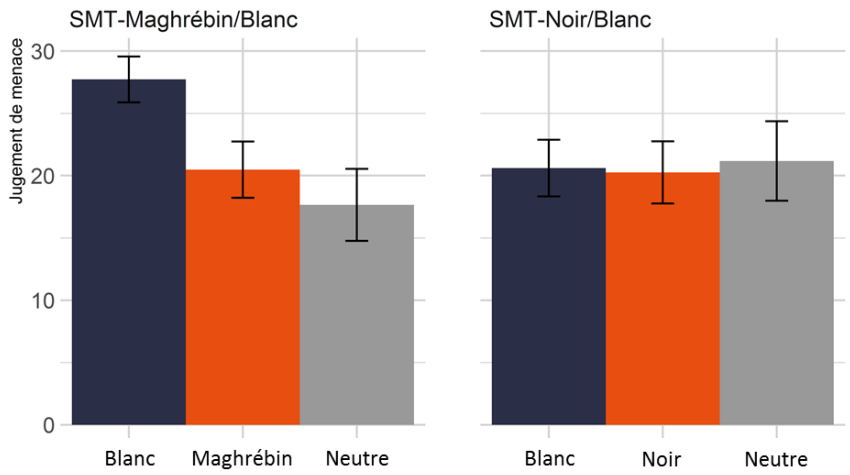


Figure 36. Moyenne des réponses « Plus menaçant » des participants hispaniques en fonction du type d'amorce (Blanc, Maghrébin/Noir, Neutre) et de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc). Les barres d'erreurs représentent les erreurs standards.

Pour les participants d'origine hispanique (51 participants) et de manière là encore surprenante, nous avons observé un effet SMT inversé pour la SMT-Maghrébin/Blanc, $t(49) = -2.32$, $p = .02$, $\eta^2_p = .10$ (Figure 36). En d'autres termes, ces participants ont davantage jugé les dessins de visages comme menaçants quand ils étaient précédés de visages blancs ($M = 27.72$, $ET = 9.92$) plutôt que de visages maghrébins ($M = 20.48$, $ET = 12.16$). L'effet SMT pour la SMT-Noir/Blanc n'était pour sa part pas significatif, $t(49) = -0.10$, $p = .92$, $\eta^2_p < .001$. Concernant les paramètres du modèle, il n'y avait pas de différence entre le paramètre SAC de la SMT-Maghrébin/Blanc ($M_{SAC} = .97$, $IC_{95\%}[.86 ; 1]$) et celui de la SMT-Noir/Blanc ($M_{SAC} = .95$, $IC_{95\%}[.71 ; .1]$), $\Delta SAC = .02$, $IC_{95\%}[-.11 ; .28]$, $p_B = .40$. Le paramètre SAP était une fois encore intéressant. Celui-ci était descriptivement moins important dans la SMT-Maghrébin/Blanc ($M_{SAP} = .39$, $IC_{95\%}[.26 ; .51]$) que dans la SMT-Noir/Blanc ($M_{SAP} = .51$, $IC_{95\%}[.40 ; .63]$), la différence n'étant pas significative, $\Delta SAP = -.12$, $IC_{95\%}[-.29 ; .05]$, $p_B = .92$. Plus intéressant pour nous, le paramètre SAP lié à la SMT-Maghrébin/Blanc était une fois encore significativement inférieur à .50, $\Delta SAP_{50} = .11$, $IC_{95\%}[.07 ; .24]$, $p_B = .04$. En d'autres

termes, face à des visages de couleur blanche et d'origine maghrébine, les participants d'origine hispanique ont davantage corrigé leur stéréotype qu'ils ne l'ont appliqué. Nous pouvons donc penser, au final, que l'effet SMT pro-Maghrébins observé chez nos participants hispaniques peut être le reflet d'une inhibition (correction) du stéréotype réactivé « Maghrébins = Menace » / « Blancs = Non menace ».

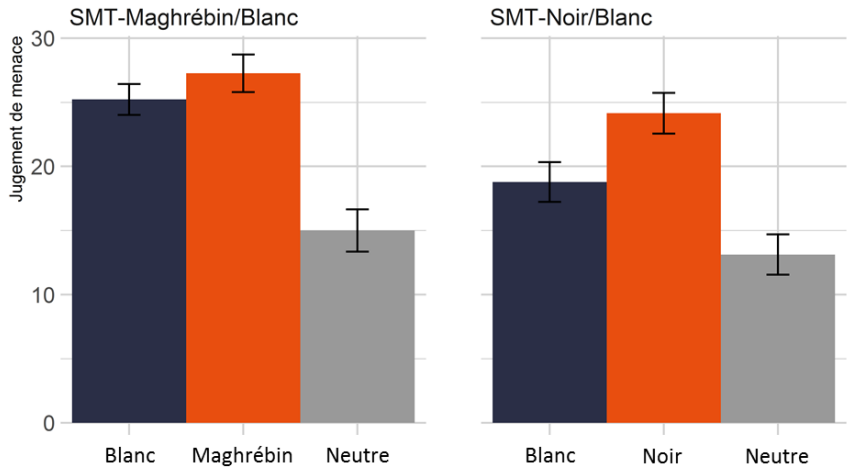


Figure 37. Moyenne des réponses « Plus menaçant » des participants asiatiques en fonction du type d'amorce (Blanc, Maghrébin/Noir, Neutre) et de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc). Les barres d'erreurs représentent les erreurs standards.

Enfin, pour les participants d'origine asiatique (95 participants), nous avons observé l'effet SMT traditionnel pour la SMT-Noir/Blanc, $t(93) = 2.66$, $p = .01$, $\eta^2_p = .07$ (Figure 37). Autrement dit, ces participants ont davantage jugé les dessins de visages comme menaçants quand ils étaient précédés de visages noirs ($M = 24.14$, $ET = 11.14$) plutôt que de visages blancs ($M = 18.78$, $ET = 10.11$). L'effet SMT pour la SMT-Maghrébin/Blanc n'était pour sa part pas significatif, $t(93) = 0.98$, $p = .33$, $\eta^2_p = .01$. Au niveau du modèle multinomial, le paramètre SAC était descriptivement plus important dans la SMT-Maghrébin/Blanc ($M_{SAC} = .86$, $IC_{95\%}$[.47 ; 1]) que dans la SMT-Noir/Blanc ($M_{SAC} = .70$, $IC_{95\%}$[.46 ; .90]), la différence n'étant pas significative, $\Delta SAC = .16$, $IC_{95\%}$[-.28 ; .45], $p_B = .16$. Le paramètre SAP était descriptivement légèrement moins important dans la SMT-Maghrébin/Blanc ($M_{SAP} = .59$, $IC_{95\%}$[.49 ; .68]) que dans la SMT-Noir/Blanc ($M_{SAP} = .62$, $IC_{95\%}$[.46 ; .90]), la différence n'étant pas significative,

ΔSAP = -.03, $IC_{95\%}$[-.19 ; .13], p_B = .65. Plus intéressant pour nous, le paramètre SAP était significativement supérieur à .50 à la fois pour la SMT-Maghrébin/Blanc, ΔSAP_{50} = .09, $IC_{95\%}$[.08 ; .18], p_B = .03, et pour la SMT-Noir/Blanc, ΔSAP_{50} = .12, $IC_{95\%}$[.08 ; .24], p_B = .03. Autrement dit, les participants d'origine asiatique ont davantage appliqué leur stéréotype « Maghrébins = Menace » / « Blancs = Non menace » pour la SMT-Maghrébin/Blanc et leur stéréotype « Noirs = Menace » / « Blancs = Non menace » pour la SMT-Noir/Blanc, expliquant en partie les effets SMT obtenus pour ces participants (i.e., l'effet SMT significatif pour la SMT-Noir/Blanc).

4.5.3 **Discussion**

La neuvième étude de ce manuscrit était particulière dans le sens où nous n'avons pas effectué cette étude au sein de l'Université Grenoble Alpes, comme nos études précédentes, mais à l'Université de Davis, en Californie, dans le laboratoire même où la SMT a été créée. Nous voulions tester notre hypothèse selon laquelle il existerait une différence au niveau du groupe ethnique principalement associé au stéréotype de menace en France et aux États-Unis. Dans l'Étude 5, nous faisions l'hypothèse qu'en France, les populations maghrébines seraient plus associées au stéréotype de menace que les populations noires. En accord avec cette hypothèse, nous avions obtenu un effet SMT plus important chez nos participants passant la tâche SMT-Maghrébin/Blanc que chez nos participants passant la tâche SMT-Noir/Blanc. Dans cette présente étude, nous faisions l'hypothèse inverse. Nous pensions en effet qu'aux États-Unis, les populations noires seraient plus associées au stéréotype de menace que les populations maghrébines. Contrairement à nos hypothèses, nous n'avons pas observé un effet SMT significativement plus important chez nos participants passant la tâche SMT-Noir/Blanc que chez nos participants passant la tâche SMT-Maghrébin/Blanc. Toutefois, ces résultats restent, sur de nombreux points, intéressants à développer.

Pour la première fois en huit études, nous n'avons pas répliqué l'effet SMT avec des visages maghrébins et des visages blancs. Bien que nous ne puissions pas conclure à une absence d'effet sur la base d'un résultat non significatif, nous pouvons émettre quelques hypothèses pouvant expliquer que nous n'ayons pas répliqué cet effet. Tout d'abord, nous pouvons naturellement revenir sur notre raisonnement théorique et aborder l'aspect culturel du stéréotype. En effet, comme nous l'avons décrit plus haut, des individus venant de zones

géographiques différentes devraient être exposés à des influences culturelles différentes (e.g., Payne et al., 2017 ; Rougier et al., 2020). Nous pensions que le stéréotype liant les Maghrébins à une menace serait avant tout un stéréotype européen (et dans notre cas, français, voire grenoblois). C'est la raison pour laquelle, de prime abord, nous pourrions penser que ces résultats aillent dans le sens de nos hypothèses.

Toutefois, nos hypothèses ne concernaient pas réellement une absence du stéréotype liant les individus maghrébins à une menace, mais un stéréotype moins important que celui liant les individus noirs à une menace. En effet, comme nous l'avons montré au Chapitre 1, plusieurs études ont mis en évidence, aux États-Unis, la présence d'attitudes raciales négatives envers les individus arabo-musulmans (Ahluwalia & Pellettiere, 2010 ; Dasgupta et al., 2009 ; Lyons et al., 2010 ; Mange et al., 2012 ; Park et al., 2007). Dans notre étude, l'activation du stéréotype de nos participants passant la SMT-Maghrébin/Blanc était particulièrement élevée (proche de 1), signe d'une association « Maghrébins = Menace » bien présente. Ainsi, pour trouver d'autres explications à cette absence d'effet significatif, il nous faut parler des résultats de l'autre condition expérimentale, à savoir l'effet SMT de nos participants ayant passé la SMT-Noir/Blanc.

De manière surprenante, nous n'avons pas répliqué l'effet SMT classique pour nos participants passant la SMT-Noir/Blanc. Ceci est d'autant plus surprenant que nous avons effectué notre étude dans le laboratoire même où cet effet est habituellement observé (Krieglmeyer & Sherman, 2012 ; Rees, Ma, et al., 2019 ; Rees, Rivers, et al., 2019 ; Rivers et al., 2020). En outre, l'attitude raciale négative envers les populations noires aux États-Unis n'est plus à démontrer, la majorité des études dans le champ des attitudes raciales ayant été effectuées sur des stimuli d'individus de couleur noire (e.g., Correll et al., 2002 ; Devine, 1989 ; Fazio et al., 1995 ; Greenwald et al., 1998 ; Nosek et al., 2002 ; Payne, 2001 ; Payne et al., 2005 ; Wittenbrink et al., 2001). Là encore, les participants prenant part à cette étude ont eu une valeur du paramètre SAC élevée, suggérant ainsi que l'activation du stéréotype a bien eu lieu. De plus, mais de manière descriptive uniquement, nous avons obtenu un paramètre SAP plus important pour les visages noirs que pour les visages maghrébins. Plus visible encore, nous avons obtenu un paramètre SAP pour les visages blancs plus important dans la SMT-Noir/Blanc que dans la SMT-Maghrébin/Blanc. En somme, et comme on pouvait s'y attendre, les participants ont descriptivement plus appliqué leur stéréotype « Noirs = Menace » et « Blancs = Non menace » dans la SMT-Noir/Blanc, sans toutefois que cela ne se reflète au

niveau des effets « bruts ». Comme nous l'avons vu, l'origine ethnique de nos participants peut peut-être nous aider à comprendre pourquoi.

Pour la première fois dans nos études, nous n'avons pas recruté majoritairement des participants se catégorisant comme d'origine ethnique européenne (individus de couleur blanche). Au contraire, l'Université de Davis étant un campus à forte composante multiculturelle, la majorité de nos participants était d'origine asiatique (95 participants), d'origine hispanique (52 participants) et enfin de couleur blanche (39 participants). Il est à noter cependant que Rivers et collaborateurs (2020) ont répliqué l'effet SMT classique avec un échantillon équivalent en termes de représentativité (58% de participants asiatiques, 21% de participants blancs et 20% de participants hispaniques). La simple présence de participants d'origine multiculturelle ne peut donc pas expliquer cette absence d'effet significatif.

Malheureusement, nous ne pouvons savoir avec ces seules données si nos participants étaient de nationalité américaine et/ou depuis un certain temps aux États-Unis (et donc sensibles à la culture américaine) ou bien s'ils étaient uniquement aux États-Unis le temps de leurs études (et donc sensibles à leur culture d'origine). Le paramètre SAC élevé chez l'ensemble de nos participants peut toutefois suggérer une connaissance du stéréotype relativement similaire chez les participants blancs et les participants d'origine asiatique ou hispanique. De manière similaire aux travaux de Payne et collaborateurs (2017), ces résultats pourraient suggérer que ces deux stéréotypes sont bien présents aux États-Unis et que l'ensemble de nos participants ont été exposés à cette influence culturelle.

Malgré leur faible puissance statistique, le pattern de résultats observé en fonction de l'origine ethnique de nos participants est particulièrement intéressant. Les participants blancs, nos participants a priori d'intérêts, ont de façon très surprenante exprimé un effet SMT inverse à l'effet traditionnel. En d'autres termes, ces participants ont jugé plus souvent les dessins de visages comme étant menaçants quand ceux-ci étaient précédés de visages noirs plutôt que de visages blancs. Cet effet semble être le fruit d'une application du stéréotype particulièrement faible : les participants ont davantage contrôlé (i.e., inhibé) leur stéréotype activé qu'ils ne l'ont appliqué. Retrouver ce résultat au sein du campus de Davis n'est finalement pas si surprenant. En effet, ce campus est connu pour être un campus multiculturel où les valeurs de tolérances sont particulièrement présentes. Le mouvement Black Live Maters était déjà à cette période l'un des mouvements les plus populaires, de nombreuses affiches promouvant ce mouvement étant affichées au sein du campus, et dans le département de psychologie en particulier. Ainsi, il est possible que les individus de couleur

blanche aient volontairement inhibé leur stéréotype activé pour ne pas paraitre comme allant à l'encontre de ces mouvements (e.g., biais de désirabilité sociale) ou parce qu'ils adhéraient eux-mêmes à ces valeurs.

De la même manière, les participants d'origine hispanique ont eu un effet SMT inversé au sein de la SMT-Maghrébin/Blanc, c'est-à-dire qu'ils ont jugé les cibles comme plus menaçantes quand celles-ci étaient précédées de visages blancs plutôt que de visages maghrébins. Là encore, cet effet semble être le fruit d'une application du stéréotype activé particulièrement faible, les participants contrôlant leur stéréotype plus qu'ils ne l'appliquent. Nous pouvons peut-être expliquer ce résultat par un biais pro-endogroupe classique au sein du groupe de bas statut social. En effet, il est envisageable que les visages maghrébins constituant notre tâche SMT aient été perçus par les individus hispaniques comme faisant partie de leur groupe social, certaines études ayant montré qu'un même visage ayant la peau mate pouvait être assimilé à différents groupes ethniques et notamment à des individus maghrébins ou espagnols (e.g., Mange et al., 2016).

Au final, seuls les participants d'origine asiatique ont exprimé l'effet SMT anti-Noirs traditionnel (et descriptivement, l'effet SMT anti-Maghrébins). Pour les deux types de SMT, ces participants ont plus appliqué leur stéréotype qu'ils ne l'ont contrôlé. Nous pouvons mettre en opposition ce résultat avec celui de nos participants de couleur blanche. Là où nos participants de couleur blanche ont en majorité contrôlé leur stéréotype pour correspondre à la mouvance actuelle plus égalitaire, les participants asiatiques, issus eux-mêmes d'une minorité ethnique, n'ont peut-être pas été soumis à la même pression pour contrôler leur stéréotype et ont donc eu tendance à appliquer celui-ci.

Bien sûr, l'ensemble de ces raisonnements ne sont que post-hoc et méritent d'être approfondis. Il semblerait tout de même que l'effet SMT traditionnellement observé ne soit pas uniquement la résultante d'une connaissance du stéréotype culturelle, mais également d'un degré d'application du stéréotype pouvant être différent (au moins descriptivement) en fonction des groupes sociaux passant cette étude. Nous pouvons ainsi nous questionner quant à l'effet SMT que nous pouvons attendre chez le groupe social associé à ce stéréotype en France (i.e., individus d'origine maghrébine) : peut-on s'attendre à une activation du stéréotype similaire chez les participants d'origine maghrébine, comme nous venons de le voir ici ? De plus, est-ce que ces participants inhiberont leur stéréotype, comme les participants hispaniques ou blancs de cette étude, ou l'appliqueront-ils et exprimeront un biais SMT classique ? La dernière étude de ce chapitre avait pour rôle de répondre à ces questions.

4.6 Étude 10 : analyse intégrative des données de participants de culture maghrébine

Cette dernière étude était pour nous l'occasion d'analyser les données de nos précédentes études en nous intéressant non plus aux participants de culture principale française, mais aux participants de culture principale maghrébine. Nos précédentes études nous ont montré que les participants de culture française exprimaient un effet SMT pro-endogroupe : les cibles étaient jugées comme plus menaçantes quand elles étaient précédées de visages maghrébins plutôt que des visages blancs. Toutefois, notamment dans nos Études 3, 5, 6 (et avec une autre mesure, l'Étude 8) nous avons vu que l'effet SMT était modéré par une mesure directe d'attitudes raciales. Plus les participants avaient une attitude raciale auto-rapporté négative envers les Maghrébins et plus ils appliquaient leur stéréotype (SAP élevé) et donc plus l'effet SMT était important. Il semblerait donc que la SMT soit, au moins à un certain niveau, sensible à l'attitude personnelle du participant. Dans une précédente étude (Rougier et al., 2020) nous avons montré, au sein de l'Université Grenoble Alpes, que des participants d'origine maghrébine passant une autre tâche indirecte (i.e., tâche d'approche-évitement, Rougier et al., 2018) avaient un biais pro-endogroupe. Notre interprétation de ces résultats était que les participants maghrébins étaient allés au-delà du leurs croyances culturelles classiques et avaient exprimé leur attitude personnelle. C'est la raison pour laquelle, suivant en cela nos précédents résultats (Rougier et al., 2020), nous faisions l'hypothèse d'un effet pro-endogroupe pour nos participants de culture maghrébine passant la SMT (i.e., effet SMT inverse) : ces participants devaient juger les cibles comme plus menaçantes quand celles-ci étaient précédées de visages blancs en comparaison à des visages maghrébins. Plus précisément et suivant en cela certains auteurs (e.g., Payne et al., 2017), nous faisions l'hypothèse d'une activation non nulle du stéréotype « Maghrébins = Menace » / « Blancs = Non menace » (i.e., SAC > 0), signe de l'influence du contexte culturel français. Toutefois, en accord avec nos précédents résultats (Rougier et al., 2020) et ceux de l'Étude 9, nous faisions l'hypothèse d'une inhibition de ce stéréotype activé (SAP < .50), signe de la sensibilité de la SMT à une attitude personnelle .

4.6.1 Méthode et résultats

Participants. Dans cette dernière étude, nous avons regroupé l'ensemble des participants de culture principale maghrébine ayant passé l'une des sept études réalisées à l'Université Grenoble Alpes. Au total, 74 participants de culture principale maghrébine ont pris part à ces études (60 femmes, 14 hommes, $M_{\text{âge}}$ = 21.52, $ET_{\text{âge}}$ = 3.02). Les participants étaient des étudiants de l'Université Grenoble Alpes et recevaient des crédits de cours ou 10 euros en compensation.

Méthode d'analyse des données. Pour analyser l'ensemble des données, nous avons entrepris une démarche d'analyses de données intégratives (Curran & Hussong, 2009). Ainsi, nous avons regroupé en un unique jeu de données l'ensemble des données des participants de culture maghrébine ayant pris part à nos études. De plus, nous n'avons gardé que les conditions expérimentales correspondant au paradigme SMT-Maghrébin/Blanc habituel. En d'autres termes, dans notre Étude 5 nous avons exclu des analyses les participants ayant passé la SMT-Noir/Blanc. De même, nous avons exclu des analyses les données des Études 6 et 7 relatives aux amorces non prototypiques des catégories Maghrébins et Blancs. Enfin, pour l'Étude 8, nous avons exclu des analyses les données correspondant aux amorces de type « prénoms ». En définitive, nous avons traité nos données à l'aide d'une ANOVA à plan mixte de type 3 (amorce : Maghrébin vs. Blanc vs. Neutre) x 2 (cible : Menaçant vs. Non menaçant) x 7 (Étude SMT : Étude 1 vs. Étude 2 vs. Étude 3 vs. Étude 5 vs. Étude 6 vs. Étude 7 vs. Étude 8), le dernier facteur étant en inter-sujet. Le pré-enregistrement de cette étude est disponible sur la plateforme OSF (https://osf.io/827da/?view_only=e0030fb02b7948bd87b64bc494aadf31).

Jugement de menace. Tout d'abord, nous avons obtenu un effet du type de cible, les participants jugeant les cibles menaçantes comme étant plus menaçantes (M = 30.95, ET = 11.19) que les cibles non menaçantes (M = 24.76, ET =11.64), $t(67) = 4.64$, $p < .001$, $\eta^2_p = .24$. À l'inverse de notre hypothèse, les participants ont jugé les dessins cibles comme plus menaçants quand ils étaient précédés de visages maghrébins (M = 24.01, ET = 11.64) plutôt que de visages blancs (M = 17.02, ET = 7.77), répliquant ainsi l'effet SMT classique, $t(67) = 4.36$, $p < .001$, $\eta^2_p = .22$ (Figure 38). De plus, les participants ont également jugé les

dessins cibles comme plus menaçants quand ils étaient précédés de visages maghrébins (M = 24.01, ET = 11.64) plutôt que de visages neutres (M = 14.66, ET = 10.18), $t(67) = 5.00$, $p < .001$, $\eta^2_p = .27$. Enfin, les participants n'ont pas jugé les dessins cibles comme significativement plus menaçants quand ils étaient précédées de visages blancs (M = 17.02, ET = 7.77) plutôt que de visages neutres (M = 14.66, ET = 10.18), $t(67) = 2.00$, $p = .051$, $\eta^2_p = .06$.

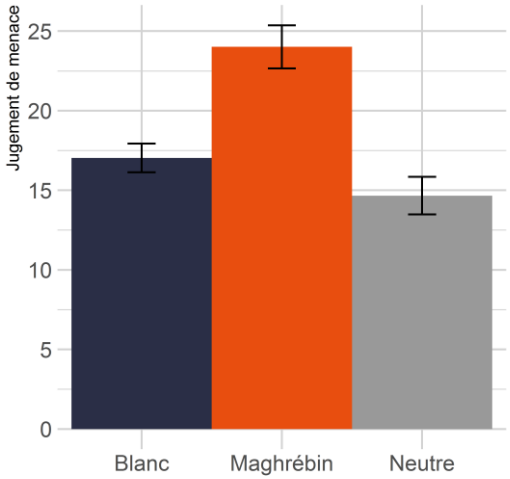


Figure 38. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin, Neutre). Les barres d'erreurs représentent les erreurs standards.

Modèle multinomial. Nous avons analysé ces données à la fois avec le modèle SAC-dominant classique à quatre paramètres (SAC, SAP, D, G) et avec le modèle SAC-dominant à six paramètres (SAC_{Mg}, SAC_{Bl}, SAP_{Mg}, SAP_{Bl}, D, G). Concernant l'ajustement du modèle classique, la différence entre les valeurs T_1 prédites (M_{pred} = 0.07) et observées (M_{obs} = 0.10) était non significative, $p = .30$, la taille d'effet associée étant acceptable, $w = .02$. Conformément à notre hypothèse, le paramètre SAC était supérieur à 0 (M_{SAC} = .67 ; $IC_{95\%}$[.39 ; .90]), ΔSAC_0 = .67, $IC_{95\%}$[.39 ; .90], $p_B < .001$, suggérant que nos participants ont effectivement activé le stéréotype « Maghrébins = Menace » / « Blancs = Non menace » (Figure 39). Toutefois, à l'inverse de notre hypothèse, le paramètre SAP était supérieur à .50 (M_{SAP} = .61 ; $IC_{95\%}$[.52 ; .69]), ΔSAP_{50} = .11, $IC_{95\%}$[.02 ; .19], p_B = .01, indiquant que nos participants ont plus appliqué ce stéréotype qu'ils ne l'ont contrôlé. Il est à noter que la valeur

du paramètre D était relativement basse (M_D = .04 ; $IC_{95\%}$[.01 ; .08]), mais supérieure à 0, ΔD_0 = .04, $IC_{95\%}$[.01 ; .08], p_B < .001. Les participants ont donc bien détecté la menace de la cible.

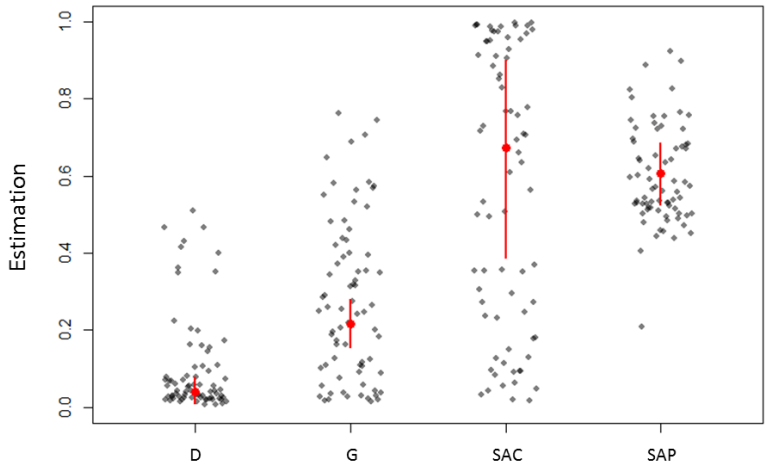


Figure 39. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) pour un modèle SAC-dominant à quatre paramètres.

