

Direct and indirect compensation

**Of consequences and antecedents of
dimensional compensation in social evaluation**

Direct and indirect compensation

Of consequences and antecedents of dimensional compensation in social evaluation

Mathias SCHMITZ

Promoteur

Vincent Yzerbyt (UCLouvain)

Thèse présentée en vue de
l'obtention du grade de Docteur
en sciences psychologiques et
de l'éducation

Comité d'encadrement

Stéphanie Demoulin (UCLouvain)
Olivier Klein (ULB)

Présidente

Prof. Stéphanie Demoulin (Université catholique de Louvain)

Promoteur

Prof. Vincent Yzerbyt (Université catholique de Louvain)

Comité d'accompagnement et jury

Prof. Stéphanie Demoulin (Université catholique de Louvain)

Prof. Olivier Klein (Université Libre de Bruxelles)

Prof. Darío Páez (University of Basque Country)

Prof. Odile Rohmer (Université de Strasbourg)

Prof. Frédéric Nils (Université catholique de Louvain)

Résumé

L'effet de compensation renvoie au compromis entre les deux dimensions fondamentales de la chaleur et de la compétence lors de la comparaison de deux cibles sociales. Les recherches sur les antécédents motivationnels demeurent limitées, en partie parce que la majorité d'entre elles se sont appuyées sur des mesures directes, malgré des études suggérant que cet effet peut être sensible aux questions de désirabilité sociale. En outre, on connaît peu des conséquences que ce phénomène pourrait avoir au-delà de son impact sur les aspects conceptuels. En recourant à des mesures directes et indirectes, notre programme de recherche révèle que des motivations distinctes peuvent être à l'origine de la compensation, notamment en fonction du point de vue des percevants en tant qu'observateurs externes ou membres d'un groupe. De plus, nous soulignons l'omniprésence de l'effet de compensation en montrant qu'il peut non seulement biaiser la conceptualisation abstraite des groupes sociaux, mais aussi leur représentation visuelle, singulièrement au niveau du visage. Globalement, nos résultats éclairent un ensemble de points importants tout en ouvrant de nouvelles pistes pour la recherche future.

Abstract

The compensation effect relates to the trade-off between the two fundamental dimensions of warmth and competence when comparing two social targets. Research on the motivational antecedents remains scarce, in part because the bulk of investigations relied on direct measure despite evidence suggesting that this effect may be susceptible to social desirability issues. Moreover, there is little knowledge about the consequences of this phenomenon beyond abstract conceptualizations. By relying on direct and indirect measures, our research program reveals that distinct motivations may drive compensation, notably as a function of the vantage point of the perceivers as external observers or group members. Additionally, we highlight the pervasiveness of the compensation effect by showing that it not only distorts the abstract conceptualization of social groups but also their visual representation, specifically in terms of their faces. As a set, our findings shed light on important issues and also offer new avenues for future research.

Table of content

Résumé	5
Abstract	5
Table of content	7
Acknowledgements	13
General introduction	17
Chapter 1: Dimensions of social and face perception	21
1. Introduction	21
2. Horizontal and vertical dimensions of social perception	21
2.1 <i>Self-other perception</i>	22
2.2 <i>Group perception</i>	22
2.3 <i>Facets and the third dimension</i>	24
3. Horizontal and vertical dimensions of face perception	25
3.1 <i>Faces as social stimuli</i>	25
3.2 <i>Face perception</i>	26
3.3 <i>Face representation</i>	27
3.4 <i>Relation between the social and facial dimensions</i>	28
4. Relation between the horizontal and vertical dimensions	29
4.1 <i>Positive relation</i>	29
4.2 <i>Orthogonal relation</i>	30
4.3 <i>Negative relation</i>	30
4.4 <i>Curvilinear relation</i>	31
4.5 <i>Relations between the facial dimensions</i>	32
5. Conclusions	33
Chapter 2: The compensation effect	37
1. Introduction	37

2.	Moderating factors and boundary conditions	37
2.1	<i>Real and fictitious groups</i>	38
2.2	<i>Groups and individuals</i>	39
2.3	<i>Uniqueness of warmth and competence</i>	39
2.4	<i>Comparison context</i>	40
2.5	<i>Low conflict and legitimate status</i>	41
2.6	<i>Observers and group members</i>	41
3.	Underlying motivations	42
3.1	<i>Harmonious intergroup relationship concerns</i>	42
3.2	<i>Justice concern</i>	43
3.3	<i>Self-affirmation concerns</i>	44
3.4	<i>Self-presentation concerns</i>	45
4.	Conclusions	47
Chapter 3: Direct and indirect compensation		51
1.	Introduction	51
2.	Direct and indirect judgments on the fundamental dimensions	52
3.	The reverse correlation as an indirect and spontaneous measure	54
4.	Compensation effect at the direct and indirect level	56
5.	Conclusions	57
Chapter 4: Direct and indirect dimensional compensation		61
1.	Introduction	61
1.1	<i>Accounting for the compensation effect</i>	62
1.2	<i>Compensation effect at the indirect level</i>	64
1.3	<i>The present research</i>	66
2.	Experiment 1	67
2.1	<i>Method</i>	67
2.2	<i>Results</i>	70
2.3	<i>Discussion</i>	72

3.	Experiment 2	73
3.1	<i>Method</i>	74
3.2	<i>Results</i>	74
3.3	<i>Discussion</i>	77
4.	General Discussion	77
5.	Open Practices Statement	81
6.	Notes	81
7.	References	81
Chapter 5: When you look competent but cold		89
1.	Introduction	89
1.1	<i>The dimensional compensation effect</i>	89
1.2	<i>Impressions from faces</i>	90
1.3	<i>Visual representation of faces</i>	91
1.4	<i>The present research</i>	93
2.	Power considerations	93
3.	Experiment 1	94
3.1	<i>Method</i>	94
3.2	<i>Results</i>	96
3.3	<i>Discussion</i>	97
4.	Experiment 2	98
4.1	<i>Method</i>	98
4.2	<i>Results</i>	99
4.3	<i>Discussion</i>	100
5.	Experiment 3	101
5.1	<i>Method</i>	101
5.2	<i>Results</i>	102
5.3	<i>Discussion</i>	102
6.	Experiment 4	103
6.1	<i>Method</i>	103

6.2	<i>Results</i>	103
6.3	<i>Discussion</i>	104
7.	General discussion	105
8.	Transparency statement	106
9.	Endnotes	106
10.	References	107
Chapter 6: Introducing the Brief Reverse Correlation		117
1.	Introduction	117
1.1	<i>The Reverse Correlation Paradigm</i>	117
1.2	<i>Current Limitations of the Traditional-RC</i>	119
1.3	<i>The Brief Reverse Correlation</i>	121
1.4	<i>The Present Research</i>	122
2.	Overview and Analytical Strategy	122
3.	Experiment 1	124
3.1	<i>Method</i>	124
3.2	<i>Results</i>	126
3.3	<i>Discussion</i>	127
4.	Experiment 2	128
4.1	<i>Method</i>	128
4.2	<i>Results</i>	131
4.3	<i>Discussion</i>	136
5.	General Discussion	137
6.	Open Practices	140
7.	References	140
Chapter 7: Facial compensation in existing groups		149
1.	Introduction	149
1.1	<i>Two fundamental dimensions of social and facial perception</i>	149
1.2	<i>The compensation effect</i>	150

1.3	<i>Relations Between Walloons and Flemings</i>	150
1.4	<i>The Present Research</i>	152
2.	Method	152
3.	Results	154
4.	General Discussion	157
5.	Open Practices	160
6.	References	160
	General discussion	167
1.	Objective of the thesis	167
2.	Summary of the empirical findings	168
3.	Discussion	170
4.	Limitations and future research	172
	Conclusion	177
	Bibliographical references	181
	Appendixes	209
1.	Bogus dot estimation task used in Chapter 4	209
2.	Behavioral statements used in Chapter 4 and 5	211
3.	Brief Implicit Association Test	214

Acknowledgements

First of all, I am extremely grateful to my thesis supervisor, Vincent Yzerbyt, for his constant optimism, support, insightful advices, and invaluable expertise. I have learned a great deal from you over the last 6 years, and you have undoubtedly deepened my passion for research, teaching, and statistics! None of this would have been possible without your support.

I am also very thankful to my thesis committee and jury—Stéphanie Demoulin, Olivier Klein, Odile Rohmer, Dario Páez, and Frédéric Nils—for taking part in this adventure, as well as for their precious feedback and support. Thank you in particular to Odile for receiving us in Strasbourg, and to Dario for all the conversations we had during the seminar in Peru.

Un immense merci à vous tous, mes amis, collègues, académiques, et équipe administrative de l'UCLouvain et d'autres universités que j'ai pu connaître tout au long de ce parcours. Merci infiniment pour votre amitié, conseils et soutien ! Merci à toi Julie Terache pour la couverture de la thèse !

Un agradecimiento especial para ti, Agustín Espinosa, por tu constante apoyo, amistad, y buen humor a lo largo de estos años; y ello a pesar de la distancia.

Last but not least, merci du fond du cœur à ma compagne Gina ainsi qu'à ma famille en Belgique et au Pérou pour leur soutien inconditionnel.

General Introduction

General introduction

Despite the wide spectrum of impressions we may build in everyday life, research suggest that their content can be structured around two fundamental dimensions, the “Big Two” (for a review, see Abele et al., in press). Specifically, we tend to think of social groups (and individuals) in terms of how warm (e.g., sociable, friendly, honest) and competent (e.g., efficient, capable, skilled) they are. If we believe groups to be (ill) well intentioned, they will likely be ascribed warmth (coldness). If they are regarded as capable of carrying out their intentions, they will tend to be perceived as competent. A similar bi-dimensional configuration also emerges when we evaluate faces or when we imagine how the prototypical face of a group would look like (Oosterhof & Todorov, 2008). Faces that are mature, look physically strong or masculine are usually perceived as dominant, whereas those that convey emotional expressions of happiness (anger) are frequently judged as (un)trustworthy.

Interestingly, many of our representations about groups exhibit a mixed content in the sense that they are either high on one dimension while being low on the other (Fiske et al., 2002). One way to account for this recurring pattern is through the so-called *compensation effect* (Yzerbyt et al., 2005; also know as *dimensional compensation*; Yzerbyt, 2018). Indeed, when comparing social targets, we often grant the advantage to one group, in contrast to the other, on one of the two fundamental dimensions (e.g., competence), whereas we do the opposite for the second dimension (e.g., warmth). For instance, in the Belgian context, Walloons are often depicted as more sociable and easy-going but less efficient and hard-working than their counterpart, the Flemings.

Although knowledge related to this effect has flourished over the last fifteen years, we still know very little about its motivational antecedents. This is hardly surprising given that compensation has been mostly investigated by straightforwardly asking individuals to report their impressions, despite the fact that this effect might be susceptible to social desirability (e.g., Cambon & Yzerbyt, 2016; Owuamalam et al., 2016; Yzerbyt & Cambon, 2017). Therefore, bypassing or reducing people’s caution when they publicly express their inner impressions (e.g., by means of indirect measures) may provide valuable insights about the underlying motivations that drive compensatory judgments. For instance, privileged groups (e.g., rich people, men, non-disable people) might publicly concede some superiority on the dimensions they care less about (e.g., warmth) to deprived groups (e.g., poor people, women, disable people) because they are concerned not to appear discriminatory. As it turns out, they might not have done so in the absence of social pressures.

Likewise, with a few exceptions (Kervyn, Yzerbyt et al., 2009), researchers have generally overlooked the consequences of compensation beyond abstract conceptualizations of social targets. For instance, is it possible for this effect to also distort people’s and groups’ facial representations? To come back to our Belgian example, it might be that Walloons facially portray themselves as more trustworthy but less dominant than the Flemings.

Evidence for the visual correlates of the compensation effect would therefore highlight the pervasiveness of this phenomenon.

In the light of the dearth of empirical work examining these issues, the present dissertation aims to shed light on some of the antecedents and consequences of the compensation effect by making use of direct and indirect measures. In Chapter 1, we review the dimensions that maps social and facial perception, as well as their relations. Chapter 2 introduces the compensation effect and examines its moderating factors and underlying motivations. In Chapter 3, we review the research regarding the fundamental dimensions and the compensation effect at both the direct and indirect level. In Chapter 4 (Schmitz & Yzerbyt, in press), we test compensation by means of direct and indirect measures and by taking into account the vantage point of view of individuals as external observers or group members. Chapter 5 (Schmitz, Vanbeneden et al., under review) presents evidence regarding the visual correlates of the compensation effect. In Chapter 6 (Schmitz, Rougier et al., under review), we develop and test an improved version of the reverse correlation method, namely the Brief-RC. In Chapter 7 (Schmitz et al., in preparation), we rely on the Brief-RC to test the visual correlates of the compensation effect in existing and meaningful groups. Finally, in the General Discussion of the dissertation, we discuss our empirical findings and offer new avenues of research. Because this dissertation rests on a series of articles, we apologize to the reader for some possible redundancies and cross-references across the presentation of our research program. Also note that the empirical articles are copied as they were submitted or accepted.

Chapter 1

Dimensions of social and face perception

Dimensions of social and face perception

1. Introduction

The first experimental studies on impression formation (for a review on the history of impression formation, see Uleman & Kressel, 2013) can be traced back six decades ago to Asch's (1946, 1952) seminal efforts. In his studies, participants were asked to form an impression about a fictitious person that was described by a series of traits (e.g., intelligent, skillful, industrious, warm, determined, practical, and cautious). His results revealed that some traits, which he called "central" (e.g., warm–cold), had a larger impact on the overall judgments in comparison to other, more peripheral, traits (e.g., polite–blunt). Moreover, he showed that first impressions acted as lenses for the interpretation of further information about the target (i.e., primacy effect; e.g., Anderson & Barrios, 1961; Hogarth & Einhorn, 1992). Building upon these findings, he promoted the idea—rooted in the Gestalt theory—that perceived personality traits about a given target are not independent from one another but rather clustered and interrelated. Taken together, Asch's innovative studies revealed some of the complexities that lie at the heart of impression formation and set the stage for the development of research on implicit personality theory (Rosenberg & Jones, 1972; Schneider, 1973), which tackles the patterns and beliefs about how perceived personality traits are structured.

Following Asch's research, investigations over the next half-century remained mainly focused on the processes underlying impression formation. In the domain of stereotypes, which can be understood as the set of beliefs and attributes associated to a group (Hilton & von Hippel, 1996; Stangor, 2009), the attention was devoted to how they are formed, activated, applied, transmitted, and modified (for a review see Fiske & Tablante, 2015; Kite & Whitley, 2016). On the one hand, investigating stereotypes processes was appealing because of their presumed universality as well as their predictability from a number of systematic principles. On the other, stereotypes content appeared to be obeying to no clear rules. Still, even though process research long dominated the field, research started to accumulate, revealing the existence of two fundamental dimensions of judgment across several domains of social perception.

2. Horizontal and vertical dimensions of social perception

The realization that two fundamental dimensions—the so-called *Big Two* of social perception (Abele & Brack, 2013; Abele & Wojciszke, 2013; Paulhus & Trapnell, 2008)—map the content of social judgment is one of the most notable achievements of research on social perception (Abele et al., 2008; Abele & Wojciszke, 2007; Abele & Wojciszke, 2014; Fiske et al., 2002; Judd et al., 2005). Even more impressive is the fact that these two dimensions seem to emerge consistently whatever the nature of the social target. That is, whether the judgments pertain to general personality, the self, others, groups, organizations,

nations, or cultures, the same two dimensions show up again and again (e.g., Blake & Mouton, 1964; Digman, 1997; Fiedler, 1967; Fiske et al., 2002; Markus & Kitayama, 1991; Phalet & Poppe, 1997; Rosenberg et al., 1968; Wiggins, 1991; Wojciszke, 1994). Because of different theoretical frameworks and methodological approaches, these two dimensions have been labeled and operationalized in slightly different ways but most researchers argue that they share a common core (Abele et al., in press; Abele & Wojciszke, 2007; Carrier, Louvet, Chauvin et al., 2014; Judd et al., 2005; Kervyn et al., 2010).

2.1 Self-other perception

In the literature on self-other perception and interpersonal attitudes, the two dimensions usually come across as *communion* and *agency* (Abele, & Brack, 2013; Abele et al., 2008; Abele et al., 2016; Abele & Wojciszke, 2007). These terms were first introduced by Bakan (1966) to denote two fundamental modalities of human existence. *Communion* refers to an individual's longing to be part of a larger organism (e.g., a group) and can be fulfilled through the establishment and maintenance of relationships with others, which requires qualities such as caring for others, emotional expressivity, and cooperativeness. In contrast, *agency* characterizes the person's need to control their environment, to individuate and expand the self. This mode of existence can be achieved through individual accomplishments and qualify persons that are ambitious, instrumental, dominant, and focused on goal-attainment (Abele et al., 2016; Abele & Wojciszke, 2007). Consequently, agency prompts people to "get ahead" whereas communion urges them to "get along" (Abele et al., 2016).

From a functional analysis (Peeters, 1992; see also Dual Perspective Model, DPM; Abele & Wojciszke, 2014; Wojciszke et al., 2011), communal qualities are said to be *other-profitable* as they directly benefit the action recipient while they remain ambiguous for the trait possessor. For instance, it is advantageous (as an action recipient) to interact with someone who is helpful; but being too nice (as a trait possessor) can result in exploitation by others. Conversely, agency traits are *self-profitable* for the holder of the characteristic but remain equivocal when evaluated in others. Indeed, it is (almost) always the case that being competent and agentic will be beneficial to achieve ones' goals. It is important to note that this is the case "by default", that is, when there is no motivational process at stake on the side of the perceiver. Agency traits may become other-profitable when perceiver's goals depend on the target's competence (i.e., interdependence contexts; see Abele & Wojciszke, 2007, 2014; Carrier et al., 2019; Cuddy et al., 2011;).

2.2 Group perception

In the domain of group perception, the Stereotype Content Model (SCM; Fiske et al., 1999; Fiske et al., 2002; see also Fiske, 2015, 2018a) posits that the content of stereotypes varies along two core dimensions. That is, when people form an impression about groups, but also individuals (Russell & Fiske, 2008) and cultures (Cuddy et al., 2009; Durante et al., 2017; Durante et al., 2013), their judgments can be summarized (as established by cluster and dimension reduction analyses) by how they are characterized in terms of warmth and competence. The *warmth* dimension translates the target's perceived intentions or motives

and is associated with traits such as friendly, sociable, sincere, or honest whereas the *competence* dimension informs about the capacity of the target to enact those intentions and encapsulates traits such as competent, intelligent, confident, or independent.

From a functional perspective of inter-individual and inter-group relations, this bi-dimensional configuration can be predicted from two socio-structural variables—*competition* (also known as interdependence) and *status* (Fiske et al., 2002). On the one hand, people like (dislike) and infer warm (cold) about groups that appear cooperative (competitive), that is, groups whose goals seem compatible (incompatible) with those of the perceiver or of the ingroup (Fiske & Ruscher, 1993). More generally, (a lack of) warmth can be forecasted from both realistic (e.g., resources, welfare, political and economic power) and symbolic (e.g., worldviews, values or beliefs) threats (Kervyn et al., 2015; see also Stephan et al., 2009). On the other hand, powerful (powerless) or high (low) status groups inspire respect (disrespect) and are depicted as competent (incompetent) in line with a meritocratic worldview (Cuddy et al., 2008; Fiske et al., 2002).

The orthogonal combination of the two core dimensions from the SCM span four quadrants that maps the content of a wide variety of stereotypes and intergroup relations. In this regard, high-status groups with cooperative goals come across as competent and warmth and elicit emotions of pride and admiration. This quadrant is often associated with allies or ingroups that drive ingroup favoritism. Conversely, the so-called derogated groups (e.g., immigrants, welfare recipients), that is, those that are powerless and disliked, are seen as both dumb and cold-hearted, and bring out emotions of disgust and contempt. Although many groups fall into the aforementioned non-mixed quadrants (i.e., univalent on both dimensions of social judgment), the majority of groups belong to the so called-mixed stereotypes (i.e., positively valenced on one dimension but negatively on the other). Specifically, collaborating but subordinate groups (e.g., traditional women, disable people) are depicted as sociable but foolish, and trigger emotions of pity and sympathy. Inversely, rival but prestigious groups (e.g., rich people, female professionals) are pictured as successful but hostile and evoke emotions of envy and jealousy. The paternalistic (i.e., incompetent but warmth) and envious (i.e., competent but cold) stereotypes are thought to justify the *status quo* as they provide a balanced view of societal groups by conferring both advantages and weakness to all groups (Fiske et al., 2002; Jost & Kay, 2005). Furthermore, an extension of this model (the Behavior from Intergroup Affect and Stereotypes Map or BIAS map; Cuddy et al., 2007, 2008) suggests that perception of warmth leads to active behaviors (active harm/facilitation) whereas competence give rise to passive behaviors (passive harm/facilitation).

From an evolutionary perspective, the bi-dimensional structure of stereotypes seems to derive from two basic survival questions that our ancestors (but also modern people) asked when encountering outgroups' members. First, are they "foes or friends", that is, do they intent to cause *harm* or not (i.e., warmth/communion)? Second, are they *capable* of carrying out their intentions (i.e., competence/agency)? The answers to these questions would therefore help individuals in their everyday life social interactions and may hold a survival value. Interestingly, the order of these inquiries does matter and reveals the primacy of warmth over competence (Wojciszke et al., 1998; Ybarra et al., 2001). It is primordial to first

discern the *intentions* of the person as they convey important information relevant for approach-avoidance tendencies, and only then if they have the *capability* to enact their intentions (Abele & Bruckmüller, 2011; Brambilla et al., 2013; Wojciszke et al., 1998; see also Oosterhof & Todorov, 2008; Todorov, Said et al., 2008). In this sense, social thinking is for doing (Fiske, 1992, 2018b) and the basic responses of doing is to approach well- and avoid ill-intentioned people (Peeters, 1992). This reasoning parallels the other-profitability quality of communal characteristics perceived in others (see above). A plethora of empirical research support the (frequent) warmth/communion over competence/agency effect by showing that information diagnostic of the former is more accessible in memory, processed faster, more consensual, and has more weight, than the later (for an overview, see Abele et al., in press).