L'ajustement (T_1) du modèle à six paramètres était lui aussi bon (M_{pred} = 0.07, M_{obs} = 0.05, p = .61, w = .02). Le paramètre SAC était supérieur à 0 pour les visages blancs (M_{SAC} = .47 ; $IC_{95\%}$[.21 ; .93]), ΔSAC_0 = .47, $IC_{95\%}$[.21 ; .93], p_B < .001, comme pour les visages maghrébins (M_{SAC} = .59 ; $IC_{95\%}$[.31 ; .92]), ΔSAC_0 = .59, $IC_{95\%}$[.31 ; .92], p_B < .001 (Figure 40). Ainsi, les participants activaient à la fois leur stéréotype « Maghrébins = Menace » et leur stéréotype « Blancs = Non menace ». Concernant le paramètre SAP, celui-ci était supérieur à .50 pour les visages maghrébins (M_{SAP} = .76 ; $IC_{95\%}$[.54 ; .97]), ΔSAP_{50} = .26, $IC_{95\%}$[.04 ; .47], p_B < .001, mais pas pour les visages blancs (M_{SAP} = .49 ; $IC_{95\%}$[.09 ; .69]), ΔSAP_{50} = -.02, $IC_{95\%}$[-.10 ; .19], p_B = .41. En d'autres termes, les participants appliquaient plus leur stéréotype « Maghrébins = Menace » qu'ils ne le contrôlaient, mais pas leur stéréotype « Blancs = Non menace ».

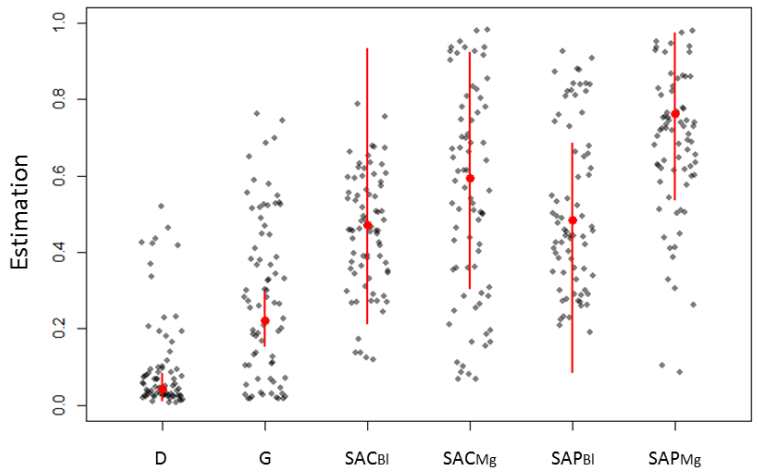


Figure 40. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) pour un modèle SAC-dominant à six paramètres. SAC_{Bl} et SAP_{Bl} = paramètres pour les visages blancs, SAC_{Mg} et SAP_{Mg} = paramètres pour les visages maghrébins.

4.6.2 **Discussion**

Cette dernière étude avait pour but d'observer si, au sein de notre paradigme SMT, des participants issus du groupe cible du stéréotype (i.e, participants de culture maghrébine) exprimaient un biais pro-exogroupe (effet SMT classique) ou un biais pro-endogroupe (effet SMT inversé). Nous faisions l'hypothèse, en accord avec l'une de nos précédentes études (Rougier et al., 2020), d'un effet SMT pro-endogroupe, celui-ci résultant notamment d'une plus forte inhibition du stéréotype activé. Contrairement à cette hypothèse, nous avons observé un effet SMT pro-exogroupe (anti-Maghrébins) de la part de nos participants se catégorisant de culture principale maghrébine.

Ces résultats sont dans la lignée des études IAT princeps ayant montré un favoritisme pro-exogroupe de la part de groupes de bas statuts (e.g., Ashburn-Nardo et al., 2003 ; Livingston, 2002 ; Nosek et al., 2002 ; Richeson et al., 2005 ; Rudman et al., 2002). Pour rappel, pour un grand nombre d'auteurs cela constitue l'une des preuves que les mesures captent effectivement les croyances culturelles dominantes au sein d'une société donnée

(Arkes & Tetlock, 2004 ; Banaji et al., 2004 ; Jost et al., 2004 ; Karpinski & Hilton, 2001 ; Rudman et al., 2002). Est-ce le cas pour notre mesure ? L'avantage d'utiliser le modèle multinomial associé à la SMT est que celui-ci nous permet d'aller plus loin dans la compréhension de nos effets. Effectivement, le paramètre d'activation du stéréotype (SAC) supérieur à 0 peut laisser penser que nos participants de culture maghrébine ont, comme nos participants de culture française, activé le stéréotype observé classiquement en France, signe peut-être d'une influence de l'environnement culturel français.

Comme nous l'avons évoqué plus haut, un point de débat concerne l'origine personnelle ou extra-personnelle de l'attitude captée au sein des mesures indirectes (Arkes & Tetlock, 2004 ; Banaji et al., 2004 ; Jost, 2019 ; Nosek & Hansen, 2008). Pour Arkes et Tetlock (2004), les résultats des mesures indirectes classiques ne seraient que le reflet d'une connaissance culturelle du stéréotype, mais ne correspondraient pas réellement à l'attitude personnelle de l'individu. Nos résultats ne semblent pas aller dans ce sens. Le paramètre SAC, comme nous l'avons montré précédemment, semble aller dans le sens de ces auteurs, c'est-à-dire peut illustrer une réactivation non intentionnelle du stéréotype culturelle. Mais qu'en est-il de l'application du stéréotype (SAP) ? Nous avons montré dans nos précédentes études (Études 3, 5 et 6) que le paramètre SAP était systématiquement corrélé à une mesure directe d'attitude raciale anti-Maghrébins. Nous pouvons donc penser que le paramètre SAP puisse être, dans une certaine mesure, sensible à l'attitude personnelle du participant. Or, dans cette présente étude, les participants de culture principale maghrébine ont davantage appliqué leur stéréotype qu'ils ne l'ont corrigé (SAP supérieur à .50). Il semblerait donc que nos participants aient en quelque sorte « validé » leur stéréotype activé. Comment expliquer l'application de ce stéréotype de la part de nos participants de culture maghrébine ?

Pour Jost et collaborateurs (2004), les préférences pro-exogroupes observées au sein des mesures indirectes seraient le reflet d'une intériorisation du stéréotype culturel de la part des individus issus de groupes de bas statut, participant ainsi à la justification du système en place. Plus particulièrement, ces individus accepteraient et ré-exprimeraient les stéréotypes négatifs associés à leur groupe afin de légitimer l'ordre social en place. Il pourra être intéressant, dans une prochaine étude, d'explorer cette théorie et de mettre en lien les effets observés dans la SMT (et notamment le paramètre SAP) avec une mesure reflétant l'adhésion au système en place (e.g., *Social Dominance Orientation*, SDO ; Pratto et al., 1994). Peut-on s'attendre à un paramètre SAP positivement corrélé à ce type d'échelle ? De plus, pour Jost et collaborateurs, cette préférence pro-exogroupe ne se retrouverait pas dans une mesure

directe d'attitude raciale, ceci afin de se conformer à la pression sociale favorisant une préférence pour son propre groupe social. Là encore, il pourrait être intéressant d'explorer, dans une prochaine étude, le lien entre les résultats de notre SMT et une mesure directe d'attitude raciale (e.g., *feeling thermometer*). Doit-on s'attendre à une corrélation négative entre le paramètre SAP, censé refléter une préférence intériorisée pour l'exogroupe et une mesure directe d'attitude raciale (e.g., *feeling thermometer*), censée refléter la préférence normative pour son propre groupe sociale ?

Essien et collaborateurs (2020) ont récemment testé sur un nombre important de participants issus de groupes sociaux de bas statut (e.g., individus noirs, individus arabo-musulmans), les modérateurs possibles au biais pro-exogroupe observé au sein de l'IAT[23]. Tout d'abord, il est intéressant de mentionner que ces auteurs ont globalement observé un biais pro-exogroupe (*dz* = 0.43) de la part des membres de groupes de bas statut. De plus, et en accord une nouvelle fois avec la théorie de justification du système, ce biais pro-exogroupe était modéré à la fois par le niveau de conservatisme du groupe, mais aussi et surtout par le niveau de stigmatisation du groupe, les groupes les plus stigmatisés étant, comme certains travaux l'ont déjà montré par le passé (e.g., Rudman et al., 2002), les groupes sociaux affichant le plus fort biais pro-exogroupe. Comme nous avons pu l'évoquer aux Chapitre 1 et 2, le groupe des Maghrébins est l'un des groupes sociaux les plus stigmatisés en France. Ce résultat pourrait donc être une nouvelle fois le reflet d'une intériorisation du biais négatif anti-Maghrébins observé en France.

Nous avons vu au Chapitre 1 qu'inhiber son stéréotype réactivé demandait entre autres des ressources motivationnelles (e.g., Devine, 1989 ; Fazio, 1990). De plus, dans notre étude précédente, (i.e., Étude 9), nous avons vu que certains participants (i.e., participants blancs et hispaniques) pouvaient inhiber dans une large mesure leur stéréotype réactivé. Dans le contexte des États-Unis où les tensions raciales sont importantes, cela peut s'expliquer, les participants pouvant être motivés de manière intrinsèque (i.e., valeurs propres) ou extrinsèque (i.e., biais de désirabilité sociale) à ne pas exprimer leur stéréotype. Cet aspect motivationnel est également compatible avec l'étude de Payne et collaborateurs (2019) ayant montré un

[23] Essien et collaborateurs se sont notamment intéressés au score IAT anti-Maghrébins de participants français d'origine maghrébine (*N* = 1455). De manière opposée aux résultats que nous avons pu observer ailleurs (Rougier et al., 2020), ainsi qu'à ce que nous avions prédit dans l'Étude 10, les participants d'origine maghrébine ont exprimé un biais IAT pro-exogroupe (*dz* = 0.33). Malheureusement, nous n'avions pas cette information à disposition au moment de la réalisation de notre Étude 10, ces résultats allant toutefois dans le sens de ce que nous avons observé dans cette étude.

biais pro-endogroupe pour des participants noirs issus d'états anciennement esclavagistes. Mais qu'en est-il de la situation en France ? Ne pourrions-nous pas imaginer que nos participants de culture maghrébine puissent-être motivés à inhiber leur stéréotype ? Cela serait en accord avec le climat actuel relatif au possible amalgame Maghrébin-terroriste développé au Chapitre 2. Est-il possible que ce climat négatif soit moins fort qu'aux Etats-Unis ? Si nous répliquons l'une de nos étude du Chapitre 2 avec une induction de contexte d'Attentats, il pourrait être intéressant d'inclure des participants de culture maghrébine afin d'observer si, dans ce contexte, nous pouvons effectivement observer une inversion du biais pro-exogroupe comme l'ont montré certaines études (Payne et al., 2019).

Comme nous l'avons évoqué plus haut, nous avons par le passé observé un biais pro-endogroupe de la part de participants Maghrébins passant une tâche d'approche-évitement dans notre université (Rougier et al., 2020). Comment expliquer, dès lors, que deux mesures indirectes (i.e., tâche d'approche-évitement vs. SMT) utilisées à trois ans d'intervalle montrent des résultats opposés (i.e., biais pro-endogroupe vs. biais pro-exogroupe) ? La tâche d'approche-évitement que nous avions utilisée est décrite comme reproduisant les mouvements d'approche et d'évitement liés au soi et donc, pour Rougier et collaborateurs (2018) moins liés aux influences culturelles extra-personnelles. Pour nous, cette mesure permettait ainsi de capter la variabilité intergroupe au-delà de l'influence culturelle. C'est là peut-être une limite à notre paradigme SMT. Moins impliquant au niveau du soi, le paradigme SMT ne permettrait peut-être pas de capter l'attitude personnelle au-delà de l'influence culturelle. Cependant, comment expliquer alors que dans une tâche relativement similaire à notre paradigme SMT (i.e., AMP), Payne et collaborateurs (2005) aient eux aussi observé un favoritisme pro-endogroupe de la part du groupe de bas statut social ? L'AMP et la tâche d'approche-évitement sont toutes deux une mesure de la composante affective de l'attitude (i.e., le préjugé) et non pas de sa composante cognitive (i.e., le stéréotype). Il est dès lors possible, une fois encore, qu'une mesure comme la nôtre, moins centrée sur une composante affective, soit moins reliée au soi et, de ce fait, nécessite moins d'investissement personnel. Aurait-on observé un paramètre SAP inférieur à .50, c'est-à-dire une plus forte inhibition du contenu mental activé, si nous avions demandé aux participants de catégoriser les visages cibles selon un caractère plaisant/déplaisant ?

Au final, cette dixième et dernière étude pose beaucoup de questions auxquelles nous ne pouvons pour l'instant pas répondre. D'une manière générale, par le biais de cette étude nous voulions mettre en lumière l'intérêt d'une prise en compte du caractère multi-processuel

des mesures indirectes dans l'explication du biais pro-endogroupe ou pro-exogroupe observé au sein de mesures indirectes. De prochaines études pourront avoir pour rôle de déterminer précisément le rôle de chaque processus dans la composante personnelle ou extra-personnelle de l'attitude. Par exemple, peut-on observer une corrélation entre le paramètre SAP et des échelles de justification du système ? Plus intéressant encore, le paramètre SAC peut-il également être corrélé à ce type d'échelle ? Des individus motivés à maintenir une attitude pro-endogroupe positive (i.e., Jesse Jackson) manifesteront-ils un biais pro-endogroupe dans notre tâche SMT ? Si oui, cela pourrait-il être lié au paramètre SAP ? Nous pensons que le paradigme SMT et plus généralement les modèles multinomiaux associés à ce type de paradigmes peuvent être des outils pertinents pour répondre à ces questions et ainsi apporter une vision nouvelle à ce débat.

4.7 Discussion du Chapitre 4 : l'effet SMT est-il généralisable ?

L'objectif de ce chapitre était double. D'abord, nous voulions étudier le caractère généralisable de l'effet SMT et ainsi investiguer plus en détail les mécanismes à l'origine de celui-ci. De plus, en écho aux chapitres précédents, nous voulions mettre en avant une fois encore certaines réflexions concernant l'interprétation des résultats relatifs aux mesures indirectes et l'apport théorique que pouvaient avoir les modèles multinomiaux dans cette réflexion. Afin d'illustrer ces deux points, rappelons nos résultats.

Nous l'avons vu par le biais d'une méta-analyse interne à nos études, l'effet SMT relatif aux visages maghrébins et blancs est fortement reproductible. Dans notre huitième étude, nous avons pu observer que cet effet n'était pas seulement reproductible avec des visages prototypiques des catégories Maghrébins et Blancs, mais aussi avec des prénoms prototypiques de ces catégories. De plus, nous avons également mis en évidence un lien entre l'effet SMT et une mesure directe des catégories Maghrébins et Blancs indépendamment du jugement des différents exemplaires composant la SMT. En somme, comme nous en faisions l'hypothèse, ces résultats sont cohérents avec une réactivation en mémoire des catégories Maghrébins et Blancs et des stéréotypes associés.

Notre neuvième étude a fait apparaitre des résultats surprenants. Tout d'abord et pour la première fois, nous n'avons pas répliqué l'effet SMT anti-Maghrébins. Ce résultat est intéressant, car notre Étude 9 prenait place dans un autre environnement culturel, aux États-

Unis, et peut donc être cohérent avec l'idée d'un stéréotype culturel différent entre France et États-Unis (Payne et al., 2017 ; Rougier et al., 2020). Cependant, le paramètre d'activation du stéréotype non nul suggérait que nos participants avaient tout de même activé ce stéréotype, signe de sa présence dans cet environnement culturel. Ensuite, et de manière encore plus surprenante, nous n'avons pas non plus répliqué l'effet SMT anti-Noirs pourtant mis en évidence à de nombreuses reprises au sein de l'université où cette étude s'est déroulée (Krieglmeyer & Sherman, 2012 ; Rees, Ma, et al., 2019 ; Rees, Rivers, et al., 2019 ; Rivers et al., 2020). En outre, l'hétérogénéité quant à l'origine de nos participants nous a permis d'observer deux choses. D'abord, ces résultats semblaient descriptivement différer en fonction de l'origine des individus. Ensuite, ces différences semblaient, une fois encore descriptivement, être le fruit d'une application différente du stéréotype activé, plutôt que d'une activation du stéréotype en lui-même.

Notre dixième et dernière étude nous a permis d'approfondir ces résultats et plus particulièrement de nous intéresser aux réponses de participants provenant du groupe cible du stéréotype négatif (i.e, les individus de culture maghrébine), habituellement décrit comme étant de bas statut social. De manière similaire à ce que l'on peut retrouver dans une partie de la littérature (e.g., Ashburn-Nardo et al., 2003 ; Livingston, 2002 ; Nosek et al., 2002 ; Richeson et al., 2005 ; Rudman et al., 2002), nous avons mis en évidence un biais pro-exogroupe chez nos participants de culture maghrébine. En d'autres termes, ces participants ont, comme nos participants de culture française, exprimé un effet SMT anti-Maghrébins. Contrairement à ce que pensent certains auteurs (Arkes & Tetlock, 2004), nos participants maghrébins n'ont pas inhibé la réactivation de ce stéréotype, ceci étant davantage en accord avec ce que peuvent avancer d'autres auteurs (Jost, 2019 ; Jost et al., 2004).

À leur manière, ces trois études nous ont donc apporté des éléments de réponses quant au caractère généralisable de nos effets. Dans nos chapitres précédents, nous avions pu mettre en évidence la robustesse de notre effet SMT. Cela est encore plus le cas à la fin de ce chapitre. Jusqu'à présent, les études SMT décrites dans la littérature utilisaient des visages en guise d'amorce (Krieglmeyer & Sherman, 2012 ; Rees, Ma, et al., 2019 ; Rees, Rivers, et al., 2019 ; Rivers et al., 2020). Pour la première fois, nous avons montré que l'effet SMT pouvait être répliqué avec un autre type de stimuli, à savoir des prénoms. De même, il n'existe à notre connaissance aucune étude SMT qui se soit réellement intéressée aux résultats relatifs au groupe cible du stéréotype. Là encore pour la première fois, nous avons montré que l'effet SMT pouvait être généralisé au groupe social de bas statut cible du stéréotype (i.e., les

Maghrébins). Enfin, notre Étude 9 a montré qu'espérer une généralisabilité de nos résultats en toute circonstance peut ne pas faire sens théoriquement. En effet, nous pensions que notre effet SMT reflétait avant tout le contexte culturel français et pouvait ne pas se retrouver ailleurs. Effectivement, la seule et unique fois où nous n'avons pas répliqué cet effet était à l'université de Davis, aux États-Unis. En somme, notre effet SMT est reproductible (Étude 8), parfois de manière inattendue (Étude 10), mais quand il ne l'est pas, cela peut s'expliquer théoriquement (Étude 9). Toutefois, comme nous l'avons vu jusqu'à présent, le principal atout de la SMT se situe ailleurs.

Si ces trois études nous ont permis de comprendre davantage l'effet SMT, c'est aussi et surtout grâce au modèle multinomial associé. L'Étude 8 est peut-être celle pour qui l'apport du modèle multinomial est le moins clair. Cela est certainement lié au fait que, pour la première fois, nos hypothèses portaient plus sur l'effet SMT en lui-même que sur les processus à l'origine de cet effet. Pourtant, les différences importantes dans l'estimation des paramètres liés aux amorces de type visages et de type prénoms nous poussent à être prudents quant à nos conclusions relatives aux mécanismes sous-tendant la SMT-visages et la SMT-prénoms, rejoignant en cela certaines critiques formulées à l'encontre de l'IAT (e.g., Foroni & Bel-Bahar, 2009). Pour les Études 9 et 10, au contraire, l'intérêt des modèles multinomiaux est plus clair. Dans une littérature comportant de nombreux résultats antagonistes, liés entre autres aux groupes de bas statut social (e.g., Nosek et al., 2002 ; Payne et al., 2005, 2019 ; Rougier et al., 2020 ; Rudman et al., 2002), il est essentiel de prendre en compte l'ensemble des principaux processus à l'œuvre et notamment des processus moins automatiques (e.g., inhibition du biais) afin de comprendre au mieux les différents résultats. Ces critiques ne sont pas nouvelles (e.g., Conrey et al., 2005 ; Fiedler & Bluemke, 2005), mais encore trop peu prises en compte dans les interprétations des résultats, notamment propre à l'IAT (Hehman et al., 2019 ; Payne et al., 2017, 2019). La discussion générale de l'ensemble de ce manuscrit nous permettra d'aborder ce point plus en détail.

Chapitre 5 – Discussion générale

Les mesures des attitudes raciales ont grandement évolué au cours du temps. À l'origine basées sur l'auto-report des individus interrogés et centrées sur leur contenu (e.g., Katz & Braly, 1933), les mesures des stéréotypes ont su se renouveler en parallèle de la prise en compte de nouvelles considérations théoriques (e.g., biais de désirabilité sociale ; Sigall & Page, 1971). Grâce notamment à l'influence de la psychologie cognitive, les mesures indirectes des attitudes se sont imposées et ont révolutionné le champ des attitudes raciales (e.g., Fazio et al., 1995 ; Greenwald et al., 1998 ; Greenwald & Banaji, 1995 ; Payne et al., 2005). Cependant, comme ce fut le cas pour les mesures directes, les mesures indirectes ont elles aussi subi leur lot de critiques (e.g., caractère multi-processuel de ces mesures ; Fiedler & Bluemke, 2005; Hütter & Klauer, 2016). Nous inscrivant dans le prolongement de ces critiques, nous pensons que les mesures indirectes peuvent encore évoluer pour correspondre aux préoccupations théoriques actuelles. Ainsi, tout au long de ce manuscrit, nous avons mis en avant l'importance d'une investigation systématique des processus à l'origine de la réponse individuelle. En cela, nous avons montré qu'utiliser une mesure indirecte couplée à un modèle multinomial pouvait s'avérer pertinent notamment pour nourrir notre réflexion théorique quant au rôle de ces processus. Plus particulièrement, nous avons défendu la thèse selon laquelle les processus et l'organisation des processus à l'origine de la réponse individuelle pouvaient varier en fonction d'un contexte particulier et du type de participants prenant part à la mesure indirecte du stéréotype. Rappelons nos principaux résultats.

5.1 Rappel des résultats : mise en évidence de la multiplicité des processus

Ce travail de thèse était composé de trois grandes parties empiriques. La première partie consistait en l'étude de l'impact des attentats sur les processus sous-tendant les réponses issues d'une mesure indirecte du stéréotype. La deuxième partie était consacrée à la prise en compte de facteurs liés à la tâche et de facteurs liés aux participants dans la relation conditionnelle entre les processus sous-tendant ces réponses. Enfin, la troisième partie avait pour objectif de tester le caractère généralisable de l'effet habituellement obtenu, l'effet SMT, à la fois au niveau du type de stimuli, mais aussi du type de participants prenant part à la

tâche. Par le biais de dix études empiriques, nous avons pu mettre en évidence certaines problématiques relatives à l'utilisation de mesures indirectes dans le champ des attitudes raciales. Détaillons ces problématiques.

Notre première partie empirique reposait sur une hypothèse double. Tout d'abord, dans la continuité du Projet Amalgame, nous avons émis l'hypothèse d'une activation plus importante du stéréotype « Maghrébins = Menace » dans un contexte d'attentats, du fait d'un possible amalgame Maghrébin-terroriste (Kepel, 2016). Mais dans le même temps, nous faisions également l'hypothèse d'une moins grande application de ce stéréotype, du fait de la volonté affichée de ne pas faire d'amalgame (e.g., Nugier et al., 2016 ; Zerhouni et al., 2016). Nous avons testé ces hypothèses une première fois dans notre Étude 2, reprenant un paradigme SMT couplé à un rappel des attentats. Contrairement à nos hypothèses, nous n'avons pas observé d'influence de notre induction sur les paramètres d'activation (SAC) et d'application (SAP) du stéréotype. Dans l'Étude 3, nous avons répliqué cette étude en utilisant une induction d'attentats nous paraissant plus optimale (Barden et al., 2004 ; Maddux et al., 2005 ; Wittenbrink et al., 2001). Une fois encore, nous n'avons pas mis en évidence les résultats que nous attendions, ceux-ci allant néanmoins descriptivement dans le sens de nos hypothèses. Enfin, dans notre Étude 4 nous avons voulu observer si la survenue de ces attaques (i.e., données récoltées au moment des attentats) pouvait être liée à une variation des processus sous-tendant les réponses dans un IAT. Une nouvelle fois, nous n'avons pas pu observer de lien entre ces évènements et la contribution des processus à l'origine des réponses individuelles. Bien que l'ensemble de ces études n'ait pas donné les résultats que nous attendions, notamment à cause de certains problèmes méthodologiques (voir Chapitre 2), nous avons voulu, par le biais de notre réflexion théorique et de la démarche que nous avons entreprise, mettre en lumière la nécessité d'une prise en compte des principaux processus à l'œuvre au sein d'une mesure indirecte des attitudes. Plus particulièrement, nous avons voulu montrer l'intérêt de ne pas arrêter nos analyses à l'influence de nos manipulations sur les réponses des participants (e.g., effet SMT en condition Contrôle vs. Attentat), mais d'avoir également conscience de l'influence de nos manipulations sur les processus cognitifs à l'origine de ces réponses.

Notre deuxième partie empirique partait du constat que les résultats de nos deux premières études étaient mieux expliqués par un modèle multinomial (i.e., modèle D-dominant) autre que le modèle multinomial classiquement associé à cette tâche (i.e., modèle SAC-dominant). L'Étude 5 a suggéré que cette différence pouvait être expliquée par les deux

types d'amorces utilisées dans nos études (i.e., visages maghrébins/blancs vs. visages noirs/blancs). Néanmoins, ce résultat n'a pas été confirmé par notre Étude 9, réplication à l'identique de l'Étude 5 dans un contexte géographique différent. Nous pensions à l'origine que les résultats de l'Étude 5 étaient le reflet d'une différence perceptive au niveau de nos amorces. L'Étude 6, où nous avons manipulé à la fois la perception de l'amorce et de la cible, n'a pas confirmé cette hypothèse. Toutefois, nous avons mis en évidence pour la première fois un possible lien entre une mesure directe d'attitudes raciales et la tendance des participants, au sein de la SMT, à utiliser en priorité les informations provenant de l'amorce (i.e., modèle SAC-dominant) ou au contraire de la cible (i.e., modèle D-dominant) pour formuler leur réponse. Dans l'Étude 7 nous avions pour objectif de répliquer ces résultats et d'approfondir les explications théoriques associées. Allant dans le sens de nos premiers résultats, nous avons pu observer qu'effectivement, plus les participants avaient une attitude raciale auto-rapportée négative envers les Maghrébins et plus ils avaient tendance à utiliser en priorité les informations de l'amorce (i.e., modèle SAC-dominant) pour formuler leur réponse. De manière exploratoire, nous avons aussi pu mettre en évidence le lien entre deux mesures motivationnelles (i.e., sous-échelles de motivation externe et interne à répondre sans préjugés) et la relation conditionnelle entre processus. De manière globale, et comme nous le détaillerons plus en détail plus bas, ces résultats nous ont non seulement montré la nécessité de prendre en compte la relation conditionnelle entre processus, mais aussi la possible influence de variables interindividuelles dans cette relation.

Enfin, dans notre dernière partie empirique, nous avons questionné le caractère généralisable de notre effet SMT anti-Maghrébins. Dans notre Étude 8, nous avons répliqué l'effet SMT à la fois avec des visages, mais aussi avec des prénoms. De plus, nous avons montré que ces effets étaient corrélés avec une mesure directe de la catégorie au-delà d'une mesure directe des exemplaires de cette catégorie. En somme, ces résultats sont cohérents avec l'explication théorique de l'effet SMT reflétant une réactivation de la catégorie associée à l'amorce (e.g., Maghrébins) et du stéréotype associé (e.g., menaçant ; Krieglmeyer & Sherman, 2012 ; Macrae & Bodenhausen, 2000, 2001). Nos Études 9 et 10, enfin, avaient pour objectif de tester le caractère généralisable de notre effet à la fois sur une population exposée à une influence culturelle différente (i.e., États-Unis ; Payne et al., 2017), mais aussi auprès du groupe social de bas statut classiquement associé au stéréotype (i.e., individus d'origine maghrébine ; Mange et al., 2016). De façon intéressante et partiellement inattendue, nous n'avons répliqué ni l'effet SMT anti-Maghrébins ni l'effet SMT anti-Noirs chez des participants d'une université californienne. Cependant, la présence d'une activation du

stéréotype non nulle nous a permis de rapprocher ces résultats de la littérature traitant de la composante culturellement partagée du stéréotype (e.g., Arkes & Tetlock, 2004 ; Devine, 1989 ; Payne et al., 2017 ; Rougier et al., 2020). De plus, nous nous sommes questionnés quant au paramètre d'application du stéréotype et plus particulièrement sur sa variabilité en fonction de l'origine ethnique des participants. Notre dixième étude nous a permis d'approfondir ces résultats sur l'ensemble de données relatives au groupe cible du stéréotype (i.e., les Maghrébins) récoltées tout au long de nos études. Contrairement à notre hypothèse, les participants se catégorisant de culture principale maghrébine ont exprimé un effet SMT pro-exogroupe, rejoignant en cela d'autres résultats classiques de la littérature (e.g., Ashburn-Nardo et al., 2003 ; Livingston, 2002 ; Nosek et al., 2002 ; Richeson et al., 2005 ; Rudman et al., 2002). Au-delà des résultats rapportés, le raisonnement théorique associé à cette dixième étude était pour nous l'occasion de montrer, une fois encore, l'importance d'une prise en compte du caractère multi-processuel des mesures indirectes dans une littérature ne prenant à nos yeux pas assez en considération cet aspect (i.e., composante personnelle ou extra-personnelle de l'attitude ; e.g., Arkes & Tetlock, 2004 ; Banaji et al., 2004 ; Jost, 2019 ; Payne et al., 2017). D'une manière générale, cette dernière partie empirique mettait une nouvelle fois en lumière l'importance du questionnement théorique dans l'interprétation de nos effets ainsi que l'apport des modèles multinomiaux dans l'éclairage de ces interprétations.

5.2 Que retenir de ces études ?

Comme nous venons de le voir, ce travail de thèse a parfois apporté des résultats empiriques mitigés (e.g., Chapitre 2). Certaines limites, notamment méthodologiques (e.g., manque de puissance), nous ont certainement été préjudiciables dans l'obtention des effets attendus. Toutefois, nous pensons que certains de ces résultats (e.g., Chapitre 3, Chapitre 4) ont apporté du crédit à la thèse que nous défendions. Plus largement, nous pensons que notre démarche générale nous a permis de mettre en évidence la nécessité de prendre en compte les principaux processus à l'œuvre au sein d'une mesure indirecte d'attitudes raciales. Le lecteur ne trouvera peut-être rien de véritablement nouveau dans ces préconisations. Nous sommes d'accord avec cela. Pourtant, force est de constater que bien que certaines de ces critiques datent maintenant de presque 20 ans (e.g., Conrey et al., 2005 ; Fiedler & Bluemke, 2005), un nombre trop important d'études ou champs théoriques ne prennent toujours pas en

considération le caractère multi-processuel des mesures indirectes. En outre, comme nous avons pu le montrer dans notre Chapitre 3, nous pouvons encore aller plus loin dans la prise en considération de ces processus. Illustrons ces deux derniers points.

5.2.1 Une prise en compte encore trop partielle du caractère multi-processuel des mesures indirectes

Les deux dernières études de ce manuscrit sont peut-être l'exemple le plus frappant pour illustrer le manque de prise en compte du caractère multi-processuel des mesures indirectes. La question de la composante personnelle ou extra-personnelle de l'attitude, déjà soulevée par certains auteurs par le passé (e.g., Arkes & Tetlock, 2004 ; Banaji et al., 2004 ; Rudman et al., 1999) est aujourd'hui encore d'actualité dans le champ des attitudes (e.g., Jost, 2019 ; Payne et al., 2017). Pourtant, il semble surprenant que peu d'auteurs (nous y compris, voir Rougier et al., 2020) fassent référence à la possibilité qu'un biais pro-endogroupe (e.g., Degner et al., 2016 ; Payne et al., 2019 ; Rougier et al., 2020) ou pro-exogroupe (e.g., Ashburn-Nardo et al., 2003 ; Livingston, 2002 ; Nosek et al., 2002 ; Richeson et al., 2005 ; Rudman et al., 2002) puisse refléter d'autres processus qu'une réactivation d'un contenu mental spécifique. Par exemple, nous avons grandement cité l'influent article de Payne et collaborateurs (2017) sur la composante situationnelle de l'attitude. Pour illustrer leur propos, ces auteurs ont pris en exemple les résultats d'études utilisant un paradigme IAT. Cependant, Payne et collaborateurs ne parlent à aucun moment de l'aspect multi-processuel de ces résultats, laissant ainsi la possibilité au lecteur d'interpréter ceux-ci comme l'influence de la situation sur les processus automatiques uniquement. D'autres auteurs avant nous ont soulevé cette critique. En effet, Rivers et collaborateurs (2017), en commentaire de cet article, ont abordé la nécessité de prendre en compte l'ensemble des processus à l'œuvre dans l'IAT (i.e., processus plus ou moins automatiques) afin de se questionner sur les processus dépendant davantage du contexte (i.e., de la situation) et ceux moins sensibles à celui-ci.

Afin d'illustrer ce dernier aspect, nous pouvons reprendre une nouvelle fois l'étude de Payne et collaborateurs (2019) montrant un lien entre le type d'état historiquement esclavagiste et le score IAT pro-endogroupe. Pour rappel, Payne et collaborateurs ont montré que les comtés et états des États-Unis ayant été les plus marqués par l'esclavagisme (e.g., états du Sud) avaient des scores IAT pro-Noirs plus importants chez les populations noires,

suggérant ainsi la réactivation d'un contenu mental anti-Blancs/pro-Noirs lié à cet héritage. Selon nous, et en accord avec Rivers et collaborateurs (2017), il aurait été intéressant de se questionner si, effectivement, ces résultats reflètent bien une réactivation de croyances situationnelles pro-Noirs (i.e., activation du contenu mental « Blancs = Négatif » et/ou « Noirs = Positif ») ou bien une plus forte inhibition de croyances situationnelles globalement pro-Blancs (inhibition du contenu mental « Blancs = Positif » et/ou « Noirs = Négatif »). Le fait que le lien entre l'état (i.e., plus ou moins historiquement esclavagiste) et le biais pro-endrogroupe disparaisse quand contrôlé par une mesure directe d'attitudes raciales peut nous faire pencher vers la seconde alternative.

Nos études ne sont pourtant pas les premières à tester empiriquement ce type de questionnements. En effet, comme nous l'avons développé dans le Chapitre 3, des études ont montré l'intérêt d'une prise en compte de ces différents processus dans l'explication de la préférence exprimée par les groupes de bas statut social (Gonsalkorale et al., 2014 ; Jin et al., 2016). Par exemple, Gonsalkorale et collaborateurs (2014) ont pu apporter une interprétation plus précise d'un résultat classique de la littérature, à savoir un score IAT traditionnellement pro-Jeunes à la fois pour les participants jeunes (biais pro-endogroupe) et pour les participants âgés (biais pro-exogroupe ; Jost et al., 2004). En appliquant un Quad Model aux données de participants jeunes et plus âgées, Gonsalkorale et collaborateurs (2014) ont montré qu'un biais pouvant sembler à première vue identique reflétait une contribution différente des processus à l'œuvre : les personnes plus âgées avaient à la fois une activation (AC) plus faible du biais pro-Jeunes que les participants jeunes, mais également une inhibition (OB) plus faible de ce biais. Ce type de résultats pourrait ainsi apporter des éléments de compréhensions supplémentaires à certaines interprétations en termes de justification du système quant au score IAT pro-Jeunes (Jost et al., 2004). Une fois encore, pourtant, il semble que le caractère multi-processuel des mesures indirectes ne soit que très faiblement abordé dans les contributions récentes relatives à ce type de théories (e.g., Jost, 2019).

À l'inverse des exemples mentionnés ci-dessus, d'autres champs théoriques proches ont su bénéficier pleinement de l'apport des modèles multinomiaux. Un des exemples les plus frappants se situe au niveau des études relatives aux interventions destinées à la réduction du biais racial (pour une revue de littérature voir, Lai et al., 2013). Nous avons vu aux Chapitres 1 et 2 que le biais racial anti-Noirs pouvait être diminué en fonction du contexte dans lequel les visages noirs étaient rencontrés (e.g., Barden et al., 2004 ; Maddux et al., 2005 ; Wittenbrink et al., 2001), ce biais étant par exemple plus important dans un contexte de prison que

d'église (Maddux et al., 2005). En réalité, un nombre important d'études ont mis en évidence diverses interventions permettant une diminution du biais racial (Lai et al., 2013). Parmi d'autres, nous pouvons citer l'exposition à un exemplaire contre-stéréotypique (e.g., Dasgupta & Greenwald, 2001) ou l'entrainement à inhiber des associations stéréotypiques (Kawakami et al., 2000). Calanchini et collaborateurs (2020) ont apporté un éclairage nouveau quant aux explications théoriques relatives à ces interventions. En réalisant une méta-analyse sur des données de plus de 20 000 participants ayant passé une des principales interventions retrouvées dans la littérature, ces auteurs ont appliqué aux données un Quad Model afin d'estimer, pour chacune de ces interventions, quel(s) processus pouvaient être impliqués dans la diminution du biais racial. Parmi les résultats principaux, ces auteurs ont montré que les interventions relatives à la présentation d'exemplaires contre-stéréotypiques (e.g., Dasgupta & Greenwald, 2001) ou aux stratégies visant à passer outre le biais (e.g., implémentation d'intention ; Stewart & Payne, 2008) influençaient à la fois le paramètre d'activation du contenu mental (AC) et le paramètre de détection de la bonne réponse (D ; mais voir Rees, Rivers, et al., 2019). À l'inverse, les interventions ciblées sur une procédure de conditionnement évaluatif (e.g., De Houwer et al., 2001) influençaient à la fois le paramètre de détection de la bonne réponse (D), mais aussi le paramètre d'inhibition du biais (OB). De plus, quand l'intervention faisait varier le paramètre d'activation (AC), cela allait davantage dans le sens d'une diminution de la réactivation du contenu mental « Blancs = Positif », que du contenu mental « Noirs = Négatif ». En somme, Calanchini et collaborateurs (2020) ont ainsi mis en avant l'intérêt d'une utilisation des modèles multinomiaux dans l'explication de l'hétérogénéité des résultats au sein de ce champ théorique.

Par le biais de ces quelques exemples, nous venons d'évoquer les limites à la non prise en compte du caractère multi-processuel des mesures indirectes, mais nous avons nous-même par le passé été trop peu précautionneux quant à certaines problématiques liées au Projet Amalgame. Effectivement, nous avons trop souvent interprété les résultats des mesures indirectes comme le reflet de processus automatiques et notamment de processus non contrôlables (e.g., Courset, 2017). À nos yeux, il n'est pas rare de trouver des auteurs faisant, encore aujourd'hui, ce type de raccourcis, c'est-à-dire interprétant les résultats d'une mesure indirecte comme étant le reflet de processus (uniquement) automatiques (voir Corneille & Hütter, 2020 ; Gawronski, 2019 ; Mitchell & Tetlock, 2017, pour des discussions similaires). Par exemple, si des individus veulent aujourd'hui passer un test IAT-Noirs sur la plateforme Project Implicit, ceux-ci sont accueillis par la description suivante : « Ce TAI [Test d'Association Implicite] nécessite la capacité de distinguer des visages d'origine Européenne

et d'origine Africaine. Il révèle que la plupart des gens ont une préférence automatique pour les personnes Blanches par rapport aux personnes Noires ». Est-on sûr que ce test révèle une préférence « automatique » ? Comme nous avons pu le voir à de nombreuses reprises au sein ce manuscrit, la réponse est non (e.g., Calanchini, 2020 ; Calanchini et al., 2018 ; Conrey et al., 2005 ; Sherman et al., 2008, 2010, 2014).

Pour nous, ce genre de raccourci ou d'omission est d'autant plus problématique, car cela influence l'image « grand public » que peuvent parfois avoir les mesures indirectes. En effet, même si dans la littérature scientifique des précautions ont pu être prises quant aux différences entre chacun des critères inclus sous le chapeau « automatique » (e.g., Melnikoff & Bargh, 2018 ; Moors, 2016) force est de constater que, hors de la sphère scientifique, ces précautions sont moins évidentes. À titre d'exemple, nous pouvons reprendre une nouvelle fois l'article de Charlie Hebdo faisant écho de certains de nos résultats (Fischetti, 2020). Dans cet article, nous pouvons trouver des phrases comme « les scientifiques préfèrent sonder les préjugés inconscients » ou même « c'est la manifestation typique d'un préjugé raciste : l'étranger est inconsciemment perçu comme une menace ». Tant que nous ne prendrons pas nous-mêmes réellement en considération le caractère multi-processuel de ces mesures, ces croyances perdureront dans le monde de la recherche et par conséquent dans le grand public.

On pourrait néanmoins argumenter qu'au final, la part de processus non automatiques à l'œuvre dans ce type de mesures est relativement faible et que les effets retrouvés reflètent effectivement des processus automatiques. En d'autres termes, il reste envisageable que cette confusion entre mesures indirectes et processus automatiques ne soit que peu préjudiciable dans la compréhension que nous avons de nos effets. Nous ne sommes pas d'accord avec cela. Tant que nous n'aurons pas quantifié réellement la contribution des différents processus et tant que nous n'aurons pas montré empiriquement que nos résultats sont effectivement le reflet d'un type de processus en particulier (e.g., non conscient, non contrôlable), nous devons nous montrer prudents dans l'interprétation de nos résultats. Dans une période où notre littérature scientifique est à la recherche d'une nouvelle crédibilité (e.g., Nelson et al., 2018), nous ne pouvons continuer à ignorer ces aspects.

5.2.2 De nouvelles avancées dans l'utilisation des modèles multinomiaux

Jusqu'à présent, l'utilisation de modèles multinomiaux dans le champ des attitudes raciales se faisait principalement pour les raisons suivantes. Premièrement, les modèles multinomiaux ont pu être utilisés pour observer, au sein d'une mesure indirecte, la contribution spécifique des différents processus à l'origine de la performance individuelle (e.g., Conrey et al., 2005 ; Krieglmeyer & Sherman, 2012 ; Nadarevic & Erdfelder, 2011). L'exemple le plus connu est naturellement celui des premières études traitant du Quad Model, ces études ayant montré l'impact non seulement d'un processus d'activation du contenu mental (AC), mais aussi d'un processus d'inhibition du contenu mental activé (OB), d'une détection de la bonne réponse (D) et d'un biais de réponse (G ; Conrey et al., 2005 ; Sherman, 2006 ; Sherman et al., 2008). Deuxièmement, et comme nous l'avons vu ci-dessus, ces modèles ont pu être utilisés pour expliquer l'impact de certaines manipulations sur ces performances. Au Chapitre 1, nous avons pris l'exemple de Allen et collaborateurs (2010) montrant que la diminution du biais racial anti-Noirs en fonction du contexte (e.g., église vs. prison) était le reflet d'un processus d'inhibition du biais négatif réactivé. De plus, ces modèles ont été utilisés pour fournir une explication à certaines différences intergroupes observées classiquement dans la littérature (e.g., Gonsalkorale, Sherman, et al., 2009, 2014). Comme nous l'avons déjà évoqué au Chapitre 1, Gonsalkorale, Sherman, et collaborateurs (2009) ont montré que le biais anti-Noirs plus important chez les individus plus âgés résultait d'une inhibition plus faible du biais réactivé. Enfin, comme nous l'avons évoqué au Chapitre 3, les modèles multinomiaux ont pu être utilisés pour mettre en évidence la relation conditionnelle entre processus aboutissant à la réponse du participant. Ce fut le cas notamment de Payne et Bishara (2009) qui ont tenté de déterminer le modèle parmi le PDP-C, le PDP-A ou le Quad Model pouvant prédire au mieux les réponses des participants passant une WIT. Toutefois, au sein de ce manuscrit de thèse, nous avons proposé une nouvelle utilisation de ces modèles.

Jusqu'à aujourd'hui, quand des chercheurs s'intéressaient à la relation conditionnelle entre processus, le modèle donnant la meilleure adéquation aux données était assimilé au modèle le plus optimal pour décrire les processus à l'œuvre (e.g., Krieglmeyer & Sherman, 2012) et utilisé de manière identique dans les études suivantes (e.,g., Rees, Ma, et al., 2019 ; Rees, Rivers, et al., 2019; Rivers et al., 2020). En d'autres termes, la relation conditionnelle classique entre processus (i.e., le modèle multinomial classiquement associé à une tâche) n'était que rarement remise en cause. De plus, même quand celle-ci pouvait être remise en

cause, les chercheurs ne se questionnaient que trop peu sur les raisons à la meilleure adéquation d'un modèle plutôt qu'un autre (e.g., Bishara & Payne, 2009). En somme, les chercheurs ne se sont que trop rarement intéressés aux facteurs pouvant expliquer la meilleure adéquation aux données d'un modèle plutôt qu'un autre. Pour la première fois, et au sein de deux études (Études 6 et 7), nous avons pu mettre en évidence que des variables individuelles (i.e., échelle d'attitude raciale, échelle de motivation) pouvaient expliquer la meilleure adéquation aux données d'un modèle plutôt qu'un autre. Nous voyons deux implications à ces résultats. Une implication pragmatique et une implication plus théorique.

D'un point de vue pragmatique, comme nous allons le voir maintenant, nous pensons que cela pourrait nous éviter des erreurs de prédiction quant à la valeur des paramètres. Prenons l'exemple concret de la différence de prédictions entre nos modèles SAC-dominant et D-dominant. Comme nous l'avons vu aux Chapitres 1 et 3, dans ces deux modèles les paramètres SAC et SAP ont pour rôle d'expliquer la différence de jugement entre les amorces Maghrébin et Blanc. Dans le modèle SAC-dominant, ces paramètres sont prioritaires, c'est-à-dire qu'ils vont orienter la réponse du participant avant l'influence de tout autre paramètre. Nous dirons donc que l'influence ou le « poids » de ces paramètres sera de 1 (i.e., la différence de jugement ne dépendra que de ces paramètres). À l'inverse, dans le modèle D-dominant les paramètres SAC et SAP sont conditionnels à l'absence du paramètre de détection (1-D). Ainsi, le poids des paramètres SAC et SAP dans le modèle D-dominant dépendra du poids du paramètre D (et plus directement du poids de son absence, 1-D).

Dans l'ensemble de nos études, le paramètre de détection était globalement bas (D = .10). Par conséquent, dans le modèle D-dominant, le poids des paramètres SAC et SAP dans l'explication des résultats (1-D = .90) était quasiment identique au poids du modèle SAC-dominant pour ces paramètres (i.e., poids de 1). Ainsi, les prédictions du modèle SAC-dominant et du modèle D-dominant quant aux paramètres SAC et SAP seront, dans ce cas précis, quasiment identiques. Pour illustrer ce premier point, simulons les prédictions des modèles SAC-dominant et D-dominant ayant tous les deux les valeurs de paramètres suivantes : D = .10 ; SAC = .60 ; SAP = .60 ; G = .20. Avec ces valeurs de paramètres, le modèle SAC-dominant prédira une valeur de jugement plus élevée lorsque les cibles sont précédées de visages maghrébins (M =21.70) que lorsqu'elles sont précédées de visages blancs (M = 16.00). Pour sa part, le modèle D-dominant prédira lui aussi une valeur de jugement plus élevée lorsque les cibles sont précédées de visages maghrébins (M =21.41) que lorsqu'elles sont précédées de visages blancs (M = 16.22). Comme nous nous y attendions, les

prédictions de ces deux modèles sont ici relativement proches. Mais nous allons voir que cela peut ne pas être le cas.

Imaginons maintenant une étude fictive où le paramètre de détection serait particulièrement élevé (D = .90) comme ce peut être le cas avec une tâche IAT (e.g., Conrey et al., 2005, Étude 4). Ici, le poids accordé aux paramètres SAC et SAP sera très différent entre les deux modèles. En effet, pour le modèle D-dominant, le poids des paramètres SAC et SAP dans l'explication des résultats sera beaucoup plus faible (1-D = .10), là où il restera identique pour le modèle SAC-dominant (i.e., poids de 1). Pour illustrer cela, simulons une nouvelle fois les prédictions des modèles SAC-dominant et D-dominant ayant tous les deux les valeurs de paramètres suivantes : D = .90 ; SAC = .60 ; SAP = .60 ; G = .20. Avec ces valeurs de paramètres, le modèle SAC-dominant prédira une fois encore une valeur de jugement plus élevée lorsque les cibles sont précédées de visages maghrébins (*M* =26.30) que lorsqu'elles sont précédées de visages blancs (*M* = 20.54). Néanmoins, le modèle D-dominant prédira pour sa part une différence beaucoup plus faible entre des cibles précédées de visages maghrébins (*M* =23.71) et celles précédées de visages blancs (*M* = 23.14). Dans le cas présent, nous voyons donc qu'un modèle prédit clairement un effet SMT (i.e., modèle SAC-dominant), alors que cela est moins évident pour l'autre (i.e., modèle D-dominant).

Dans ces deux exemples, nous connaissions la valeur des différents paramètres et nous regardions l'estimation des modèles en termes d'effet SMT. En réalité, comme nous l'avons vu tout au long de ce manuscrit, les modèles fonctionnent de manières opposées à cela : l'effet SMT est connu et ce sont les modèles qui vont estimer, par itération successive de la valeur de leurs paramètres, les meilleures valeurs de paramètres rendant compte de ces effets. Par conséquent, si nous reprenons l'exemple ci-dessus, pour un même effet SMT et pour une tâche où la détection de la cible serait importante (D = .90), les valeurs des paramètres SAC et SAP du modèle D-dominant devraient être beaucoup plus élevées que celles du modèle SAC-dominant. En somme, cet exemple nous montre de manière simpliste que, selon le type de modèle multinomial choisi en amont, le chercheur pourra avoir des estimations potentiellement très différentes dans la valeur des paramètres associés.

Encore plus que par cet exemple pratique, montrer que des participants différents (e.g., différences interindividuelles en termes d'attitude raciale) peuvent utiliser diverses « stratégies » pour répondre (i.e., utiliser en priorité l'information de l'amorce ou de la cible), nous force une nouvelle fois à nous poser la question des processus à l'œuvre au sein de la tâche et de la relation entre ces processus. Plus globalement, cela nous pousse à être plus

vigilants quant aux interprétations que nous faisons de nos résultats. Pour nous, cela s'inscrit dans le prolongement naturel des améliorations apportées aux mesures des attitudes (raciales) en psychologie. Reprenons l'évolution de ces mesures.

Les premiers auteurs demandaient directement aux participants de reporter leur jugement ou évaluation envers un objet attitudinal particulier (e.g., Adorno et al., 1950 ; Brigham, 1973 ; Katz & Braly, 1933 ; Lambert, 1952). Avec l'évolution de la société, d'autres auteurs ont mis en évidence que les participants répondant à ces mesures directes pouvaient contrôler leur réponse (e.g., Crosby et al., 1980 ; Crowne & Marlowe, 1960 ; Sigall & Page, 1971). Des mesures indirectes sont apparues permettant l'étude des processus à l'origine de l'attitude raciale et censées capter au plus près des processus de Type 1 (Fazio et al., 1995 ; Greenwald & Banaji, 1995, 1995). Là encore, d'autres auteurs ont mis en évidence les précautions à prendre dans l'utilisation de ces mesures, celles-ci reflétant une multitude de processus, chacun d'entre eux plus ou moins automatiques (Calanchini et al., 2018 ; Hütter & Klauer, 2016 ; Sherman et al., 2010, 2014). Nous avons montré avec ce travail de thèse que nous pouvions aller encore plus loin.

Non seulement les réponses des participants sont le fruit d'une multitude de processus, mais la relation entre ces processus peut varier d'un participant à un autre. Les participants ne sont pas tous homogènes face à la tâche qu'ils doivent effectuer. Des participants sont plus sensibles à certains aspects de la tâche (i.e., amorces), alors que d'autres peuvent être sensibles à d'autres composantes de cette tâche (i.e., cibles). Même si les consignes sont identiques d'un participant à un autre, que les conditions de passations sont standardisées au maximum (e.g., box expérimental, passation sur ordinateur) et que le type de participants peut être relativement homogène (i.e., étudiants se catégorisant de culture principale française), des différences importantes peuvent subsister dans la façon qu'ont ces participants d'appréhender la tâche. Certains auteurs ont questionné la possibilité que, au sein de l'AMP, des participants puissent juger non pas l'idéogramme chinois présenté, mais l'amorce en elle-même (Bar-Anan & Nosek, 2012 ; Cummins et al., 2019 ; Gawronski & Ye, 2015 ; Mann et al., 2019 ; Payne et al., 2013). Par le biais de nos études, nous nous inscrivons dans cette dynamique, à savoir la prise en compte de l'interaction entre le participant et la tâche dans l'explication des résultats. Avec l'utilisation de modèles multinomiaux, nous pouvons avoir des informations supplémentaires. Effectivement, grâce à une formalisation précise des principaux processus à l'œuvre, nous avons pu mettre en évidence un lien entre des différences interindividuelles (i.e., niveau de préjugé, motivation à répondre sans préjugés)

et la relation conditionnelle entre ces processus (priorité aux informations provenant de l'amorce ou de la cible). Avec cette volonté d'avoir une explication des résultats toujours plus précise, nous pensons que nos études participent à un éclaircissement nouveau quant à la passation d'une mesure indirecte d'attitudes raciales ainsi qu'aux différences d'un participant à l'autre face à ces mesures.