The two fundamental dimensions of social and other- as well as self-perception are sometimes called in more neutral terms as the *horizontal* (i.e., communion/warmth) and *vertical* (i.e., agency/competence) dimensions (e.g., Abele et al., in press; Carrier, Louvet, & Rohmer, 2014; Hall et al., 2005; Locke, 2003). A large body of evidence supports the universality of this bi-dimensional configuration, which has been replicated across different nations and cultures, in experimental but also in more ecological settings (Abele & Wojciszke, 2014; Cuddy et al., 2009; Durante et al., 2017; Durante et al., 2013; Fiske et al., 2007).

2.3 Facets and the third dimension

The two dimensions have more recently been further broken down in two facets or sub-dimension each (Abele et al., 2016; see also Carrier, Louvet, & Rohmer, 2014). The vertical dimension—“getting ahead”—comprises both motivation, *agency/assertiveness* (e.g., ambitious, assertive, self-confident), and resources, *ability/capability* (also known as *competence*; e.g., capable, efficient, competent). The horizontal dimension—“getting along”—calls for social abilities, *friendliness/sociability* (e.g., caring, friendly, warm)—and traits that foster principled relations with others, *morality* (e.g., honest, fair, reliable).

Interestingly, theoretical and empirical research over the last decade seem to put forward two conceptions of the vertical dimension. Indeed, competence denotes a set of available, fixed, resources (e.g., intelligent, skilled, organized) that can be exploited to achieve a goal, and therefore belongs to the domain of the *means*. In contrast, agency has a motivational component geared toward self-expansion (e.g., ambitious, determined, instrumental) and thus pertains to the domain of the *ends* (see Carrier, Louvet, & Rohmer., 2014; see also Imhoff & Koch, 2017; Kervyn et al., 2013; Oliveira, Garcia-Marques, Garcia-Marques et al., 2019).

In the same vein, *morality* can sometimes be distinguished from *sociability*. According to the Behavior Regulatory Model (BRM; Ellemers, 2017; Ellemers et al., 2008; see also Leach et al., 2007), morality, in comparison to sociability and competence, is primary for ingroup identification and pride as well as the self-image projected to the ingroup (i.e., moral group members are “good” group members). This applies whether the ingroup is successful or not and even whether competence is context-relevant or -irrelevant. Indeed, morality (i.e.,

trustworthiness, loyalty) is essential for group coordination and survival as it fosters pro-group behaviours over selfish ones (i.e., to avoid free riders).

Although the present review focused on the models building or at least inspired by the Big Two dimensions (for a review and comparison of the different models see Abele et al., in press), a recent model has also been proposed that involves the addition of a third dimension. According to the ABC model (Koch et al., 2016), people also distinguish groups (and more generally societies) in terms of the conservative–progressive beliefs dimension (e.g., traditional-modern, religious-science-oriented, conventional-alternative, conservative-liberal). Findings suggest that the similarity between the self in terms of Agency (A) and Beliefs (B) predicts perceived group’s Communion/warmth (C). The larger (smaller) the similarity between the self and the target group in terms of A and (even more so) B, the more (less) people will ascribe C to the target group (Koch et al., 2020).

3. Horizontal and vertical dimensions of face perception

When we think about a group (or an individual), for instance, the elderly people, not only do we think of them as caring and sociable, but we also spontaneously picture their wrinkled-smiling-looking faces in our minds. In other words, we think of them in terms of *abstract* personality traits together with a vivid image of how they *physically* look-like. This example illustrates the almost century-old definition of stereotypes by Lippman (1922) as “pictures in our heads”, emphasizing the visual component of stereotypes that has long been overlooked in the literature. In this sense, stereotypes can incorporate multimodal vectors of information, such as verbal or abstract information (e.g., traits; Fiske et al., 2002), visual information (e.g., face or body physical appearance; Dotsch et al., 2008; Lick et al., 2013) or auditory information (e.g., sexually oriented voices and vocal attractiveness; Fasoli et al., 2017; Zuckerman & Driver, 1989).

3.1 Faces as social stimuli

Faces are amongst the most frequent and rich source of visual information from our social environment as they bear vital information for everyday life social interactions (Jack & Schyns, 2017; Todorov et al., 2015). As a case in point, an aggressive or dominant looking face may indicate a potential threat and thus prompt the perceiver to avoid the target, whereas a trustworthy looking face may signal approach (Oosterhof & Todorov, 2008; Todorov, 2008). It is therefore of little surprise that humans have evolved a remarkable cognitive ability (Bar et al., 2006; Todorov et al., 2013; Willis & Todorov, 2006) and a specialized neural system (Grill-Spector et al., 2017; Haxby et al., 2000; Yovel & Duchaine, 2006) to process visual information from faces.

People are inherently drawn to infer dispositional traits from facial features. For instance, they can infer traits as quickly as 34 to 100ms (Todorov et al., 2009; Willis & Todorov, 2006) and can reliably guess (visually evident) demographic variables such as gender, age, or ethnicity from minimal facial cues (Bruce & Young, 2012). However, not all deductions are valid. Since the 19th century, physiognomy disseminated the idea that character could be read

from faces (Hassin & Trope, 2000). For example, thieves were believed to have expressive faces and small wandering eyes (Todorov, 2008). Although this pseudoscience has been long discredited, social attributions from faces remain pervasive despite little evidence for their exactitude (for a review, see Todorov et al., 2015). Scientist are still debating on the accuracy of the link between personality and facial appearance (i.e., the kernel of truth hypothesis; see Berry, 1990; Todorov et al., 2015). Paradoxically, there is abundant correlational and causal evidence that face-based social attributions can have direct consequential on a host of social outcomes, such as political preferences and voting intentions (Olivola et al., 2018; Olivola & Todorov, 2010; Todorov et al., 2005), sentencing decisions (Porter et al., 2010), financial and careers opportunities (Rezlescu et al., 2012; Rule & Ambady, 2009), or mating preferences (Karremans et al., 2011; Little et al., 2006).

3.2 Face perception

Paralleling the research in the domain of social perception, researchers started investigating if fundamental dimensions could underlie the content of facial impressions. In an illustrative work, Oosterhof and Todorov (2008) asked their participants to rate neutral faces on a series of traits that had previously been selected from unconstrained descriptions of these faces. Although people spontaneously express their evaluation about faces on several dimensions (e.g., masculinity-femininity, self-resemblance, attractiveness, maturity), these are often highly correlated. Nonetheless, a principal component analysis (PCA) revealed two underlying factors. The first axis, labelled *trustworthiness* (also interpreted as *valence* or *approachability*), accounted for 63.3% of the variance and comprised traits such as trustworthy, emotionally stable, sociable, or caring. The second axis, labeled *dominance*, accounted for 18.3% of the variance and was characterized by traits such as dominant, aggressive, confident, or mean. The trustworthiness dimension strongly resembles emotional expression of anger (negative pole) and happiness (positive pole) which are associated with behavioral intentions of the target (i.e., help vs. harm). In contrast, the dominance dimension captures the capability of the target to implement these intentions and is associated to cues of physical strength, facial maturity, and masculinity (Flowe, 2012; Jones et al., 2009; Oosterhof & Todorov, 2008; Sutherland et al., 2013; Toscano et al., 2016).

From a functional perspective, impressions of neutral faces would derive from the overgeneralization of facial features that provide adaptative information. In other words, “resemblance of neutral faces to emotional expressions is perceived as indicating the trait attributes or behavioral tendencies associated with these emotions” (Oosterhof & Todorov, 2009, p. 128). This is known as the overgeneralization hypothesis (Knutson, 1996; Montepare & Dobish, 2003; Zebrowitz et al., 2003; Zebrowitz & Montepare, 2008), and can be thought of as a face-specific fundamental attribution error (Albohn & Adams, 2020). Specifically, trustworthy judgments would stem from overgeneralizations of emotional expressions such as anger (negative pole) and happiness (positive pole) that prompt approach and avoidance responses (Oosterhof & Todorov, 2008; Said et al., 2009; Todorov, 2008; Todorov, Baron et al., 2008). Conversely, face-driven impressions of dominance would arise from the overgeneralization of facial features associated with physical strength and maturity

(Oosterhof & Todorov, 2008; Toscano et al., 2016). Although these two (character-based) dimensions consistently emerge, subsequent research that relied on broader-age samples incorporated a third (appearance-based) *youthful-attractiveness* dimension (e.g., youthful, healthy, baby-faced, beautiful) that may serve sexual selection functions (Hehman et al., 2017; Lin et al., 2019; Sutherland et al., 2018; Sutherland et al., 2013; Sutherland et al., 2016). More recently, a fourth dimension, namely *female-stereotype* (e.g., feminine, prudish, emotional, self-critical), was revealed after submitting a very large database of face ratings to a deep neural network algorithm (Lin et al., 2019). Nevertheless, most of the work on face perception revolves around the first two dimensions of trustworthiness and dominance.

3.3 Face representation

People can not only form impressions from faces but can also picture with ease an individual or a typical group member facial appearance in their mind (Hassin & Trope, 2000). To study and capture the visual representation of a given target's face (e.g., a Moroccan), many studies relied on an innovative technique, namely the reverse correlation method (RC; Dotsch et al., 2008; Mangini & Biederman, 2004). In a typical RC task, participants are asked, over several trials, to choose among two randomly generated noisy faces the one that best matches the intended target (e.g., a "Moroccan-looking face"). The average of all the selected noisy faces yields what is known as the individual (i.e., for a given participant) or group-wise/average (e.g., for participants within a given condition) classification image (CI)—a proxy of the internal representation of the target (for the interpretation of the CI see Brinkman et al., 2017; Dotsch & Todorov, 2012). One of the main strength of the RC is that it is a data-driven technique which makes no prior assumption about what those mental templates might look like (see Dotsch & Todorov, 2012; Brinkman et al., 2017; for other techniques see Jack & Schyns, 2017; Todorov et al., 2011; Todorov et al., 2015). One of the main caveats, however, is that it requires a large amount of trials to produce reliable individual CIs (Brinkman et al., 2017).

In an illustrative example, Dotsch et al. (2008) captured the prototypical visual representation of a Moroccan-looking face and showed that high-prejudiced (low-prejudiced) participants had a more negative (positive) CIs (as determined by external judges). Further studies demonstrated that people are able to build a facial representation from minimal information, such as behavioral information (Dotsch et al., 2013), minimal group membership (Ratner, Dotsch et al., 2014), or social categories (e.g., racial groups; Dotsch et al., 2008; Krosch & Amodio, 2014; national groups; Imhoff, & Dotsch, 2013; Imhoff et al., 2011; or occupational stereotypes; Brown-Iannuzzi et al., 2017; Hehman et al., 2015). Crucially, investigations have also shown that people can successfully and spontaneously encode the dimensions of trustworthiness-by-dominance and warmth-by-competence models into a face. These visual templates (i.e., CIs) were highly diagnostic of the trait they originated from (e.g., the trustworthy CI was evaluated by external judges as highly trustworthy), although there was less agreement when mapping the vertical dimensions onto the face content (Dotsch & Todorov, 2012; Imhoff et al., 2013; Oliveira, Garcia-Marques, Dotsch et al., 2019; Oliveira, Garcia-Marques, Garcia-Marques et al., 2019). Furthermore,

and echoing Asch's (1946) seminal work, Oliveira et al. (Oliveira, Garcia-Marques, & Dotsch, 2019) provided convincing evidence that the integration of traits into a face follows a *gestaltic* pathway. Specifically, a CI that integrates simultaneously multiple traits (e.g., a trustworthy-and-dominant-looking face) carries relevant information (e.g., facial emotional expressions) that goes beyond the mere linear combination of the single-trait CIs that compose it (e.g., a trustworthy-looking face and a dominant-looking face). Also, in this process, trustworthiness is taken into account more heavily than dominance (see also Oosterhof & Todorov, 2008), mirroring the primacy of warmth/communion found in social perception (see above).

3.4 Relation between the social and facial dimensions

Clearly, the conceptualization of the trustworthiness and dominance dimensions proposed in the face perception model bears a striking resemblance with the fundamental dimensions of warmth and competence in the domain of social perception, although the former is based on judgments situated at the visual level (e.g., faces) whereas the latter is based on conceptual evaluation (e.g., groups labels, descriptions, traits). Sutherland et al. (2016) investigated the correspondence between the two models when evaluating faces and found that the two horizontal dimensions (i.e., trustworthiness and warmth) were closely related (for similar results between trustworthiness and communion, see Walker & Vetter, 2016) whereas this was less the case between the vertical dimensions (i.e., dominance and competence). Their findings held whether considering facial judgments at the face-averaged (Exp. 1–2) or individual level (Exp. 3), as well as with more direct measures (i.e., similarity ratings between faces diagnostic of each dimension) and experimental manipulations (Exp. 4). Oliveira et al. (Oliveira, Garcia-Marques, Dotsch et al., 2019) conceptually replicated and extended these findings by examining the overlap between the two models in the *face space*—the space defined by physical facial features (e.g., large jaw, hairy eyebrows, long nose) as captured in classification images (CIs) generated from reverse correlations tasks. In other words, they examined if the way people imagine in their mind a friendly-looking face (i.e., warmth) would be similar to how people imagine a trustworthy-looking face, and similarly for a dominant- and competent-looking face. Their results revealed that the horizontal CIs (i.e., warmth and trustworthiness) were physically (i.e., as assessed by pixel-wise correlations between the two CIs; see Brinkman et al., 2017) and subjectively (i.e., as determined by independent judges) more similar than the vertical CIs (i.e., dominance and competence/intelligence).

The substantial overlap between warmth and trustworthiness is rather straightforward as they both unambiguously relate to the target's intentions and strongly correlate with valence (Fiske et al., 2002; Kervyn et al., 2013; Oosterhof & Todorov, 2008; Rosenberg et al., 1968). This suggests that they refer to the same underlying social evaluative dimension. Conversely, there exist several possible explanations to account for the rather intriguing divergence between competence and dominance. For instance, the evolutionary-based theory of Dominance-Prestige Account of Social Rank (Chapais, 2015; Cheng & Tracy, 2014; Cheng et al., 2013; Henrich & Gil-White, 2001) holds that prestige or competence (i.e., sharing

expertise to achieve genuine persuasion and gain respect) and dominance (i.e., the use of physical strength or psychological threats to induce fear, intimidation and coercion) represent two dissimilar strategies—the former being more beneficial (vs. harmful) to others than the latter—to achieve status in hierarchical systems. This is supported by the fact that dominant faces are threatening and negatively valenced whereas competent faces are positively valenced (Chen et al., 2014; Oliveira, Garcia-Marques, Dotsch et al., 2019; Oliveira, Garcia-Marques, Garcia-Marques et al., 2019). Interestingly, the use of each approach may further depend on the specific context. For example, facial competence is more relevant than facial dominance in a political context (Chiao et al., 2008), but the reverse is true in a military context (Mueller & Mazur, 1996).

In the same vein, previous studies also emphasized the distinction between the sub-dimensions of competence: Ability has been characterized as a set of stable resources to achieve a goal whereas assertiveness has been linked to a motivation for self-enhancement—that may be detrimental to others (see previous section; Carrier, Louvet, Chauvin et al., 2014; see also Abele et al., 2016; Imhoff & Koch, 2017). An alternative rationale is that cues signaling competence (i.e., ability) may vary between the conceptual (i.e., abstract traits) and facial level. Specifically, some components that are diagnostic of competence may be more readily encoded in faces (e.g., physical strength and dominance) than others (e.g., intelligence and ability) (Sutherland et al., 2016). Last but not least, the vertical dimension discrepancy may also arise from the high (low) inter-individual variability in the expectations regarding the facial cues signaling competence (dominance) (Dotsch et al., 2019; Hehman et al., 2017; Oliveira, Garcia-Marques, Dotsch et al., 2019; Sutherland et al., 2018).

4. Relation between the horizontal and vertical dimensions

An orthogonal, positive, negative, or curvilinear relationship between the horizontal and vertical dimensions have been proposed throughout the literature.

4.1 Positive relation

Early work on social perception emphasized the importance of evaluative and cognitive consistency. This refers to the tendency to infer traits about someone in such a way that they match the valence of the previously established overall impression about this person (Asch, 1946; Felipe, 1970; Festinger & Carlsmith, 1959; Rosenberg & Olshan, 1970). This means that people will be inclined to attribute favorable (unfavorable) qualities to a person to whom they have already attributed positive (negative) attributes. This phenomenon closely matches the so-called *halo effect* (Kelley, 1950; Thorndike, 1920) that takes place when the evaluation of some trait influences the perception of another, apparently unrelated, trait in the same evaluative direction (i.e., positive vs. negative). For instance, a soldier with desirable physical qualities will also tend to be evaluated as intelligent by his commanding officer, even though physical prowess is unrelated to wisdom (Thorndike, 1920).

In line with this reasoning, some studies have found a positive relationship between the two fundamental dimensions of social perception (e.g., Abele & Wojciszke, 2007; Durante

et al., 2013; Judd et al., 2005; Richetin et al., 2012; Rosenberg et al., 1968; Wojciszke & Abele, 2008). As a case in point, Judd et al. (2005) found in their pilot study that behaviors that were positively (negatively) diagnostic of a one dimension (e.g., competence) were judge as positively (negatively) diagnostic of the other dimension (e.g., warmth). In general, a positive relation emerges when judging a single target (Judd et al., 2005) or ingroup (Fiske et al., 2002).

4.2 Orthogonal relation

Another conception, championed by the Stereotype Content Model (SCM), is that that competence and warmth are in fact orthogonal (Fiske et al., 2007; Fiske et al., 2002). In principle, information about a person's intention—which translates into perceived warmth—should not confer any information about this person's capacity to carry out those intentions—which informs about the competence of the person—and vice versa. In a similar vein, the Big Two of communion and agency were theorized as perpendicular axes (Osgood et al., 1957; Wiggins, 1979) and have been shown to be independent in the perception of traits, behaviors, self, and groups (Cislak & Wojciszke, 2008; Abele, 2003; Cuddy et al., 2009; Kervyn et al., 2013; see Abele et al., in press; Imhoff & Koch, 2017).

Despite the orthogonality between the two axes assumed by the SCM, a significant portion of social groups (e.g., elderly people, rich people, disable people, poor people) fall within the so-called ambivalent quadrants. These quadrants characterize groups that are positively valenced on one dimension but negatively on the other (i.e., high on warmth and low on competence, or low on warmth and high on competence; Durante et al., 2017; Durante et al., 2013; Fiske et al., 2002). Such configuration reflects a negative relationship between the vertical and horizontal dimensions. An illustrative example comes from previous work by Glick and Fiske (1996, 2001a, 2001b) on ambivalent sexism. On the one hand, hostile sexism portrays women, such as businesswomen, competing in the so-called battle of the sexes for power and control over men through marriage, sexual wiles, or even feminist ideology. They are seen as competent but cold. On the other hand, benevolent sexism fosters traditional gender roles, such as homemakers, that depicts women as caring, naïve, "pure", and nurtured, but especially fit for low status and power domestic labors such as cleaning or laundering. These women come across as incompetent but warmth (Kite & Whitley, 2016). A similar pattern can be found for racial mixed stereotypes about African American that are depicted as more good-hearted, family oriented, but less fit for high-status roles in comparison to White Americans (Whitley, 1999).

4.3 Negative relation

The evidence for a negative relation is not only found at a correlational level but also seems to emerge in experimental designs. As a case in point, Cuddy et al. (2004) found that consultant women (but not men) who were presented in a cover story as mothers to participants playing the role of clients looking to hire a consultant were evaluated as less competent, warmer and more prone to be discriminated than when they were introduced as childless. Along these lines, stereotype-inconsistent targets of elderly people competent (e.g.,

with a sharp memory) were perceived as, unsurprisingly, more competent, but, importantly, less warm than stereotype-consistent targets of elderly people with memory issues (Cuddy et al., 2005).

Findings such as these can be interpreted as evidence for what is known as the *compensation effect* (Yzerbyt et al., 2005). Compensation characterizes the trade-off between the two fundamental dimensions of warmth and competence that takes place in the context of social intergroup comparisons. Specifically, the compensation effect gives rise to a pattern in which one of two groups is perceived as higher on one dimension relative to the other group but as lower on the other dimension. In their seminal work, Yzerbyt et al. (2005) showed that participants from both French and Belgian speaking communities acknowledged the superiority of the formers over the later in terms of linguistic skills (i.e., competence). Crucially, both parties compensated for this gap by granting Belgians speakers a superiority over French people on the other dimension, namely warmth. Since these initial findings, a large number of correlational and experimental studies have provided evidence for the robustness of this effect (see Chapter 2 of the present thesis; for a review see Yzerbyt, 2018).

4.4 Curvilinear relation

A curvilinear relationship between the two fundamental dimensions has also been proposed. The main argument follows the Aristotelian philosophy about the “golden mean” which states that midway qualities are more desirable than the two extremes of deficiency and excess (see Grant & Schwartz, 2011). In social psychology, the idea echoes Peabody’s (1967; see also Peeters, 2008) tetrads model of social judgments which shows that the two moderate values (e.g., thrifty and generous) of a dimension described by a four traits-continuum (tetrads) tend to be more positive than the two flanks values (e.g., stingy and extravagant). Applied to the vertical dimension of agency, excessively low levels may transmit lack of vigor and high levels of it (e.g., overconfident, ambitious) may convey aggressiveness (Ames et al., 2017; Carré et al., 2009) or egocentrism (Bakan, 1966; Carrier, Louvet, & Rohmer, 2014) whereas moderate levels may be more socially valued (Dufner et al., 2013; Oliveira, Gracías-Marques, García-Marques et al., 2019).