5.3 Nécessité d'une théorisation forte préalable à l'utilisation des modèles multinomiaux

Dans ce manuscrit, nous avons tenté de répondre au mieux à un ensemble de questions relatives à nos objets de recherche. Cependant, à la fin de celui-ci, certaines interrogations peuvent rester en suspens. Dans les pages qui vont suivre, nous allons tenter d'apporter des réponses à certaines de ces questions. D'autres resteront cependant toujours sans réponses. De manière générale, par le biais de l'ensemble de ces questionnements, nous voulons une nouvelle fois mettre en évidence la nécessité, pour le chercheur, de faire preuve d'une prudence dans l'interprétation des résultats. Cependant, nous prendrons cette fois comme illustration nos études et notamment nos résultats liés aux modèles multinomiaux.

5.3.1 Réactivation du stéréotype ou d'un préjugé négatif ?

Comme nous l'avons spécifié au Chapitre 1, la SMT est construite à partir de l'AMP (Payne et al., 2005). Dans l'AMP, les participants expriment un jugement évaluatif à propos d'idéogrammes chinois, ceux-ci étant précédés d'amorces de visages blancs ou de visages noirs. À l'inverse de l'AMP, les participants prenant part à la SMT ne doivent pas exprimer de jugement évaluatif, mais un jugement basé sur un trait caractéristique des visages cibles, à savoir la menace associée à ces visages. Pour Krieglmeyer et Sherman (2012) « demander aux participants d'exprimer des jugements basés sur des traits (au lieu de jugements évaluatifs) amorce l'utilisation de stéréotypes plutôt que d'attitudes [préjugés] dans la construction des jugements » p. (207). Toutefois, nous pouvons nous questionner sur la nature réellement sémantique du jugement. En effet, nous sommes ici en présence d'une variable confondue, le trait « menaçant » et les visages de couleur noire (d'origine maghrébine) étant classiquement

associés à une valence négative, le trait « non menaçant » et les visages de couleur blanche étant eux associés à une valence positive. Comment être sûr, dès lors, que l'effet SMT est effectivement le fruit d'une réactivation d'un contenu de croyances (stéréotype) et pas plutôt d'un contenu affectif (préjugé) ? Deux études peuvent nous conforter dans notre première interprétation.

Tout d'abord, la tâche SMT a été répliquée avec d'autres stimuli liés au genre (i.e., femmes vs. hommes) et donc moins antagonistes en termes de valence (Rees, Rivers, et al., 2019). Rees, Rivers et collaborateurs (2019) ont montré, conformément à leurs attentes, que des participants allaient davantage juger la cible comme menaçante quand celle-ci était précédée de visages d'hommes par rapport à des visages de femmes. Pour ces auteurs, ce résultat est le fruit d'une réactivation d'un stéréotype de menace associé aux visages d'hommes contrastant avec le stéréotype de non menace associé aux visages de femmes. Ensuite, et cette fois directement lié au champ des attitudes raciales, Krieglmeyer et Sherman (2012) ont répliqué l'effet SMT classique avec un stéréotype de valence positive. En effet, dans l'une de leurs études, ces auteurs ont montré que des participants allaient davantage juger des cibles comme étant athlétiques (i.e., sportives) quand celles-ci étaient précédées de visages de couleur noire plutôt que de visages de couleur blanche. Pour ces auteurs, cela suggère que le paradigme SMT réactive bien un contenu sémantique (i.e., association stéréotypique positive entre les individus de couleur noire et le sport) et non pas un contenu affectif (i.e., individus de couleur noire liés à une évaluation négative).

Bien sûr, ces études ne peuvent pas être entièrement transposées à notre propre SMT, celle-ci utilisant des visages maghrébins et blancs. Il est possible que, contrairement aux deux études mentionnées ci-dessus, notre effet SMT soit le reflet d'un préjugé négatif (préjugé positif) réactivé à la vision de visages maghrébins (blancs). Par conséquent, Il serait pertinent de répliquer nos effets avec des stéréotypes positifs caractéristiques de la catégorie des Maghrébins, ceci présentant toutefois comme contrainte la nécessité de présenter aux participants des cibles variant sur la dimension de ce trait stéréotypique (comme cela a été le cas pour les dessins menaçants ou sportifs).

5.3.2 La SMT est-elle une mesure directe ?

Comme nous l'avons évoqué plus haut, le paradigme AMP est critiqué parce que potentiellement le reflet d'une évaluation directe des amorces (Bar-Anan & Nosek, 2012 ; Cummins et al., 2019, mais voir Gawronski & Ye, 2015 ; Mann et al., 2019 ; Payne et al., 2013). Pour Bar-Anan et Nosek (2012), l'AMP fonctionne principalement sur les participants qui pensent avoir intentionnellement jugé l'amorce. Pour ces auteurs, les participants passant l'AMP peuvent être divisés en deux groupes : ceux qui ne rapportent pas avoir été influencés par les amorces et ceux qui reportent une influence consciente des amorces et une intention de juger directement celles-ci ; les effets obtenus dans AMP seraient le reflet de ces derniers participants. Pour Cummins et collaborateurs (2019), l'AMP aurait notamment des problèmes méthodologiques : « presque toutes les études AMP donnent au participant comme instruction de ne pas tenir compte de l'amorce comme source d'information et de se concentrer uniquement sur la cible. Le cœur de l'effet de fausse attribution est l'hypothèse que les participants respectent ces consignes et que tout effet qui en résulte est dû au fait que les amorces influencent les évaluations des cibles sans intentionnalité ni conscience. Cependant, une alternative plausible est que les individus ne tiennent tout simplement pas compte de ces consignes et se fient à la première source d'information pour évaluer la cible » (p. 7). Le paradigme SMT étant structurellement proche de l'AMP, peut-on transposer ces questionnements à la SMT ?

Krieglmeyer et Sherman (2012), pour expliquer l'effet SMT, s'appuient eux aussi sur ce phénomène de fausse attribution : l'application du contenu mental réactivé par l'amorce pour déterminer la menace du visage cible pourrait être le reflet d'une mauvaise attribution du contenu mental activé par l'amorce sur la cible (Loersch & Payne, 2011). En outre, les consignes de la SMT sont très proches des consignes de l'AMP décrites ci-dessus : les participants doivent se former une impression par rapport à des dessins de visages cibles sur la dimension de menace. Il leur est demandé de ne pas se concentrer sur les photos de visages présentées avant les dessins cibles, mais de toutefois leur prêter attention, car des questions doivent leur être posées à la fin de l'étude. Se pourrait-il alors que l'effet SMT reflète un jugement direct de l'amorce ?

Nous avons commencé à nous poser cette question au moment de la comparaison du modèle SAC-dominant et du modèle D-dominant. Pour nous, le modèle reflétant au mieux les consignes données au participant était le modèle D-dominant. En effet, la tâche du participant

étant de juger les dessins cibles, il semblerait logique que le participant utilise en priorité les informations de la cible (détection de la menace) pour répondre. Ce ne serait que lorsque ces participants n'ont pas détecté la menace de la cible qu'ils pourraient utiliser d'autres informations pour répondre, comme le contenu mental réactivé par l'amorce. Que penser alors du modèle SAC-dominant ? Celui-ci reflète-t-il un phénomène de fausse attribution du contenu réactivé par l'amorce sur la cible ou bien un jugement direct de l'amorce ? En effet, il est en théorie possible que l'architecture du modèle SAC-dominant reflète plus simplement un jugement direct de l'amorce, et ce n'est que lorsque le stéréotype n'est pas activé que le participant donne sa réponse avec l'information provenant de la cible. Ceci pourrait apporter une interprétation alternative à nos résultats des Études 6 et 7, le paramètre λ pouvant alors mesurer la tendance des participants à juger directement la cible (modèle D-dominant) ou bien directement l'amorce (modèle SAC-dominant).

Notons tout d'abord que nous n'avons pas de réponse définitive à cette question. Toutefois, certains indices peuvent laisser envisager le fait que les participants passant notre SMT ont, au moins en partie, jugé les cibles. D'abord, dans l'intégralité de nos études, nous avons obtenu un effet principal de la cible, c'est-à-dire que les participants ont jugé les cibles menaçantes comme effectivement plus menaçantes que les cibles non menaçantes, preuve que ceux-ci ne s'intéressent pas (uniquement) à un jugement direct de l'amorce. En outre, dans l'Étude 6, cet effet était modéré par le caractère discernable de la cible. Plus le niveau de menace de la cible était facile à discerner et plus l'effet de la cible était important. De plus, au niveau des paramètres associés au modèle et dans l'ensemble de nos études, le paramètre de détection (D) était supérieur à zéro, celui-ci différant également en fonction du caractère discernable de la cible au sein de l'Étude 6.

L'interrogation la plus forte se situe certainement au niveau de la mesure directe d'attitude raciale. Nous avons vu dans les Études 6 et 7 que plus les participants avaient une attitude négative envers les Maghrébins, plus ils suivaient le modèle SAC-dominant (plutôt que le modèle D-dominant). Une explication alternative pourrait être que ces participants jugent directement l'amorce et s'intéressent moins au jugement de la cible que les participants ayant une attitude moins négative envers les Maghrébins. Allant dans le sens de cette interprétation, dans l'Étude 7 nous avons obtenu un lien négatif faible, mais significatif, entre l'attitude raciale auto-rapportée et l'effet de la cible. Autrement dit, plus les participants avaient une attitude auto-rapportée négative envers les Maghrébins et moins ils jugeaient la cible menaçante quand celle-ci était effectivement menaçante. Toutefois, dans notre

réplication de cette étude (i.e., Étude 8), comme dans notre Étude 3, nous n'avons pas reproduit cet effet. Une possibilité peut être que l'Étude 7 comportait des cibles à la fois fortement et faiblement discernables. Les participants ayant un fort préjugé se sont peut-être plus désintéressés des cibles, ces dernières donnant moins d'information qu'habituellement.

En résumé, comme pour l'AMP, la question du jugement direct de l'amorce ou de la cible reste en suspens. Il sera intéressant d'approfondir ce questionnement dans une prochaine étude et d'avoir une idée plus précise de la signification du modèle SAC-dominant. Correspond-il, comme Krieglmeyer et Sherman (2012) le supposent, à une utilisation prioritaire du stéréotype lié à l'amorce pour juger la cible ou bien au contraire à un jugement direct de l'amorce en elle-même ? Afin de tester ce raisonnement, réutiliser des conditions expérimentales spécifiques à l'AMP (e.g., proposer aux participants de passer l'essai s'ils ont l'impression de juger directement l'amorce ; Cummins et al., 2019 ; Payne et al., 2013) pourrait s'avérer pertinent.

5.3.3 Indice de fiabilité de la SMT et de validité des modèles associés

Nous avons vu au Chapitre 1 que les mesures indirectes pouvaient être hétérogènes en termes de fiabilité, certaines ayant montré des indices satisfaisants (e.g., AMP ; IAT), là où d'autres ont souvent été critiquées pour leur faible fiabilité (e.g., EPT ; Gawronski & De Houwer, 2014 ; Gawronski & Hahn, 2019). Pourtant, une bonne fiabilité de mesure est primordiale afin de ne pas sous-estimer les possibles corrélations entre notre mesure et d'autres outils de mesures (e.g., corrélation entre mesures indirecte et directe), mais aussi pour maximiser nos chances d'observer les effets de nos différentes manipulations (e.g., De Schryver et al., 2018). En d'autres termes, il faut nous assurer qu'une erreur de mesure importante n'ait pas été préjudiciable à l'observation des effets que nous attendions.

Krieglmeyer et Sherman (2012), afin de s'assurer de la bonne fiabilité du paradigme SMT, ont corrélé entre eux la moitié des essais composant la SMT avec l'autre moitié de ces essais (i.e., *split half correlation*). Plus précisément, ces auteurs ont tout d'abord calculé un score (i.e., effet SMT) différent pour chacun des deux blocs constituant la tâche, et ce, pour leurs deux premières études. Ainsi, pour chacun de ces blocs, ces auteurs ont calculé la différence de jugement entre les essais comportant des amorces Noir et les essais comportant des amorces Blanc. Krieglmeyer et Sherman ont ensuite effectué une corrélation entre ces

deux scores en appliquant une correction de Spearman-Brown, celle-ci ayant pour rôle de corriger la division par deux du nombre d'items (*split half*), entrainant une sous-estimation de la fiabilité réelle de la mesure dans son ensemble (e.g., De Schryver et al., 2018). Ces auteurs ont observé que la fiabilité de la SMT était forte, la corrélation entre ces deux mesures étant élevée dans leur première étude (*r* = .92) comme dans leur seconde (*r* = .90).

Afin d'estimer la fiabilité de notre propre SMT (i.e., SMT-Maghrébin/Blanc) nous avons entrepris, nous aussi, de calculer un indice de fiabilité pour chacune de nos études. La littérature n'étant pas homogène sur la manière la plus optimale de calculer l'indice de fiabilité (voir De Schryver et al., 2018), nous avons suivi la démarche de Krieglmeyer et Sherman (2012) afin de pouvoir comparer notre tâche SMT à la tâche SMT classique. Le Tableau 1 reporte ainsi, pour nos huit études (i.e., l'Étude 4 utilisant un paradigme IAT et l'Étude 10 étant une analyse intégrative), la corrélation entre effets SMT de chacun des deux blocs avant et après avoir appliqué une correction de Spearman-Brown. Comme nous pouvons le voir, bien que la fiabilité de notre mesure semble être en deçà de celle de la tâche originale, notre mesure présente toutefois une relativement bonne fiabilité.

Tableau 1

Indice de fiabilité (r) des Études 1 à 9 avec et sans correction de Spearman-Brown.

	Indice de fiabilité (r) Sans correction de Spearman-Brown	Indice de fiabilité (r) Avec correction de Spearman-Brown
Étude 1	.70	.82
Étude 2	.75	.86
Étude 3	.57	.73
Étude 5	.67	.81
Étude 6	.60	.75
Étude 7	.67	.80
Étude 8	.51	.69[24]
Étude 9	.85	.92

[24] L'Étude 8 comprenant à la fois des visages et des prénoms comme stimuli, nous avons aussi décomposé l'indice de fiabilité en fonction du type d'amorce. Ainsi, pour les visages la fiabilité était *r* = .51 sans correction et *r* = .68 avec correction de Spearman-Brown. Pour les prénoms, la fiabilité était *r* = .61 sans correction et *r* = .76 avec correction de Spearman-Brown.

Comme nous l'avons décrit en détail au Chapitre 1, un point important dans l'utilisation des modèles multinomiaux est la validité des paramètres du modèle. Pour être utilisé, un modèle multinomial doit passer par une étape de validation rigoureuse de l'ensemble de ses paramètres, entre autres au niveau de la validité convergente (i.e., est-ce que ma manipulation expérimentale influence bien le paramètre concerné) et de la validité divergente (i.e., est-ce que ma manipulation expérimentale n'influence pas les autres paramètres du modèle).

Habituellement, les auteurs analysent leurs données de manière agrégée entre participants, c'est-à-dire sur la somme totale de réponses correctes ou incorrectes de l'ensemble des participants pour chaque catégorie (e.g., Moshagen, 2010 ; Singmann & Kellen, 2013). Cependant, comme nous l'avons expliqué en détail, utiliser ce type d'analyse présuppose que l'estimation des paramètres est homogène d'un participant à l'autre (Heck et al., 2018 ; Klauer, 2010). Une violation de cette homogénéité peut entrainer une estimation biaisée des différents paramètres ainsi que des intervalles de confiances associés (Heck et al., 2018 ; Klauer et al., 2011 ; Smith & Batchelder, 2010). Comme nous l'avons vu, la SMT ne satisfait pas à cette homogénéité, raison pour laquelle nous avons analysé nos données selon la méthode *latent-trait*. Que penser alors de l'étude originale de Krieglmeyer et Sherman (2012) et de l'étape de validation des paramètres réalisées sur des données agrégées ? La validité des paramètres du modèle serait-elle la même avec des analyses prenant en compte l'hétérogénéité d'un participant à l'autre ?

Afin de répondre à cette question, nous avons réanalysé les données de Krieglmeyer et Sherman (2012) et plus particulièrement des quatre études de validation des paramètres liés à la SMT (i.e., D, G, SAC et SAP). Nous avons effectué nos analyses selon la méthode *latent-trait*, à la fois pour le modèle SAC-dominant (modèle d'intérêt dans les étapes de validation originales) et pour le modèle D-dominant (modèle largement réutilisé dans nos études). Afin de ne pas alourdir la lecture du manuscrit, l'intégralité des estimations des différents paramètres est détaillée en annexe (Annexe 7). La première étude concernait la validation du paramètre D, celui-ci devant être plus important lorsque les cibles étaient fortement discernables que lorsqu'elles étaient faiblement discernables. Effectivement, nous avons obtenu une différence significative du paramètre D entre ces deux conditions pour le modèle SAC-dominant ($p_B < .001$) et une différence non significative, mais proche du seuil pour le modèle D-dominant ($p_B = .06$). De plus, ces manipulations n'ont pas impacté les autres

paramètres de ces modèles[25]. La deuxième étude concernait la validation du paramètre G, celui-ci devant être plus important lorsque les participants devaient répondre « Plus menaçant » quand ils n'étaient pas certains de leur réponse que lorsqu'ils devaient répondre « Moins menaçant ». Effectivement, à la fois pour le modèle SAC-dominant (p_B < .001) et pour le modèle D-dominant (p_B < .001) le paramètre G variait significativement d'une condition à l'autre. Cependant, un autre paramètre (i.e., SAC) variait lui aussi significativement d'une condition à l'autre, et ce, à la fois pour le modèle SAC-dominant (p_B = .02), comme pour le modèle D-dominant (p_B = .002). La troisième étude était spécifique à la validation du paramètre SAC, celui-ci devant être plus important quand les amorces étaient prototypiques de la catégorie associée que lorsqu'elles ne l'étaient pas. Dans le modèle SAC-dominant (p_B < .001), comme dans le modèle D-dominant (p_B = .02), le paramètre SAC variait significativement d'une condition à l'autre. Cependant, de manière surprenante, le paramètre SAP variait lui aussi significativement dans le modèle SAC-dominant (p_B < .001) et dans le modèle D-dominant (p_B < .001). De même, le paramètre G variait lui aussi significativement à la fois pour le modèle SAC-dominant (p_B < .001) et pour le modèle D-dominant (p_B = .03). Enfin, la dernière étude était consacrée à la validation du paramètre SAP, celui-ci devant être plus important quand il était demandé aux participants d'appliquer leur stéréotype par rapport à une condition sans instruction, et plus faible quand il était demandé aux participants de corriger leur stéréotype par rapport à la condition sans instructions. De manière similaire à l'analyse originale, le paramètre SAP variait uniquement de manière significative pour le contraste opposant la condition de correction et celle sans instruction, et ce, pour le modèle SAC-dominant (p_B < .001) comme pour le modèle D-dominant (p_B < .001). Le contraste opposant la condition d'application et celle sans instruction n'était pas significatif à la fois dans le modèle SAC-dominant (p_B = .13) et dans le modèle D-dominant (p_B = .13). En outre, l'ensemble des autres paramètres ne variaient pas en fonction des conditions.

À la vue de ces résultats, nous pouvons tout d'abord avancer que, dans la plupart des cas, les preuves d'une validité convergente des paramètres du modèle sont relativement solides (ou seulement quelque peu affaiblies). A l'inverse, et c'est là un bémol dans l'utilisation de ce paradigme, les preuves d'une validité divergente peuvent être remises en causes dans

[25] Nous pourrions ici critiquer le type de méthode employée pour conclure à une validité divergente des paramètres. En effet, montrer comme ici un effet de la manipulation sur un paramètre « A » (i.e., effet significatif) mais pas sûr un autre « B » (i.e., effet non significatif) peut masquer une différence non significative entre l'effet de la manipulation sur A et l'effet de la manipulation sur B (Gelman & Stern, 2006). Cependant, comme nous avons pu le voir dans le Chapitre 1, ceci constitue la méthode classique de validation des paramètres.

une certaine mesure (e.g., paramètre SAC). Ainsi, ces résultats nous poussent à être d'autant plus prudents dans l'interprétation des paramètres du modèle et de l'impact de nos manipulations sur ces paramètres. En outre, ce problème n'est certainement pas limité à la SMT, mais peut-être appliqué plus largement dans la littérature. En effet, la majorité des phases de validation de ces modèles l'ont été sur des données agrégées, donc potentiellement soumises aux mêmes interrogations (e.g., Quad Model, Conrey et al., 2005). Des modèles multinomiaux comme le Quad Model ont cependant pu montrer de nombreuses preuves de validité convergente, divergente ou prédictive, multipliant ainsi les apports empiriques quant à la validité de ce modèle (e.g., Conrey et al., 2005; Sherman et al., 2008). De même, mais dans une moindre mesure, le modèle SMT a aussi été réutilisé plusieurs fois et a pu montrer des résultats cohérents dans l'ensemble de ces études (Rees, Ma, et al., 2019 ; Rees, Rivers, et al., 2019 ; Rivers et al., 2020). Comme nous allons le voir, nous avons nous-même pu participer à l'apport de preuves empiriques participant à la validation de ce modèle.

Tout au long de ce manuscrit, nous avons testé la validité empirique du modèle associé au paradigme SMT. Ainsi, nous avons montré que manipuler le caractère discernable de la cible influençait effectivement la valeur du paramètre D sans influencer les autres paramètres du modèle (Étude 6). Concernant le caractère prototypique de l'amorce, nous avons montré que cette manipulation semblait influencer la valeur du paramètre SAC, en accord avec Krieglmeyer et Sherman (2012), mais aussi et surtout la valeur du paramètre SAP, ceci allant à l'encontre des prédictions de Krieglmeyer et Sherman (2012 ; Études 6 et 7). De plus, nous avons pu mettre en avant la validité prédictive de nos paramètres. Le paramètre SAP était ainsi corrélé plusieurs fois avec une mesure directe d'attitudes raciales (Dambrun & Guimond, 2001 ; Études 3, 6 et 7) ou une mesure directe du stéréotype (Étude 8). Pour sa part, le paramètre D a montré des résultats plus mitigés, celui-ci étant corrélé à la mesure directe d'attitudes raciales dans notre Étude 6, mais pas dans notre Étude 7. Afin d'être complet, il nous faudrait toutefois apporter d'autres éléments de validation des paramètres, et ce notamment, pour le modèle utilisant le paramètre λ. Une prochaine étude aura pour rôle de valider empiriquement ce paramètre. Effectivement, en manipulant les consignes données aux participants, nous tenterons de montrer que le fait d'orienter nos participants à une utilisation prioritaire de l'amorce influence de manière positive le paramètre λ, alors qu'orienter nos participants à une utilisation prioritaire de la cible influence de manière négative le paramètre λ.

5.3.4 Et si l'activation du stéréotype était inversée ?

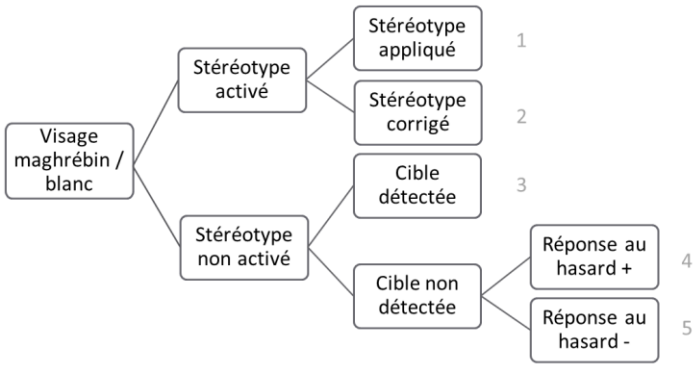
L'exemple illustrant peut-être le plus la nécessité d'avoir une théorisation forte dans l'utilisation des modèles multinomiaux est l'exemple suivant. Dans l'ensemble de nos modèles, et dans la plupart des modèles multinomiaux utilisés dans le champ des attitudes raciales, le paramètre d'activation (e.g., SAC, AC) correspond à l'activation d'un contenu négatif lié au groupe de bas statut social (e.g., « Maghrébins = Menace ») et/ou à l'activation d'un contenu positif lié au groupe de haut statut social (e.g., « Blancs = Non menace »). Ne peut-on toutefois pas penser que chez certains participants un stéréotype inverse soit activé (e.g., « Maghrébins = Non menace ») ? En réalité, comme nous allons le montrer maintenant, la compatibilité entre une catégorie (e.g., Maghrébins) et le stéréotype associé (e.g., menaçant) relève ici d'un postulat basé sur la littérature (e.g., Mange et al., 2016), mais ne peut pas dans un modèle en l'état être vérifiée. Nous allons prendre un exemple pour expliquer notre propos.

Dans nos études, nos participants de culture majoritaire française ont obtenu un effet SMT classique (i.e., jugement de la cible comme plus menaçante quand précédée d'un visage Maghrébin par rapport à un visage Blanc). Le modèle multinomial associé a systématiquement expliqué ces résultats par deux processus : une activation du stéréotype non nulle (SAC > 0) et une application de ce stéréotype plus importante qu'une éventuelle inhibition (SAP > .50). Suivant en cela la littérature en cognition sociale sur la perception négative des individus maghrébins (e.g., Dasgupta et al., 2009 ; Essien et al., 2017 ; Lyons et al., 2010 ; Mange et al., 2012 ; Unkelbach et al., 2008), nous avons interprété ces résultats comme étant le reflet d'une activation du stéréotype « Maghrébins = Menace » / « Blancs = Non menace » et d'une application de ce même stéréotype.

Dans notre dernière étude, nous avons analysé les données SMT de nos participants se catégorisant de culture principale maghrébine. Ces participants ont eux aussi exprimé un effet SMT classique (effet SMT anti-Maghrébins). Le modèle multinomial a expliqué ces résultats par une activation du stéréotype non nulle (i.e., SAC > 0) et une application de ce stéréotype plus importante qu'une éventuelle inhibition (SAP > .50). Une fois encore, nous avons interprété ces résultats comme étant le reflet d'une activation du stéréotype « Maghrébins = Menace » / « Blancs = Non menace » chez nos participants de culture maghrébine et d'une application de ce stéréotype. Plus particulièrement, nous avons interprété ces résultats comme pouvant suggérer une sensibilité de nos participants cibles du stéréotype à ce même stéréotype. En d'autres termes, nos participants de culture maghrébine,

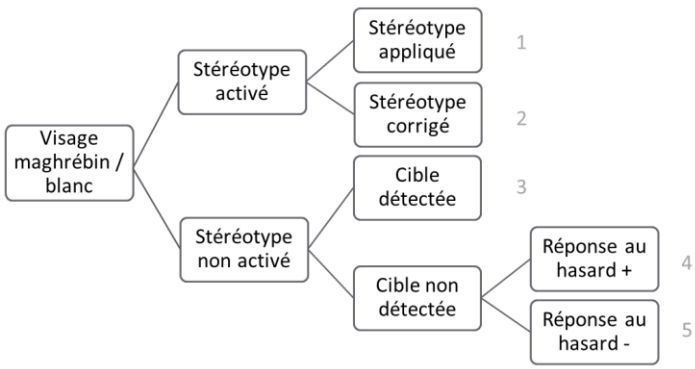
comme nos participants de culture française, réactiveraient le stéréotype culturel classique liant les Maghrébins à une menace.