Imhoff and Koch (2017) investigated and found evidence for an inverted U-shaped relationship between the Big Two of agency and communion, such that excessive or insufficient levels of agency are negatively related with communion, whereas the relation tend to be positive or orthogonal for moderate levels of agency. To account for some inconsistencies in the literature, the authors call upon the idea of restrictive sampling on the agency-continuum. For instance, a negative relation is likely to be found if the evaluating targets belong to the mid-to-upper tail of the agency-continuum.

Oliveira et al. (Oliveira, García-Marques, García-Marques et al., 2019) extended Imhoff and Koch’s (2017) findings by considering both facets of the vertical dimension (i.e., ability and assertiveness). Their results show that although (most) assertiveness-related-traits (i.e., confidence, power) had a significant inverted U-shaped relation with valence, it was not a crucial nor consistent feature. Dominance was the only assertiveness-related-trait that

unambiguously manifested an inverted U-shape. Indeed, both submissiveness and high dominance convey negative impressions of vulnerability/exploitability/fearfulness (Oosterhof & Todorov, 2008; Richards et al., 1991) and threat/aggressiveness/anger (Oosterhof & Todorov, 2008; Toscano et al., 2016) respectively. In sharp contrast, competence-related traits (i.e., competence, intelligence) clearly exhibited a linear relation with valence (i.e., a halo effect). Furthermore, the linear vs. curvilinear relationship of traits with valence was modulated, among other things, by the domain of judgment, such that a linear (vs. curvilinear) relationship became more prominent in the domain of face (vs. conceptual trait) perception. In addition, the authors emphasize that the valence/warmth-agency relation can be highly variable, context-dependent (e.g., Carrier et al., 2019; Chen et al., 2014), and may also change from individual to individual (see Hehman et al., 2019).

4.5 Relations between the facial dimensions

In the domain of face perception, the two dimensions of dominance and trustworthiness/valence were statistically modeled as orthogonal on the basis of data-driven statistical analyses of facial judgments (Oosterhof & Todorov, 2008; Todorov, Said et al., 2008; see Todorov et al., 2015). However, a negative relation could also be expected since facial cues moving toward the high pole of dominance (i.e., dominant faces) and the low pole of trustworthiness (i.e., untrustworthy faces) are both associated with negative evaluations (Todorov et al., 2013; Toscano et al., 2016). Interestingly, although empirical evidence points to an overall negative relation between the two dimensions, this link is highly variable and sometimes relatively low or non-significant (Dotsch & Todorov, 2012; Flowe, 2012; Kleisner et al., 2013; Oliveira, Garcia-Marques, Dotsch et al., 2019; Oliveira, Garcia-Marques, Garcia-Marques et al., 2019; Oosterhof & Todorov, 2008; Ramos et al., 2016; Todorov et al., 2009).

One explanation may come from gender biases in face perception. Specifically, research has shown that dominance and trustworthiness are negatively correlated for female faces, but not for males faces (Oh et al., 2020; Sutherland, Young et al., 2015); and that this effect is further accentuated for people that endorse stronger gender stereotypes (Oh et al., 2020). These findings are consistent with ambivalent sexism theory (Glick & Fiske, 1996, 2011) and the backlash effect (Rudman & Fairchild, 2004; Rudman & Glick, 2001) which posits that, unlike man, counter-stereotypical women (e.g., dominant- or masculine-looking) that deviates from the traditional stereotype (e.g., feminine- and trustworthy-looking) are evaluated more negatively (and endure more social and economic sanctions) than those who endorse the prescriptive stereotype.

The relation between the facial dimensions of warmth/trustworthiness/valence and competence tells a different story. Early evidence suggests that a positive relation could be expected between the two. Indeed, a halo effect has been found between valence-laden traits, such as likeability (Todorov, Said et al., 2008) or attractiveness (Thorndike, 1920; see also Moore et al., 2011), and competence. More recent investigations confirm the positive relation between the two dimensions whether it is at the level of face judgment or face production (Imhoff et al., 2013; Oliveira, Garcia-Marques, Dotsch et al., 2019; Oliveira, Garcia-

Marques, Garcia-Marques et al., 2019). These results thus highlight the divergence between the facets of the vertical dimension, especially in the domain of face perception, and thus, prompts further research to disentangle these two facets (see discussion in Chapter 1, section 3).

5. Conclusions

Chapter 1 reveals that despite the wild variety of traits involved in impression formation, two fundamental dimensions underpin social judgments across both social and facial perception. The horizontal dimensions (i.e., warmth, communion, valence, trustworthiness) convey information about the intention of the target, and the vertical dimensions (i.e., competence, agency, dominance) relates to the capacity of the target to carry out these intentions. Whereas the sub-dimensions of the horizontal dimension tend to strongly overlap, recent evidence suggests a valence-driven divergence between the sub-dimensions of competence and dominance/agency that form the vertical dimension, in the domain of traits but especially so in the domain of face perception. Finally, there is evidence for an orthogonal, positive, negative, and even curvilinear relationship between the horizontal and vertical dimensions. In Chapter 2, we focus more specifically on one of these relationships, namely the negative one, as this is the one that is at the heart of the so-called compensation effect.

Chapter 2

The compensation effect

The compensation effect

1. Introduction

In comparison to poor people, the wealthiest are often depicted as more knowledgeable and educated but they are also seen as less empathic and supportive. Likewise, traditional role genres foster a vision of men as cold-and-logical thinkers, and women as naïve but well-intentioned. In the same vein, disabled people are thought to be weak and helpless, but they are seen as more friendly and warm than non-disabled people. Still in the same vein, Latin Americans are portrayed as friendly and easy-going in comparison to cold but efficient Europeans. These are only a handful of the commonly held views about groups that comprise contemporary societies. As can be seen, they all share a common core: They imply comparisons between two social targets that are counterbalanced with respect to their relative advantage in terms of warmth and competence. The process that gives rise to this pattern has been labeled the *compensation effect* (Yzerbyt et al., 2005; also known as *dimensional compensation*, Yzerbyt, 2018). Specifically, compensation characterizes the negative relationship between the two fundamental dimensions of social perception—warmth and competence—when comparing two social targets. In the domain of intergroup relationship, this effect materializes when one group that enjoys a superior standing on one of the two dimensions (e.g., competence) in comparison to another group is perceived as inferior on the other dimension (e.g., warmth).

Compensation can be assessed at both the mean, i.e., group or nomothetic, level and at the correlational, i.e., respondent or idiothetic, level. Compensation takes place at the mean level when participants view—on average—one group as superior to another on one dimension, but as inferior on the other dimension. At the correlational level, the differences between targets on one dimension should be negatively correlated with the differences between them on the other dimension (i.e., within-dimension correlations). Alternatively, the ratings of one group on the first dimension should correlate positively with ratings of the other group on the second dimension (i.e., correlation across dimensions and across groups). Although both analytical strategies have been used (e.g., Judd et al., 2019; Judd et al., 2005; Kervyn et al., 2008) the mean level analysis is more common as it provides more consistent results. For this reason, we adopted the mean analysis in the present dissertation.

In Chapter 2, we first examine the moderating factors and boundary conditions that shape the compensation effect. In the second part, we explore the underlying motivations that could account for this compensation pattern.

2. Moderating factors and boundary conditions

Over the years, an impressive amount of empirical evidence accumulated in support of the so-called Dimensional Compensation Model (DCM; Yzerbyt, 2018). In this section we

review several moderating factors and boundary condition related to the emergence of the compensation effect. Specifically, we review factors related to the characteristics of the evaluated social targets, the context, and the nature of the dimensions.

2.1 Real and fictitious groups

The first reported study that examined the compensation effect did so by relying on existing groups. As mentioned above, Yzerbyt et al. (2005) found a compensatory pattern in which Belgians came across as less competent but more social than French people independently of whether participants were members of the low- (Belgians) or high-standing group (French people), or even members of a third party national group (Swiss people). These findings were further extended with other national groups (e.g., Italian, Germans, Brazilians, Canadians; Kervyn et al., 2008) and less strongly stereotyped, naturally occurring, occupational groups (e.g., truck drivers, production senior manager, foremen; Cambon & Yzerbyt, 2016) or even majors at university (e.g., psychologists, sociologists, nurses, medical doctors; Cambon et al., 2015; Yzerbyt & Cambon, 2017). Taken together, these studies support the external validity of compensation by showing that it is not limited to a handful of groups but can be generalized to a wider range of social categories.

One shortcoming of the work based on real groups, however, is that participants come with pre-existing stereotyped characteristics that are difficult to control and manipulate. Also, it may thus be the case that in Yzerbyt et al.'s (2005) study, participants had *a priori* impressions about Belgian and French people that happened to match a compensatory pattern, which might have not been caused by a compensatory process *per se* (Cambon & Yzerbyt, 2016). To address this limitation, Judd et al. (2005) relied on an experimental design that had participants learn about two unknown and fictitious groups (i.e., the “Greens” and “Blues” groups) that were introduced by the means of behaviors allegedly performed by their group members. Each behavior was pilot-tested such that they were diagnostic of either competence (e.g., high competence: “X worked hard on the extra-credit assignment in linear algebra”; low competence: “X did poorly on the exam because of mixing up the chapters that needed to be studied”) or warmth (e.g., high warmth: “X loves to be with other people”, low warmth: “X yelled at the driver who took the empty parking space.”). Despite their best effort to keep the diagnosticity of the behaviors limited to a single dimension while remaining uninformative about the other, their pretests revealed a positive correlation between the two dimensions (i.e., a halo effect; Rosenberg et al., 1968).

Information about the novel groups was manipulated in such a way that one group (i.e., the “high group”) was relatively higher on one dimension (i.e., the “manipulated dimension”) than the other group (i.e., the “low group”), while they were both equal and ambiguous on the other dimension (i.e., the “unmanipulated dimension”). Specifically, in Experiment 1, the manipulated dimension was competence. Therefore, the high group was presented as relatively more competent, but equally warmth than the low group. Conversely, in Experiment 2, the manipulated dimension was warmth so that the difference between the high and low group resided on the warmth dimension while they were equal in terms of competence.

After reading the behaviors, participants were asked to convey their impressions of the two groups on a series of traits encompassing the two fundamental dimensions. In line with the manipulation, the high group was depicted as higher than the low group on the manipulated dimension (e.g., the high group was perceived as more competent than the low group in Experiment 1). Importantly, and in line with the compensation effect, the high group was rated as lower on the unmanipulated dimension (e.g., the high group was rated as colder than the low group in Experiment 1) even though the groups were initially presented as being equal on this dimension and despite the halo effect reported in the pilot study. Several follow-up experiments relied on artificial social targets to investigate the compensation effect with similar results (e.g., Kervyn et al., 2016; Kervyn, Judd et al., 2009; Kervyn et al., 2011; Kervyn, Yzerbyt et al., 2009; Yzerbyt et al., 2008).

2.2 Groups and individuals

Although most investigations on compensation took place in the context of intergroup relations, some examined whether the effect may also take place when considering interpersonal relations. As a case in point, Judd et al. (2005, Study 3) experimentally manipulated the target entity such that participants had to form an impression of either two fictitious groups or two fictitious individuals that initially differed in terms of competence (see methodology described above). In line with their expectations, compensatory ratings emerged at the mean level and were nearly identical irrespective of whether the social targets were groups or individuals. However, this was not the case at the correlational level where the relation was negative for the former and near zero for the latter. Having said this, further studies also found compensatory judgments in interpersonal contexts (Kervyn et al., 2016; Kervyn, Yzerbyt et al., 2009; Terache et al., 2020; Yada & Ikegami, 2014).

2.3 Uniqueness of warmth and competence

An important question regarding the compensation effect is whether it is limited to the two dimensions of warmth and competence or if it concerns any pair of dimensions. Yzerbyt et al. (2008; see also Cambon et al., 2015, Study 1c) examined this question by introducing a third, unrelated, dimension, namely healthiness. Based on Judd et al.'s (2005) methodology, their participants had to learn about two fictitious groups that initially differed on one of the two fundamental dimensions (e.g., competence) but were equal on the other fundamental dimension (e.g., warmth) and/or a third dimension (e.g., healthiness). Ratings of the two groups revealed, as in previous studies, a compensation effect between warmth and competence (only when competence was manipulated; Exp. 3). Importantly, a classic halo effect emerged between the manipulated dimension and the third, healthiness dimension. In other words, the high group was perceived as superior on the manipulated dimension (e.g., more competent) but also healthier than the low group. Holoien and Fiske (2013) found similar results in the domain of impression management and further showed that the "political interest" dimension failed to behave in a compensatory manner. Taken together, these studies highlight the special status and ubiquity of the fundamental dimensions of warmth and competence when it comes to social judgments. Nevertheless, and given that these efforts

were limited to a few dimensions (i.e., healthiness, political interest), future studies should seek to generalize these results by examining if other, relatively uncorrelated dimensions (e.g., attractiveness-youthfulness, Sutherland et al., 2013; female-stereotype, Lin et al., 2019), exhibit a specific relation with the two fundamental dimensions.

2.4 Comparison context

Another defining feature of compensatory judgments is that they seem to be promoted by social comparisons. Indeed, when Judd et al. (2005) presented participants with a single group (instead of two as in previous experiments) that was either highly competent or incompetent, the compensation effect did not emerge. Unsurprisingly, the high-competence group came across as more competent than the low-competence group but, more importantly, there was no difference in terms of warmth. In fact, at a descriptive level, the high-competence group was perceived as slightly warmer than its counterpart, revealing a propensity for a halo effect.

The tendency of social comparison contexts to elicit compensation has not only been shown when contrasting two groups when people are observers, but also when they are part of one of the entities being evaluated, either as group members in intergroup comparisons (e.g., Judd et al., 2005; Kervyn et al., 2008; Kervyn, Yzerbyt et al., 2009; Yzerbyt et al., 2005;) or as the self-versus-other in interpersonal comparisons (e.g., Terache et al., 2020; Yada & Ikegami, 2014). Furthermore, studies have shown that manipulating the vantage point can affect compensation (Kervyn et al., 2008; see also Cambon et al., 2015). For instance, Canada's warmth did not significantly differ from its attributed competence when evaluated in isolation, whereas the country came as more competent than warmth when compared to Brazil, and conversely, as less competent than warmth when compared to Japan. Similar results also emerged when participants were members of the evaluated group (i.e., Belgians).

A recent study by Judd et al. (2019) suggests that not only a comparative context may be required for compensation to emerge, but that targets should also be judged simultaneously on both dimensions. Specifically, a halo effect materialized when participants were asked to evaluate several targets on a single dimension, and then evaluating them all a second time on the other dimension (Study 1), whereas they displayed compensation when asked to rate all the target on the two dimensions concurrently (Study 2). This pattern is in line with Yzerbyt and Cambon's (2017) finding that compensation emerged more systematically when participants were well aware that they were to evaluate the groups on both dimensions. In contrast, when participants believed that only one dimension would be used to evaluate the groups, the compensation pattern tended to disappear and ingroup bias emerged instead.

The pervasiveness of social comparisons in everyday life (Hornsey & Hogg, 2000) together with the possibility to make judgments on any dimension may thus explain, at least in part, the widespread of compensatory stereotypes and judgments.

2.5 Low conflict and legitimate status

Socio-structural factors such as status asymmetry between groups and the nature of intergroup relations also seem to condition the emergence of the compensation effect. Past efforts from the Social Identity Theory tradition (SIT; Tajfel & Turner, 1979; Turner & Brown, 1978) have shown that when the status gap between groups is perceived as illegitimate (legitimate) and unstable (stable) it leads to conflicting (harmonious) intergroup relations and intergroup competition (cooperation). Building on these efforts, Cambon et al. (2015; see also Cambon & Yzerbyt, 2016; Dang et al., 2017) experimentally manipulated the status gap between an ingroup and outgroup (asymmetrical vs. symmetrical on either warmth or competence), which acted as a proxy of perceived status legitimacy, and they also varied the level of intergroup conflict (present vs. absent). As predicted by the authors, compensation emerged in the judgments solely when intergroup relation where non-conflictual and asymmetrical (whether the ingroup was very inferior or very superior to the outgroup), whereas an ingroup bias (i.e., ingroup favoritism) took over in the other conditions. These results held whether participants were members of minimal groups (Study 1) or real groups (Study 2).

The above findings nicely echo large-scale cross-national and cross-cultural studies. On the one hand, they show that ambivalent stereotypes are most prevalent in societies with unequal incomes (Durante et al., 2013). On the other, they also reveal that the relation between the degree of conflict and compensatory judgments is more nuanced. As a matter of fact, mixed stereotypes appear to flourish in nations with intermediate level of conflict (e.g., United States, South Africa, Peru) as they provide stability to the system, whereas univalent impressions dominate in highly peaceful countries with only have a few outcasts (e.g., Denmark, Norway, Finland) or highly conflictual regions that require unequivocal distinctions between friends and foes (e.g., Pakistan, Israel, Afghanistan) (Durante et al., 2017).

2.6 Observers and group members

Interestingly, the trade-off between warmth and competence has been reported not only when judges are external observers, but also when they are themselves part of the evaluated social targets. For instance, in Yzerbyt et al.'s (2005), the Swiss (i.e., external observers) perceived the Belgians as warmer but less competent than the French (Study 2). The same pattern emerged when participants were either member of the low-competence group (Belgians) or high-competence group (French). In other words, both parties seem keen to acknowledge their inferiority (competence for the Belgians, warmth for the French) on one dimension and their advantage on the other (warmth for the Belgians, competence for the French). Judd et al. (2005) replicated these results in an experimental setting using minimal groups and in which participants had randomly been assigned to either the high- or low-competence group by the means of a bogus test (i.e., a dot estimation task). Once again, compensation emerged independently of whether participants were external observers (Experiment 1 and 2) or group members (Experiment 5). This general pattern has been conceptually replicated in more recent investigations for external observers (e.g., Kervyn et

al., 2008, Exp 2; Kervyn et al., 2011; Kervyn, Yzerbyt et al., 2009), group members (e.g., Kervyn et al., 2008, Exp 1; Cambon & Yzerbyt, 2018; Cambon et al., 2015; Yzerbyt & Cambon, 2017), or individuals directly involved in the interaction (e.g., Collange et al., 2009; Kervyn, Judd, et al., 2009; but see Terache et al., 2020).

Interestingly, when participants belonged to one of the evaluated social group, they tended to exhibit what has been labeled as *overcompensation* (Yzerbyt et al., 2005), that is, they exaggerated their relative advantage on the dimension that they could shine on and that proved more relevant to the ingroup (Tajfel & Turner, 1979) while minimizing their weakness on the dimension that disfavored them. As a case in point, Kervyn et al. (2008; see also Oldmeadow & Fiske, 2010) found that participants that most identified to their national ingroups evaluated their country as warmer and more competent than those you felt less identified. Such a tendency to favor one's ingroup suggest that motivations may be driving this effect.

3. Underlying motivations

According to compensation theorists, several motivations may account for the emergence of this hydraulic relationship between warmth and competence when characterizing social targets (Kervyn et al., 2010; Yzerbyt, 2018; Yzerbyt et al., 2008). In the present dissertation, we focus on four core motivations, namely the need to promote harmonious intergroup relationships, justice concerns, self-affirmation concerns, and self-presentation concerns.

3.1 Harmonious intergroup relationship concerns

A first motivation is to *promote harmonious intergroup relationships*. According to Stangor and Leary (2006), people seek affiliation and social harmony with other outgroups. Indeed, collaboration with other groups can be rewarding for both parties. In this context, compensation would work as an ideal strategy to reduce social competition while promoting peaceful relations and collaboration given that each group can claim some advantage on non-corresponding fundamental dimensions (Mummendey & Schreiber, 1983; Tajfel & Turner, 1979; Yzerbyt et al., 2005). Antagonism between groups would therefore be curtailed given that the two sides can ensure some degree of positive distinctiveness (Hornsey & Hogg, 2000), and thus come across as “different” but not as “better” (Mummendey & Schreiber, 1983).

This pattern of assigning qualities, albeit different ones, to the groups in presence is reminiscent of Hewstone and Brown's (1986) model of mutual differentiation as well as recent work on multiculturalism (Hahn et al., 2015; Wolsko et al., 2000). Both perspectives are loosely rooted in social identity approaches (Tajfel & Turner, 1979; Turner et al., 1987) and question the view that achieving positive intergroup relations requires people to see groups on similar terms and to avoid ingroup bias if they wish to foster social harmony and intergroup understanding. Rather, these two frameworks consider that groups are likely associated with distinct features that are at the heart of stereotypes. In other words, assigning a higher standing on different attributes to each one of the groups may offer a promising

pathway to constructive relations and mutual respect (for a review, see Yzerbyt & Demoulin, 2010).

Although seemingly positive, the motivation to promote cooperation may be Janus-faced in some instances. This seems to be particularly the case with powerful groups that ascribe favorable traits on status-irrelevant dimensions (e.g., warmth, physical appearance) to flatter subordinate groups and incite them to willingly cooperate and maintain the hierarchical system (i.e., cooptation) (Eckes, 2002; Fiske et al., 2002; Glick & Fiske, 2001a; Jackman, 1994; Jost & Banaji, 1994;).

3.2 Justice concern

In line with the above, a second motivation that may drive compensation rests on *justice concerns*. According to System Justification Theory (SJT; Jost & Banaji, 1994), “there is a general psychological tendency to justify and rationalize the *status quo*; in other words, there is a motivation to see the system as good, faire, legitimate, and desirable” (Kay et al., 2007, p. 308). There exists a number of psychological reasons underlying this tendency, for instance, it provides a sense of closure, consistency, and control; it reduces uncertainty, fear, and anxiety; and, more importantly, it serves as a rationale for social inequalities that strengthen the belief in a just world (Jost et al., 2009; Jost et al., 2003; Lerner, 1980; Rubin & Peplau, 1975). As surprising as it may seem, the desire to bolster the *status quo* is not only present among stakeholders that directly benefit from the system (e.g., rich people), but also by those who most suffer from it (e.g., poor people). One of the most noxious consequences of the system justification is that marginalized groups can exhibit outgroup favoritism and a low collective or self-esteem (e.g., Jost et al., 2004; Jost et al., 2002; Jost & Thompson, 2000). Moreover, those that try to defy the establishment are prone to the backlash effect (Rudman & Glick, 1999).

One potent way to vindicate the system is by endorsing and promoting stereotypes that match the current social arrangement (Fiske et al., 2002; Jost & Banaji, 1994; Sidanius & Pratto, 1999). For instance, people’s higher (lower) status can be rationalized by ascribing them high (low) levels of competence (Fiske et al., 2002). This works as a form of correspondence bias (Gilbert & Malone, 1995) in which people are held accountable for their outcomes. Complementary stereotypes are especially convenient to professedly justify inequalities (caused by the current system; e.g., capitalism) by conferring both the powerless and powerful a distinct and offsetting set of strengths and weaknesses, thus fostering the “illusion of equality” and that “everyone gets his share” (Jost et al., 2005; Kay & Jost, 2003; for a review, see Kay et al., 2007). As a case in point, the gender status gap that has women overrepresented in subordinate positions in contrast to men is partly sustained by the belief that women’s social skills but low capability makes them best suited for low-ranked interpersonal-oriented occupations, whereas man’s cold-hearted but rational mind makes them fit for high-ranked goal-achieving occupations (Eagly, 1987; Glick & Fiske, 2001a, 2001b; Jackman, 1994).

Given that dimensional compensation engenders and perpetuates complementary stereotypes, such a process could be expected to relate to justice concerns (Kervyn et al., 2010; Yzerbyt et al., 2008; Yzerbyt et al., 2005). In line with this idea, Kay and Jost (2003) showed that exposure to non-complementary stereotypes (e.g. “poor but miserable”) implicitly prompted the activation of justice concerns, and that exposure to complementary stereotypes (e.g. “poor but happy”) translated into an increase in explicit system-justification measures (for similar findings see also Cichocka et al., 2015; Jost & Kay, 2005). Furthermore, the chronic tendency to justify the system (e.g., individuals that score higher on the economic system justification scale) or the temporary activation of system justification motives (i.e., by threatening the social order) lead to an increased endorsement of mixed stereotypes which subsequently boost the perceived legitimacy of the system (Jost et al., 2005; see also Jost & Kay, 2005; Yada & Ikegami, 2017). Having said this, Owuamalam et al. (2016) did not find any significant relationship between system justification motives and compensatory stereotypes when evaluating overweight versus ideally weighted people. More research is therefore required to further clarify this relationship.

3.3 Self-affirmation concerns

A third motivation to compensate is *self-affirmation*. When people belong to a meaningful group, they develop what is known as a social identity—the portion of the self-concept associated with the ingroup which holds emotional and value significance (Tajfel & Turner, 1979). According to Social Identity Theory (SIT; Tajfel & Turner, 1979), individuals strive to achieve or to maintain a distinct and positive social identity.

Satisfactory social identities mostly derive from advantageous comparisons between the ingroup and a meaningful outgroup. In his famous experiments, Tajfel (see Tajfel & Turner, 1979) repeatedly showed that the mere—random—categorization of individuals into minimal groups (i.e., a minimal group paradigm) led them to try to maximize the difference in points (or economic gain) in favor of their ingroup at the expenses of the outgroup, even if conceding more to the later would not have trumped ingroup gains. This phenomenon, called ingroup favoritism, reflects the tendency to favor one’s ingroup over an outgroup in evaluation and behavior. In terms of (self)stereotyping, the domains on which groups seek contrast to gain positive distinctiveness broadly maps the two fundamental dimensions of social perception, namely warmth and competence (Kervyn et al., 2015; Oldmeadow & Fiske, 2010; Richetin et al., 2012). Interestingly, high- and low-status groups pursue social distinctiveness primarily on the dimension more readily associated with their group (i.e., competence and warmth respectively; Fiske et al., 2002; Oldmeadow & Fiske, 2010). Nonetheless, impression management research has shown that they also seek to gain positiveness on their less favored dimension (e.g., by downplaying their advantage on their favorite dimension; e.g., Kervyn, Judd et al., 2009; Lindholm & Yzerbyt, 2018; Swencionis et al., 2017). This means that both advantaged and disadvantaged group members would be eager to claim their superiority on the two dimensions. However, given structural and social constraints, it is not always possible for people to discriminate on all dimensions all the time.

Because of their underprivileged position, low-status group members are likely to experience a negative social identity. SIT (Tajfel & Turner, 1979) holds that, depending on a series of factors, people can rely on a variety of strategies to deal with these situations. If group boundaries are permeable, people may opt for an individual upward mobility strategy by dissociating themselves from their ingroup and migrating to another, allegedly worthier, group (e.g., moving from the lower socio-economic class to the middle class). Another strategy is social competition which usually takes place when group boundaries are sealed and coupled with an unstable status gap. This more conflictual approach has parties directly compete against each other to improve their respective standing (e.g., by competing over scarce resources or political power).

Of particular interest, when groups boundaries are considered impermeable and the hierarchical distance is judged stable and legitimate (Cambon et al., 2015; Dang et al., 2017; Durante et al., 2013; Jackson et al., 1996), individuals can rely on social creativity by altering elements of the comparative situation. One way to achieve this involves shifting the dimension of comparison from one that profits the outgroup to one on which the ingroup can shine on (Lemaine, 1974; Mummendey & Schreiber, 1983). This strategy is tantamount to the so-called compensation effect (Yzerbyt et al., 2005). Specifically, for low-status group members it would mean acknowledging their inferiority on competence, while shifting the comparison to the warmth dimension in order to redeem some positive distinctiveness. In line with this idea, low-standing (but also high-standing) group members did not display compensation when they were initially denied the possibility to first claim their superiority on their preferred dimension of warmth (competence) (Yzerbyt & Cambon, 2017). Specifically, low-standing group members challenged, at least to some extent, the superiority of the outgroup on competence. Also, compensation failed to emerge when disadvantaged (but not advantaged) group members had the chance to boost their self-esteem by completing a self-affirmation task prior to the groups' evaluation (Cambon & Yzerbyt, 2018). Following self-affirmation, low-standing group members conceded equal warmth to the high-standing group members.

3.4 Self-presentation concerns

A fourth motivation that fuels the emergence of compensatory judgments has to do with *self-representation concerns*. Indeed, compensation sometimes appears to come across as a socially desirable strategy when dealing with strong normative pressures against the expression of discrimination. This is highly relevant in contemporary societies in which social norms that condemn discrimination—especially toward the so-called “protected groups”, i.e. those that rank low on the social ladder—coexist with prejudiced beliefs (e.g., racism, sexism, classism; Crandall et al., 2002; Dambrun & Guimond, 2004; Jeffries et al., 2012; Monteith et al., 1996).

Cambon et al. (2015) reasoned that compensation as a self-presentation strategy should be particularly at work among individuals belonging to high-status groups, a phenomenon known as magnanimity or noblesse oblige effect in the power literature (Sachdev & Bourhis, 1985; Vanbeselaere et al., 2006). Because of their privileged position, advantaged groups

should be aware of their blatant superiority which would likely activate their knowledge about the non-discrimination norm and consequently prevent them from displaying an ingroup bias on both dimensions. Instead, they should be willing to acknowledge some superiority to the low-status group on the dimension they care less about (i.e. warmth) in order to come across as tolerant and fair (i.e., politically correct; e.g. Mummendey & Schreiber, 1983). In line with their prediction, the emergence of a compensation pattern was mediated by pressures toward non-discrimination among privileged group members (for similar results, see Bergsieker et al., 2012; Cambon & Yzerbyt, 2016; Jeffries et al., 2012; Owumalam et al., 2016; Terache et al., 2020; Yzerbyt & Cambon, 2017).

Cambon and Yzerbyt (2018; Expt 2) went a step further by experimentally manipulating the activation of the norm. Participants who had randomly been assigned (by the means of a minimal group paradigm) to the high- or low-status group, were either presented with a fake survey on political correctness or honesty. The key message of the political correctness (honesty) survey suggested that it was bad (good) and unfair (fair) to express criticism toward other groups. As predicted, the norm manipulation only affected high-status group members. Specifically, ingroup bias emerged on both dimensions when the honesty norm was activated whereas compensatory judgment emerged when the political correctness norm was activated. These results are in line with Cambon and Yzerbyt's (2017) previous work showing that pressures toward nondiscrimination are more prevalent among the high-competence groups and that they are strongly correlated with a bias on the warmth dimension in favor of the low-competence group.

As discussed previously, one crucial aspect for the compensation pattern to emerge in this context is the legitimacy of the status gap between the groups. Indeed, as previous studies have shown, *noblesse oblige* only takes place as long as advantaged group members feel secure about their group standing (Sachdev & Bourhis, 1985; Vanbeselaere et al., 2006; Yzerbyt & Cambon, 2017). This should be even more the case if the intergroup difference concerns the competence—rather than warmth—dimension arguably because the former is perceived as more objective, less malleable, and thus less disputable than the latter (Yzerbyt & Cambon, 2017). Moreover, Yzerbyt and Cambon (2017) revealed that magnanimity among high-competence groups only materialize to the extent that they can first secure their advantage on their preferred dimension, namely competence. Specifically, compensation only occurs if they are asked to rate themselves and the low-competence group on competence first without previously knowing that they would have to subsequently do the same on the warmth dimension (i.e., “comfortable” situation in which they can first secure a high standing), but not the inverse (i.e., “uncomfortable” situation).

Compensation has also proven to be an effective impression management strategy to convey a desired image of the self or the ingroup to others that goes beyond coming across as tolerant or fair. Although people are generally driven to present themselves under a good light, they may sometimes seek to transmit a specific impression to attain their goals (Holoien & Fiske, 2013; Leary, 1995; Schlenker, & Leary, 1982). For instance, they may want to highlight their competence over their social skills in a job interview to better match the desired profile, whereas at a party they may want to emphasize their sociability over their

skills. Interestingly, research has shown that people take advantage of the negative relation between the two dimensions in impression formation and actively use it to shape their self-image or to disconfirm status-based stereotypes of their ingroup. Specifically, they tend to downplay their warmth so that they can come across as more competent, and vice versa (Lindholm & Yzerbyt, 2018; Swencionis & Fiske, 2016). These findings suggest that people are aware and spontaneously make use of the relations of the two fundamental dimensions of social judgment.

4. Conclusions

In this chapter, we showed that the compensation effect (i.e., the negative relationship between warmth and competence) is both widespread and robust. This pattern emerged for social targets that were either real or fictitious, groups or individuals, and considering different vantage viewpoints (i.e., external observers, high- or low-group members). Research also revealed that a comparison context between two targets on the unique dimension of warmth and competence, coupled with a low level of conflict and legitimate status gap favored the emergence of compensatory judgments. Finally, studies have shown that compensation can be driven by motivations to promote harmonious intergroup relations between parties, to justify the current social arrangements, to achieve positive social distinctiveness, or to convey a specific impression about the self or the ingroup to others. Interestingly, some motives may conflict with each other, leading to the public expression of dimensional compensation that may depart from the judges' convictions. As most, if not all, of the work reviewed above rests on direct measures, it would be important to examine what can be obtained by more indirect, indeed unobstructive, methods. In Chapter 3, we examine studies assessing judgments on the two dimensions at both the direct and indirect level.

Chapter 3

Direct and indirect compensation

Direct and indirect compensation

1. Introduction

Although a large body of evidence points to the robustness of the compensation effect, researchers have almost exclusively relied on so-called explicit or direct measures with almost no attempt to make use of implicit or indirect measures. Indeed, it is often the case in compensation studies that, after participants have been introduced to the to-be-rated social targets, they are simply asked to report their impression of each target by indicating their degree of (dis)agreement with a series of handpicked traits encompassing the two dimensions of social judgment on Likert-type scales (e.g., Cambon & Yzerbyt, 2016; Cambon & Yzerbyt, 2018; Dang et al., 2017; Judd et al., 2019; Judd et al., 2005; Kervyn et al., 2008; Lindholm & Yzerbyt, 2018; Owuamalam et al., 2016; Terache et al., 2020; Yada & Ikegami, 2014; Yada & Ikegami, 2017; Yzerbyt & Cambon, 2017; Yzerbyt et al., 2005). This is slightly surprising because direct measures are susceptible to social desirability issues and self-presentation concerns that may thus be driving compensation (Cambon & Yzerbyt, 2018; Cambon et al., 2015).

The terms *implicit* and *explicit* initially came from the field of memory research (see Schacter, 1987). Individuals were said to display implicit memory when they remained unable to report explicitly, or were unaware of, past events (e.i., unable to tell from a list which words they had been previously presented with) even though their performance on selected measures suggested the opposite (e.i., they were faster in a fragment completion task to complete words they had previously been exposed to). A few years later, Greenwald and Banaji (1995) suggested that “the signature of implicit cognition is that traces of past experience affect some performance, even though the influential earlier experience is not remembered in the usual sense—that is, it is unavailable to self-report or introspection” (p. 4); while pointing out that the study of this construct required *indirect measures*, “which neither inform the subject of what is being assessed nor request self-report concerning it” (p. 5). Since then, there has been an ever-growing body of research in the field of social psychology that made use of indirect measures. However, the variety of methods and conceptualizations led to a terminological confusion of the term *implicit* as highlighted by recent reviews (see Corneille & Hütter, 2020; Greenwald & Lai, 2020; Van Dessel et al., in press; see also Fazio & Olson, 2003; Jost, 2019). Moreover, these reviews have come to question, among other things, the *unconscious* nature of the phenomenon. These researchers advocated for the description of each measure in functional behavioral terms as well as for abandoning the terminology “implicit” and “explicit” in favor of “direct” and “indirect” (Corneille & Hütter, 2020; Fazio & Olson, 2003; Van Dessel et al., in press; but see Greenwald & Lai, 2020). In light of these recent efforts, we will report previous research adopting the terms that they originally used and make use of the “direct” and “indirect” terminology in our own research.

Despite the heated debate surrounding what indirect measures actually tap, a large proportion of the research in social psychology, including those reviewed in the current chapter, relied on these types of measurement not as a mean to assess unconscious cognitions but rather for empirical and practical reasons. The key goal is to diminish demand characteristics or social desirability issues (Fazio & Olson, 2003; Greenwald & Banaji, 1995; Nosek, 2005). In this context, the most prominent issue with direct measures is that plainly asking participants to assess the construct of interest undoubtedly makes them aware of what is being measured. Participants may thus refrain from expressing their genuine impressions, attitudes, or emotions insofar there exist some form of experimental demand, self-presentation concerns, or when the topic under investigations is sensitive to social desirability, as it is the case when evaluating “protected groups” (e.g., Crandall et al., 2002; Dovidio & Gaertner, 2004; Jeffries et al., 2012).

It is assumed that indirect measures may be less prone to these issues because responses are presumably less controllable (e.g., Banse et al., 2001; Kim, 2003). Consequently, correlations between direct and indirect measures should be larger (and positive) the lower the social desirability. Evidence suggests that this is indeed the case, but only when considering specific motivations (e.g., motivation to control prejudice) in conjunction with sensitive topics (e.g., when evaluating stigmatized groups) rather than when measuring social desirability in general (see Gawronski et al., 2007; Greenwald & Lai, 2020). However, caution is advised when interpreting these findings given that “motivational factors are neither necessary nor sufficient to moderate the relation between self-reports and indirectly assessed representations” (Gawronski et al., 2007, p. 185). Indeed, confounding variables may account for these relations (e.g., differences between measures in degree of reliability, structural task, target object; see Gawronski & De Houwer, 2014; Greenwald & Lai, 2020; Hofmann et al., 2005). As for controllability, there is evidence that the Implicit Association Test’s score (and probably other indirect measures as well which have not been investigated) can be deliberately faked (e.g., Banse et al., 2001; De Houwer et al., 2007; Kim, 2003; Röhner et al., 2013), although this seems more likely when effectively instructing participants to fake the test and when providing them with faking strategies (Greenwald & Lai, 2020).

In Chapter 3 we first review the work that has assessed mixed stereotypes at both the direct and indirect level. In the second section, we review the reverse correlation technique and examine how this technique could serve as an indirect measure. In the last section, we present the only study that, to our knowledge, has tested compensation at an indirect level.

2. Direct and indirect judgments on the fundamental dimensions

Wade and Brewer (2006) were among the first to probe mixed stereotypes—about subgroups of women—at the implicit and explicit level on both warmth and competence dimensions. To do so, they relied on the Lexical Decision Task (LDT; Wittenbrink, Judd, & Park, 1997) in order to assess how quickly participants discriminate between non-words (e.g., “cetopment”, “bleciaso”) and words (e.g., “competent”, “sociable”) after the subliminal

presentation of a prime. They evaluated facilitation by comparing response times when recognizing a word after the presentation with a prime of interest (e.g., “businesswomen” or “homemaker”) relative to a neutral prime (e.g., “XXXX”). Results in explicit measures indicated mixed stereotypes of businesswomen as more competent but colder than homemaker by both men and women participants. However, the self-reported ambivalence was overridden by general affect (positive versus negative) on the LDT such that female participants more readily associated businesswomen than homemakers with positive adjectives regardless of their dimensional content (i.e., competence vs. warmth), whereas men did the opposite. Similarly, Ebert et al. (2014; see also Richeson & Ambady, 2001) found implicit self-favoring stereotypes on both dimensions when female participants assessed their ingroup (i.e., women) relative to men. Moreover, male participants more strongly associated competence with their gender but displayed no preferential responses in terms of warmth.

Rohmer and Louvet (2012) assessed implicit and explicit stereotypes of a strongly normatively protected group, namely people with disabilities. Like Wade and Brewer (2006), their findings show discrepancies between responses measured at the implicit versus explicit level. Indeed, self-report measures revealed mixed stereotypes about people with disabilities and without disabilities, with the former being perceived as warmer but less competent than the latter. As for LDT’s time reactions, they suggested that both competence- and warmth-related traits were more readily associated with people without disabilities than with disabilities. Clément-Guillotin et al. (2018) conceptually replicated these results by making use of another time-reaction-based measure, namely the Implicit Association Test (IAT; Greenwald et al., 1998) which compares reaction times (i.e., facilitation) when categorizing stimuli in stereotype-congruent (e.g., Black-negative and White-positive) versus stereotype-incongruent (e.g., Black-positive and White-negative) blocks. Also, Rohmer and Louvet (2018) found that people systematically display implicit derogation of disabled people on the warmth dimension but that derogation on competence seemed to be bounded to competence-relevant situations (e.g., work environment).

In contrast with the aforementioned literature which found divergence between self-reported versus reaction-time-based responses, Imhoff et al. (2013) observed a concordance between the two. That is, their participants more readily associated nursery teachers with warmth and managers with competence at both the explicit and implicit level (the latter being assessed by a Brief IAT; BIAT; Sriram & Greenwald, 2009) as well as at the facial level (as assessed by a reverse correlation technique; see next section). Carlsson and Björklund (2010) also found ambivalent stereotypes about lawyers and preschool teachers on both Likert-type scales ratings and IAT reaction times. Interestingly, a follow-up study revealed that when preschool teachers were used as participants, they actually displayed ingroup favoritism at the implicit level (i.e., they mapped more quickly preschool teachers vs. lawyers with both warmth and competence) while professing mixed stereotypes at the explicit level.

The above review reveals some discrepancies between findings when comparing explicit and implicit measurements of mixed stereotypes. Inconsistencies may be accounted, for instance, by the different measures that were used (e.g., LDT, IAT, BIAT) as each of them

may tap into different implicit processes (Van Dessel et al., in press). Also, it seems that the overlap between implicit and explicit responses is greater when the social targets evaluated are not protected groups (e.g., Carlsson & Björklund, 2010; Imhoff et al., 2013; but see Hofmann et al., 2005). For instance, Carlsson and Björklund (2010) carefully pre-selected the two social groups such that they would not be considered as socially sensitive in order for participants not to feel normatively pressured to refrain from expressing criticisms about these groups (e.g., Crandall et al., 2002; Dovidio & Gaertner, 2004; Jeffries et al., 2012). In contrast, it seems that implicit and explicit measure did not match when the social targets were socially sensitive such as women (Ebert et al., 2014; Wade & Brewer, 2006) and disabled people (Clément-Guillotin et al., 2018; Rohmer & Louvet, 2012, 2018;). Another interesting pattern that emerges is that self-serving associations (e.g., women as both competent and warm) arise when people implicitly assess their ingroup although they express mixed stereotypes on self-reported measures. The latter result suggests that some of the motivations discussed in Chapter 2 (e.g., self-affirmation and self-presentation concerns) may indeed play a role in shaping people's public expression in contrast to their more private beliefs.

3. The reverse correlation as an indirect and spontaneous measure

The two-forced-choice-image-reverse-correlation method introduced by Dotsch et al. in 2008 has been increasingly popular among social psychologists to tap into a wide array of visual representations of social targets or facial expressions (see Brinkman et al., 2017). One of the most widely praised advantage of this data-driven method is that it captures spontaneous mental templates of a given target while making no prior assumptions about what those mental representations should look like (Brinkman et al., 2017; Dotsch & Todorov, 2012; Jack & Schyns, 2017). Interestingly, several researchers consider this technique as an implicit or indirect measure allowing to bypass social desirability or to tap into internal representations outside the scope of awareness (Brinkman et al., 2017; Degner et al., 2019; Gunaydin & DeLong, 2015; Hinzman & Maddox, 2017; Moon et al., 2020; Walker & Keller, 2019). Surprisingly, however, only but a few provide a clear rationale for why this would be the case.