Toutefois, comment être sûr que nos participants de culture maghrébine aient effectivement activé ce stéréotype et pas un stéréotype inverse ? Autrement dit, et comme nous allons le développer plus bas, ne peut-on théoriquement pas aussi envisagé que nos participants de culture maghrébine aient plutôt eu une activation d'un contenu mental « Maghrébins = Non menace » / « Blancs = Menace » ? Regardons les prédictions d'un tel modèle (Figure 41b) et comparons-le au modèle classique (Figure 41a). En observant les prédictions de ces deux modèles, nous voyons que les seuls changements se situent au niveau de la signification associée aux Branches 1 et 2. Prenons un visage maghrébin. Dans un modèle classique, si le stéréotype est activé (SAC) et qu'il est appliqué (SAP), le modèle prédira une réponse « Plus menaçant » (Branche 1). Nous voyons que ces prédictions correspondent à la Branche 2 du modèle inversé, c'est-à-dire si le participant active son stéréotype (SAC) et qu'il contrôle ce stéréotype (1-SAP). Dans un modèle classique, si le stéréotype est activé (SAC) et que le participant contrôle ce stéréotype (1-SAP), le modèle prédira une réponse « Moins menaçant » (Branche 2). Nous voyons que, cette fois-ci, ces prédictions correspondent à la Branche 1 du modèle inversé, c'est-à-dire si le participant activé son stéréotype (SAC) et qu'il applique son stéréotype (SAP). En somme, même si le paramètre SAC n'a pas la même signification entre le modèle classique (i.e., SAC = « Maghrébins = Menace » / « Blancs = Non menace ») et le modèle inversé (i.e., « Maghrébins = Non menace » / « Blancs = Menace »), les deux modèles font les mêmes prédictions, inversant juste la Branche 1 et la Branche 2. Voyons comment cela se matérialise avec un exemple chiffré.



Maghrébin		Blanc	
Bas	Haut	Bas	Haut
+	+	-	-
-	-	+	+
-	+	-	+
+	+	+	+
-	-	-	-

Figure 41a. Représentation en arbre du modèle SAC-dominant classique.



Maghrébin		Blanc	
Bas	Haut	Bas	Haut
-	-	+	+
+	+	-	-
-	+	-	+
+	+	+	+
-	-	-	-

Figure 41b. Représentation en arbre du modèle SAC-dominant inversé.

Le Tableau 2 compare les estimations du modèle SAC-dominant classique et du modèle SAC-dominant inversé pour notre Étude 10, c'est-à-dire pour nos participants de culture maghrébine. Comme nous pouvons le voir, et si nous laissons de côté des différences minimes liées à l'incertitude dans l'estimation des paramètres, l'estimation des paramètres D, G et SAC sont identiques. En réalité, l'estimation du paramètre SAP est elle aussi la même, mais juste inversée (symétrique par rapport à .50). Cela est logique, comme nous l'avons montré, les modèles classiques et inversés font les mêmes prédictions, le paramètre SAP étant

simplement opposé d'un modèle à l'autre. En d'autres termes, la prédiction d'une activation du stéréotype « Maghrébins = Menace » / « Blancs = Non menace » (i.e., SAC = .67) et d'une application de ce stéréotype (i.e., SAP = .61) dans le modèle classique sera « exactement » la même prédiction que celle de l'activation du stéréotype « Maghrébins = Non menace » / « Blancs = Menace » (i.e., SAC = .68) et du contrôle de ce stéréotype (i.e., 1 – SAP = .61) dans le modèle inversé. En somme, ces deux modèles font exactement les mêmes prédictions pour l'ensemble des différents paramètres, seules la signification donnée à ces paramètres change. Cela est confirmé par l'adéquation du modèle aux données, qui présente une estimation T_1 similaire des modèles SAC-dominant classique (M_{pred} = 0.07, M_{obs} = 0.10, p = .30) et SAC-dominant inversé (M_{pred} = 0.07, M_{obs} = 0.10, p = .31). Tout cela est encore une fois logique : les deux modèles ont la même structure, la signification du paramètre SAP étant simplement inversée d'un modèle à l'autre.

Tableau 2

Valeurs de paramètres prédites par les modèles SAC-dominant classique et SAC-dominant inversé.

	SAC-dominant (classique)	SAC-dominant (inversé)
D	.04 [.01; .08]	.04 [.01; .08]
G	.22 [.16 ;.28]	.22 [.16 ;.28]
SAC	.67 [.39 ; .90]	.68 [.38 ; .91]
SAP	.61 [.52 ; .69]	.39 [.31 ; .47]

Note. Les valeurs entre crochets représentent les intervalles de crédibilité à 95%.

Ainsi, si l'on se fie à ces modèles, il est tout autant envisageable que nos participants d'origine maghrébine aient activé leur stéréotype « Maghrébins = Menace » / « Blancs = Non menace » et majoritairement appliqué ce stéréotype, que ces participants aient en réalité activé le stéréotype « Maghrébins = Non menace » / « Blancs = Menace » et majoritairement contrôlé ce stéréotype. Cela nous amène à notre mise en garde : pour utiliser ces modèles, les auteurs doivent avoir une théorisation claire des processus à l'œuvre et envisager la probabilité qu'une hypothèse alternative soit plus vraisemblable. Dans nos études, et surtout pour des individus se catégorisant comme de culture française, l'hypothèse alternative semblait moins probable que notre hypothèse originale. Mais est-ce toujours le cas ? Pour nos

participants de culture maghrébine, l'hypothèse originale est-elle beaucoup plus probable que l'hypothèse alternative ? Nous n'en sommes pas sûrs.

Reprenons une dernière fois l'exemple de Payne et collaborateurs (2019) montrant un biais pro-Noirs pour des participants Noirs au sein de l'IAT. Pour expliquer ces résultats, serait-il plus probable que ces participants aient activé un contenu mental « Noirs = Négatif » / « Blancs = Positif » et inhibé ce contenu mental réactivé (Hypothèse 1), ou au contraire qu'ils aient activé un contenu mental « Noirs = Positif » / « Blancs = Négatif » et appliqué ce contenu mental réactivé (Hypothèse 2) ? Si un premier chercheur défendant l'hypothèse 1 appliquait un modèle classique et qu'un deuxième chercheur défendant l'hypothèse 2 appliquait un modèle inversé, ceux-ci obtiendraient « exactement » les mêmes résultats, confirmant pour chacun leur hypothèse. Vraisemblablement pourtant, au moins un des deux chercheurs interpréterait ses résultats de manière erronée.

Nous invitons ainsi le lecteur à la plus grande prudence dans l'utilisation de ces modèles multinomiaux pour démêler l'impact de différents processus aboutissant à la réponse individuelle. Pour Calanchini (2020), « un modèle multinomial peut seulement apporter des résultats sensés si la théorie est correcte » (p. 5). De plus, cet auteur ajoute « et l'adéquation du modèle aux données apporte un indice quantitatif pour juger de la validité de la théorie sous-tendant le modèle » (p. 5). Nous sommes entièrement d'accord avec la première partie du message : un modèle multinomial n'a de sens que si la théorie sous-jacente est correcte. Toutefois, nous avons montré que l'adéquation d'un modèle aux données ne nous donnait pas forcément d'informations quant à la validité de la théorie sous-jacente au modèle. Afin de ne pas interpréter de manière erronée les paramètres du modèle, les auteurs voulant utiliser ces modèles doivent au préalable s'assurer qu'une hypothèse alternative dans l'interprétation de ces processus ne soit pas plus vraisemblable.

5.3.5 Choix d'un modèle à quatre paramètres ou d'un modèle à six paramètres

Enfin, revenons sur le nombre de paramètres à inclure dans le modèle et la réflexion théorique associée. Dans le Chapitre 2, nous avons mis en évidence la pertinence théorique d'utiliser un modèle SAC-dominant à six paramètres, c'est-à-dire avec un paramètre SAC et un paramètre SAP différents en fonction des visages maghrébins et des visages blancs. Comme nous l'avons vu, ce modèle est différent du modèle classique à quatre paramètres, validé par

Krieglmeyer et Sherman (2012), qui comportait un paramètre SAC et un paramètre SAP uniques pour les visages maghrébins et les visages blancs. Krieglmeyer et Sherman ont validé et réutilisé un modèle à quatre paramètres, car ces auteurs utilisaient l'indice d'ajustement G^2, indice nécessitant un plus grand nombre de catégories (i.e., visage blanc et cible menaçante) que de paramètres. Or, un des avantages à l'utilisation d'un indice d'ajustement T_1 (issu de la méthode *latent-trait*) est que celui-ci nous permet d'utiliser des modèles ayant un nombre identique de catégories et de paramètres, donc un modèle à six paramètres. Comme nous l'avons montré tout au long de ce manuscrit, utiliser un tel modèle peut être un avantage. Mais cela entraine aussi une réflexion supplémentaire.

Un modèle à six paramètres nous a permis de faire des prédictions plus précises quant à l'impact de nos manipulations (i.e., contexte d'attentats) sur les paramètres du modèle. Pour rappel, nous faisions l'hypothèse qu'en contexte d'attentats, les participants réactiveraient davantage leur stéréotype « Maghrébins = Menace » mais appliqueraient moins ce stéréotype. C'est la raison pour laquelle il nous semblait plus cohérent d'opérationnaliser ces hypothèses par des prédictions spécifiques aux paramètres reliés aux visages maghrébins (i.e., SAC_{Mg} plus important en contexte d'attentats, mais SAP_{Mg} plus faible) indépendamment des paramètres reliés aux visages blancs (SAC_{Bl} et SAP_{Bl}). Mais, avec ces mêmes hypothèses, ne pouvait-on pas également envisager une influence de notre manipulation sur les paramètres reliés aux visages blancs ?

Pour le paramètre SAC, nous pourrions de prime abord penser que si effectivement en contexte d'attentats les individus maghrébins sont plus associés au stéréotype de menace, induire un tel contexte devrait influencer le paramètre SAC lié aux visages maghrébins sans influencer le paramètre SAC lié aux visages blancs. Plus généralement, nous pourrions en effet imaginer qu'une manipulation censée impacter la réactivation d'un contenu mental lié à un type de stimulus (e.g., visages maghrébins) n'impacte pas ou en tout cas moins la réactivation du contenu mental lié à un autre type de stimulus (e.g., visages blancs). Néanmoins, la littérature a pu montrer des résultats différents. En effet, Calanchini et collaborateurs (2020) ont mis en évidence, de façon surprenante, que des interventions relatives à la diminution du biais anti-Noirs classique avaient davantage impacté la réactivation du contenu mental positif lié aux visages blancs que la réactivation du contenu mental négatif lié aux visages noirs. En réalité, ces résultats sont aussi cohérents avec certains travaux (et notre Étude 4) ayant montré que le biais IAT anti-Noirs (et dans notre cas, anti-Maghrébins) habituellement observé dans ce type de tâche semblait davantage être le reflet d'une association entre les

stimuli blancs et les mots positifs que d'une association entre les stimuli noirs (maghrébins) et les mots négatifs (Calanchini, en préparation). Nous voyons donc ici tout l'intérêt de séparer le paramètre d'activation (SAC, AC) en fonction du type de stimuli (e.g., Maghrébins vs. Blancs), ceci nous permettant une interprétation encore plus précise de l'impact de nos manipulations. Mais qu'en est-il du paramètre SAP ?

Pour le paramètre SAP, nous pensions que si un participant avait la volonté d'inhiber son stéréotype « Maghrébins = Menace », celui-ci répondrait « Moins menaçant » dans un essai composé d'une amorce Maghrébin. La proportion de jugement de menace associée aux visages maghrébins serait donc moins importante en condition Attentat qu'en Condition Contrôle et se rapprocherait donc de celle des visages blancs. Cependant, ne peut-on pas envisager une autre « stratégie » de réponse ? En effet, ne peut-on pas envisager qu'un participant voulant inhiber son stéréotype « Maghrébins = Menace » puisse répondre « Plus menaçant » dans un essai composé d'une amorce Blanc ? La différence de jugement entre les essais composés de visages maghrébins et ceux composés de visages blancs diminuerait également, mais cette fois parce que la proportion de jugement de menace associé aux visages blancs augmenterait pour se rapprocher de celles des visages maghrébins. En somme, l'inhibition du stéréotype activé sera-t-elle principalement le fruit d'une diminution du paramètre SAP lié aux visages maghrébins, d'une diminution du paramètre SAP lié aux visages blancs ou bien d'une combinaison des deux ? Ainsi, la séparation du paramètre d'inhibition (SAP, OB) en fonction du type de stimuli (e.g., Maghrébins vs. Blancs) peut, dans le cas présent, être discutable. Il peut être tout autant envisageable d'imaginer un paramètre d'inhibition (ou de contrôle) séparé entre les amorces, comme ce fut le cas pour certaines de nos analyses, qu'un paramètre d'inhibition général, comme c'est souvent le cas dans les études utilisant le Quad Model (e.g., Gonsalkorale et al., 2011, 2014).

En résumé, et comme ce ne fut peut-être pas assez le cas lors de nos études, nous pensons qu'une spécification a priori du type de modèle utilisé, c'est-à-dire du nombre précis de paramètres à inclure dans le modèle, peut permettre une interprétation des résultats au plus proche des hypothèses des chercheurs. Avant d'utiliser ces modèles, les chercheurs devraient avoir une perception relativement claire de l'impact de leurs manipulations sur les processus cognitifs à l'origine de la réponse et de la différence ou non de cet impact en fonction des stimuli présentés (e.g., visage maghrébin vs. visage blanc). Nous pouvons toutefois envisager une alternative à cette approche théorique. Une autre approche pourrait en effet consister à sélectionner le modèle s'ajustant le mieux aux données (e.g., Rees, Rivers,

et al., 2019), c'est-à-dire à tester l'ensemble des modèles envisageables (e.g., modèle avec SAC et SAP vs. modèle avec SAC_{Bl}, SAC_{Mg} et SAP vs. modèle avec SAC_{Bl}, SAC_{Mg}, SAP_{Bl} et SAP_{Mg}) et de sélectionner le modèle présentant le meilleur indice d'ajustement aux données. Dans le cas où les chercheurs privilégieraient ce type d'approche, ceux-ci devront toutefois veiller à utiliser un indice d'ajustement prenant en compte le nombre possiblement différent de paramètres au sein de chaque modèle, comme nous avons pu le faire dans ce manuscrit.

5.4 Pour conclure : des perspectives futures dans l'utilisation de ces modèles

À travers l'exemple d'une mesure indirecte du stéréotype, nous avons défendu la thèse selon laquelle les processus et l'organisation des processus à l'origine de la réponse individuelle pouvaient varier en fonction d'un contexte particulier et du type de participants prenant part à cette mesure. Plus généralement, nous avons suggéré qu'une investigation systématique des processus à l'œuvre au sein de ces mesures pouvait nourrir notre réflexion théorique quant aux interprétations que nous faisions des réponses individuelles. Nous avons montré que les modèles multinomiaux pouvaient être un outil intéressant dans l'investigation de ces processus. Plus particulièrement, les modèles multinomiaux associés à la *Stereotype Misperception Task* nous ont permis d'investiguer la possibilité que, théoriquement, un même contexte (i.e., contexte d'attentats) puisse influencer de manière contraire certains des processus jouant un rôle dans la réponse du participant (i.e., activation et application du stéréotype). De plus, grâce à ces modèles nous avons mis en évidence l'importance de la prise en compte de la relation conditionnelle entre ces processus (i.e., modèle SAC-dominant vs. D-dominant) et de leur lien avec des facteurs individuels (i.e., mesure directe d'attitudes raciales). Enfin, nous avons mis en avant l'utilité que pouvaient avoir ces modèles dans des réflexions théoriques liées à l'interprétation des résultats issus de ces mesures (e.g., influence culturelle, attitude personnelle vs. extra-personnelle). En somme, nous avons voulu présenter tout l'intérêt d'une utilisation des modèles multinomiaux dans la littérature traitant des attitudes raciales.

Apportant toutefois quelques bémols à l'usage de ces modèles, nous avons évoqué certaines contraintes ou interrogations relatives à l'utilisation des modèles multinomiaux et à l'interprétation que l'on pouvait faire de leurs résultats. Comme ce fut le cas pour les mesures indirectes, nous avons mis en lumière l'importance d'un raisonnement théorique fort associé à

l'utilisation de ces modèles afin de ne pas interpréter nos résultats de manière erronée. Nous avons de plus montré qu'en l'état, ces modèles pouvaient être limités et ne permettaient pas de répondre à l'ensemble de nos questionnements théoriques. Ces modèles n'ont pas pour vocation d'expliquer l'intégralité des processus cognitifs à l'œuvre, mais simplement l'impact des principaux processus reliés à la réponse du participant. Enfin, nous avons évoqué la validation parfois incomplète des paramètres du modèle et les précautions à avoir quant à la méthode d'analyse des résultats. De manière globale, nous invitons le lecteur à se montrer prudent quant à l'utilisation de ces modèles et aux conclusions qu'ils permettent de tirer. En ce sens, nous nous plaçons dans le prolongement de notre message principal. Toutefois, cela ne doit en aucun cas masquer l'apport de ces modèles.

Arrivés à la conclusion de ce manuscrit, nous espérons avoir répondu à la question suivante : au final, qu'apportent réellement les modèles multinomiaux à nos recherches ? Même si ces modèles ne sont pas parfaits, même si certaines limites relatives aux mesures indirectes s'appliquent également à ces modèles, ceux-ci ont le mérite de participer au développement d'un questionnement toujours plus fin quant à nos mesures, ainsi qu'à une perpétuelle remise en question de nos interprétations. Les travaux présentés dans ce manuscrit, nos propres travaux ou ceux tirés de la littérature, ont illustré une petite partie de l'utilisation possible de ces modèles. Mais nous sommes persuadés que ce type de modèles constitue, à l'heure actuelle, un outil encore trop inexploité. Pourtant, les revues de littérature ne manquent pas (e.g., Calanchini, 2020 ; Calanchini et al., 2018 ; Hütter & Klauer, 2016 ; Payne & Bishara, 2009 ; Sherman et al., 2010), des logiciels relativement faciles d'accès sont à disposition (e.g., Heck et al., 2018 ; Mosaghen, 2010 ; Singmann & Kellen, 2013) et de nouveaux modèles exploitant les temps de réponse font peu à peu leur apparition (Hartmann et al., 2020 ; Heck & Erdfelder, 2016 ; Klauer & Kellen, 2018). Au regard de la popularité dont jouissent aujourd'hui les mesures indirectes, les perspectives d'utilisations de ces modèles sont nombreuses. Nous invitons le lecteur à envisager l'utilisation des modèles multinomiaux dans leurs recherches ou, du moins, à entreprendre une réflexion quant aux processus pouvant être mis en lumière par ces modèles. Même s'ils ont déjà montré leur utilisé par le passé, les modèles multinomiaux ont encore beaucoup à apporter à la littérature traitant des stéréotypes et préjugés.

Bibliographie

Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., Kenrick, D. T., Becker, D. V., Griskevicius, V., Maner, J. K., & Schaller, M. (2006). They all look the same to me (unless they're angry) from out-group homogeneity to out-group heterogeneity. *Psychological Science*, *17*(10), 836–840. doi: 10.1111/j.1467-9280.2006.01790.x

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harpers.

Ahluwalia, M. K., & Pellettiere, L. (2010). Sikh men post-9/11: Misidentification, discrimination, and coping. *Asian American Journal of Psychology*, *1*(4), 303-314. doi: 10.1037/a0022156

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179-211. doi: 10.1016/0749-5978(91)90020-T

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*(5), 888-918. doi: 10.1037/0033-2909.84.5.888

Ajzen, I., & Fishbein, M. (2000). Attitudes and the attitude-behavior relation: Reasoned and automatic processes. *European Review of Social Psychology*, *11*(1), 1-33. doi: 10.1080/14792779943000116

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723. doi: 10.1109/tac.1974.1100705

Al Ramiah, A., Hewstone, M., Dovidio, J. F., & Penner, L. A. (2010). The social psychology of discrimination: Theory, measurement and consequences. In L. Bond, F. McGinnity, & H. Russel (Eds). *Making Equality Count: Irish and International Research Measuring Equality and Discrimination*. Dublin, Ireland: Liffety Press.

Allen, T. J., Sherman, J. W., & Klauer, K. C. (2010). Social context and the self-regulation of implicit bias. *Group Processes & Intergroup Relations*, *13*(2), 137-149. doi: 10.1177/1368430209353635

Allport, G. W. (1935). *Attitudes: A handbook of social psychology*. Clark University Press, Worcester.

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

Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. San Francisco: Jossey-Bass.

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*(4), 652-661. doi: 10.1037/0022-3514.91.4.652

Arciszewski, T., Verlhiac, J.-F., Goncalves, I., & Kruglanski, A. (2009). From psychology of terrorists to psychology of terrorism. *International Review of Social Psychology*, *Tome 22*(3), 5-34.

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or" would Jesse Jackson 'fail' the Implicit Association Test?". *Psychological Inquiry*, *15*(4), 257–278. doi: 10.1207/s15327965pli1504_01

Arnoult, M., Selimbegović, L., & Chatard, A. (2019, Février). « You will not have my hate. or will you ? ». When terror in paris spreads implicit anti-Muslim attitudes around the world. Poster présenté à la *Society for Personality and Social Psychology (SPS*P), Portland, États-Unis.

Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. (2003). Black Americans' implicit racial associations and their implications for intergroup judgment. *Social Cognition*, *21*(1), 61-87. doi: 10.1521/soco.21.1.61.21192

Atkinson, A. C. (1970). A method for discriminating between models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *32*(3), 323-345. doi:10.1111/j.2517-6161.1970.tb00845.x

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*(3), 136-141. doi: 10.1111/j.1467-9280.1996.tb00346.x

Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2004). No Place for Nostalgia in Science: A Response to Arkes and Tetlock. *Psychological Inquiry*, *15*(4), 279-289. doi: 10.1207/s15327965pli1504_02

Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the Affective Misattribution Procedure. *Personality and Social Psychology Bulletin*, *38*(9), 1194-1208. doi: 10.1177/0146167212446835

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology*, *87*(1), 5-22. doi: 10.1037/0022-3514.87.1.5

Bargh, J. A. (1994). The four horsemen of automaticity: intention, awareness, efficiency, and control as separate issues. In R. Wyer & T. Srull (Eds.), *Handbook of Social Cognition*: *Vol 1*., *Basic Processes* (2nd edn, pp. 1-40), Erlbaum.

Bar-Tal, D., & Labin, D. (2001). The effect of a major event on stereotyping: Terrorist attacks in Israel and Israeli adolescents' perceptions of Palestinians, Jordanians and Arabs. *European Journal of Social Psychology*, *31*(3), 265-280. doi: 10.1002/ejsp.43

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57-86. doi: 10.3758/BF03210812

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). Parsimonious Mixed Models. *arXiv:1506.04967* http://arxiv.org/abs/1506.04967

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*(5), 1252-1265. doi: 10.1037/0022-3514.74.5.1252

Benson, P. L., Karabenick, S. A., & Lerner, R. M. (1976). Pretty pleases: The effects of physical attractiveness, race, and sex on receiving help. *Journal of Experimental Social Psychology*, *12*(5), 409-415. doi: 10.1016/0022-1031(76)90073-1

Bianchi, M., Carnaghi, A., & Shamloo, S. E. (2018). Intergroup attitudes accessibility and motor approach-avoidance responses in White and Black individuals in Portugal. *Psicologia sociale*, *2*, 147-164. doi: 10.1482/90780

Bishara, A. J., & Payne, B. K. (2009). Multinomial process tree models of control and automaticity in weapon misidentification. *Journal of Experimental Social Psychology*, *45*(3), 524-534. doi: 10.1016/j.jesp.2008.11.002

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of afrocentric facial features in criminal sentencing. *Psychological Science*, *15*(10), 674-679. doi: 10.1111/j.0956-7976.2004.00739.x

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology*, *83*(1), 5-25. doi: 10.1037//0022-3514.83.1.5

Bodenhausen, G. V., & Richeson, J. A. (2010). Prejudice, stereotyping, and discrimination. In R. F. Baumeister & E. J. Finkel (Eds.), *Advanced social psychology: The state of the science* (pp. 341–383). Oxford University Press.

Bogardus, E. S. (1947). Measurement of Personal-Group Relations. *Sociometry*, *10*(4), 306. doi: 10.2307/2785570

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition (*Vol. 1, pp. 1-36). Lawrence Erlbaum Associates, Inc.

Brigham, J. C. (1971a). Ethnic stereotypes. *Psychological Bulletin*, *76*(1), 15-38. doi: 10.1037/h0031446

Brigham, J. C. (1971b). Racial stereotypes, attitudes, and evaluations of and behavioral intentions toward Negroes and Whites. *Sociometry*, *34*(3), 360. doi: 10.2307/2786204

Brigham, J. C. (1973). Ethnic stereotypes and attitudes: A different mode of analysis. *Journal of Personality*, *41*(2), 206-223. doi: 10.1111/j.1467-6494.1973.tb00089.x

Brown, R. (2011). *Prejudice: Its social psychology*. John Wiley & Sons.

Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General*, *124*(2), 137-160. doi: 10.1037/0096-3445.124.2.137

Calanchini, J. (en préparation). The contributions of positive outgroup and negative ingroup evaluation to implicit bias favoring outgroups.

Calanchini, J. (2020). How multinomial processing trees have advanced, and can continue to advance, research using implicit measures. *Social Cognition.* Manuscrit accepté pour publication.

Calanchini, J., Lai, C. K., & Klauer, K. C. (2020). Reducing implicit racial preferences: III. A process-level examination of changes in implicit preferences. *Journal of Personality and Social Psychology*. doi: 10.1037/pspi0000339

Calanchini, J., Rivers, A. M., Klauer, K. C., & Sherman, J. W. (2018). Multinomial processing trees as theoretical bridges between cognitive and social psychology. *Psychology of Learning and Motivation*, 39-65. doi: 10.1016/bs.plm.2018.09.002

Calanchini, J., Sherman, J., Klauer, K. C., & Ferrer, E. (2017). *An empirical validation of the Quadruple Process Model of implicit attitudes against alternate, theoretically defensible specifications.* PsyArXiv. doi: 10.31234/osf.io/7b6nz

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, *16*(4), 330-350. doi: 10.1177/1088868312440047

Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science*, *9*(1), 40-48. doi: 10.1177/1745691613513470

Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds), *Social influence: The Ontario symposium* (Vol. 5, pp. 3-39). Hillsdale, NJ: Erbaum.

Cheon, B. K., Melani, I., & Hong, Y. (2020). How USA-centric is psychology? An archival study of implicit assumptions of generalizability of findings to human nature based on origins of study samples. *Social Psychological and Personality Science*, *11*(7), 928-937. doi: 10.1177/1948550620927269

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155-159. doi: 10.1037/0033-2909.112.1.155

Cohu, M., Maisonneuve, C., & Testé, B. (2016). The "Charlie-Hebdo" effect: Repercussions of the january 2015 terrorist attacks in France on prejudice toward immigrants and North-Africans, social dominance orientation, and attachment to the principle of *Laïcité*. *International Review of Social Psychology*, *29*(1), 50. doi: 10.5334/irsp.59

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The Quad Model of implicit task performance. *Journal of Personality and Social Psychology*, *89*(4), 469-487. doi: 10.1037/0022-3514.89.4.469

Corneille, O., & Hütter, M. (2020). Implicit? What do you mean? A comprehensive review of the delusive implicitness construct in attitude research. *Personality and Social Psychology Review*,. doi: 10.1177/1088868320911325

Correll, J., Judd, C. M., Park, B., & Wittenbrink, B. (2010). Measuring prejudice, stereotypes and discrimination. In J. F. Dovidio, M. Hewstone, P. Glick, & V.M. Esses (Eds), *The SAGE handbook of prejudice, stereotyping, and discrimination* (pp. 45–62). SAGE Publications doi: 10.4135/9781446200919.n3

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*(6), 1314-1329. doi: 10.1037//0022-3514.83.6.1314

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to « prejudice ». *Journal of Personality and Social Psychology*, *88*(5), 770-789. doi: 10.1037/0022-3514.88.5.770

Courset, R. (2017). *Amalgame Maghrébin-terroriste : Conséquences des attentats sur les processus implicites* [Manuscrit de mémoire non publié]. Université Grenoble Alpes.

Courset, R., Rougier, M., Palluel-Germain, R., Smeding, A., Jonte, J. M., Chauvin, A., & Muller, D. (2018). The Caucasian and North African French Faces (CaNAFF): A face database. *International Review of Social Psychology*, *31*(1). doi: 10.5334/irsp.179

Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review. *Psychological Bulletin*, *87*(3), 546-563. doi: 10.1037/0033-2909.87.3.546

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349-354. doi: 10.1037/h0047358

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from Intergroup Affect and Stereotypes. *Journal of Personality and Social Psychology*, *92*(4), 631-648. doi: 10.1037/0022-3514.92.4.631

Cummins, J., Hussey, I., & Hughes, S. (2019). *The AMPeror's New Clothes: Performance on the Affect Misattribution Procedure is Mainly Driven by Awareness of Influence of the Primes*. PsyArXiv. doi: 10.31234/osf.io/d5zn8

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*(2), 81-100. doi: 10.1037/a0015914

Dambrun, M., & Guimond, S. (2001). La théorie de la privation relative et l'hostilité envers les Nord-Africains. *International Review of Social Psychology*, *14*(1), 57-89.

Das, E., Bushman, B. J., Bezemer, M. D., Kerkhof, P., & Vermeulen, I. E. (2009). How terrorism news reports increase prejudice against outgroups: A terror management account. *Journal of Experimental Social Psychology*, *45*(3), 453-459. doi: 10.1016/j.jesp.2008.12.001

Dasgupta, N., DeSteno, D., Williams, L. A., & Hunsinger, M. (2009). Fanning the flames of prejudice: The influence of specific incidental emotions on implicit prejudice. *Emotion*, *9*(4), 585-591. doi: 10.1037/a0015961

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*(5), 800-814. doi: 10.1037/0022-3514.81.5.800

De Houwer, J. (2003). The Extrinsic Affective Simon Task. *Experimental Psychology*, *50*(2), 77-85. doi: 10.1026//1618-3169.50.2.77

De Houwer, J., & Moors, A. (2010). Implicit measures: Similarities and differences. In B. Gawronski & B. K. Payne (Eds), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 176-193). New York, NY: Guilford Press.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*(6), 853-869. doi: 10.1037/0033-2909.127.6.853

De Schryver, M., Hughes, S., De Houwer, J., & Rosseel, Y. (2018). On the reliability of implicit measures: Current practices and novel perspectives. *A psychometric analysis of choice reaction time measures*. PsyArXiv. doi: 10.31234/osf.io/w7j86

Degner, J., Essien, I., & Reichardt, R. (2016). Effects of diversity versus segregation on automatic approach and avoidance behavior towards own and other ethnic groups: Diversity and automatic intergroup behavior. *European Journal of Social Psychology*, *46*(6), 783-791. doi: 10.1002/ejsp.2234

Deutsch, R., & Strack, F. (2006). Duality models in social psychology: From dual processes to interacting systems. *Psychological Inquiry*, *17*(3), 166-172. doi: 10.1207/s15327965pli1703_2

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5-18. doi: 10.1037//0022-3514.56.1.5

Devine, P. G., & Monteith, M. J. (1999). Automaticity and control in stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 339–360). The Guilford Press.

Dion, K. L., & Dion, K. K. (1987). Belief in a just world and physical attractiveness stereotyping. *Journal of Personality and Social Psychology*, *52*(4), 775-780. doi: 10.1037/0022-3514.52.4.775

Doosje, B., Zimmermann, A., Küpper, B., Zick, A., & Meertens, R. (2009). Terrorist threat and perceived islamic support for terrorist attacks as predictors of personal and institutional out-group discrimination and support for anti-immigration policies – evidence from 9 European countries. *International Review of Social Psychology*, *22*(3), 203-233.

Dotsch, R., Wigboldus, D. H. J., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, *19*(10), 978-980. doi: 10.1111/j.1467-9280.2008.02186.x

Dovidio, J. F., Brigham, J. C., Johnson, B. T., & Gaertner, S. L. (1996). Stereotyping, prejudice, and discrimination: Another look. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 276-319). New York: Guilford Press.

Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping and discrimination: Theoretical and empirical overview. In *The SAGE Handbook of Prejudice, Stereotyping and Discrimination* (pp. 3-28). London, England: Sage. doi: 10.4135/9781446200919.n1

Dumont, M., Yzerbyt, V., Wigboldus, D., & Gordijn, E. H. (2003). Social categorization and fear reactions to the september 11th terrorist attacks. *Personality and Social Psychology Bulletin*, *29*(12), 1509-1520. doi: 10.1177/0146167203256923

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*(3), 316–326. doi: 10.1177/0146167297233009

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* Fort Worth, TX: Harcourt brace.

Eagly, A. H., & Diekman, A. B. (2005). What is the problem? Prejudice as an attitude-in-context. In J. F. Dovidio, P. Glick, & L. Rudman (Eds.), *On the nature of prejudice: Fifty years after Allport* (pp. 19-35). Malden, MA: Blackwell.

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., … Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, *67*, 68-82. doi: 10.1016/j.jesp.2015.10.012

Echebarria-Echabe, A., & Fernández-Guede, E. (2006). Effects of terrorism on attitudes and ideological orientation. *European Journal of Social Psychology*, *36*(2), 259-265. doi: 10.1002/ejsp.294

Essien, I., Calanchini, J., & Degner, J. (2020). Moderators of intergroup evaluation in disadvantaged groups: A comprehensive test of predictions from system justification theory. *Journal of Personality and Social Psychology*. doi: 10.1037/pspi0000302

Essien, I., Stelter, M., Kalbe, F., Koehler, A., Mangels, J., & Meliß, S. (2017). The shooter bias: Replicating the classic effect and introducing a novel paradigm. *Journal of Experimental Social Psychology*, *70*, 41-47. doi: 10.1016/j.jesp.2016.12.009

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*(1), 255-278. doi: 10.1146/annurev.psych.59.103006.093629

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223-241. doi: 10.1177/1745691612460685

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. doi: 10.3758/BF03193146

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE Model as an integrative framework. *Advances in Experimental Social Psychology*, 23, 75-109. doi: 10.1016/S0065-2601(08)60318-4

Fazio, R. H. (2007). Attitudes as object–evaluation associations of varying strength. *Social Cognition*, *25*(5), 603-637. doi: 10.1521/soco.2007.25.5.603

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*(6), 1013-1027. doi: 10.1037/0022-3514.69.6.1013

Fazio, R. H., & Olson, M. A. (2014). The MODE model: Attitude-behavior processes as a function of motivation and opportunity. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 155–171). New York, NY: Guilford Press.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*(2), 229-238. doi: 10.1037/0022-3514.50.2.229

Fetzer, J. S., & Soper, J. C. (2003). The roots of public attitudes toward state accommodation of European Muslims' religious practices before and after september 11. *Journal for the Scientific Study of Religion*, *42*(2), 247-258. doi: 10.1111/1468-5906.t01-1-00176

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology*, *27*(4), 307-316. doi: 10.1207/s15324834basp2704_3

Fischer, P., Greitemeyer, T., & Kastenmüller, A. (2007). What do we think about Muslims? The validity of westerners' implicit theories about the associations between Muslims' religiosity, religious identity, aggression potential, and attitudes toward terrorism. *Group Processes & Intergroup Relations*, *10*(3), 373-382. doi: 10.1177/1368430207078697

Fischetti, A. (2020, Septembre 16). Le racisme ne fleurit pas sur le fumier des attentats. *Charlie Hebdo*.

Fishbein, H. D. (2002). *Peer prejudice and discrimination: The origins of prejudice* (2nd ed). L. Erlbaum.

Fiske, S. T. (1989). Examining the role of intent: Toward understanding its role in stereotyping and prejudice. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 253-283). New York, NY: Guilford Press

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 357–411). New York: McGraw-Hill.

Fiske, S. T. (2000). Stereotyping, prejudice, and discrimination at the seam between the centuries: Evolution, culture, mind, and brain. *European Journal of Social Psychology*, *30*(3), 299–322. doi: 10.1002/(SICI)1099-0992(200005/06)30:3<299::AID-EJSP2>3.0.CO;2-F

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878-902. doi: 10.1037/0022-3514.82.6.878

Fiske, S. T., & Neuberg, S. L. (1990). A Continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, *23*, 1-74. doi: 10.1016/S0065-2601(08)60317-2

Fiske, S. T., & Taylor, S. E. (2013). *Social cognition: From brains to culture* (2nd ed.) New York, NY: McGraw-Hill.

Foroni, F., & Bel-Bahar, T. (2009). *Picture* -IAT versus *Word* -IAT: Level of stimulus representation influences on the IAT. *European Journal of Social Psychology*, *40*(2), 321-337. doi: 10.1002/ejsp.626

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*(2), 247-279. doi: 10.1037/a0022327

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in Motion Reveals Mind in Motion. *Frontiers in Psychology*, *2*. doi: 10.3389/fpsyg.2011.00059

Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: Social status cues shape race perception. *PLOS ONE*, *6*(9), e25107. doi: 10.1371/journal.pone.0025107

Gaertner, S., & Bickman, L. (1971). Effects of race on the elicitation of helping behavior: The wrong number technique. *Journal of Personality and Social Psychology*, *20*(2), 218-222. doi: 10.1037/h0031681

Gauthier, J. (2015). Origines contrôlées: Police et minorités en France et en Allemagne. *Sociétés contemporaines*, *97*(1), 101. doi: 10.3917/soco.097.0101

Gawronski, B. (2019). Six lessons for a cogent science of implicit bias and its criticism. *Perspectives on Psychological Science*, *14*(4), 574-595. doi: 10.1177/1745691619826015

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692-731. doi: 10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE model. *Social Cognition*, *25*(5), 687–717. doi: 10.1521/soco.2007.25.5.687

Gawronski, B., & Bodenhausen, G. V. (2011). The Associative–Propositional Evaluation Model. *Advances in Experimental Social Psychology*, *44*, 59-127. doi: 10.1016/B978-0-12-385522-0.00002-0

Gawronski, B., & Bodenhausen, G. V. (2014). Implicit and explicit evaluation: A brief review of the Associative-Propositional Evaluation Model: APE Model. *Social and Personality Psychology Compass*, *8*(8), 448-462. doi: 10.1111/spc3.12124

Gawronski, B., & Bodenhausen, G. V. (2018). Evaluative conditioning from the perspective of the Associative-Propositional Evaluation Model. *Social Psychological Bulletin*, *13*(3). doi: 10.5964/spb.v13i3.28024

Gawronski, B., & Creighton, L. A. (2013). Dual-process theories. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition* (pp. 282–312). Oxford, England: Oxford University Press.

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis C. M. & Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 283–310). New York: Cambridge University Press.

Gawronski, B., & Hahn, A. (2019). Implicit measures: Procedures, use, and interpretation. In H. Blanton, J. M. LaCroix, & G. D. Webster (Eds.), *Measurement in social psychology* (pp. 29–55). Taylor & Francis.

Gawronski, B., Peters, K. R., Brochu, P. M., & Strack, F. (2008). Understanding the relations between different forms of racial prejudice: A cognitive consistency perspective. *Personality and Social Psychology Bulletin*, *34*(5), 648-665. doi: 10.1177/0146167207313729

Gawronski, B., Sherman, J. W., & Trope, Y. (2014). Two of what? A conceptual analysis of dual-process theories. In J. W. Sherman, B. Gawronski & Y. Trope (Eds.), *Dual-Process Theories of the Social Mind* (pp. 3–19). New York: Guilford Press.

Gawronski, B., & Ye, Y. (2015). Prevention of intention invention in the Affect Misattribution Procedure. *Social Psychological and Personality Science*, *6*(1), 101-108. doi: 10.1177/1948550614543029

Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician*, *60*(4), 328-331. doi: 10.1198/000313006X152649

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*(4), 509-517. doi: 10.1037/0022-3514.60.4.509

Gilbert, G. M. (1951). Stereotype persistence and change among college students. *The Journal of Abnormal and Social Psychology*, *46*(2), 245-254. doi: 10.1037/h0053696

Gonsalkorale, K., Hippel, W. von, Sherman, J. W., & Klauer, K. C. (2009). Bias and regulation of bias in intergroup interactions: Implicit attitudes toward Muslims and interaction quality. *Journal of Experimental Social Psychology*, *45*(1), 161-166. doi: 10.1016/j.jesp.2008.07.022

Gonsalkorale, K., Sherman, J. W., Allen, T. J., Klauer, K. C., & Amodio, D. M. (2011). Accounting for successful control of implicit racial bias: The roles of association activation, response monitoring, and overcoming bias. *Personality and Social Psychology Bulletin*, *37*(11), 1534-1545. doi: 10.1177/0146167211414064

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2009). Aging and prejudice: Diminished regulation of automatic race bias among older adults. *Journal of Experimental Social Psychology*, *45*(2), 410-414. doi: 10.1016/j.jesp.2008.11.004

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2014). Measures of implicit attitudes may conceal differences in implicit associations: The case of antiaging bias. *Social Psychological and Personality Science*, *5*(3), 271-278. doi: 10.1177/1948550613499239

Greenberg, J., Solomon, S., & Pyszczynski, T. (1997). Terror management theory of self-esteem and cultural worldviews: Empirical assessments and conceptual refinements. *Advances in Experimental Social Psychology*, *29*, 61‑139. doi: 10.1016/S0065-2601(08)60016-7

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4‑27. doi: 10.1037/0033-295X.102.1.4

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464‑1480. doi: 10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197‑216. doi: 10.1037/0022-3514.85.2.197

Haddock, G., Zanna, M. P., & Esses, V. M. (1993). Assessing the structure of prejudicial attitudes: The case of attitudes toward homosexuals. *Journal of Personality and Social Psychology*, *65*(6), 1105‑1118. doi: 10.1037/0022-3514.65.6.1105

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., Brand, R., Brandt, M. J., Brewer, G., Bruyneel, S., Calvillo, D. P., Campbell, W. K., Cannon, P. R., Carlucci, M., Carruth, N. P., Cheung, T., Crowell, A., De Ridder, D. T. D., Dewitte, S., … Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, *11*(4), 546‑573. doi: 10.1177/1745691616652873

Hahn, A., & Gawronski, B. (2019). Facing one's implicit biases: From awareness to acknowledgment. *Journal of Personality and Social Psychology*, *116*(5), 769‑794. doi: 10.1037/pspi0000155

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General*, *143*(3), 1369‑1392. doi: 10.1037/a0035028

Harabi, S. (2018). *Représentations et partitions sociales : Le rapport aux français d'origine maghrébine en France* [Manuscrit de thèse non publié]. Université Bourgogne Franche-Comté.

Hartmann, R., Johannsen, L., & Klauer, K. C. (2020). rtmpt : An R package for fitting response-time extended multinomial processing tree models. *Behavior Research Methods*, *52*(3), 1313‑1338. doi: 10.3758/s13428-019-01318-x

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1), 264‑284. doi: 10.3758/s13428-017-0869-7

Heck, D. W., & Erdfelder, E. (2016). Extending multinomial processing tree models to measure the relative speed of cognitive processes. *Psychonomic Bulletin & Review*, *23*(5), 1440‑1465. doi: 10.3758/s13423-016-1025-6

Hehman, E., Calanchini, J., Flake, J. K., & Leitner, J. B. (2019). Establishing construct validity evidence for regional measures of explicit and implicit racial bias. *Journal of Experimental Psychology: General*, *148*(6), 1022‑1040. doi: 10.1037/xge0000623

Hilton, J. L., & von Hippel, W. (1996). STEREOTYPES. *Annual Review of Psychology*, *47*(1), 237-271. doi: 10.1146/annurev.psych.47.1.237

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A Meta-Analysis on the correlation between the Implicit Association Test and explicit self-Report measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369-1385. doi: 10.1177/0146167205275613

Horry, R., & Wright, D. B. (2009). Anxiety and terrorism: Automatic stereotypes affect visual attention and recognition memory for White and Middle Eastern faces. *Applied Cognitive Psychology*, *23*(3), 345-357. doi: 10.1002/acp.1465

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, *59*(1), 21–47. doi: 10.1007/BF02294263

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology*, *27*(1), 116-159. doi: 10.1080/10463283.2016.1212966

IJzerman, H., Lewis, N. A. J., Przybylski, A. K., Weinstein, N., DeBruine, L. M., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*. doi: 10.1038/s41562-020-00990-w

Ito, T. A., & Tomelleri, S. (2017). Seeing is not stereotyping: The functional independence of categorization and stereotype activation. *Social Cognitive and Affective Neuroscience*, *12*(5), 758-764. doi: 10.1093/scan/nsx009

Ito, T. A., & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(1), 21-36. doi: 10.3758/CABN.5.1.21

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513-541. doi:10.1016/0749-596X(91)90025-F

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, *122*(2), 139-154. doi: 10.1037/0096-3445.122.2.139

James, W. (1890). *The principle of psychology*. New York: Holt.

Jin, Z., Rivers, A., Sherman, J. W., & Chen, R. (2016). Measures of implicit gender attitudes may exaggerate differences in underlying associations among Chinese urban and rural women. *Psychologica Belgica*, *56*(1), 13-22. doi: 10.5334/pb.308

Jobard, F., Lévy, R., Lamberth, J., & Névanen, S. (2012). Mesurer les discriminations selon l'apparence: Une analyse des contrôles d'identité à Paris. *Population*, *67*(3), 423-451.

Jost, J. T. (2019). The IAT is dead, long live the IAT: Context-sensitive measures of implicit attitudes are indispensable to social and political psychology. *Current Directions in Psychological Science*, *28*(1), 10-19. doi: 10.1177/0963721418797309

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, *33*(1), 1-27. doi: 10.1111/j.2044-8309.1994.tb01008.x

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*(6), 881-919. doi: 10.1111/j.1467-9221.2004.00402.x

Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, *100*(1), 109-128. doi: 10.1037/0033-295X.100.1.109

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54-69. doi: 10.1037/a0028347

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Karlins, M., Coffman, T. L., & Walters, G. (1969). On the fading of social stereotypes: Studies in three generations of college students. *Journal of personality and social psychology*, *13*(1), 1. doi: 10.1037/h0030839

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*(5), 774-788. doi: 10.1037//0022-3514.81.5.774

Karpinski, A., & Steinman, R. B. (2006). The Single Category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*(1), 16-32. doi: 10.1037/0022-3514.91.1.16

Kastenmüller, A., Greitemeyer, T., Aydin, N., Tattersall, A. J., Peus, C., Bussmann, P., Fischer, J., Frey, D., & Fischer, P. (2011). Terrorism threat and networking: Evidence that terrorism salience decreases occupational networking. *Journal of Organizational Behavior*, *32*(7), 961-977. doi: 10.1002/job.745

Katz, D., & Braly, K. W. (1935). Racial prejudice and racial stereotypes. *The Journal of Abnormal and Social Psychology*, *30*(2), 175-193. doi: 10.1037/h0059800

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *The Journal of Abnormal and Social Psychology*, *28*(3), 280. doi: 10.1037/h0074049

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, *78*(5), 871-888. doi: 10.1037/0022-3514.78.5.871

Kepel, G. (2016). *La fracture*. Paris : Gallimard.

Kidder, C. K., White, K. R., Hinojos, M. R., Sandoval, M., & Crites, S. L. (2018). Sequential stereotype priming: A meta-analysis. *Personality and Social Psychology Review*, *22*(3), 199-227. doi: 10.1177/1088868317723532

Kite, M. E., & Whitley Jr, B. E. (2016). *Psychology of prejudice and discrimination*. Psychology Press.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*(1), 70-98. doi: 10.1007/s11336-009-9141-0

Klauer, K. C., & Kellen, D. (2018). RT-MPTs: Process models for response-time distributions based on multinomial processing trees with applications to recognition memory. *Journal of Mathematical Psychology*, *82*, 111-130. doi: 10.1016/j.jmp.2017.12.003

Klauer, K. C., & Meiser, T. (2000). A source-monitoring analysis of illusory correlations. *Personality and Social Psychology Bulletin*, *26*(9), 1074-1093. doi: 10.1177/01461672002611005

Klauer, K. C., Stahl, C., & Voss, A. (2011). Multinomial models and diffusion models. In K. C. Klauer, A. Voss, & C. Stahl (Eds.): Cognitive methods in social psychology (pp. 367- 390). New York: Guilford Press.

Klein, Richard A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., … Nosek, B. A. (2014). Investigating Variation in Replicability: A "Many Labs" Replication Project. *Social Psychology*, *45*(3), 142-152. doi: 10.1027/1864-9335/a000178

Klein, R. A., Cook, C. L., Ebersole, C. R., Vitiello, C. A., Nosek, B. A., Chartier, C. R., Christopherson, C. D., Clay, S., Collisson, B., Crawford, J., Cromar, R., Vidamuerte, D., Gardiner, G., Gosnell, C., Grahe, J. E., Hall, C., Joy-Gaba, J. A., Legg, A. M., Levitan, C., … Ratliff, K. A. (2019). Many Labs 4: Failure to replicate mortality salience effect with and without original author involvement. PsyArXiv. doi: 10.31234/osf.io/vef2c

Krieglmeyer, R., & Sherman, J. W. (2012). Disentangling stereotype activation and stereotype application in the stereotype misperception task. *Journal of Personality and Social Psychology*, *103*(2), 205-224. doi: 10.1037/a0028764

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin*, *129*(4), 522-544. doi: 10.1037/0033-2909.129.4.522

Lai, C. K., Hoffman, K. M., & Nosek, B. A. (2013). Reducing implicit prejudice. *Social and Personality Psychology Compass*, *7*(5), 315–330. doi: 10.1111/spc3.12023

Lambert, W. E. (1952). Comparison of French and American modes of response to the Bogardus Social Distance Scale. *Social Forces*, *31*(2), 155-160. doi: 10.2307/2573400

Langner, O., Ouwens, M., Muskens, M., Trumpf, J., Becker, E. S., & Rinck, M. (2010). Faking on direct, indirect, and behavioural measures of spider fear: Can you get away with it? *Cognition & Emotion*, *24*(3), 549-558. doi: 10.1080/02699930802652412

Ledgerwood, A., Eastwick, P. W., & Smith, L. K. (2018). Toward an integrative framework for studying human evaluation: Attitudes toward objects and attributes. *Personality and Social Psychology Review*, *22*(4), 378-398. doi: 10.1177/1088868318790718

Légal, J.-B., & Delouvée, S. (2015). *Stéréotypes, préjugés et discriminations*. Paris: Dunod.

Lemm, K. M., Lane, K. A., Sattler, D. N., Khan, S. R., & Nosek, B. A. (2008). Assessing implicit cognitions with a paper-format Implicit Association Test. In T. Morrison & M. Morrison (Eds.), *The psychology of modern prejudice* (pp. 123-146). Hauppauge, NY: Nova Science Publishers

Leyens, J.-P., Yzerbyt, V., & Schadron, G. (1996). *Stéréotypes et cognition sociale*. Bruxelles : Mardaga.

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 219. doi: 10.1037/0096-1523.20.2.219

Lippmann, W. (1922). *Public opinion*. New Yok: Harcourt, Brace and Company.

Livingston, R. W. (2002). The role of perceived negativity in the moderation of African Americans' implicit and explicit racial attitudes. *Journal of Experimental Social Psychology*, *38*(4), 405-413. doi: 10.1016/S0022-1031(02)00002-1

Loersch, C., & Payne, B. K. (2011). The situated inference model: An integrative account of the effects of primes on perception, behavior, and motivation. *Perspectives on Psychological Science*. doi: 10.1177/1745691611406921

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of personality and social psychology*, *81*(5), 842-855. doi: 10.1037/0022-3514.81.5.842

Lyons, P. A., Kenworthy, J. B., & Popan, J. R. (2010). Ingroup identification and group-level narcissism as predictors of U.S. citizens' attitudes and behavior toward Arab immigrants. *Personality and Social Psychology Bulletin*, *36*(9), 1267-1280. doi: 10.1177/0146167210380604

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*(1), 93-120. doi: 10.1146/annurev.psych.51.1.93

Macrae, C. N., & Bodenhausen, G. V. (2001). Social cognition: Categorical person perception. *British Journal of Psychology*, *92*(1), 239-255. doi: 10.1348/000712601162059

Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of personality and Social Psychology*, *66*(1), 37.

Maddux, W. W., Barden, J., Brewer, M. B., & Petty, R. E. (2005). Saying no to negativity: The effects of context and motivation to control prejudice on automatic evaluative responses. *Journal of Experimental Social Psychology*, *41*(1), 19-35. doi: 10.1016/j.jesp.2004.05.002

Mange, J., Chun, W. Y., Sharvit, K., & Belanger, J. J. (2012). Thinking about Arabs and Muslims makes Americans shoot faster: Effects of category accessibility on aggressive responses in a shooter paradigm. *European Journal of Social Psychology*, *42*(5), 552-556. doi: 10.1002/ejsp.1883

Mange, J., Sharvit, K., Margas, N., & Sénémeaud, C. (2016). Do I shoot faster because I am thinking about an outgroup or a threatening outgroup?: Shooter bias, perceived threat, and intergroup processes. *Social Psychology*, *47*(1), 29-37. doi: 10.1027/1864-9335/a000255

Mann, T. C., Cone, J., Heggeseth, B., & Ferguson, M. J. (2019). Updating implicit impressions: New evidence on intentionality and the affect misattribution procedure. *Journal of Personality and Social Psychology*, *116*(3), 349-374. doi: 10.1037/pspa0000146

Mayer, N., & Tiberj, V. (2016). Who were the « Charlie » in the Streets? A Socio-Political Approach of the January 11 Rallies. *International Review of Social Psychology*, *29*(1), 59-68. doi: 10.5334/irsp.63

McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, *36*(9), 929-940. doi: 10.1037/0022-3514.36.9.929

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91-125). Orlando, FL: Academic Press.