As described in Chapter 1, in a typical reverse correlation (RC) participants go through several trials comprising two noisy faces. Their task is to select, at each trial, the face that best matches the expected target (e.g., a more Moroccan-looking face; Dotsch et al., 2008). The noisy faces are generated by superimposing a random sinusoidal noise layer (Mangini & Biederman, 2004) to a constant base image which is typically a morph of several real faces. The visual representation of the target, the so-called classification image (CI), is obtained by averaging all the noise layers selected from the noisy faces (i.e., the stimuli) by a given participant (the individual CI) or a group of participants (the group or condition CI) and then combining it with the base image. Particularly, because CIs are averages of all the selected noise layers, they will tend to capture patterns that are consistent (i.e., the signal; e.g., a smiling mouth) and cancel out those that are inconsistent (i.e., the noise) across trials.

The method is said to be unconstrained for two reasons. First, it allows capturing spontaneous use of information given that participants are free to adopt whatever criteria they wish to make their choice (Brinkman et al., 2017; Degner et al., 2019; Dotsch & Todorov, 2012; Jack & Schyns, 2017). For example, when selecting the most dominant-looking face, some participants may rely on how angry the face appears to make their judgment, while others may privilege masculinity traits. This feature clearly contrasts with more traditional measures that constrain participants' responses. In the case of the compensation literature for instance, participants can only convey their impressions on the traits pre-selected by the researcher. This issue also affects indirect measures in which participants react to associations between specific a pre-established categories/group labels/pictures and traits/adjectives/words (e.g., IAT, LDT). Second, this method allows for unconstrained variation of the stimuli (i.e., the noisy faces) thus making no prior assumptions about what the final product may look like. In other words, the randomness of stimuli provides the researcher with the opportunity to sample an immensely large array of hypotheses (i.e., the face space of the CIs) without even requiring them to define beforehand the dimensions along which the stimuli should vary.

We agree with previous researchers (Brinkman et al., 2017; Degner et al., 2019; Gunaydin & DeLong, 2015; Hinzman & Maddox, 2017; Moon et al., 2020; Walker & Keller, 2019) that the reverse correlation method may serve as an indirect measure. In line with this idea, empirical findings show that implicit measures vary positively along facial representations (e.g., more implicitly prejudiced individuals have a more negative CI of Moroccans; Dotsch et al., 2008; see also Dotsch et al., 2013). However, this relation is not a consistent one (e.g., Degner et al., 2019; Imhoff et al., 2013).

One argument in favor of the indirectness of the RC is that it makes the ineffable (i.e., what cannot be expressed in words) explicit (Mangini & Biederman, 2004). For instance, "people can readily recognize celebrities but if asked to describe the difference between Tom Cruise's and John Travolta's faces, they typically resort to naming individual features, such as Travolta's cleft chin, which fails to capture the rich differences in their appearance. Although readily perceived, such differences remain ineffable" (Mangini & Biederman, 2004, p. 210). Indeed, the tiny variations produced from random noise in the stimuli (i.e., the noisy faces) can sometimes match (by pure coincidence) subtle facial features for which participants may be unable to provide accurate verbal reports or that may even be outside the scope of awareness. Anecdotal evidence from our own online and laboratory experiments shows that many participants who underwent a RC task report that "all faces look the same" and that they feel like they were "picking at random". Nonetheless, participants' CIs do end-up with clearly defined facial characteristics (that are markedly distinct from the set of stimuli), thus suggesting that their choices were in fact not entirely random, although participants had this subjective impression.

More importantly, it may also be argued that the RC procedure can take care of social desirability issues. Imagine a prejudiced individual who is sensitive to social desirability concerns. Let us assume that his representation of a typical Moroccan corresponds to an angry-, criminal-, and untrustworthy-looking face with a dark skin tone. If he had to convey

his representation of a Moroccan directly, whether by means of a verbal report, trait ratings, or a selection among real faces, the one that best matches his prototype, he would likely be reluctant to provide an accurate response because he would not want to appear prejudiced. Now imagining that instead he is asked to perform a RC task. First, he will never be exposed to the final measurement outcome, that is, his own CI, so it is unlikely that he will know in advance or even while performing the task what the final product will be. Second, he may not even be aware that the task is actually measuring his mental representation given that he only has to select images and not to convey his mental representation in full detail. Third, if we conceive social desirability as a continuum, we may suppose that he would be more willing to display responses that show a small bias than a large bias. Under this assumption, he might be willing to select noisy faces with only a few (e.g., angry, dark skin) among all his prototype defining features, or noisy faces that do display these features but in an attenuated fashion (e.g., slightly dark skin tone). As previously mentioned, the differences between noisy faces in the RC are usually so subtle that they may thus not activate social desirability concerns, and yet capture the visual presentation from the participant. In other words, “these methods are so subtle that they do not reveal the dimensions under investigation and, therefore, they are not susceptible for correction strategies” (Walker & Keller, 2019, p. 497).

Certainly, the aforementioned arguments remain speculative and require empirical corroboration. To our knowledge, there exist no study to date that has systematically and empirically investigated if this noise-based technique fulfills the criteria to be classified as an indirect measure. Further research in this direction is therefore desirable.

4. Compensation effect at the direct and indirect level

Although quite informative, the above efforts did not aim to probe compensation per se. To our knowledge, Kervyn et al. (2011) conducted the only research that tested the compensation effect at an indirect level.

In Experiment 1, Kervyn et al. (2011) introduced their participants to two fictitious groups whose behaviors initially differed on one of the two fundamental dimensions while being ambiguous on the other (as in Judd et al., 2005; see Chapter 2). Participants were then presented with a rating task (i.e., explicit measure) comprising traits from both dimensions to assess their impressions about the two groups. Next, they were presented with a set of four new written behaviors for each of the two groups. Each behavior in the set was either diagnostic of positive warmth, negative warmth, positive competence, or negative competence. The attribution task was to indicate, for each new behavior, the extent to which they thought it had been caused by the actor’s dispositions (e.g., because of his personality traits). The authors reasoned, based on causal attribution research (see Gilbert, 1998; Trope, 1986), that dispositional attribution of new behaviors would indirectly signal that participants believe these behaviors to be diagnostic of stable and internal dispositions of the target. In accordance with their expectations, results revealed direct compensation (i.e., traits ratings) as well as indirect compensation as new behaviors in line with compensation were attributed

more dispositionally than those that were inconsistent with this effect, whether warmth or competence was the manipulated dimension. For instance, in the condition in which groups were contrasted in terms of competence (i.e., competence manipulated), new competence-related behaviors of the high-competence groups were attributed more dispositionally than for the low-competence group, whereas the opposite pattern was observed for new warmth-diagnostic behaviors.

In Experiment 2, the authors went a step further by relying on an even more unobtrusive method. Specifically, after introducing participants to the two fictitious group (as in Experiment 1), they presented them with 16 pictures of new behaviors that were diagnostic of either high/low warmth or competence (e.g., high warmth: A mother hugging her child; high competence: A student at his graduation; low warmth: A wife yelling at her husband; low competence: A person that crashed her car into a pole). Each picture came accompanied by four sentences describing the main character's behavior. Participants task was to pick the sentence that best described the behavior illustrated in the picture. The descriptions for each picture (e.g., a person that crashed her car into a pole) varied from very abstract to more concrete (e.g., "X is reckless", "X looks at the damage to his car", "X drives badly", "X provoked an accident"). Importantly, the Linguistic Category Model (LCM; Semin & Fiedler, 1988) holds that people tend to describe behaviors more (less) abstractly to transmit the idea that the behaviors are more (less) diagnostic of stable and long-lasting traits of the person performing the action. Their findings revealed that descriptions of behaviors congruent with compensation were described in more abstract terms than those incongruent with this effect. However, compensation only emerged when competence (but not warmth) was the manipulated dimension, consistent with previous findings that suggest that this effect is stronger when competence, rather than warmth, is manipulated (Judd et al., 2005). Taken together, these two experiments highlight the pervasiveness of compensation by suggesting that they not only show at a direct level but also at an indirect level.

5. Conclusions

In Chapter 3 we examined the available intergroup perception literature that assessed mixed stereotypes at both the direct and indirect level. Findings suggest that the pattern found in explicitly measured mixed stereotypes does not always emerge in implicit measures, possibly because of underlying motivations such as self-presentation or self-affirmation concerns that can distort public expressions of inner beliefs. Moreover, we examined an innovative data-driven technique, namely the reverse correlation, and argued that it may come handy as a spontaneous and indirect measure. Finally, we reviewed the only work available that, to our knowledge, tested compensation per se at both the explicit and implicit level.

Chapter 4

Direct and indirect dimensional compensation

Is there a difference between observers and group members?

Dimensional compensation takes place when perceivers judge one of two social targets higher on one of the two fundamental dimensions while judging the other target higher on the second dimension. Interestingly, the majority of studies on the compensation effect focused on direct measures, with almost no attempt to rely on more indirect measures. We tested whether dimensional compensation also takes place at a more indirect level (Brief-IAT). In Experiment 1, observers presented with unknown groups compensated both directly and indirectly. Experiment 2 had participants assigned to one of two novel groups. Whereas low-competence group members compensated on both direct and indirect measures, high-competence group members compensated at the direct level but did not concede any advantage to the low-competence group at the indirect level. As a set, our findings shed new light on direct and indirect compensatory judgments as a function of perceivers' vantage point as observers and group members.

Keywords: compensation effect; indirect measures; direct measures; stereotype content model; minimal group paradigm

Reference

Schmitz, M., & Yzerbyt, V. (in press). Direct and indirect dimensional compensation: Is there a difference between observers and group members? *Group Processes & Intergroup Relations*. <https://doi.org/10.1177/1368430220963176>.

Direct and indirect dimensional compensation

Is there a difference between observers and group members?

1. Introduction

Dimensional compensation takes place when a contrasting pattern emerges between warmth and competence in people's judgments about two social targets (Yzerbyt, Provost, & Corneille, 2005; for a review, see Yzerbyt, 2018). Although a considerable body of evidence supports the robustness of the compensation effect, the vast majority of the research relied on direct, also called explicit, measures (e.g., traits ratings). This raises the possibility that social desirability concerns contribute to shaping the specific pattern reported in previous work. Indeed, Yzerbyt, Kervyn, and Judd (2008) suggested that dimensional compensation might emerge partly because of people's desire to manifest a balanced evaluation of the social targets under comparison, particularly among high-status groups (Cambon & Yzerbyt, 2018). The goal of the present research is to investigate whether dimensional compensation emerges when measured both directly and indirectly, and when people are in a position of external observers or group members.

Social perceivers rely on two dimensions to deal with their social environment (Abele, Cuddy, Judd, & Yzerbyt, 2008; Fiske, 2015; Yzerbyt, 2016). When it comes to group stereotypes, the Stereotype Content Model (SCM; Fiske, 2018; Fiske, Cuddy, Glick, & Xu, 2002) holds that warmth refers to the target's intentions and encompasses traits such as kindness, friendliness, and sociability, whereas competence translates the capacity of the target to materialize these intentions and comprises characteristics such as intelligence, skillfulness, and organization (Fiske, 2015). Although the SCM considers stereotypes corresponding to all possible values on these two dimensions, a great many groups are in fact linked to so-called mixed stereotypes, that is, that are positively valenced on one dimension but negatively on the other (e.g., rich, lawyers, homemakers, disabled people; for a recent review, see Yzerbyt, 2016).

Building on social identity approaches (Tajfel & Turner, 1979), Yzerbyt et al. (2005) argued that one possible account for ambivalent stereotypes resides in the negative relationship between the two fundamental dimensions that results whenever perceivers approach groups in comparative terms. That is, intergroup relations often operate such that a group higher than another on one of the two fundamental dimensions comes across as lower on the other dimension, a phenomenon that Yzerbyt et al. (2005) called the compensation effect. Initial work on French and Belgians, two groups enjoying different prestige and power, confirmed this conjecture: Respondents from both groups agreed that French were more competent but less warm than Belgians (see also, Kervyn, Yzerbyt, Demoulin, & Judd,

2008). A follow-up study also revealed that the compensation pattern proved stronger for group members (i.e., Belgians or French people) than for external observers or third-party members (i.e., Swiss citizens).

The compensation effect occurs not with any pair of dimensions. Indeed, a halo effect materializes when considering a third dimension (e.g., healthiness; Yzerbyt et al., 2008). Dimensional compensation is thus not a mere artefact emerging with any two dimensions in a comparative context (i.e., a contrast effect; Mussweiler, 2007) or solely the product of undifferentiated justice concerns (Kay et al., 2007). Rather, it builds on the unique structural and functional components of the two fundamental dimensions of social perception (Kervyn, Yzerbyt, & Judd, 2010).

Empirical evidence of the so-called dimensional compensation model (DCM; Yzerbyt, 2018) is plenty. For instance, Judd, James-Hawkins, Yzerbyt, and Kashima (2005) asked their participants to form an impression about two novel groups from behaviors allegedly performed by their members. The authors manipulated behaviors in such a way that one group came across as superior than the other on one dimension (the manipulated dimension; e.g., competence), while both groups were ambiguous on the other dimension (the non-manipulated dimension; e.g., warmth). Explicit ratings revealed dimensional compensation, although the effect was somewhat stronger when the manipulation concerned competence rather than warmth. Interestingly, dimensional compensation also emerged in a variant of the scenario that manipulated participants' membership. Both low- and high-competence group members dimensionally compensated. Replicating Yzerbyt et al.'s (2005) pattern, group members also maximized intergroup differences on their preferred dimensions (i.e., competence for the high-competence group and warmth for the low-competence group) and reduced differences on the other, less crucial, dimension (i.e., warmth for the high-competence group and competence for the low-competence group).

1.1 Accounting for the compensation effect

Several psychological motivations possibly account for a negative relation between the two fundamental dimensions (Kervyn et al., 2010; Yzerbyt, 2018; Yzerbyt et al., 2008). Here we identify four core motivations that could give rise to dimensional compensatory judgments. For each, we consider a moderating factor, namely, the vantage points of the perceiver as a member of either the advantaged or the disadvantaged group or as an external observer.

A first motivation is the *desire to promote harmonious intergroup relationships*. Building on Mummendey and Schreiber (1983) and the social identity approach (Tajfel & Turner, 1979), Yzerbyt et al. (2005) reasoned that when group members find themselves in a comparative context, granting each group with superiority on a different aspect stands as an ideal strategy to promote collaboration (for a review, see Yzerbyt & Demoulin, 2010). This motivation may be especially relevant for group members who are willing to overcome conflictual intergroup relationships (e.g., over scarce resources) or would like to promote cooperation. Conversely, this motivation is presumably absent when there is no status

asymmetry between groups as well as in highly conflictual contexts (Cambon, Yzerbyt, & Yakimova, 2015; Durante et al., 2017; Durante et al., 2013), or when there are few incentives to interact, as often the case in minimal group scenarios.

A second motivation is *justice concerns*. According to system justification theory (SJT; Jost & Banaji, 1994), people strive to justify and rationalize the *status quo* and are keen to perceive the current system as fair, legitimate, and desirable (Kay et al., 2007). Indeed, stereotypes often play an important role in maintaining and bolstering the hierarchical social system that people navigate (Jost & Kay, 2005; Tajfel, 1981). Complementary stereotypes appear particularly suited for such a purpose as they confer both advantaged and disadvantaged groups a set of weaknesses and strengths that even out, thus boosting the sense that the system is balanced, fair, and justified (Jost, Kivetz, Rubini, Guermandi, & Mosso, 2005; Kay & Jost, 2003). Given that dimensional compensation engenders and perpetuates complementary stereotypes, such a process might indeed satisfy a concern for justice (Kervyn et al., 2010; Yzerbyt et al., 2008). Interestingly, system-justification appears particularly insidious, as it would seem not only to operate among those who benefit from the *status quo* (i.e., high-status groups) but also among those who suffer from it (i.e., low-status groups). Moreover, this tendency also seems to take place under less controlled conditions (i.e., as assessed by implicit measures), presumably as the product of the internalization and acceptance of consensual stereotypes. Indeed, derogated group members can exhibit implicit outgroup favoritism (Jost, Pelham, & Carvallo, 2002) and exposition to non-complementary (i.e., non-compensatory) stereotypes implicitly activates justice concerns (Kay & Jost, 2003).

A third motivation stems from the need for *self-affirmation*. In the context of intergroup relations, this motivation is at the heart of Social Identity Theory (SIT; Tajfel, 1981; Tajfel & Turner, 1979). Indeed, SIT holds that group members are motivated to maintain and achieve positive ingroup distinctiveness and a concomitant positive social identity for the self. Although group members would readily affirm themselves on both fundamental dimensions of social perception (i.e., warmth and competence), inescapable structural factors often restrain such a desire. A common structural constraint in modern societies is that groups rank according to competence-based outcomes (Oldmeadow & Fiske, 2007; Son Hing et al., 2011). When the social hierarchy comes across as legitimate and stable, and when group boundaries are impermeable, low-status groups may then adopt a social creativity strategy (Tajfel & Turner, 1979) in order to cope with their disadvantaged position. Because competence is perceived as more objective and more constrained by reality than warmth is (Yzerbyt & Cambon, 2017), low-status group members will likely acknowledge their inferiority on the status-relevant dimension (i.e., competence) while focusing on a status-irrelevant dimension (i.e., warmth) that is more malleable (Yzerbyt & Cambon, 2017) and can be turned in their favor (Lemaine, 1974; Mummendey & Schreiber, 1983). Doing so offers them the opportunity to redeem positive distinctiveness and self-esteem (Yzerbyt et al., 2008). In line with this reasoning, depriving low-status group members from the opportunity to shine on the warmth dimension (Yzerbyt & Cambon, 2017) or giving them the opportunity to complete a self-reaffirmation task prior to making judgments (Cambon & Yzerbyt, 2018) prevented dimensional compensation.

A fourth motivation is *self-presentation*. Indeed, when making judgments about disadvantaged groups, displaying dimensional compensatory judgments may allow people to avoid appearing discriminatory. Being “politically correct” may thus come at hand in contemporary societies where expressing criticisms, at least toward the so-called “protected” groups, is hardly tolerated (Bergsieker, Leslie, Constantine, & Fiske, 2012; Cambon & Yzerbyt, 2018; Cambon et al., 2015; Crandall, Eshleman, & O'Brien, 2002; Dambrun & Guimond, 2004; Jeffries, Hornsey, Sutton, Douglas, & Bain, 2012; Owuamalam, Wong, & Rubin, 2016). Clearly, this motivation should be more at work among high-status group members, leading advantaged groups to feel embarrassed to express ingroup favoritism (i.e., self-affirmation) on both fundamental dimensions. As a consequence, they should be willing to concede outgroup superiority on the status-irrelevant dimension (i.e., warmth) while ascertaining their advantage on the relevant one (i.e., competence). Research confirms that dimensional compensation emerges to the extent that high-status group members can secure their dominance on their favorite dimension (i.e., competence; Yzerbyt & Cambon, 2017) and that the non-discrimination norm remains activated (Cambon & Yzerbyt, 2018), a pattern known as magnanimity or noblesse oblige (Vanbeselaere, Boen, van Avermaet, & Buelens, 2006). Dimensional compensation may thus reflect a more thoughtful motivation or strategy fueled by self-presentation concerns that would deceptively appear to others as justice concerns.

In sum, distinct motivations may lead people to compensate on the two dimensions and their relevance and strengths may differ both as a function of the context and the social position of the perceiver. Specifically, the desire to promote harmonious relations would seem more relevant for group members than for uninvolved observers. As to justice concerns, these may stand high on the agenda of both observers and group members, although the latter may also have more pressing concerns in mind. In particular, self-affirmation seems highly relevant for advantaged and disadvantaged group members as they would both readily enhance or secure a positive identity by affirming the value of their group on the two fundamental dimensions. This motive may take the form of social creativity (i.e., dimensional compensation) for low-status group members because of structural constraints (e.g., status gap and the objectivity of competence). For high-status group members, self-affirmation may conflict with self-presentation concerns due to non-discrimination pressures. Instead of displaying ingroup bias on all dimensions, high-status group members may come across as just and humble by publicly granting superiority to the low-status group on the less relevant dimension (i.e., by displaying dimensional compensation). In other words, although not necessarily absent among external observers and low-status group members, self-presentation concern may be particularly at work among high-status group members.

1.2 Compensation effect at the indirect level

Clearly, the above-mentioned motives may sometimes conflict with each other, potentially leading to the public expression of dimensional compensation that departs from their announcer's convictions. This is most interesting insofar as dimensional compensation research predominantly used such measures as self-report questionnaires with almost no

attempt to rely on indirect measures¹ (Corneille & Hütter, 2020). The few intergroup relations studies assessing judgments on the two dimensions at both the direct and indirect level reveal not only discrepancies between these measures but also inconsistencies across findings.

Wade and Brewer (2006) investigated implicit stereotypes of businesswomen and homemakers using a lexical decision task (LDT). Although both male and female participants refer to the mixed stereotype of these subgroups of women on direct measures, indirect measures were driven by respondents' general affect (i.e., non-ambivalent). Specifically, female participants evaluated businesswomen more positively than homemakers whereas the opposite was true for male participants. Similarly, Rohmer and Louvet (2012, 2018) found mixed stereotypes of disabled people on self-report measures whereas LDT measures revealed non-mixed stereotypes (low competence/low warmth). Clément-Guillotin et al. (2018) replicated these results using the Implicit Association Test (IAT).

In addition, Imhoff, Woelki, Hanke and Dotsch (2013) found complementary stereotypes of managers and nursery teachers in self-report and in reaction-time-based measures, namely a Brief Implicit Association Test (BIAT). The same mixed pattern also emerged in prototypical facial representations of these groups, as assessed by what admittedly counts as an indirect measure, that is, a reverse correlation method. In a similar vein, Carlsson and Björklund (2010) observed mixed stereotypes of preschool teachers and lawyers on both direct and indirect measures (IAT). Importantly, their participants did not belong to either of these two groups. In a follow-up study, preschool teachers implicitly evaluated their own group as warmer and as more competent than lawyers, suggesting ingroup favoritism. Ebert, Steffens, and Kroth (2014) obtained similar results when investigating gender stereotypes. The gender-warmth IAT revealed an overall association between women and warmth, consistent with SCM predictions. In contrast, an ingroup-favoring pattern emerged for the gender-competence IAT among both male and female respondents.