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of Personality and Social Psychology*, *104*(1), 45-69. doi: 10.1037/a0030734

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*(4), 280-293. doi: 10.1016/j.tics.2018.02.001

Mitchell, G., & Tetlock, P. E. (2017). Popularity as a poor proxy for utility: The case of implicit prejudice. In S. O. Lilienfeld & I. D. Waldman (Éds.), *Psychological Science Under Scrutiny* (pp. 164-195). John Wiley & Sons, Inc. doi: 10.1002/9781119095910.ch10

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*(3), 455-469. doi: 10.1037/0096-3445.132.3.455

Monteith, M. J., Woodcock, A., & Gulker, J. E. (2013). Automaticity and control in stereotyping and prejudice: The revolutionary role of social cognition across three decades of research. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition* (pp. 74–94). New York, NY: Oxford University Press

Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology*, *67*(1), 263-287. doi: 10.1146/annurev-psych-122414-033550

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297-326. doi: 10.1037/0033-2909.132.2.297

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42-54. doi: 10.3758/BRM.42.1.42

Moskalenko, S., McCauley, C., & Rozin, P. (2006). Group identification under conditions of threat: College students' attachment to country, family, ethnicity, religion, and university before and after september 11, 2001. *Political Psychology*, *27*(1), 77-97. doi: 10.1111/j.1467-9221.2006.00450.x

Moskowitz, G. B. (2005). *Social cognition: Understanding self and others*. New York: Guilford Press.

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of personality and social psychology*, *77*(1), 167. doi: 10.1037/0022-3514.77.1.167

Nadarevic, L., & Erdfelder, E. (2011). Cognitive processes in implicit attitude tasks: An experimental validation of the Trip Model. *European Journal of Social Psychology*, *41*(2), 254-268. doi: 10.1002/ejsp.776

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226-254. doi: 10.1037/0096-3445.106.3.226

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, *69*(1), 551-534. doi: 10.1146/annurev-psych-122216-011836

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516-524. doi: 10.3758/BF03195598

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231-259. doi: 10.1037/0033-295X.84.3.231

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-go Association Task. *Social cognition*, *19*(6), 625–666. doi: 10.1521/soco.19.6.625.20886

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*(1), 101-115. doi: 10.1037/1089-2699.6.1.101

Nosek, B. A., & Hansen, J. J. (2008). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition & Emotion*, *22*(4), 553-594. doi: 10.1080/02699930701438186

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36-88. doi: 10.1080/10463280701489053

Nugier, A., & Guimond, S. (2016). « Je suis Charlie »: New findings on the social and political psychology of terrorism. *International Review of Social Psychology*, *29*(1), 45. doi: 10.5334/irsp.60

Nugier, A., Roebroeck, E., Anier, N., Kleinlogel, E. P., Chatard, A., & Guimond, S. (2016). The psychological effects of terrorism are moderated by cultural worldviews. *International Review of Social Psychology*, *29*(1), 77-84. doi: 10.5334/irsp.61

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087-11092. doi: 10.1073/pnas.0805664105

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. doi: 10.1126/science.aac4716

Orehek, E., Fishman, S., Dechesne, M., Doosje, B., Kruglanski, A. W., Cole, A. P., Saddler, B., & Jackson, T. (2010). Need for closure and the social response to terrorism. *Basic and Applied Social Psychology*, *32*(4), 279-290. doi: 10.1080/01973533.2010.519196

Oswald, D. L. (2005). Understanding anti-Arab reactions post-9/11: The role of threats, social categories, and personal ideologies. *Journal of Applied Social Psychology*, *35*(9), 1775–1799. doi: 10.1111/j.1559-1816.2005.tb02195.x

Park, B., & Judd, C. M. (2005). Rethinking the link between categorization and prejudice within the social cognition perspective. *Personality and Social Psychology Review*, *9*(2), 108-130. doi: 10.1207/s15327957pspr0902_2

Park, J., Felix, K., & Lee, G. (2007). Implicit attitudes toward Arab-Muslims and the moderating effects of social information. *Basic and Applied Social Psychology*, *29*(1), 35–45. doi: 10.1080/01973530701330942

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*(2), 181-192. doi: 10.1037/0022-3514.81.2.181

Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology*, *20*(1), 272-314. doi: 10.1080/10463280903162177

Payne, B. K., Brown-Iannuzzi, J., Burkley, M., Arbuckle, N. L., Cooley, E., Cameron, C. D., & Lundberg, K. B. (2013). Intention invention and the Affect Misattribution Procedure: Reply to Bar-Anan and Nosek (2012). *Personality and Social Psychology Bulletin*, *39*(3), 375-386. doi: 10.1177/0146167212475225

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, *94*(1), 16-31. doi: 10.1037/0022-3514.94.1.16

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277-293. doi: 10.1037/0022-3514.89.3.277

Payne, B. K., & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 1-15). New York: Guilford Press.

Payne, B. K., Hall, D. L., Cameron, C. D., & Bishara, A. J. (2010). A process model of affect misattribution. *Personality and Social Psychology Bulletin*, *36*(10), 1397-1408. doi: 10.1177/0146167210383440

Payne, B. K., Vuletich, H. A., & Brown-Iannuzzi, J. L. (2019). Historical roots of implicit bias in slavery. *Proceedings of the National Academy of Sciences*, *116*(24), 11693-11698. doi: 10.1073/pnas.1818816116

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233‑248. doi: 10.1080/1047840X.2017.1335568

Pelletier, P., & Drozda-Senkowska, E. (2016). The Charlie Hebdo terror attack in Paris: Follow-up of French citizens' terrorist threat perception and its aftermath. *International Review of Social Psychology*, *29*(1), 85. doi: 10.5334/irsp.51

Persson, A. V., & Musher-Eizenman, D. R. (2005). College students' attitudes toward Blacks and Arabs following a terrorist attack as a function of varying levels of media exposure. *Journal of Applied Social Psychology*, *35*(9), 1879–1892. doi: 10.1111/j.1559-1816.2005.tb02200.x

Perugini, M., Gallucci, M., & Costantini, G. (2018). A Practical primer to power analysis for simple experimental designs. *International Review of Social Psychology*, *31*(1). doi: org/10.5334/irsp.181

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–205). New York: Academic.

Phills, C. E., Kawakami, K., Tabi, E., Nadolny, D., & Inzlicht, M. (2011). Mind the gap: Increasing associations between the self and blacks with approach behaviors. *Journal of Personality and Social Psychology*, *100*(2), 197‑210. doi: 10.1037/a0022159

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*(3), 811‑832. doi: 10.1037/0022-3514.75.3.811

Popa-Roch, M.-A., & Delmas, F. (2011). Evidence for the salience asymmetries producing IAT effects: Replication and extension with a French-North African IAT. *International Review of Social Psychology*, *Tome 24*(1), 37‑61.

Poskocil, A. (1977). Encounters between Blacks and White liberals: The collision of stereotypes. *Social Forces*, *55*(3), 715. doi: 10.2307/2577465

Posner, M. I., & Snyder, C. R. R. (2004). *Attention and Cognitive Control*. In D. A. Balota & E. J. Marsh (Eds.), *Key readings in cognition. Cognitive psychology: Key readings* (pp. 205–223). Psychology Press.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*(4), 741‑763. doi: 10.1037/0022-3514.67.4.741

Ratliff, K. A., Lofaro, N., Howell, J. L., Conway, M. A., Lai, C. K., O'Shea, B. A., Smith, C. T., Jiang, C., Redford, L., Pogge, G., Umansky, E., Vitiello, C., & Zitelny, H. (2020). Documenting bias from 2007–2015: Pervasiveness and correlates of implicit attitudes and stereotypes II. https://osf.io/rfzhu/

R Core Team (2013). R: A language and environment for statistical computing. Available from: https://www.R-project.org/.

Rees, H. R., Ma, D. S., & Sherman, J. W. (2019). Examining the relationships among categorization, stereotype activation, and stereotype application. *Personality and Social Psychology Bulletin*, *46*(4), 499-513. doi: 10.1177/0146167219861431

Rees, H. R., Rivers, A. M., & Sherman, J. W. (2019). Implementation intentions reduce implicit stereotype activation and application. *Personality and Social Psychology Bulletin*, *45*(1), 37-53. doi: 10.1177/0146167218775695

Reichardt, R., Rivers, A. M., Reichardt, J., & Sherman, J. (2019). *Further validation of measures of target detection and stereotype activation in the Stereotype Misperception Task*. PsyArXiv. doi: 10.31234/osf.io/9h2pd

Ric, F., & Muller, D. (2017). *La cognition sociale*. Presses universitaires de Grenoble.

Richeson, J. A., & Trawalter, S. (2008). The threat of appearing prejudiced and race-based attentional biases. *Psychological Science*, *19*(2), 98-102. doi: 10.1111/j.1467-9280.2008.02052.x

Richeson, J. A., Trawalter, S., & Shelton, J. N. (2005). African Americans' implicit racial attitudes and the depletion of executive function after interracial interactions. *Social Cognition*, *23*(4), 336–352. doi: 10.1521/soco.2005.23.4.336

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318-339. doi: 10.1037/0033-295X.95.3.318

Rivers, A. M., Rees, H. R., Calanchini, J., & Sherman, J. W. (2017). Implicit bias reflects the personal and the social. *Psychological Inquiry*, *28*(4), 301-305. doi: 10.1080/1047840X.2017.1373549

Rivers, A. M., Sherman, J. W., Rees, H. R., Reichardt, R., & Klauer, K. C. (2020). On the roles of stereotype activation and application in diminishing implicit bias. *Personality and Social Psychology Bulletin*, *46*(3), 349-364. doi: 10.1177/0146167219853842

Rohmer, O., & Louvet, E. (2018). Implicit stereotyping against people with disability. *Group Processes & Intergroup Relations*, *21*(1), 127-140. doi: 10.1177/1368430216638536

Rosenberg, M. J., & Hovland, C. I. (Eds.). (1960). *Attitude organization and change: An analysis of consistency among attitude components*. Yale University Press.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*(2), 139-165. doi: 10.1037/0096-3445.133.2.139

Rougier, M., Muller, D., Courset, R., Smeding, A., Devos, T., & Batailler, C. (2020). Toward the use of approach/avoidance tendencies as attitude measures: Individual- and group-level variability of the ingroup bias. *European Journal of Social Psychology*, *50*(4), 857-875. doi: 10.1002/ejsp.2653

Rougier, M., Muller, D., Ric, F., Alexopoulos, T., Batailler, C., Smeding, A., & Aubé, B. (2018). A new look at sensorimotor aspects in approach/avoidance tendencies: The role of visual whole-body movement information. *Journal of Experimental Social Psychology*, *76*, 42-53. doi: 10.1016/j.jesp.2017.12.004

Rudman, L. A., Feinberg, J., & Fairchild, K. (2002). Minority members' implicit attitudes: Automatic ingroup bias as a function of group status. *Social Cognition*, *20*(4), 294–320. doi: 10.1521/soco.20.4.294.19908

Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition*, *17*(4), 437–465. doi: 10.1521/soco.1999.17.4.437

Schneider, D. J. (2004). *The psychology of stereotyping. Distinguished contributions in psychology*. New York: Guilford Publications.

Schuster, M. A., Stein, B. D., Jaycox, L. H., Collins, R. L., Marshall, G. N., Elliott, M. N., Zhou, A. J., Kanouse, D. E., Morrison, J. L., & Berry, S. H. (2001). A national survey of stress reactions after the September 11, 2001, terrorist attacks. *New England Journal of Medicine*, *345*(20), 1507–1512. doi: 10.1056/nejm200111153452024

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464. doi: 10.1214/aos/1176344136

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social cognition*, *25*(5), 638–656. doi: 10.1521/soco.2007.25.5.638

Shaheen, J. G. (2003). Reel bad Arabs: How Hollywood vilifies a people. *The ANNALS of the American Academy of Political and Social Science*, *588*(1), 171-193. doi: 10.1177/0002716203588001011

Sherman, J. W. (2006). On building a better process model: It's not only how many, but which ones and by which means? *Psychological Inquiry*, *17*(3), 173-184. doi: 10.1207/s15327965pli1703_3

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*(2), 314-335. doi: 10.1037/0033-295X.115.2.314

Sherman, J. W., Klauer, K. C., & Allen, T. J. (2010). Mathematical modeling of implicit social cognition: The machine in the ghost. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and application* (pp. 156-175). New York: Guilford Press.

Sherman, J. W., Krieglmeyer, R., & Calanchini, J. (2014). Process models require process measures. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 121-138). New York: Guilford Press.

Sherman, J. W., Macrae, C. N., & Bodenhausen, G. V. (2000). Attention and stereotyping: cognitive constraints on the construction of meaningful social impressions. *European Review of Social Psychology*, *11*(1), 145-175. doi: 10.1080/14792772043000022

Shiffrin, R. M., & Schneider, W. (1984). Automatic and controlled processing revisited. *Psychological Review*, *91*(2), 269‑276. doi: 10.1037/0033-295X.91.2.269

Sigall, H., & Page, R. (1971). Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology*, *18*(2), 247‑255. doi: 10.1037/h0030839

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359‑1366. doi: 10.1177/0956797611417632.

Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models in R. *Behavior Research Methods*, *45*(2), 560‑575. doi: 10.3758/s13428-012-0259-0

Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of Personality and Social Psychology*, *70*(5), 893‑912. doi: 10.1037//0022-3514.70.5.893

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and social psychology review*, *4*(2), 108–131. doi: 10.1207/s15327957pspr0402_01

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*, *54*(1), 167‑183. doi: 10.1016/j.jmp.2009.06.007

Spruyt, B., & van der Noll, J. (2017). The “Islamized stranger”: On “chronic” versus “contextual” salience in the measurement of anti-Muslim prejudice: The “Islamized stranger”. *Political Psychology*, *38*(6), 977‑989. doi: 10.1111/pops.12383

Stangor, C. (2016). The study of stereotyping, prejudice, and discrimination within social psychology: A quick history of theory and research. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 3–27). Psychology Press.

Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. *Personality and Social Psychology Bulletin*, *34*(10), 1332‑1345. doi: 10.1177/0146167208321269

Stewart, B. D., von Hippel, W., & Radvansky, G. A. (2009). Age, race, and implicit prejudice: Using process dissociation to separate the underlying components. *Psychological Science*, *20*(2), 164–168. doi: 10.1111/j.1467-9280.2009.02274.x

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220‑247. doi: 10.1207/s15327957pspr0803_1

Stroebe, W., & Insko, C. A. (1989). Stereotype, prejudice, and discrimination: Changing conceptions in theory and research. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Stroebe (Éds.), *Stereotyping and Prejudice* (pp. 3‑34). Springer New York. doi: 10.1007/978-1-4612-3582-8_1

Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, *66*(1), 21‑36.

doi: 10.1037/0022-3514.66.1.21

Taylor, D. G., Sheatsley, P. B., & Greeley, A. M. (1978). Attitudes toward racial integration. *Scientific American*, *238*(6), 42-49. doi: 10.1038/scientificamerican0678-42

Teige-Mocigemba, S., Becker, M., Sherman, J. W., Reichardt, R., & Christoph Klauer, K. (2017). The Affect Misattribution Procedure: In search of prejudice effects. *Experimental Psychology*, *64*(3), 215-230. doi: 10.1027/1618-3169/a000364

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion*, *13*(4), 724. doi: 10.1037/a0032335

Todorov, A., & Oosterhof, N. N. (2011). Modeling social perception of faces [social sciences]. *IEEE Signal Processing Magazine*, *28*(2), 117–122. doi: 10.1109/msp.2010.940006

Trawalter, S., Todd, A. R., Baird, A. A., & Richeson, J. A. (2008). Attending to threat: Race-based patterns of selective attention. *Journal of Experimental Social Psychology*, *44*(5), 1322-1327. doi: 10.1016/j.jesp.2008.03.006

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological review*, *76*(3), 282. doi: 10.1037/h0027242

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, *93*(3), 239-257. doi: 10.1037/0033-295X.93.3.239

Ullrich, J., & Cohrs, J. C. (2007). Terrorism salience increases system justification: Experimental evidence. *Social Justice Research*, *20*(2), 117-139. doi: 10.1007/s11211-007-0035-y

Unkelbach, C., Forgas, J. P., & Denson, T. F. (2008). The turban effect: The influence of Muslim headgear and induced affect on aggressive responses in the shooter bias paradigm. *Journal of Experimental Social Psychology*, *44*(5), 1409-1413. doi: 10.1016/j.jesp.2008.04.003

Unkelbach, C., Goldenberg, L., Müller, N., Sobbe, G., & Spannaus, N. (2009). A shooter bias in Germany against people wearing Muslims headgear. *International Review of Social Psychology*, *Tome 22*(3), 181-201. doi: 10.1016/j.jesp.2008.04.003

Van de Vyver, J., Houston, D. M., Abrams, D., & Vasiljevic, M. (2016). Boosting belligerence: How the July 7, 2005, London bombings affected liberals' moral foundations and prejudice. *Psychological science*, *27*(2), 169–177. doi: 10.1177/0956797615615584

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

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *The Journal of Machine Learning Research*, *11*, 3571–3594.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95-I 16). New York: Guilford Press.

Wittenbrink, B. (2007). Measuring attitudes through priming. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 17– 58). New York: Guilford Press

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*(2), 262-274. doi: 10.1037/0022-3514.72.2.262

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*(5), 815-827. doi: 10.1037//0022-3514.81.5.815

Wojnowicz, M. T., Ferguson, M. J., Dale, R., & Spivey, M. J. (2009). The self-organization of explicit attitudes. *Psychological Science*, *20*(11), 1428–1435. doi: 10.1111/j.1467-9280.2009.02448.x

Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology*, *10*(2), 109-120. doi: 10.1016/0022-1031(74)90059-6

Zerhouni, O., Rougier, M., & Muller, D. (2016). “Who (really) is Charlie?” French cities with lower implicit prejudice toward Arabs demonstrated larger participation rates in Charlie Hebdo rallies. *International Review of Social Psychology*, *29*(1), 69-76. doi: 10.5334/irsp.50

Index des figures

Figure 1. Représentation en arbre de la *Process Dissociation Procedure* (PDP). **p. 49**

Figure 2. Représentation en arbre du *Quadruple Process Model* (Quad Model). **p. 54**

Figure 3. Représentation en arbre du modèle SAC-dominant. **p. 65**

Figure 4. Représentation visuelle d'un essai SMT. **p. 72**

Figure 5. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin, Neutre). (Étude 1). **p. 74**

Figure 6. Évolution de l'intérêt de recherche Google pour le mot « Amalgame » au cours du temps (entre janvier 2004 et septembre 2020). **p. 84**

Figure 7. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin, Neutre) et de la condition expérimentale (Contrôle vs. Attentat). (Étude 2). **p. 88**

Figure 8. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un modèle SAC-dominant à quatre paramètres. (Étude 2). **p. 90**

Figure 9. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres (incluant l'intervalle de crédibilité à 95% de la moyenne groupale) en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un modèle SAC-dominant à six paramètres. (Étude 2). **p. 91**

Figure 10. Manipulation expérimentale de l'arrière-plan. À gauche, devanture d'un musée (condition Contrôle). À droite, devanture du Bataclan (condition Attentat). (Étude 3). **p. 95**

Figure 11. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin, Neutre) et de la condition expérimentale (Contrôle vs. Attentat). (Étude 3). **p. 96**

Figure 12. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un modèle SAC-dominant à quatre paramètres. (Étude 3). **p. 98**

Figure 13. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un modèle SAC-dominant à six paramètres. (Étude 3). **p. 99**

Figure 14. Évolution de la moyenne individuelle (gris) et groupale (rouge) des Scores IAT au cours du temps. La zone bleue correspond à la zone post-attentat de Charlie Hebdo (janvier 2015). (Étude 4). **p. 105**

Figure 15. Prédictions pour l'IAT-Charlie. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un Quad Model. (Étude 4). **p. 106**

Figure 16. Prédictions pour l'IAT-Bataclan. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un Quad Model. (Étude 4). **p. 107**

Figure 17. Prédictions pour l'IAT-Nice. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un Quad Model. (Étude 4). **p. 108**

Figure 18. Prédictions pour l'ensemble des participants. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres en fonction de la condition expérimentale (Contrôle vs. Attentat) et pour un Quad Model. (Étude 4). **p. 110**

Figure 19. Représentation en arbre du modèle D-dominant. **p. 122**

Figure 20. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin/Noir, Neutre) et de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc). (Étude 5). **p. 128**

Figure 21. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres en fonction de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc) et pour un modèle SAC-dominant à quatre paramètres. (Étude 5). **p.130**

Figure 22. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres en fonction de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc) et pour un modèle SAC-dominant à six paramètres. (Étude 5). **p. 131**

Figure 23. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin, Neutre) et du caractère prototypique de l'amorce (Haut vs. Bas). (Étude 6). **p. 137**

Figure 24. Représentation en arbre du modèle englobant à la fois le modèle SAC-dominant (en haut) et le modèle D-dominant (en bas). **p. 140**

Figure 25. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres en fonction du caractère prototypique de l'amorce (Haut vs. Bas) et du caractère discernable de la cible (Haut vs. Bas). (Étude 6). **p. 141**

Figure 26. λ moyen en fonction du score de préjugé (Étude 6). **p. 143**

Figure 27. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin, Neutre) et du caractère prototypique de l'amorce (Haut vs. Bas). (Étude 7). **p. 149**

Figure 28. λ moyen en fonction du score de préjugé (Étude 7). **p. 152**

Figure 29. Représentation graphique de la méta-analyse interne à nos six études. **p. 162**

Figure 30. Moyenne des réponses « Plus menaçant » en fonction de l'amorce (Blanc, Maghrébin/Noir, Neutre) et du type d'amorce (visages vs. prénoms). (Étude 8). **p. 172**

Figure 31. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres en fonction du type d'amorce (visages vs. prénoms) et pour un modèle SAC-dominant à quatre paramètres. (Étude 8). **p. 173**

Figure 32. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin/Noir, Neutre) et de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc). (Étude 9). **p. 187**

Figure 33. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres en fonction de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc) et pour un modèle SAC-dominant à quatre paramètres. (Étude 9). **p. 188**

Figure 34. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres en fonction de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc) et pour un modèle SAC-dominant à six paramètres. (Étude 9). **p. 190**

Figure 35. Moyenne des réponses « Plus menaçant » des participants blancs en fonction du type d'amorce (Blanc, Maghrébin/Noir, Neutre) et de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc). (Étude 9). **p. 191**

Figure 36. Moyenne des réponses « Plus menaçant » des participants hispaniques en fonction du type d'amorce (Blanc, Maghrébin/Noir, Neutre) et de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc). (Étude 9). **p. 192**

Figure 37. Moyenne des réponses « Plus menaçant » des participants asiatiques en fonction du type d'amorce (Blanc, Maghrébin/Noir, Neutre) et de la condition expérimentale (SMT-Maghrébin/Blanc vs. SMT-Noir/Blanc). (Étude 9). **p. 193**

Figure 38. Moyenne des réponses « Plus menaçant » en fonction du type d'amorce (Blanc, Maghrébin, Neutre). (Étude 10). **p. 200**

Figure 39. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres pour un modèle SAC-dominant à quatre paramètres. (Étude 10). **p. 201**

Figure 40. Estimation de la moyenne individuelle (gris) et groupale (rouge) des différents paramètres pour un modèle SAC-dominant à six paramètres. (Étude 10). **p. 202**

Figure 41a. Représentation en arbre du modèle SAC-dominant classique. **p. 232**

Figure 41b. Représentation en arbre du modèle SAC-dominant inversé **p. 232**

Index des tableaux

Tableau 1. *Indice de fiabilité (r) des Études 1 à 9 avec et sans correction de Spearman-Brown.* **p. 226**

Tableau 2. *Valeurs de paramètres prédites par les modèles SAC-dominant classique et SAC-dominant inversé.* **p. 233**

Tableau 3. *Moyennes, écarts-types et corrélations (avec intervalles de confiance à 95%) entre les différents paramètres du modèle et le score de préjugé (PREJ).* (Étude 6). **p. 273**

Tableau 4. *Moyennes, écarts-types et corrélations (avec intervalles de confiance à 95%) entre les différents paramètres du modèle, le score de préjugé et les sous-échelles de motivation à réduire l'expression du préjugé.* (Étude 7). **p. 274**

Tableau 5. *Amorces de prénoms (français, maghrébins et neutres) utilisés dans l'Étude 8.* **p. 278**

Tableau 6. *Validation du paramètre D (Krieglmeyer & Sherman, 2012 ; Étude 2).* **p. 279**

Tableau 7. *Validation du paramètre G (Krieglmeyer & Sherman, 2012 ; Étude 3).* **p. 279**

Tableau 8. *Validation du paramètre SAC (Krieglmeyer & Sherman, 2012 ; Étude 4).* **p. 280**

Tableau 9. *Validation du paramètre SAP (Krieglmeyer & Sherman, 2012 ; Étude 5).* **p. 280**

Annexes

Annexe 1 : inductions expérimentales du contexte (Étude 3)

Condition Attentat

21 h 20 : le premier commando entre en action près du stade de France, pendant le match de football amical France-Allemagne, auquel assistent le président de la République et plusieurs personnalités politiques. Ce commando est composé de trois hommes. Aux abords du stade, après le début du match, quand les lieux sont relativement peu fréquentés, le premier des trois hommes actionne sa ceinture d'explosifs, suivi quelques minutes plus tard par le deuxième terroriste. La troisième explosion, rue de la Cokerie, survient une demi-heure plus tard. Les trois explosions font au total, outre les terroristes, un mort, tué par l'explosion et une dizaine de blessés graves. Les trois hommes avaient l'intention de pénétrer dans le stade pour y perpétrer leurs actes, mais ont échoué à quatre reprises. Compte tenu de l'armement dont ils disposaient, ils auraient pu causer des pertes beaucoup plus nombreuses.

21 h 25 : le second groupe de trois terroristes intervient dans plusieurs rues des 10e et 11e arrondissements à bord d'une Seat noire. Plusieurs éléments identifient notamment l'un des terroristes comme étant suspecté par la police d'être le principal organisateur des attentats et trouvera la mort cinq jours plus tard à Saint-Denis lors d'une action de police. Ils ouvrent tout d'abord le feu sur le bistrot Le Carillon, puis sur le restaurant qui lui fait face, Le Petit Cambodge. Ils tirent ensuite sur la brasserie Café Bonne Bière, et près de la terrasse d'un restaurant italien, Casa Nostra. À 21 h 36, ils attaquent le restaurant La Belle Equipe. Enfin à 21 h 43, ayant été déposé par la Seat au café-brasserie Comptoir Voltaire, l'un des trois terroristes se fait exploser au moment où il passe commande en utilisant le même dispositif explosif que les terroristes du stade de France.

Ses deux complices disparaissent. La Seat sera retrouvée vingt-quatre heures plus tard à Montreuil, avec trois Kalachnikov, cinq chargeurs pleins et onze vides abandonnés à bord. Plus de 400 coups ont été tirés en 20 minutes et 116 cartouches de kalachnikov sont retrouvées sur les terrasses du Petit Cambodge et du Carillon.

21 h 40 : le troisième groupe (qui s'est déplacé à bord d'une Volkswagen Polo noire) arrive au théâtre du Bataclan (11e arrondissement), et abat des personnes à l'extérieur du bâtiment. À 21 h 42, ils envoient un SMS vers un numéro localisé en Belgique dans lequel ils indiquent « On est parti, on commence ». Ils pénètrent alors dans le bâtiment et commencent à tirer sur les spectateurs par rafale, en criant « Allahu akbar ». Pendant une vingtaine de minutes, les trois hommes assassinent froidement les spectateurs, un par un, cherchant également à abattre les membres du groupe Eagles of Death Metal. Des otages, utilisés comme boucliers humains, sont placés devant les portes et les

fenêtres de la salle. La tuerie fait 90 morts et des dizaines de blessés graves avant que les forces de police n'interviennent.

22 h : Après qu'un commissaire de la BAC a abattu l'un des terroristes, peu après 22 heures, les deux terroristes restants se retranchent à l'étage dans une pièce, emmenant avec eux une vingtaine d'otages. À 22 h 15, soixante hommes de la BRI prennent le relais de leurs collègues de la BAC restés à l'extérieur du Bataclan et qui ont essuyé le feu des terroristes. La BRI investit les lieux, appuyés par dix homologues du RAID. Entre 22 h 15 et 22 h 30, elle sécurise le bas de la salle puis quarante hommes de la BRI formant deux colonnes montent au premier étage vers 23 h. À 23 h 15, ils repèrent les deux terroristes retranchés derrière leurs otages dans un couloir. Le policier négociateur va alors entamer pendant près d'une heure un dialogue depuis l'extérieur du bâtiment, installé dans un véhicule banalisé avec un collègue et la psychologue de la BRI. Cinq appels sont passés entre 23 h 27 et 0 h 18. Les terroristes demandent à la BRI de les laisser sortir. Un policier raconte : « Ils étaient très énervés, très speed et confus ».

0h18 : Au terme de tractations infructueuses, l'assaut est donné par les fonctionnaires de la BRI de Paris, soutenus par les effectifs du RAID, positionnés au rez-de-chaussée. Progressant, dans un couloir très étroit, sous le feu des Kalachnikov, la colonne enjambe les otages, bouclier tactique en tête, et les exfiltre les uns après les autres jusqu'à parvenir à neutraliser les deux terroristes et à libérer tous les otages, vivants. Au cours de l'assaut, l'un des kamikazes, réussit à actionner son gilet explosif. La déflagration blessera grièvement un policier de la BRI. L'assaut aura duré cinq minutes. Tous les otages sont évacués sains et saufs lors de cette ultime action.

Au final ces trois attentats auront fait 130 morts, et 413 blessés.

Condition Contrôle

Le mardi 20 avril, les deux étudiants arrivent à l'école à 11h10 dans deux voitures différentes. Ils se garent dans deux parkings différents sur des emplacements non-réservés. Ils s'équipent et se rejoignent peu après. Les deux tireurs se postent en deux points stratégiques offrant une bonne vue sur la cafétéria de l'université et contrôlant les deux principales sorties.

À 11h14 une explosion se produit à 800m de l'école. Le but était de créer une diversion et de mobiliser pompiers et policiers à ce moment. Les pompiers parviennent rapidement à éteindre le feu créé par l'explosion. Les deux étudiants pénètrent alors dans la cafétéria de l'établissement pour y placer deux bonbonnes de gaz propane de 9 kg, chacune censée exploser à 11h17. Une fois les charges placées, ils ressortent et reprennent leurs places respectives près de leurs voitures. Les tireurs ont prévu de garder les sorties afin d'éliminer les étudiants survivants qui tenteraient de s'échapper. À 11h17, les bombes de la cafétéria n'ayant pas explosé, ils s'arment pour donner l'assaut. Ils se dirigent vers l'escalier de l'entrée ouest qui est le point le plus élevé du campus.

Un témoin entend crier « Go ! Go ! ». Les deux adolescents ouvrent le feu et tirent sur deux étudiants qui étaient en train de déjeuner assis sur un monticule herbeux à côté de l'entrée ouest de l'école. L'un des tueurs enlève alors son manteau, sort son 9 mm semi-automatique et tire sur un groupe de

trois étudiants vers l'escalier ouest. Ils tirent ensuite sur des étudiants assis dans l'herbe vers le sud. Ils descendent alors l'escalier et tirent une fois de plus. Ils se dirigent vers la cafétéria, certainement pour voir pourquoi les bombes au propane n'ont pas explosé. Une professeur d'arts plastiques, pensant que les deux américains tournaient une vidéo amateur, se dirige vers eux et les interpelle en leur demandant d'arrêter. Ils tirent alors en sa direction, mais celle-ci réussit à se réfugier dans la bibliothèque, où elle parvient à prévenir la police par téléphone.

À 11h24, un premier officier de police se rend rapidement sur les lieux et échange des tirs avec les deux étudiants meurtriers. L'alerte est alors donnée. Lorsque l'officier est à court de munitions, les deux étudiants se dirigent vers la cafétéria en traversant les couloirs, tirant sur ceux qu'ils rencontrent sans distinction. Plusieurs mourront.

Les deux élèves pénètrent alors dans la bibliothèque où se cachent sous les tables 52 étudiants et 2 professeurs. L'un des tueurs crie : « Tous ceux avec des casquettes blanches, levez-vous ! C'est pour tout ce que vous nous avez fait subir durant ces quatre dernières années ». Il était de coutume de porter une casquette blanche chez les sportifs de l'école. Les deux étudiants tirent sur les élèves cachés sous les tables, discutent avec certains avant de les abattre ou de les laisser partir. Ils commettent dans ce lieu le plus grand nombre de victimes : 10 élèves périssent en 7 minutes. Ils lancent également plusieurs bombes, bien que nombre d'entre elles ne soient pas fonctionnelles. Ils évoquent alors l'idée de commencer à poignarder les élèves : « Peut-être qu'on devrait commencer à poignarder les gens, ça serait plus drôle ».

À 11h42, les deux tueurs sortent de la bibliothèque, mettant ainsi fin au massacre. Ils se rendent à la cafétéria qu'ils vandalisent puis errent dans les couloirs de l'établissement, se moquant des élèves cachés dans leurs salles de classe ou dans les toilettes en les menaçant de mort, mais sans toutefois se servir à nouveau de leurs armes. Enfin, les deux meurtriers retournent à la bibliothèque d'où toutes les victimes, excepté deux blessés, ont fui. Après quelques échanges de coup de feu avec les policiers situés à l'extérieur, ils se suicident à 12h08, en se tirant une balle dans la tête.

Au final, la tuerie de Columbine aura fait 15 morts, ainsi que 24 blessés.

Annexe 2 : description des résultats pour le modèle D-dominant (Étude 5)

Modèle D-dominant à quatre paramètres. Pour rappel, l'ajustement (T_1) aux données du modèle D-dominant quant à la SMT-Noir/Blanc (M_{pred} = 0.05, M_{obs} = 0.16, p =.02, w = 0.03) et la SMT-Maghrébin/Blanc (M_{pred} = 0.05, M_{obs} = 0.09, p = .12, w = 0.03) était bon. Contrairement au modèle SAC-dominant, le paramètre SAC était significativement plus important dans la SMT-Maghrébin/Blanc (M_{SAC} = .87, $IC_{95\%}$[.69 ; .98]) que dans la SMT-Noir/Blanc (M_{SAC} = .54, $IC_{95\%}$[.31 ; .76]), ΔSAC = .34, $IC_{95\%}$[.06 ; .60], p_B = .01. Ainsi, les participants ont plus activé leur stéréotype dans la SMT-Maghrébin/Blanc que dans la SMT-Noir/Blanc. De même, le paramètre SAP était significativement plus important dans la SMT-Maghrébin/Blanc (M_{SAP} = .55, $IC_{95\%}$[.50 ; .61]) que dans la SMT-Noir/Blanc (M_{SAP} = .43, $IC_{95\%}$[.36 ; .50]), ΔSAP = .12, $IC_{95\%}$[.04 ; .21], p_B = .003. Ainsi, les participants ont plus appliqué leur stéréotype dans la SMT-Maghrébin/Blanc que dans la SMT-Noir/Blanc. Enfin, nous n'avons pas observé de différences pour les paramètres D et G entre la SMT Maghrébin/Blanc (M_D = .11, $IC_{95\%}$[.01 ; .15] ; M_G = .26, $IC_{95\%}$[.22 ; .31]) et la SMT-Noir/Blanc (M_D = .12, $IC_{95\%}$[.01 ; .16] ; M_G = .30, $IC_{95\%}$[.24 ; .35]), ces différences étant non significatives pour le paramètre D, ΔD = -.01, $IC_{95\%}$[-.06 ; .04], p_B = .65, comme pour le paramètre G, ΔG = -.03, $IC_{95\%}$[-.10 ; .04], p_B = .83.

Modèle D-dominant à six paramètres. L'ajustement (T_1) aux données du modèle D-dominant quant à la SMT-Noir/Blanc (M_{pred} = 0.05, M_{obs} = 0.10, p =.09, w = 0.03), et la SMT-Maghrébin/Blanc (M_{pred} = 0.05, M_{obs} = 0.09, p = .12, w = 0.03) était bon. Intéressons-nous à la différence entre les paramètres SAC et SAP associés aux visages maghrébins ou aux visages noirs. Tout d'abord, nous n'avons pas obtenu de différence significative quant au paramètre SAC entre les visages maghrébins (M_{SAC} = .54, $IC_{95\%}$[.34 ; .93]) et les visages noirs (M_{SAC} = .69, $IC_{95\%}$[.21 ; .99]), ΔSAC = -.15, $IC_{95\%}$[-.57 ; .40], p_B = .72, bien que ce paramètre ait été descriptivement plus important dans cette dernière condition. En revanche, nous avons obtenu un SAP lié aux visages maghrébins (M_{SAP} = .81, $IC_{95\%}$[.55 ; .99]) significativement plus élevé que le SAP lié aux visages noirs (M_{SAP} = .40, $IC_{95\%}$[.28 ; .54]), ΔSAP = .41, $IC_{95\%}$[.12 ; .65], p_B = .001. En d'autres termes, les participants passant la SMT-Maghrébin/Blanc ont significativement plus appliqué leur stéréotype « Maghrébins = Menace » que les participants passant la SMT-Noir/Blanc n'ont appliqué leur stéréotype « Noirs = Menace ». Comme on pouvait s'y attendre, nous n'avons pas obtenu de différence quant au paramètre SAC entre les

visages blancs de la SMT-Maghrébin/Blanc (M_{SAC} = .60, $IC_{95\%}$[.35 ; .93]) et les visages blancs de la SMT-Noir/Blanc (M_{SAC} = .62, $IC_{95\%}$[.25 ; .94]), ΔSAC = -.02, $IC_{95\%}$[-.50 ; .55], p_B = .55. De la même manière, nous n'avons pas obtenu de différence significative entre le paramètre SAP lié aux visages blancs de la SMT-Maghrébin/Blanc (M_{SAP} = .39, $IC_{95\%}$[.14 ; .55]) et celui lié aux visages blancs de la SMT-Noir/Blanc (M_{SAP} = .41, $IC_{95\%}$[.08 ; .55]), ΔSAP = -.02, $IC_{95\%}$[-.32 ; .36], p_B = .57.

Annexe 3 : tableaux de corrélations entre les paramètres du modèle et les mesures directes (Études 6 et 7)

Étude 6

Tableau 3

Moyennes, écarts-types et corrélations (avec intervalles de confiance à 95%) entre les différents paramètres du modèle et le score de préjugé (PREJ).

Variable	*M*	*ET*	SAC	SAP	D	G	λ
SAC	0.97	0.48					
SAP	0.58	0.08	-.25** [-.42, -.07]				
D	0.15	0.16	-.02 [-.20, .17]	-.23* [-.40, -.05]			
G	0.28	0.22	-.10 [-.28, .08]	.25** [.07, .41]	.06 [-.12, .24]		
λ	0.46	0.23	-.30** [-.46, -.12]	.90** [.86, .93]	-.08 [-.27, .10]	.40** [.23, .54]	
PREJ	2.92	0.99	-.03 [-.21, .16]	.37** [.20, .52]	-.25** [-.41, -.07]	.12 [-.07, .30]	.24** [.06, .41]

*Note. M et ET représentent respectivement la moyenne et l'écart type. Les valeurs entre crochets représentent les intervalles de confiance à 95% pour chaque corrélation. * indique p < .05. ** indique p < .01.*

Étude 7

Tableau 4

Moyennes, écarts-types et corrélations (avec intervalles de confiance à 95%) entre les différents paramètres du modèle, le score de préjugé et les sous-échelles de motivation à réduire l'expression du préjugé.

Variable	*M*	*ET*	SAC	SAP	D	λ	PREJ	RAD	CAP	EMS
SAC	0.67	0.30								
SAP	0.56	0.08	.21** [.09, .32]							
D	0.15	0.13	-.21** [-.33, -.10]	-.29** [-.39, -.17]						
λ	0.49	0.24	.39** [.28, .48]	.86** [.82, .89]	-.44** [-.53, -.34]					
PREJ	0.02	0.95	.11 [-.01, .23]	.21** [.09, .32]	.10 [-.02, .21]	.23** [.12, .34]				
RAD	0.03	0.82	.09 [-.03, .20]	.04 [-.08, .16]	-.03 [-.15, .09]	.07 [-.05, .19]	.00 [-.12, .12]			
CAP	-0.01	0.85	.15* [.03, .26]	.09 [-.03, .20]	-.09 [-.20, .03]	.10 [-.02, .21]	-.27** [-.38, -.16]	.29** [.18, .40]		
EMS	0.01	1.53	.17** [.05, .28]	.25** [.13, .35]	-.02 [-.14, .10]	.21** [.10, .32]	.20** [.08, .31]	.24** [.13, .35]	.33** [.22, .43]	
IMS	0.08	1.27	-.01 [-.13, .11]	-.11 [-.23, .01]	.01 [-.11, .13]	-.10 [-.22, .02]	-.55** [-.63, -.46]	.13* [.01, .25]	.50** [.41, .58]	.08 [-.04, .19]

*Note. M et ET représentent respectivement la moyenne et l'écart type. Les valeurs entre crochets représentent les intervalles de confiance à 95% pour chaque corrélation. PREJ = score de préjugé, RAD = score à la sous-échelle « retenue pour éviter les différents », CAP = score à la sous-échelle « soucieux d'agir sans préjugé », EMS = score à la sous-échelle « motivation externe à répondre sans préjugés », IMS = score à la sous-échelle « motivation interne à répondre sans préjugés ». * indique p < .05. ** indique p < .01.*

Annexe 4 : Échelle de Préjugés Généralisés (Études 3, 6 et 7)

Vous trouverez ci-dessous une liste de propositions portant sur la diversité et l'immigration. Vous pouvez être en total accord, en total désaccord ou plus neutre vis-à-vis de chacune de ces propositions. Lisez-les attentivement et donnez votre propre avis par rapport à chacune d'entre elles, en entourant le chiffre correspondant, sachant que :

1 = Pas du tout d'accord	**2** = Pas d'accord	**3** = Plutôt pas d'accord	**4** = Ni d'accord ni pas d'accord	**5** = Plutôt d'accord	**6** = D'accord	**7** = Tout à fait d'accord

1- On peut facilement comprendre la colère que ressentent les Maghrébins en France.	1	2	3	4	5	6	7
2- Les Français devraient être prioritaires en matière de logement.	1	2	3	4	5	6	7
3- Les étrangers qui vivent en France devraient avoir le droit de vote.	1	2	3	4	5	6	7
4- Il est normal que les étrangers en situation irrégulière soient renvoyés dans leur pays d'origine.	1	2	3	4	5	6	7
5- La diversité qu'apportent les étrangers est un enrichissement pour le pays.	1	2	3	4	5	6	7
6- Les Français devraient être prioritaires en matière d'augmentation des prestations sociales.	1	2	3	4	5	6	7
7- Le port du Tchador au sein des écoles devrait être autorisé.	1	2	3	4	5	6	7
8- Les Français devraient être prioritaires en matière d'emploi.	1	2	3	4	5	6	7
9- Il est insensé d'attribuer aux Algériens et aux Marocains les problèmes économiques de la France.	1	2	3	4	5	6	7
10- On devrait limiter plus strictement l'entrée des familles étrangères en France.	1	2	3	4	5	6	7
11- Je considère que la société est injuste à l'égard des Arabes.	1	2	3	4	5	6	7
12- S'il y a beaucoup de chômage en France, c'est parce que les étrangers prennent le travail des Français.	1	2	3	4	5	6	7
13- Il faudrait donner plus de droits aux immigrants.	1	2	3	4	5	6	7
14- On ne devrait pas accorder aussi facilement la nationalité française.	1	2	3	4	5	6	7
15- Je ne serais pas inquiet si la plupart de mes amis à l'université étaient d'origine arabe.	1	2	3	4	5	6	7

Annexe 5 : échelles et sous-échelles de motivation à réduire l'expression de ses préjugés (Étude 7)

Échelle de motivation à contrôler ses réactions teintées de préjugé (MCPR)

Vous trouverez ci-dessous une liste de propositions portant sur vos préjugés et la motivation que vous pouvez avoir à ne pas les utiliser. Vous pouvez être en total accord, en total désaccord ou plus neutre vis-à-vis de chacune de ces propositions. Lisez-les attentivement et donnez votre propre avis par rapport à chacune d'entre elles, en entourant le chiffre correspondant, sachant que :

1 = Pas du tout d'accord	**2** = Pas d'accord	**3** = Plutôt pas d'accord	**4** = Ni d'accord ni pas d'accord	**5** = Plutôt d'accord	**6** = D'accord	**7** = Tout à fait d'accord

1- Dans notre société actuelle il est important de ne pas être perçu comme quelqu'un ayant des préjugés de quelques manières que ce soit.	1	2	3	4	5	6	7
2- J'exprime toujours mes pensées et sentiments, aussi controversés soient-ils.	1	2	3	4	5	6	7
3- Je suis en colère contre moi-même quand j'ai une pensée ou un sentiment qui pourrait être considéré comme étant de l'ordre du préjugé.	1	2	3	4	5	6	7
4- Si je participais à une discussion dans un cours et qu'un étudiant d'origine maghrébine exprimait une opinion avec laquelle je ne suis pas d'accord, j'hésiterai à exprimer mon propre point de vue.	1	2	3	4	5	6	7
5- Passer sa vie à se demander si on ne risque pas d'offenser quelqu'un provoque plus d'ennui que ça n'en vaut la peine.	1	2	3	4	5	6	7
6- Il est important pour moi que les autres ne pensent pas que j'ai des préjugés.	1	2	3	4	5	6	7
7- Je pense qu'il est important de se comporter selon les normes de la société.	1	2	3	4	5	6	7
8- Je fais attention à ne pas offenser mes amis, mais je ne me soucie pas d'offenser les personnes que je ne connais pas ou que je n'aime pas.	1	2	3	4	5	6	7
9- Je pense qu'il est important de dire ce que l'on pense plutôt que de se soucier du fait que l'on puisse offenser quelqu'un.	1	2	3	4	5	6	7
10- Il n'est jamais acceptable d'exprimer ses préjugés.	1	2	3	4	5	6	7
11- Je me sens coupable quand j'ai des pensées ou sentiments négatifs à propos d'une personne d'origine maghrébine.	1	2	3	4	5	6	7
12- Quand je parle à une personne d'origine maghrébine, il est important pour moi qu'elle ne pense pas que j'ai des préjugés.	1	2	3	4	5	6	7
13- Cela me dérange beaucoup quand j'ai l'impression que j'ai offensé quelqu'un, donc je fais toujours attention à prendre en considération les sentiments des autres.	1	2	3	4	5	6	7
14- Si j'ai une pensée ou un sentiment qui correspond à un préjugé, je le garde pour moi.	1	2	3	4	5	6	7
15- Je ne ferais jamais des blagues qui pourraient offenser les autres.	1	2	3	4	5	6	7
16- Je n'ai pas peur de dire aux autres ce que je pense, même si je sais qu'ils ne sont pas d'accord avec moi.	1	2	3	4	5	6	7
17-Si une personne qui me met mal à l'aise s'asseyait à côté de moi dans le bus, je n'hésiterais pas à changer de place.	1	2	3	4	5	6	7

Échelle de motivation interne et externe à répondre sans préjugés (MRWP)

Vous trouverez ci-dessous une liste de propositions portant sur vos préjugés et la motivation que vous pouvez avoir à ne pas les utiliser. Vous pouvez être en total accord, en total désaccord ou plus neutre vis-à-vis de chacune de ces propositions. Lisez-les attentivement et donnez votre propre avis par rapport à chacune d'entre elles, en entourant le chiffre correspondant, sachant que :

1 = Pas du tout d'accord	**2** = Fortement pas d'accord	**3** = Modérément pas d'accord	**4** = Légèrement pas d'accord	**5** = Ni d'accord ni pas d'accord	**6** = Légèrement d'accord	**7** = Modérément d'accord	**8 =** Fortement d'accord	**9 =** Tout à fait d'accord

1- J'essaie d'agir sans préjugés envers les personnes d'origine maghrébine car c'est important pour moi sur le plan personnel.	1	2	3	4	5	6	7	8	9
2- Selon mes valeurs personnelles, il est acceptable d'utiliser les stéréotypes à l'égard des personnes d'origine maghrébine.	1	2	3	4	5	6	7	8	9
3- Si j'agissais de manière conforme aux préjugés envers les personnes d'origine maghrébine, je serais préoccupé par le fait que les gens pourraient m'en vouloir.	1	2	3	4	5	6	7	8	9
4- Mes convictions personnelles me motivent à ne pas utiliser mes préjugés envers les personnes d'origine maghrébine.	1	2	3	4	5	6	7	8	9
5- J'essaye de cacher toutes les pensées négatives que j'ai envers les personnes d'origine maghrébine afin d'éviter que les gens ne réagissent de manière négative à mon encontre.	1	2	3	4	5	6	7	8	9
6- Me conduire sans préjugés à l'égard des personnes d'origine maghrébine est important pour l'image que j'ai de moi.	1	2	3	4	5	6	7	8	9
7- J'essaie de ne pas passer pour quelqu'un ayant des préjugés envers les personnes d'origine maghrébine afin d'éviter la désapprobation des gens.	1	2	3	4	5	6	7	8	9
8- En raison de mes valeurs personnelles, je pense qu'il n'est pas acceptable d'utiliser les stéréotypes envers les personnes d'origine maghrébine.	1	2	3	4	5	6	7	8	9
9- J'essaie de ne pas agir de manière conforme aux préjugés envers les personnes d'origine maghrébine en raison de la pression sociale.	1	2	3	4	5	6	7	8	9
10- En raison de la norme du "politiquement correct", j'essaie d'apparaître comme n'ayant pas de préjugés à l'égard des personnes d'origine maghrébine.	1	2	3	4	5	6	7	8	9

Annexe 6 : amorces (prénoms) maghrébins, français et neutres (Étude 8)

Tableau 5

Amorces de prénoms (français, maghrébins et neutres) utilisés dans l'Étude 8.

Prénoms français	Prénoms maghrébins	Prénoms neutres
Alexandre	Mustafa	Gervase
Julien	Amin	Brynmor
Laurent	Hakim	Gebhard
Philippe	Karim	Dieter
Pierre	Malik	Sesto
Nicolas	Abdul	Jochem
Louis	Nadir	Ernie
Antoine	Farouk	Paderau
Adrien	Aziz	Launce
Baptiste	Munir	Lammert
Benoit	Farid	Haggai
Thibault	Omar	Gomer
Theo	Zahir	Bernt
Bastien	Youssef	Oskari
Luc	Kader	Cree
Etienne	Hassan	Patsy
Mathieu	Ahmed	Joord
Fabrice	Mohammed	Elbert
Sylvain	Hamid	Bartel
Clement	Ali	Carmi
Maxime	Salah	Nibori
Olivier	Habib	Ebbe
Marc	Jamal	Imre
Romain	Samir	Darrel

Annexe 7 : réanalyse des étapes de validation des modèles SAC-dominant et D-dominant (Krieglmeyer & Sherman, 2012 ; méthode *latent-trait*)

Tableau 6

Validation du paramètre D (Krieglmeyer & Sherman, 2012 ; Étude 2).

	SAC-dominant	D-dominant
D1	.13 [.01 ; .29]	.02 [.00 ; .06]
D2	.003 [.00 ; .01]	.00 [.00 ; .01]
G1	.05 [.00 ; .19]	.16 [.05 ; .30]
G2	.17 [.07 ; .28]	.17 [.07 ; .29]
SAC1	.80 [.64 ; .93]	.82 [.67 ; .99]
SAC2	.86 [.72 ; .97]	.85 [.69 ; .98]
SAP1	.62 [.54 ; .71]	.62 [.53 ; .72]
SAP2	.62 [.55 ; .70]	.62 [.54 ; .71]

Tableau 7

Validation du paramètre G (Krieglmeyer & Sherman, 2012 ; Étude 3).

	SAC-dominant	D-dominant
D1	.03 [.002 ; .08]	.02 [.00 ; .06]
D2	.03 [.00 ; .09]	.02 [.00 ; 05]
G1	.06 [.01 ; .15]	.08 [.02 ; .16]
G2	.41 [.23 ; .56]	.42 [.25 ; .56]
SAC1	.76 [.55 ; .93]	.76 [.55 ; .94]
SAC2	.96 [.87 ; .99]	.98 [.93 ; .99]
SAP1	.61 [.52 ; .69]	.62 [.53 ; .70]
SAP2	.62 [.54 ; .70]	.62 [.55 ; .73]

Tableau 8

Validation du paramètre SAC (Krieglmeyer & Sherman, 2012 ; Étude 4).

	SAC-dominant	D-dominant
D1	.15 [.03 ; .28]	.02 [.00 ; .06]
D2	.08 [.02 ; .14]	.03 [.00 ; 07]
G1	.01 [.00 ; .08]	.13 [.03 ; .23]
G2	.17 [.07 ; .30]	.20 [.09 ; .33]
SAC1	.71 [.57 ; .85]	.78 [.58 ; .93]
SAC2	.96 [.87 ; .99]	.96 [.83 ; 1]
SAP1	.58 [.48 ; .68]	.57 [.47 ; .67]
SAP2	.70 [.62 ; .78]	.71 [.60 ; .82]

Tableau 9

Validation du paramètre SAP (Krieglmeyer & Sherman, 2012 ; Étude 5).

	SAC-dominant	D-dominant
D1	.08 [.01 ; .17]	.00 [.00 ; .02]
D2	.04 [.00 ; .11]	.03 [.00 ; 06]
D3	.07 [.02 ; .13]	.04 [.01 ; 09]
G1	.37 [.20 ; .53]	.38 [.20 ; .55]
G2	.40 [.23 ; .58]	.40 [.22 ; .59]
G3	.29 [.16 ; .43]	.30 [.16 ; 44]
SAC1	.99 [.94 ; 1]	.98 [.92 ; 1]
SAC2	.90 [.74 ; .99]	.95 [.82 ; 1]
SAC3	.86 [.65 ; .98]	.87 [.68 ; 1]
SAP1	.96 [.89 ; .99]	.92 [.60 ; 1]
SAP2	.49 [.28 ; .67]	.47 [.28 ; 69]
SAP3	.63 [.45 ; .81]	.64 [.45 ; 83]