To sum up, the available intergroup perception literature reveals inconsistencies when testing the relation between the fundamental dimensions with direct and indirect measures as well as among group members and observers. Interestingly, none of the studies reviewed above aimed to examining the dimensional compensation *per se*. Additionally, all of them relied on existing group stereotypes and it remains difficult to evaluate whether previous knowledge and prevailing structural factors (e.g., groups status and power, group size, competence-relevant context) play a role in the observed pattern. It is thus not possible to evaluate the extent to which dimensional compensation emerges differently at the direct level and in indirect measures and whether the vantage point of the judge influences this pattern.

To our knowledge, only one experiment looked at dimensional compensation at the indirect level. Kervyn, Yzerbyt, and Judd (2011) presented their participants with pictures illustrating various behaviors allegedly performed by members of one of two novel groups differing on one of the two fundamental dimensions. For each picture, participants had to select one of four descriptions varying in degree of abstractness, from very concrete to very abstract. According to the Linguistic Category Model (LCM; Semin & Fiedler, 1988), judges should rate behaviors in line with their expectations in more abstract than concrete terms

(Wigboldus, Semin, & Spears, 2000). As predicted, behaviors supporting dimensional compensation led to more abstract terms than the other behaviors, suggesting stronger trait inferences. Even though this research provides a first look into dimensional compensation on indirect measures, its message remains tentative and requires further research. Beyond checking whether these findings with observers replicate, it would also be important to look at what happens when people belong to one of the two groups, as their underlying motivations to dimensionally compensate may differ.

1.3 The present research

The present research aimed to examine the above questions in a more systematic manner. We used indirect measures (i.e., based on reaction times) and relied on a minimal group paradigm to manipulate participants' social position. In two experiments, participants read about two fictitious groups (i.e., the "green" and "blue" group). Groups initially differed on one of the two fundamental dimensions (i.e., either competence or warmth in Experiment 1; competence in Experiment 2). Whereas participants were external observers in Experiment 1, they believed to be members of one of the two groups in Experiment 2. We assessed participants' impressions about these groups at both the direct and indirect level to help disentangle the motives that give rise to dimensional compensation by capitalizing on responses believed to be less sensitive to social desirability and normative pressures (Greenwald & Lai, 2020; Nosek, 2005; for a critical evaluation, see Corneille & Hütter, 2020).

We used the Brief Implicit Association Test (BIAT; Sriram & Greenwald, 2009) as this measure offers several benefits over other reaction-time-based measures. First, it has good psychometric properties together with a high versatility to measure stereotypes of all kinds (Nosek, Bar-Anan, Sriram, Axt, & Greenwald, 2014), including those revolving around warmth and competence (e.g., Imhoff et al., 2013). Secondly, it has proven to be a valid and sensitive tool to assess implicit ingroup favoritism in a minimal group paradigm (Yang, Shi, Luo, Shi, & Cai, 2014). Thirdly, the structure of the task fosters a comparative context because the to-be-classified targets in each block include stimuli pertaining to both groups and dimensions (Gawronski & De Houwer, 2014). This is an important feature given that a comparison context between the social targets on the two dimensions simultaneously facilitates the emergence of a compensation effect (Judd, Garcia-Marques, & Yzerbyt, 2019; Judd et al., 2005). Finally, it is considerably shorter than the traditional Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998).

Compared to real groups that bring about a vast array of emotions and attitudes that are hardly malleable, fictitious groups ensure that a compensation effect emerges because participants did dimensionally compensate. They also greatly facilitate the manipulation of the vantage point of participants (as high- or low-competence group members or external observers).

In Experiment 1, we expected observers to compensate at both the direct and indirect level, in line with several compensation studies (see Yzerbyt, 2018). Such a pattern is

consistent with the idea that dimensional compensation among observers is not merely driven by self-presentation concerns but rather by justice concerns that do not only manifest publicly but that have also been internalized and should thus show up at an indirect level as well (Jost et al., 2002; Kay & Jost, 2003; Yzerbyt et al., 2008).

A different picture should emerge in Experiment 2 in which participants believed to be members of one of the two novel groups. Turning to low-status, i.e., low-competence, group members first, self-affirmation should be especially high on their agenda because of their threatened social identity (Tajfel, 1981; Tajfel & Turner, 1979). In line with previous studies (e.g., Cambon & Yzerbyt, 2018; Judd et al., 2005; Yzerbyt & Cambon, 2017), we predicted a compensatory pattern (i.e., social creativity) on direct measures. Such a pattern allows members of disadvantaged groups to build on the malleability of the warmth dimensions in order to affirm some positivity as a means to deal with the rather objective disadvantage on competence (Cambon et al., 2015; Yzerbyt & Cambon, 2017). Given the presence of these self-affirmation concerns, we expected this pattern to show on indirect measures as well.

As for of high-status, i.e., high-competence, group members, and in line with previous studies (e.g., Cambon & Yzerbyt, 2018; Cambon et al., 2015; Judd et al., 2005; Yzerbyt & Cambon, 2017), we expected self-presentation concerns to result in explicit dimensional compensation. In sharp contrast, we expected these group members' responses on indirect measures to tell a different story: Self-affirmation concerns would lead them to acknowledge their upper hand on competence and, more importantly, to dispute the superiority of low-competence group members on warmth.

For both experiments, the data and R scripts to analyze the data are publicly available on Open Science Framework (https://osf.io/uwrma/?view_only=a29ec230f5ac4d2cb62e2a83022b34a0).

2. Experiment 1

In Experiment 1, participants were observers and learned about two novel groups, namely the “blue group” and the “green group”, that initially differed on one of the two fundamental dimensions (i.e., the manipulated dimension) but were equal on the other (i.e., the unmanipulated dimension). We expected uninvolved observers to display dimensional compensation both on direct and indirect measures.

2.1 Method

Participants and design. We conducted a power analysis with the PANGEA web app (<https://jakewestfall.shinyapps.io/pangea/>) to estimate the sample size required to find a small effect size ($d = 0.20$) for the main effect of target group in the context of a 2×2 design with one factor varying within participants and the other between them. The analysis revealed that 351 participants would be required to achieve a power of 90% with a .05 alpha two-tailed criterion. Because the dropout is usually higher in crowdsourcing platforms than in laboratory

settings (Peer, Brandimarte, Samat, & Acquisti, 2017), we decided to collect the data from 500 participants.

We recruited 511 participants from the United States using Prolific Academic—an online crowdsourcing platform (<https://www.prolific.co/>). Participants who had an invariant response style on the self-report measures ($n = 23$), responded faster than 300ms to more than 10% of their trials ($n = 31$), or showed an error rate greater than 30% ($n = 43$) during the indirect measure were excluded from the analyses, leaving a sample of 414 ($M_{\text{age}} = 33.65$, $SD_{\text{age}} = 11.85$; 226 males). The experiment involved a 2 (Manipulated Dimension: Competence vs. warmth) $\times$ 2 (Target Group: High vs. low group on the manipulated dimension) $\times$ 2 (Judged Dimension: Manipulated vs. non-manipulated) design, with the first factor varying between participants and the others within them.

Measures. Participants' impressions about the two groups were assessed by direct (i.e., self-reported ratings of traits) and indirect (i.e., BIAT) measures.

Direct measures. We evaluated participants' impressions about the target groups by means of traits rated on a 7-point scale (0 = strongly disagree; 6 = strongly agree). Efficient, competent, and hardworking stood for the competence dimension ($\alpha = .87-.89$ for the fictitious blue and green group, respectively), whereas warm, sociable, and friendly tapped the warmth dimension ($\alpha = .92-.93$ for the fictitious blue and green group, respectively).

Indirect measures. We assessed the association between target group and dimension (i.e., competence or warmth) by means of two Brief Implicit Association Tests (BIAT; Sriram & Greenwald, 2009). This reaction time task maps two focal categories onto a focal key, and two non-focal categories onto a non-focal key. Participants learned that they had to press as fast and accurately as possible the focal key whenever the stimuli belonged to either one of the focal categories and the non-focal key otherwise. Each BIAT task comprised four blocks of 20 trials with the first two blocks being identical to the last two to increase reliability. The focal dimension remained constant (i.e., warmth or competence), whereas the focal category (i.e., "blue" or "green" group) was distinct for each block. For the warmth-BIAT ($\alpha = .49$; for similar levels of internal consistency see Imhoff et al., 2013), we used *warmth* as the focal attribute and *competence* as the non-focal attribute. The opposite held for the competence-BIAT ($\alpha = .50$). We computed a d-score for each BIAT (for scoring algorithm, see Greenwald, Nosek, & Banaji, 2003; Nosek et al., 2014), with a positive value indicating a stronger association between the advantaged vs. disadvantaged group and the dimension of interest. We used traits (i.e., warmth dimension: sociable, caring, friendly, and kind; competence dimension: capable, skilled, smart, and organized) and figures (i.e., four green and four blue face shaped figures presented during Step 1) as the to-be-classified targets.

Procedure. The experiment comprised three steps. Step 1 relied upon Judd et al.'s (2005) procedure whereby subjects had to form an impression about two fictitious groups—the green group and the blue group—based on a series of behaviors allegedly performed by their members. Twenty-four behaviors, 12 for each group, were adapted from this study (i.e., the accountability of the actor was made more explicit for some behaviors; e.g., we changed "X's electricity was turned off because the bill hadn't been paid" to "X's electricity was turned off

because he/she didn't paid the bill"). Judd et al. (2005) pretested each behavior to be either diagnostic of positive competence (e.g., "X wrote a little computer program that solved a tough calculus integration problem"), negative competence (e.g., "X's bicycle was stolen several times because he/she forgot to set the lock"), positive warmth (e.g., "X loves to be with other people"), or negative warmth (e.g., "X couldn't be bothered to give directions to a stranger").

Initial differences between groups on the manipulated dimension (e.g., competence) were created by assigning 6 positive diagnostic behaviors of the manipulated dimensions (e.g., positive competence) to the group high on the manipulated dimension (e.g., high-competence group), and by assigning 6 negative diagnostic behaviors of the manipulated dimensions (e.g., negative competence) to the group low on the manipulated dimension (e.g., low-competence group). Moreover, both groups received 2 positive and 2 negative behaviors related to the non-manipulated dimension (e.g., warmth). In other words, when competence was the manipulated dimension, the high-competence group was ascribed more positive and less negative competence-diagnostic behavior but an equal number of positive and negative warmth-diagnostic behaviors than the low-competence group. Conversely, when warmth was the manipulated dimension, the high-warmth group was ascribed more positive and less negative warmth-diagnostic behavior but an equal number of positive and negative competence-diagnostic behaviors than the low-warmth group. Each participant saw a random sample from the pool of behaviors within the constraints of the aforementioned ratio. We also randomized the identity of the high group on the manipulated dimension between participants (i.e., for some participants the high group on the manipulated dimension was the blue group and the green group was the low group on the manipulated dimension whereas for other participants it was the inverse).

Behaviors appeared one at a time and in random order at the center of the screen for 5500ms with an inter-stimulus delay of 400ms. Each of them came with a colored background rectangle as well as four colored face shaped figures to indicate group membership (blue vs. green). Participants were then paced a second time through the randomly sorted statements and had to read and classify them one by one according to their author's group membership. They then went through the entire list of behaviors a third time with the behaviors sorted by group identity (e.g., all the blue group statements came first). Next, participants had to convey in a few lines their general impression about each group in order to solidify their impression. In this last step, we randomized the identity of the first group (blue vs. green) across participants (e.g., some participants first provided their impression of the blue group and then for the green group whereas it was the reverse order for other participants).

In Step 2, we collected the indirect measures. Participants first completed a 20-trial practice BIAT with birds/angled as focal categories and mammals/curved as non-focal categories. Next, two BIATs tapped each fundamental dimension. We randomized the identity of the first group (blue vs. green group) and the first focal dimension (i.e., warmth vs. competence) across participants.

In Step 3, we collected the direct measures. We assessed participants' impression about each group by means of a series of traits ratings presented in a random order and covering the two fundamental dimensions of judgment. Finally, we collected demographic information and debriefed participants. For steps 1 and 2, we relied on Psytoolkit (Stoet, 2010, 2017). We implemented Step 3 using Qualtrics.

2.2 Results

Direct measures. We first examined participants' ratings on the manipulated dimension to assess the success of our manipulation. There was a main effect of Target Group on the score relative to the manipulated dimension, $F(1, 411) = 366.69, p < .001, \eta^2_p = .471$, such that the high group on the manipulated dimension ($M = 5.71, SD = 1.09$) received a higher score than the low group ($M = 3.79, SD = 1.38$) on this dimension. The Target Group $\times$ Manipulated Dimension interaction was not significant, $F(1, 411) = 0.35, p = .557, \eta^2_p = .001$, confirming that this difference was not affected by whether warmth or competence was the manipulated dimension.

Of more interest, and in line with our hypothesis, there was a main effect of Target Group on the score relative to the unmanipulated dimension, $F(1, 411) = 11.37, p < .001, \eta^2_p = .027$, such that the high group on the manipulated dimension ($M = 4.69, SD = 1.28$) received lower ratings than the low group ($M = 5.05, SD = 1.34$). Again, the Target Group $\times$ Manipulated Dimension interaction was not significant, $F(1, 411) = 0.58, p = .447, \eta^2_p = .001$, signaling that this difference was not affected by whether warmth or competence was the manipulated dimension. Taken together, these results signal the presence of a dimensional compensatory pattern on the direct measures (see Figure 1 and Table 1).

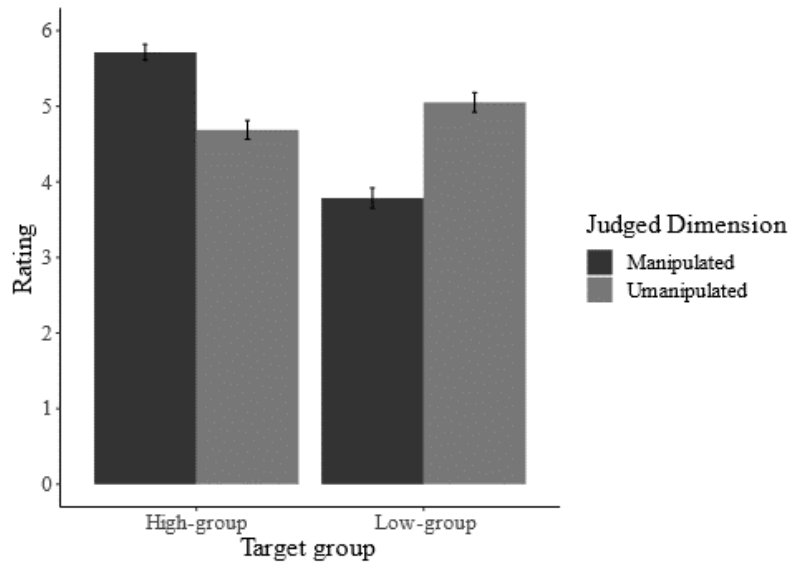


Figure 1. Ratings as a function of Target Group (high- vs. low-group on the manipulated dimension) and Judged Dimension (manipulated vs. unmanipulated). Error bars represent 95% confidence intervals.

Table 1

Ratings as a function of Target Group, Manipulated Dimension and Judged Dimension

Judged Dimension	Manipulated Dimension			
	Competence manipulated		Warmth manipulated	
	High-competence group	Low-competence group	High-warmth group	Low-warmth group
Competence	5.71 (0.94)	3.84 (1.25)	5.72 (1.23)	3.73 (1.49)
Warmth	4.51 (1.29)	4.96 (1.40)	4.87 (1.26)	5.15 (1.27)

Note. Standard deviations are in parenthesis.

Indirect measures. Again, we analyzed the d-scores for the manipulated vs. non-manipulated dimension separately. Specifically, we found a significant positive d-score for the manipulated dimension ($M = 0.11$, $SD = 0.38$), $F(1, 412) = 34.22$, $p < .001$, $\eta_p^2 = .077$. The Manipulated Dimension, i.e., whether we manipulated warmth or competence, did not qualify this main effect, $F(1, 412) = 0.29$, $p = .588$, $\eta_p^2 = .001$, confirming that participants

were faster at mapping the manipulated dimension with the high vs. low group on the manipulated dimension.

More importantly, we found a negative d-score for the non-manipulated dimension ($M = -0.04$, $SD = 0.39$), $F(1, 412) = 4.68$, $p < .05$, $\eta^2_p = .011$, which was not affected by the Manipulated Dimension, i.e., by whether warmth or competence was the manipulated dimension, $F(1, 412) = 0.14$, $p = .706$, $\eta^2_p < .001$, confirming that participants were faster at mapping the non-manipulated dimension with the low vs. high group on the manipulated dimension. As a whole, these results confirm the presence of a dimensional compensatory pattern at the indirect level (see Table 2).

Table 2

D-scores as a function of Manipulated Dimension and Dimension

Dimension	Manipulated Dimension	
	Competence manipulated	Warmth manipulated
Competence	0.12 (0.35)	-0.03 (0.42)
Warmth	-0.05 (0.36)	0.10 (0.40)

Note. A positive d-score indicates a stronger association between the high vs. low group on the manipulated dimension and the dimension of interest. Standard deviations are in parenthesis.

2.3 Discussion

Experiment 1 tested whether observers would show dimensional compensation both on direct and on indirect measures. Whether we manipulated competence or warmth, both participants' ratings on Likert-type scales and their reaction times via a BIAT confirmed the success of our manipulation such that the high group on the manipulated dimension was superior to the low group on this dimension. More importantly, and in line with our predictions, both types of measures revealed the presence of dimensional compensatory evaluations of the groups on the non-manipulated dimension.

One distinct feature of the present experiment is that participants were observers rather than members of either group. The question is whether this pattern would replicate for participants who see themselves as members of one of the two groups involved in the comparison. Admittedly, and building upon Yzerbyt and Cambon's (2017) findings that reality constraints play a larger role in judgments of competence than in judgments of warmth, one would expect all participants in this situation to acknowledge the superiority of the high-competence group. Moreover, this differential should emerge on both direct and

indirect measures. In sharp contrast, judgments of warmth should be more open to disagreement.

Specifically, self-affirmation concerns on the part of low-competence group members, coupled with the lesser flexibility of competence, would generate dimensional compensation on warmth. As for high-competence group members, and in line with findings showing that members of more privileged groups not only derogate stigmatized groups on competence but also tend to question their superiority on warmth (Clément-Guillotin et al., 2018; Rohmer & Louvet, 2012, 2018), the more subjective nature of warmth rating should render dimensional compensation less likely. That is, when it comes to the dimension of warmth, high-competence group members' magnanimity judgments should be apparent on direct measures but less so on indirect measures. We designed Experiment 2 to test the difference in pattern for members of low- and high-competence groups.

3. Experiment 2

In Experiment 2, we assigned participants to the high- or the low-competence group. We only manipulated the competence dimension because previous studies have shown that the compensation effect is usually stronger than when warmth is manipulated (e.g., Judd et al., 2005; Kervyn et al., 2011; Yzerbyt et al., 2008). Although this asymmetry may be accounted for by the primacy of warmth over competence (see Abele & Wojciszke, 2007; Fiske, Cuddy, & Glick, 2007), an additional factor resides in the constraints imposed by the more "objective" quality of competence over warmth (Yzerbyt & Cambon, 2017). Indeed, it is generally difficult to deny the reality of a power, status, or resource differential between groups and this holds even for low-competence group members. In contrast, warmth is more "negotiable", and all parties involved can claim a high level on this dimension. In light of this reasoning, we deemed it important to start by manipulating the dimensions that is most likely to remain undisputed and by checking for the emergence of differences between direct and indirect measures on the more malleable dimension.

Our decision to manipulate competence thus works in favor of our hypothesis for the low-competence group members but against it for the high-competence group members. Indeed, we expected participants to face a different set of motivations that would give rise to a compensation effect as a function of their distinct social status as high- or low-competence group members. Based on previous research (e.g., Cambon & Yzerbyt, 2018; Rohmer & Louvet, 2012, 2018; Yzerbyt & Cambon, 2017), we predicted both groups to display a dimensional compensation pattern on a direct level. At the indirect level, we expected high- and low-competence group members to acknowledge the superiority of the high-competence group on competence. In contrast, we predicted that participants would diverge with respect to warmth, such that low-competence group members would preferentially associate warmth dimension with their group but that high-competence group members would not go along with such a characterization.

3.1 Method

Participants and design. As in Experiment 1, we conducted a power analysis with the PANGEA web app to estimate the sample size required. Because previous studies on dimensional compensation have shown that ingroup members tend to exaggerate the difference on their favored dimension compared to external observers (Judd et al., 2005; Yzerbyt et al., 2005), we expected a medium effect size ($d = 0.50$) for Target Group $\times$ Group Membership interaction. The analysis revealed that 94 participants would be required to achieve a power of 90% with a .05 alpha two-tailed criterion. We decided to involve about 100 participants.

Ninety-five participants from a large European university took part in this experiment in exchange for credits. We applied the same exclusion criteria from Experiment 1. Two participants with an error rate greater than 30% during the indirect measure were excluded from the analyses, leaving a sample of 93 ($M_{\text{age}} = 20.41$, $SD_{\text{age}} = 2.69$; 83 females). The experiment involved a 2 (Group Membership: High- vs. low-competence group) $\times$ 2 (Target Group: High- vs. low-competence group) $\times$ 2 (Judged Dimension: Competence vs. warmth) design, with the first factor varying between participants and the others within them.

Measures. The measures collected in Experiment 2 were almost identical to those of Experiment 1, except that we added an identification measure at the end of the survey as a manipulation check and that the BIAT tasks did not repeat blocks in order to reduce the length of the experiment. The internal consistencies of the direct measures (warmth traits rating $\alpha = .89-.91$; competence traits rating $\alpha = .91-.92$; for the blue and green group, respectively) and indirect measures (warmth-BIAT $\alpha = .52$; competence-BIAT $\alpha = .50$) were similar to those in Experiment 1. We assessed identification with the ingroup versus outgroup on two 7-point scales (0 = not at all; 6 = very much) with the following instruction: “indicate your degree of identification with each of the two groups”. We computed an overall identification score by subtracting outgroup from ingroup identification.

Procedure. The experiment took place in a laboratory at the university. We manipulated participants’ membership using a bogus personality test similar to Judd et al. (2005, Study 5). Participants completed a dot estimation task in which they had to indicate, over the course of several trials, whether there were more yellow or orange dots on the screen. After completion, they learned that the task was a new unobtrusive personality test aimed to categorizing people in two groups, namely the “blues” and “greens”. The remaining part of the procedure was the same as in Experiment 1.

3.2 Results

Identification with the ingroup. Confirming the success of our group membership manipulation, ingroup identification was positive ($M = 0.45$, $SD = 1.58$) and significant, $F(1, 91) = 7.46$, $p < .01$, $\eta^2_p = .076$, and did not differ as a function of group membership, $F(1, 91) = 0.05$, $p = .818$, $\eta^2_p = .001$, suggesting that both high- and low-competence group members felt more identified with their ingroup than with the outgroup.

Direct measures. Again, we first examined participants' ratings on the manipulated dimension (i.e., competence). Confirming the success of our manipulation, there was a main effect of Target Group on the competence ratings, $F(1, 91) = 109.81$, $p < .001$, $\eta^2_p = .547$, such that participants judged the high-competence group ($M = 5.09$, $SD = 0.87$) more competent than the low-competence group ($M = 3.17$, $SD = 1.33$). The Target Group $\times$ Group Membership interaction was not significant, $F(1, 91) = 0.19$, $p = .663$, $\eta^2_p = .002$, suggesting that this difference was not affected by Group Membership.

More importantly, and in line with our prediction of dimensional compensation on direct measures, there was a main effect of Target Group on the warmth ratings, $F(1, 91) = 30.25$, $p < .001$, $\eta^2_p = .249$, such that the high-competence group ($M = 3.11$, $SD = 1.40$) came across as less warm than the low-competence group ($M = 4.48$, $SD = 1.30$). Again, the Target Group $\times$ Group Membership interaction was not significant, $F(1, 91) = 1.47$, $p = .229$, $\eta^2_p = .016$, suggesting that this difference was not affected by Group Membership. These results confirm that both high- and low-competence group members dimensionally compensated on direct measures (see Figure 2 and Table 3).

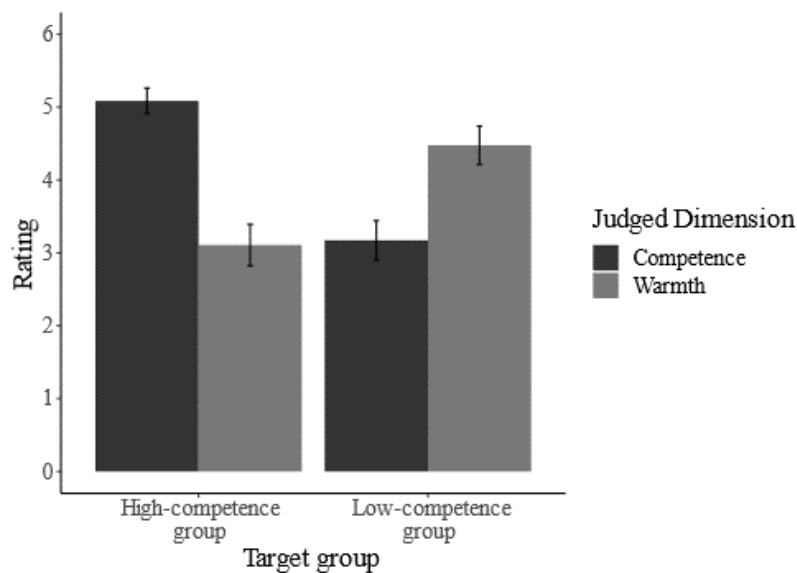


Figure 2. Ratings as a function of Target Group (high- vs. low-competence group) and Judged Dimension (competence vs. warmth). Error bars represent 95% confidence intervals.

Table 3

Ratings as a function of Target Group, Group Membership and Judged Dimension

Judged Dimension	Group Membership			
	High-competence group members		Low-competence group members	
	High-competence group	Low-competence group	High-warmth group	Low-warmth group
Competence	5.05 (0.73)	3.22 (1.27)	5.12 (0.99)	3.13 (1.40)
Warmth	3.25 (1.27)	4.32 (1.39)	2.96 (1.51)	4.63 (1.20)

Note. Standard deviations are in parenthesis.

Indirect measures. As in Experiment 1, we analyzed the d-score for the competence vs. warmth dimension separately. The positive d-score for the competence dimension ($M = 0.15$, $SD = 0.55$) confirmed that participants were faster at mapping the high-competence group than the low-competence group with the competence dimension, $F(1, 91) = 6.75$, $p < .05$, $\eta_p^2 = .069$, and this difference was not affected by Group Membership, $F(1, 91) = 1.12$, $p = .293$, $\eta_p^2 = .012$.

Turning to the warmth dimension, the negative d-score ($M = -0.07$, $SD = 0.57$) was not significant, $F(1, 91) = 1.32$, $p = .254$, $\eta_p^2 = .014$. Importantly, and in line with our hypothesis, there was a significant effect of Group Membership, $F(1, 91) = 4.58$, $p < .05$, $\eta_p^2 = .048$, revealing that high- and low-competence group members disagreed as to the ascription of warmth to the groups. As expected, this score was significantly negative for low-competence group members ($M = -0.19$, $SD = 0.51$), $F(1, 91) = 5.46$, $p < .05$, $\eta_p^2 = .057$, but not for high-competence group members ($M = 0.06$, $SD = 0.61$), $F(1, 91) = 0.49$, $p = .487$, $\eta_p^2 = .005$. In other words, low-competence group members saw warmth as more associated with the low-competence group (i.e., ingroup) than with the high-competence group (i.e., outgroup) whereas no preferential association emerged among high-competence group members (see Table 4).

Table 4

D-scores as a function of Group Membership and Judged Dimension

Dimension	Group Membership	
	High-competence group members	Low-competence group members
Competence	0.21 (0.53)	0.09 (0.57)
Warmth	0.06 (0.61)	-0.19 (0.51)

Note. A positive d-score indicates a stronger association between the high- vs. low-competence group and the dimension of interest. Standard deviations are in parenthesis.

3.3 Discussion

In Experiment 2, we tested whether dimensional compensation would take place for participants who find themselves to be members of the target groups rather than mere observers. We manipulated participants' membership via a bogus personality test and randomly assigned them to either the high- or low-competence group, while both groups were ambivalent on the warmth dimension.

As expected, a compensation pattern emerged at the direct level for both high- and low-competence group members. This pattern confirms earlier findings that relied on existing groups (e.g., Cambon & Yzerbyt, 2016; Yzerbyt et al., 2005) or minimal groups (e.g., Cambon et al., 2015; Judd et al., 2005). Crucially, we conjectured that dimensional compensation would emerge at the indirect level only for low- but not high-competence group members. In line with our hypothesis, high-competence group members did not concede their inferiority on the warmth dimension. In contrast, and in agreement with our prediction, low-competence group members exhibited dimensional compensation by showing ingroup bias on warmth.

These data confirm the presence of a compensation pattern for low-competence group members on both direct and indirect measures. This was not the case for high-competence group members, who proved reluctant to concede outgroup superiority on warmth on indirect measures even though they professed dimensional compensation on direct measures.

4. General Discussion

An increasing body of research has established the robustness of dimensional compensation, i.e., a negative relationship between the two fundamental dimensions when judging two social targets (for a recent review, see Yzerbyt, 2018). As of today, the available empirical evidence rests almost exclusively on direct measures such as traits ratings. To the

extent that dimensional compensation may in some cases reflect a response strategy fueled by self-presentation concerns to cope with normative pressures (Cambon & Yzerbyt, 2018; Cambon et al., 2015; Yzerbyt & Cambon, 2017), the sole use of direct measures remains a limitation. In this context, the present research aimed to test dimensional compensation on both self-reports, i.e., direct, as well as less controlled reaction-time-based, indeed indirect, measures (Greenwald & Lai, 2020; Nosek, 2005). Additionally, we examined whether judges' social position as observers or as group members shapes people's responses.

In Experiment 1, participants were third-party observers and read about two novel groups that initially differed only on one of the two dimensions, i.e., the manipulated dimension. Not surprisingly, participants ascribed higher (lower) ratings to the high (low) group on the manipulated dimension. Crucially, however, they manifested lower (higher) ratings to the high (low) group on the non-manipulated dimension. Furthermore, this compensation pattern held whether we manipulated warmth or competence. These findings at the direct level replicate earlier work on fictitious groups (e.g., Judd et al., 2005; Kervyn, Yzerbyt, Judd, & Nunes, 2009; Yzerbyt et al., 2008) and further supports the idea that this effect goes beyond the mere covariation between the fundamental dimensions from existing stereotypes.

More importantly, the predicted compensation pattern also emerged at the indirect level, in line with Kervyn et al.' (2011) earlier findings using the Linguistic Category Model (Semin & Fiedler, 1988). Because such indirect responses are less likely to fall under participant's intention or control (Gawronski & De Houwer, 2014), the obtained findings suggest that dimensional compensation among external observers is likely driven by justice or harmony concerns (rather than, for instance, self-presentation concerns) that may have been internalized. In other words, their judgments seem consistent with a need of justice and harmony that promotes a hydraulic relationship between warmth and competence when forming an impression about social targets that are initially asymmetrical on one of the two dimensions.

In Experiment 2, participants were either members of the high- or low-competence group. In agreement with previous investigations (e.g., Cambon et al., 2015; Judd et al., 2005; Yzerbyt & Cambon, 2017), a compensation pattern emerged at the direct level. Not only did participants from both groups acknowledge the superiority of the high-competence group over the low-competence group on the competence dimension but the reverse pattern also emerged for the non-manipulated warmth dimension. Crucially, we expected the driving forces behind this compensatory pattern to be essentially different between the two groups, producing distinct response at the indirect level. Specifically, we reasoned that members of both groups would strive to achieve positive distinctiveness (i.e., self-affirmation concerns) as previous studies have found an evaluative preference for the ingroup (i.e., ingroup favoritism) at an indirect level even with minimal groups (e.g., Greenwald, Pickrell, & Farnham, 2002; Otten & Moskowitz, 2000; Pinter & Greenwald, 2004). However, we posited that the differential reality constraints of the two dimensions and the existence of normative pressures should join forces to give way to different indirect judgments as a function of group membership.

The “more objective” and “less negotiable” nature of competence compared to warmth (Yzerbyt & Cambon, 2017) makes it difficult for low-competence group members to dispute the outgroup superiority on this dimension. Consequently, they would likely seek to redeem positive distinctiveness by shining on the status-irrelevant dimension (i.e., warmth). This ought to produce a compensation effect at both the direct and indirect level. The present results support our hypothesis. Of course, to the extent that research has shown that even low-status group members may endorse stereotypes that disfavor them and perpetuate the *status quo* (Jost et al., 2002; Kay & Jost, 2003), we cannot totally exclude the possibility that internalized justice concerns shaped this compensatory effect.

More importantly, we conjectured that the situation would likely be different for high-status group members. Although these group members would be tempted to claim their superiority on both dimensions (i.e., self-affirmation concerns), such a behavior would infringe the non-discrimination norm (Jeffries et al., 2012). Actually, several pieces of evidence emphasize the presence of self-presentation concerns among high-status group members (e.g., Cambon & Yzerbyt, 2018; Owuamalam et al., 2016; Yzerbyt & Cambon, 2017). To cope with these normative pressures, high-status group members may readily concede some advantage to the low-status group on the dimension they care the less about (i.e., warmth) even though they may not fully adhere to the idea. Experiment 2 confirmed that they dimensionally compensated at the direct level. In contrast, and in line with our rationale, we found no outgroup bias on the warmth dimension at the indirect level. This incongruence between direct and indirect measures suggests that dimensional compensation as manifested by advantaged members may indeed result from self-presentation concerns.

The present experiments shed interesting light on earlier research (see Rohmer & Louvet, 2012). Some of these efforts attributed the discrepancies between direct and indirect measures to the social norms that protect the evaluated groups, i.e., it may be less acceptable to express criticism towards disabled people than towards lawyers. Other invoked the difficulty to access in memory the different social categories used in the indirect measure, i.e., the social category of lawyers may be more accessible in memory than that of disabled people. Clearly, the use of unknown groups allows ruling out these alternative explanations. At the same time, because these categories are novel, they might also make participants more thoughtful about their responses. Indeed, the IAT might alert participants of what is being measured (Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013). Future research may want to investigate dimensional compensation with novel groups using alternative, possibly less controllable, indirect measures.

As such, the present work also provides a new framework to interpret past findings. For instance, Carlsson and Björklund (2010; Study 1), and Imhoff et al.’ (2013) participants (i.e., students) can be seen as observers who, just as in Experiment 1, showed direct and indirect mixed stereotypes about two outgroups (i.e., preschool teachers and lawyers, and nursery teachers and managers, respectively). In stark contrast, and just like in Experiment 2, Rohmer and Louvet’s (2012, 2018; see also Clément-Guillotin et al., 2018) participants (i.e., students) were in fact members of the high-status group (i.e., non-disabled people) advocating mixed stereotypes while portraying themselves as superior to the low-status outgroup (i.e., disabled

people) on both dimensions under less direct measures. Interestingly, when participants were members of the disadvantaged group (i.e., preschool teachers as in Carlsson & Björklund, 2010; women as in Ebert et al., 2014), they dimensionally compensated on direct measures but favored their ingroup on indirect measures regardless of the dimension. Although we predicted and found an indirect compensation pattern for low-competence group members, one could speculate that the competence asymmetry between existing groups is less strict and thus more malleable than with artificial groups produced under an experimental context. Future studies need to be conducted to nail down these boundary conditions.

One possible caveat of the present research is that, although Experiment 1 evidenced that direct and indirect dimensional compensation held whether competence or warmth was the manipulated dimension, we only manipulated competence in Experiment 2. Actually, several studies reported a weaker effect when warmth is the manipulated dimension (e.g., Judd et al., 2005; Kervyn et al., 2011; Yzerbyt et al., 2008). Clearly, it would be interesting for further research to investigate the direct and indirect response pattern when manipulating warmth rather than competence.

Although the current research offers valuable insights about the motivations that underlie dimensional compensation, our results constitute an initial step, as they do not allow disentangling every motive at work. This is because the compensation pattern may be accounted by more than one motive. Thus, new research could examine each motive in more detail. For instance, testing whether dimensional compensation among low-status group derives from self-affirmation concerns, justice concerns, or both. Similarly, research could manipulate the intergroup relations (e.g., competitive vs cooperative) to test the impact of the desire to promote harmonious intergroup relations on the emergence of dimensional compensatory responses.

An additional limitation concerns the spectrum of traits that we used as stimuli. Different lines of research suggest that the two fundamental dimension of social perception could be broken down into facets: ability and assertiveness for the competence/agency dimension, and sociability and morality for the warmth/communion dimension (Abele et al., 2008; Abele et al., 2016). Our experimental material comprised the facets of capability and sociability, but further studies should seek to generalize our results to the other facets (Abele, Ellemers, Fiske, Koch, & Yzerbyt, in press).

As a final remark, the reported effect sizes of the indirect measures may come as statistically small. At the same time, numerous studies have shown that such effects can have a societally large and significant impact as they may affect a large number of people or affect a single individual on a regular basis (Greenwald, Banaji, & Nosek, 2015; but see Oswald et al., 2013; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015). Future work should allow clarifying these issues by examining the impact of compensation on a variety of phenomena down the line.

To sum up, the current efforts contribute to extend our understanding of the dimensional compensation effect. To our knowledge, the present experiments constitute the first attempt to assess dimensional compensation between two fundamental dimensions on both direct and

indirect measures and to do so with participants either as observers or as members of fictitious groups. Our results emphasize the discrepancies between indirect and direct responses as a function of people's social position and shed new light on the different motives underlying dimensional compensation. As such, they contribute to enriching the research on the way people navigate their social world.

5. Open Practices Statement

The data and R scripts to analyze the data are publicly available on Open Science Framework (https://osf.io/uwrma/?view_only=a29ec230f5ac4d2cb62e2a83022b34a0).

6. Notes

¹ We used the term *indirect* following the recommendations from Corneille and Hütter (2020) to discontinue the usage of the *implicit* terminology to avoid conceptual ambiguities associated with this nomenclature.

7. References

- Abele, A. E., Cuddy, A. J., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, 38(7), 1063–1065. doi: 10.1002/ejsp.574
- Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (in press). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*. doi: 10.31234/osf.io/b5nq6
- Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness: Communion with warmth and morality. *Frontiers in Psychology*, 7, 1–17. doi: 10.31234/osf.io/b5nq6
- Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, 93(5), 751–763. doi: 10.1037/0022-3514.93.5.751
- Bergsieker, H. B., Leslie, L. M., Constantine, V. S., & Fiske, S. T. (2012). Stereotyping by omission: Eliminate the negative, accentuate the positive. *Journal of Personality and Social Psychology*, 102(6), 1214–1238. doi: 10.1037/a0027717
- Cambon, L., & Yzerbyt, V. Y. (2016). Compensation is for real: Evidence from existing groups in the context of actual relations. *Group Processes & Intergroup Relations*, 20(6), 745–756. doi: 10.1177/1368430215625782
- Cambon, L., & Yzerbyt, V. Y. (2018). Two routes toward compensation: An investigation into the mechanisms of compensation for high-and low-status groups. *Journal of Experimental Social Psychology*, 77, 24–35. doi: 10.1016/j.jesp.2018.04.002

- Cambon, L., Yzerbyt, V. Y., & Yakimova, S. (2015). Compensation in intergroup relations: An investigation of its structural and strategic foundations. *British Journal of Social Psychology, 54*(1), 140–158. doi: 10.1111/bjso.12067
- Carlsson, R., & Björklund, F. (2010). Implicit stereotype content: Mixed stereotypes can be measured with the implicit association test. *Social Psychology, 41*(4), 213–222. doi: 10.1027/1864-9335/a000029
- Clément-Guillotin, C., Rohmer, O., Forestier, C., Guillotin, P., Deshayes, M., & d'Arripe-Longueville, F. (2018). Implicit and explicit stereotype content associated with people with physical disability: Does sport change anything? *Psychology of Sport and Exercise, 38*, 192–201. doi: 10.1016/j.psychsport.2018.06.014
- 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, 24*(3), 212–232. doi: 10.1037/10.1177/1088868320911325
- Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology, 82*(3), 359–378. doi: 10.1037/0022-3514.82.3.359
- Dambrun, M., & Guimond, S. (2004). Implicit and explicit measures of prejudice and stereotyping: Do they assess the same underlying knowledge structure? *European Journal of Social Psychology, 34*(6), 663–676. doi: 10.1002/ejsp.223
- Durante, F., Fiske, S. T., Gelfand, M. J., Crippa, F., Suttora, C., Stillwell, A., . . . Teymoori, A. (2017). Ambivalent stereotypes link to peace, conflict, and inequality across 38 nations. *Proceedings of the National Academy of Sciences, 114*(14), 669–674. <https://doi.org/10.1073/pnas.1611874114>
- Durante, F., Fiske, S. T., Kervyn, N., Cuddy, A. J., Akande, A. D., Adetoun, B. E., ... Storari, C. C. (2013). Nations' income inequality predicts ambivalence in stereotype content: How societies mind the gap. *British Journal of Social Psychology, 52*(4), 726 –746. <http://dx.doi.org/10.1111/bjso.12005>
- Ebert, I. D., Steffens, M. C., & Kroth, A. (2014). Warm, but maybe not so competent? — Contemporary implicit stereotypes of women and men in Germany. *Sex roles, 70*(9-10), 359–375. doi: 10.1007/s11199-014-0369-5
- Fiske, S. T. (2015). Intergroup biases: A focus on stereotype content. *Current Opinion in Behavioral Sciences, 3*, 45–50. doi: 10.1016/j.cobeha.2015.01.010
- Fiske, S. T. (2018). Stereotype Content: Warmth and Competence Endure. *Current Directions in Psychological Science, 27*(2), 67–73. doi: 10.1177/0963721417738825
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences, 11*(2), 77–83. doi: 10.1016/j.tics.2006.11.005
- 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

- Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 283–310). New York, NY, USA: Cambridge University Press. doi: 10.1017/CBO9780511996481.016
- Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, 15(4), 553–561. doi: 10.1037/pspa0000016
- Greenwald, A. G., & Lai, C. K. (2020). Implicit Social Cognition. *Annual Review of Psychology*, 71(1), 1–25. doi: 10.1146/annurev-psych-010419-050837
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (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
- Greenwald, A. G., Pickrell, J. E., & Farnham, S. D. (2002). Implicit partisanship: Taking sides for no reason. *Journal of Personality and Social Psychology*, 83(2), 367–379. <https://doi.org/10.1037/0022-3514.83.2.367>
- Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in psychology*, 4, 386. doi: 10.3389/fpsyg.2013.00386
- Jeffries, C. H., Hornsey, M. J., Sutton, R. M., Douglas, K. M., & Bain, P. G. (2012). The David and Goliath principle. Cultural, ideological, and attitudinal underpinnings of the normative protection of low-status groups from criticism. *Personality and Social Psychology Bulletin*, 38(8), 1053–1065. doi: 10.1177/0146167212444454.
- 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., & Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: consequences for specific and diffuse forms of system justification. *Journal of Personality and Social Psychology*, 88(3), 498–509. doi: 10.1037/0022-3514.88.3.498
- Jost, J. T., Kivetz, Y., Rubini, M., Guermendi, G., & Mosso, C. (2005). System-justifying functions of complementary regional and ethnic stereotypes: Cross-national evidence. *Social Justice Research*, 18(3), 305–333. doi: 10.1007/s11211-005-6827-z
- Jost, J. T., Pelham, B. W., & Carvallo, M. R. (2002). Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups. *Journal of Experimental Social Psychology*, 38(6), 586–602. doi: 10.1016/S0022-1031(02)00505-X
- Judd, C. M., Garcia-Marques, T., & Yzerbyt, V. Y. (2019). The complexity of relations between dimensions of social perception: Decomposing bivariate associations with crossed random factors. *Journal of Experimental Social Psychology*, 82, 200–207. <https://doi.org/10.1016/j.jesp.2019.01.008>

- Judd, C. M., James-Hawkins, L., Yzerbyt, V. Y., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, 89(6), 899–913. doi: 10.1037/0022-3514.89.6.899
- Kay, A. C., & Jost, J. T. (2003). Complementary justice: effects of "poor but happy" and "poor but honest" stereotype exemplars on system justification and implicit activation of the justice motive. *Journal of Personality and Social Psychology*, 85(5), 823–837. doi: 10.1037/0022-3514.85.5.823
- Kay, A. C., Jost, J. T., Mandisodza, A. N., Sherman, S. J., Petrocelli, J. V., & Johnson, A. L. (2007). Panglossian ideology in the service of system justification: How complementary stereotypes help us to rationalize inequality. *Advances in experimental social psychology*, 39, 305–358. doi: 10.1016/S0065-2601(06)39006-5
- Kervyn, N., Yzerbyt, V. Y., Demoulin, S., & Judd, C. M. (2008). Competence and warmth in context: The compensatory nature of stereotypic views of national groups. *European Journal of Social Psychology*, 38(7), 1175–1183. doi: 10.1002/ejsp.526
- Kervyn, N., Yzerbyt, V. Y., & Judd, C. M. (2010). Compensation between warmth and competence: Antecedents and consequences of a negative relation between the two fundamental dimensions of social perception. *European Review of Social Psychology*, 21(1), 155–187. doi: 10.1080/13546805.2010.517997
- Kervyn, N., Yzerbyt, V. Y., & Judd, C. M. (2011). When compensation guides inferences: Indirect and implicit measures of the compensation effect. *European Journal of Social Psychology*, 41(2), 144–150. doi: 10.1002/ejsp.748
- Kervyn, N., Yzerbyt, V. Y., Judd, C. M., & Nunes, A. (2009). A question of compensation: the social life of the fundamental dimensions of social perception. *Journal of Personality and Social Psychology*, 96(4), 828–842. doi: 10.1037/a0013320
- Lemaine, G. (1974). Social differentiation and social originality. *European Journal of Social Psychology*, 4(1), 17–52. doi: 10.1002/ejsp.2420040103
- Mummendey, A., & Schreiber, H.-J. (1983). Better or just different? Positive social identity by discrimination against, or by differentiation from outgroups. *European Journal of Social Psychology*, 13(4), 389–397. doi: 10.1002/ejsp.2420130406
- Mussweiler, T. (2007). Assimilation and Contrast as Comparison Effects: A Selective Accessibility Model. In D. A. Stapel & J. Suls (Eds.), *Assimilation and contrast in social psychology* (p. 165–185). Psychology Press.
- Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, 134(4), 565–584. doi: 10.1037/0096-3445.134.4.565
- Nosek, B. A., Bar-Anan, Y., Sriram, N., Axt, J., & Greenwald, A. G. (2014). Understanding and using the brief implicit association test: Recommended scoring procedures. *PLoS One*, 9(12), e110938. doi: 10.1371/journal.pone.0110938
- Oldmeadow, J., & Fiske, S. T. (2007). System-justifying ideologies moderate status = competence stereotypes: Roles for belief in a just world and social dominance orientation. *European Journal of Social Psychology*, 37(6), 1135–1148. doi: 10.1002/ejsp.428

- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, 105(2), 171–192. <https://doi.org/10.1037/a0032734>
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology*, 108(4), 553-561. doi: 10.1037/pspa0000016
- Otten, S., & Moskowitz, G. B. (2000). Evidence for implicit evaluative in-group bias: Affect-biased spontaneous trait inference in a minimal group paradigm. *Journal of Experimental Social Psychology*, 36(1), 77–89. <https://doi.org/10.1006/jesp.1999.1399>
- Owuamalam, C. K., Wong, K. X., & Rubin, M. (2016). Chubby but cheerful? Investigating the compensatory judgments of high, medium, and low status weight groups in Malaysia. *Cogent Psychology*, 3(1), 1188441. doi: 10.1080/23311908.2016.1188441
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 15–163. doi: 10.1016/j.jesp.2017.01.006
- Pinter, B., & Greenwald, A. G. (2004). Exploring Implicit Partisanship: Enigmatic (But Genuine) Group Identification and Attraction. *Group Processes & Intergroup Relations*, 7(3), 283–296. <https://doi.org/10.1177/1368430204046112>
- Rohmer, O., & Louvet, E. (2012). Implicit measures of the stereotype content associated with disability. *British Journal of Social Psychology*, 51(4), 732–740. doi: 10.1111/j.2044-8309.2011.02087.x
- Rohmer, O., & Louvet, E. (2018). Implicit stereotyping against people with disability. *Group Processes & Intergroup Relations*, 21(1), 127–140. doi: 10.1177/1368430216638536
- Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, 54(4), 558–568. doi: 558. 10.1037/0022-3514.54.4.558
- Son Hing, L. S., Bobocel, D. R., Zanna, M. P., Garcia, D. M., Gee, S. S., & Oraziatti, K. (2011). The merit of meritocracy. *Journal of Personality and Social Psychology*, 101, 433–450. doi: 10.1037/ a0024618
- Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology*, 56(4), 283–294. doi: 10.1027/1618-3169.56.4.283
- Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096–1104. doi: 10.3758/BRM.42.4.1096
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24–31. doi: 10.1177/0098628316677643
- Tajfel, H. (1981). *Human groups and social categories*. Cambridge: Cambridge University Press.

- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup relations. In W. G. Austin & S. Worchel (Eds.), *The psychology of intergroup relations*. Monterey, CA: Brooks-Cole.
- Vanbeselaere, N., Boen, F., van Avermaet, E., & Buelens, H. (2006). The Janus face of power in intergroup contexts: A further exploration of the noblesse oblige effect. *The Journal of Social Psychology*, 146(6), 685–699. doi: 10.3200/SOCP.146.6.685-699
- Wade, M. L., & Brewer, M. B. (2006). The structure of female subgroups: An exploration of ambivalent stereotypes. *Sex Roles*, 54(11-12), 753–765. doi: 10.1007/s11199-006-9043-x
- Wigboldus, D. H. J., Semin, G. R., & Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. *Journal of Personality and Social Psychology*, 78(1), 5–18. doi: 10.1037/0022-3514.78.1.5
- Yang, J., Shi, Y., Luo, Y. L. L., Shi, J., & Cai, H. (2014). The Brief Implicit Association Test is valid: Experimental evidence. *Social Cognition*, 32(5), 449–465. doi: 10.1521/soco.2014.32.5.449
- Yzerbyt, V. Y. (2016). Intergroup stereotyping. *Current Opinion in Psychology*, 11, 90–95. doi: 10.1016/j.copsyc.2016.06.009
- Yzerbyt, V. Y. (2018). The dimensional compensation model: Reality and strategic constraints on warmth and competence in intergroup perceptions. In A. E. Abele & B. Wojciszke (Eds.), *Agency and communion in social psychology* (pp. 126–141). Routledge. doi: 10.4324/9780203703663
- Yzerbyt, V. Y., & Cambon, L. (2017). The dynamics of compensation: When ingroup favoritism paves the way for outgroup praise. *Personality and Social Psychology Bulletin*, 43(5), 587–600. doi: 10.1177/0146167216689066
- Yzerbyt, V. Y., & Demoulin, S. (2010). Intergroup relations. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *The handbook of social psychology* (5th Edition, Vol. 2, pp. 1024–1083). Hoboken, NJ: Wiley. doi: 10.1002/9780470561119
- Yzerbyt, V. Y., Kervyn, N., & Judd, C. (2008). Compensation versus halo: The unique relations between the fundamental dimensions of social judgment. *Personality and Social Psychology Bulletin*, 34(8), 1110–1123. doi: 10.1177/0146167208318602
- Yzerbyt, V. Y., Provost, V., & Corneille, O. (2005). Not Competent but Warm... Really? Compensatory stereotypes in the French-speaking world. *Group Processes & Intergroup Relations*, 8(3), 291–308. doi: 10.1177/1368430205053944

Chapter 5

When you look competent but cold

Taking compensation at the face value

Previous research shows that stereotypes can distort the visual representation of groups in a top-down fashion. The present experiments tested if the compensation effect—the negative relationship that emerges between the social dimensions of warmth and competence when judging two social targets—would bias the visual representations of these targets in a compensatory way. We captured participants' near spontaneous facial prototypes of social targets by means of an unconstrained technique, namely the reverse correlation. In four experiments ($N = 869$), we find that the expectations of the facial content of two novel groups that differed on one of the two social dimensions are biased in a compensatory manner on the facial dimensions of trustworthiness, warmth and dominance but not competence. The present research opens new avenues by showing that compensation not only manifests itself on abstract ratings but that it also orients the visual representations of social targets.

Keywords: compensation, stereotype content model, social perception, face perception, reverse correlation, visual representations

Reference

Schmitz, M., Vanbeneden, A., & Yzerbyt, V. (under review). When you look competent but cold: Taking compensation at the face value. Manuscript under review in *European Journal of Social Psychology*.

When you look competent but cold

Taking compensation at the face value

1. Introduction

Dimensional compensation takes place when a target comes across as superior to another on one of the two fundamental dimensions of social perception, i.e., warmth or competence, but as inferior on the other dimension (Yzerbyt, Provost, & Corneille, 2005; for a recent review, see Yzerbyt, 2018). Notwithstanding the fact that this effect has proven robust at an explicit and verbal level, no study has tested if this phenomenon could also emerge in people's representation of faces. This is all the more important given the key role of facial cues in the context of social interactions and the top-down impact of stereotypes on facial representations (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). On this basis, we tested if compensation could bias the visual rendering of social targets in a compensatory manner. Importantly, we relied on an unconstrained and near spontaneous data-driven technique to capture participants' facial representations of fictitious target groups.

1.1 The dimensional compensation effect

Almost a century ago, Lippman (1922) referred to stereotypes as “pictures in our heads.” This conceptualization emphasizes the visual component of stereotypes while remaining closely related to the contemporary definition of stereotypes as beliefs or opinions about the characteristics, attributes or behaviors shared by individuals from the same group (Hilton & Von Hippel, 1996). A long tradition of research has shown that, although the collection of traits used to describe groups could in principle be vast and diverse, the content of stereotypes revolves around two fundamental dimensions (Abele & Wojciszke, 2007, 2014; Fiske, Cuddy, Glick, & Xu, 2002; Yzerbyt, 2016; for a recent review, see Abele, Ellemers, Fiske, Koch, & Yzerbyt, 2019). Different labels have been proposed for these two dimensions depending on the theoretical frameworks, and sub-dimensions or facets have also been identified (Abele, Cuddy, Judd, & Yzerbyt, 2008; Abele et al., 2016; Carrier, Louvet, Chauvin, & Rohmer, 2014). On the one hand, the warmth/communion dimension conveys information about the target's intention (i.e., “is it a friend or a foe?”) and comprises the sub-dimensions of friendliness (e.g., nice, friendly, and caring) and morality (e.g., honest, moral, and trustworthy). On the other, the competence/agency dimension is inferred from the target's capacity to concretize his or her intentions and encompasses the sub-dimensions of ability (e.g., skilled, efficient, and organized) and assertiveness (e.g., determined, ambitious, and self-confident).

Interestingly, most stereotypes came across as positively valenced on one dimension but negatively on the other—the so-called mixed or ambivalent stereotypes (e.g., gender stereotypes; Glick & Fiske, 2001). Yzerbyt et al. (2005) have suggested that such an

arrangement may stem from the compensation effect that characterizes the negative relationship between the two dimensions when judging two targets. For instance, in the context of national groups, Italians were perceived as warmer but less competent than Belgians (Kervyn, Yzerbyt, Demoulin, & Judd, 2008; see also Jost, Kivetz, Rubini, Guermendi, & Mosso, 2005). In addition, in a more controlled setting, when asked to judge two novel groups that initially differed only on one dimension (e.g., one group was initially presented as more competent than the other), participants compensated for this difference in the opposite direction on the other dimension (e.g., warmth; Judd, James-Hawkins, Yzerbyt, & Kashima, 2005).

Through the years, a large and growing body of research has shown not only the robustness of the compensation effect but also a series of boundary conditions for its emergence (e.g., Cambon & Yzerbyt, 2017; Judd et al., 2005; Kervyn, Bergsieker, Grignard, & Yzerbyt, 2016; Yzerbyt & Cambon, 2017; for a review, see Yzerbyt, 2018). Nonetheless, the bulk of the research has tested the compensatory relation between warmth and competence under constrained conditions, that is, participants only had a limited set of possible responses options that were established beforehand by the researchers (e.g., list of traits, labels, and possible answers to questions). Moreover, compensation was only tested at a verbal level, that is, by the means of traits' ratings (e.g., Judd, Garcia-Marques, & Yzerbyt, 2019), by having people choose specific content- and valence-oriented questions (Kervyn, Yzerbyt, Judd, & Nunes, 2009), select behavioral descriptions (Kervyn, Yzerbyt, & Judd, 2011), or react to associations between group labels and traits (i.e., Brief-IAT; Schmitz & Yzerbyt, *in press*). Hence, the question remains if non-verbal stimuli, such as visual information conveyed by faces, may also be biased in a compensatory way and under unconstrained conditions. This issue becomes even more relevant when considering the role and importance of facial information in everyday life social interactions (Imhoff, Woelki, Hanke, & Dotsch, 2013; Todorov et al., 2015).

1.2 Impressions from faces

Faces are one of the most common visual and rich components of our social environment and are imbued with essential information for navigating everyday life social interactions. It is therefore not surprising that people are strongly drawn to make a myriad of social inferences from facial features (Bruce & Young, 2012; Hassin & Trope, 2000) despite little evidence for their accuracy (for a review, see Todorov et al., 2015). These facial impressions can have important outcomes such as political preferences and voting intentions (Olivola, Tingley, & Todorov, 2018; Olivola & Todorov, 2010; Todorov, Mandisodza, Goren, & Hall, 2005) or sentencing decisions (Porter, ten Brinke, & Gustaw, 2010).

Earlier research on the structure underlying social judgments from faces revealed that two core dimensions could be identified (Oosterhof & Todorov, 2008; Todorov, Baron, & Oosterhof, 2008). The first axis was best associated with valence/trustworthiness judgments pointing to the target's intention (i.e., harmless vs. harmful), while the second axis was approximated by dominance judgments indicating the target's capability to carry out these intentions. According to the overgeneralization hypothesis (Knutson, 1996; Montepare &

Dobish, 2003; Zebrowitz, Fellous, Mignault, & Andreoletti, 2003; Zebrowitz & Montepare, 2008), impressions from neutral faces are overgeneralization of cues that provide adaptive information. Trustworthiness would be an overgeneralization of emotional cues (e.g., anger or happiness) signaling approach or avoidance tendencies (Oosterhof & Todorov, 2008; Said, Sebe, & Todorov, 2008; Todorov, 2008; Todorov et al., 2008), whereas dominance judgments would stem from overgeneralizations of facial indicators of physical strength and facial maturity (Oosterhof & Todorov, 2008; Toscano, Schubert, Dotsch, Falvello, & Todorov, 2016). This two-dimensional model was replicated more recently with the addition of a third dimension related to youthful-attractiveness judgments that could serve sexual selection functions (Sutherland et al., 2015; Sutherland et al., 2018; Sutherland et al., 2013; Vernon, Sutherland, Young, & Hartley, 2014). More recent efforts have begun to investigate the universality of this model (see Hehman, Stoller, Freeman, Flake, & Xie, 2019; Lin, Keles, & Adolphs, 2019; Stoller, Hehman, & Freeman, 2020; Stoller, Hehman, Keller, Walker, & Freeman, 2018; Sutherland et al., 2018).

The eye-catching resemblance between the competence-by-warmth social model and the dominance-by-trustworthiness facial model was investigated by Sutherland et al. (Sutherland, Oldmeadow, & Young, 2016). Their results pointed to substantial overlap between judgments on the two horizontal dimensions of trustworthiness and warmth while the relation between the vertical dimensions of dominance and competence was weaker. These authors suggested different explanations for the intriguing discrepancy regarding the second axis. For instance, cues signaling competence (i.e., ability) may vary between the conceptual and facial level, that is, some components that are diagnostic of competence may be more readily encoded in faces (e.g., physical strength and dominance) than others (e.g., intelligence and ability). Another possibility, based on an evolutionary perspective, is that competence/prestige (i.e., sharing expertise to gain respect) and dominance (i.e., the use physical strength to induce fear) represent two different strategies—the former being beneficial and the later harmful for others—to climb the social ladder (Chapais, 2015; Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013).

1.3 Visual representation of faces

The above review reveals that people are inclined to make social inferences from faces. An interesting question is whether the reverse process also takes place. Specifically, can our previous knowledge (e.g., stereotypes), beliefs (e.g., political ideologies), or sympathy (e.g., prejudice) for a given group bias the way we imagine their faces? A first response comes from Eberhardt et al.' (Eberhardt, Dasgupta, & Banaszynski, 2003) study in which participants were presented with a picture of a racially ambiguous target before drawing the target's face. The drawings were clearly biased by the racial label attached to the target. Building on these initial findings, several studies showed that visual representation of social targets can be biased in a top-down fashion by *a priori* beliefs, such as racial stereotypes (e.g., Dotsch, Wigboldus, Langner, & van Knippenberg, 2008; Dotsch, Wigboldus, & van Knippenberg, 2011; Krosch & Amodio, 2014), minorities stereotypes (Brown-Iannuzzi, Dotsch, Cooley, & Payne, 2017; Hinzman & Maddox, 2017), national stereotypes (e.g.,

Imhoff & Dotsch, 2013; Imhoff, Dotsch, Bianchi, Banse, & Wigboldus, 2011), gender stereotypes (Brooks, Stoler, & Freeman, 2018), or occupational stereotypes (Brown-Iannuzzi, Dotsch, Cooley, & Payne, 2017; Hehman, Flake, & Freeman, 2015; Imhoff et al., 2013; Oldmeadow, Sutherland, & Young, 2012). Moreover, behavioral information (Dotsch, Wigboldus, & van Knippenberg, 2013) or minimal group membership (Ratner, Dotsch, Wigboldus, van Knippenberg, & Amodio, 2014) proved sufficient to bias the representation of facial trait-diagnostic-features of novel groups.

The vast majority of the above efforts relied on a variant of the reverse correlation technique (see Dotsch & Todorov, 2012; Dotsch et al., 2008; Mangini & Biederman, 2004) in order to capture the visual representation of the target and identify the specific visual diagnostic features that drive social inferences. In a typical reverse correlation task (RC), participants are presented with several trials for which they have to choose between a pair of noisy faces the face that looks the most like the specified target (e.g., “the most Moroccan-looking face”). These noisy faces are, in fact, random variations of a single face (i.e., the base-face). The average of all selected noise patterns superimposed to the base-face is then used to produce the classification image (CI)—a visual read-out mental representation of the target (Brinkman, Todorov, & Dotsch, 2017).

This procedure offers several advantages (for a review, see Brinkman et al., 2017; but also Jack & Schyns, 2017). First and foremost, it is unconstrained and data-driven. Indeed, the variation in the stimulus set—the noisy faces generated from a random noise pattern superimposed to be base-face—is immensely large and can thus span an entire space of hypotheses rather than a single one. Secondly, it allows visualizing “near *spontaneous use of information* because participants are free to adopt whatever criteria they want for their judgments (in fact, participants might not even be aware of the criteria they adopt)” (Brinkman et al., 2017, p. 336). In sum, this method clearly contrasts with more traditional social judgment paradigms or indirect measures that force participants to make judgments or categorizations along pre-established dimensions or categories rooted in researchers’ *a priori* hypotheses that may not be germane to real-life scenarios (Brinkman et al., 2017; Imhoff et al., 2013; Jack & Schyns, 2017).

Using this RC approach, previous studies successfully established the encoding and decoding of the trustworthiness-by-dominance and the warmth-by-competence dimensions’ content onto physical facial cues. Specifically, the trustworthiness and dominance CIs were negatively correlated, whereas the warmth and competence CIs were positively correlated (e.g., Dotsch & Todorov, 2012; Imhoff et al., 2013; Oliveira, Garcia-Marques, & Dotsch, 2019; Oliveira, Garcia-Marques, Dotsch, & Garcia-Marques, 2019). In addition, trustworthiness and warmth CIs were highly similar, and both overlapped with judgments of valence (Oliveira, Garcia-Marques, Dotsch, & Garcia-Marques, 2019; see also Oosterhof & Todorov, 2008). Interestingly, there was a dissimilarity between the dominance and competence CIs driven by the negative evaluation of the former and positive evaluation of the later. In particular, Oliveira et al. (Oliveira, Garcia-Marques, Garcia-Marques, & Dotsch, 2019) found that valence (which overlaps with warmth/sociability) had a negative linear relationship with dominance in the domain of face representation but had a U-shaped relation

when judging trait concepts (see also Imhoff & Koch, 2017; for a more general argument, see Peabody, 1967).

1.4 The present research

Research on the compensation effect has mainly taken place at a verbal level (e.g., traits' ratings) and under constrained conditions (e.g., a limited number of items or stimuli). In light of these limitations, we aimed to examine whether compensation would also emerge at a visual level and in a more spontaneous way. Previous work in the domain of face perception suggests that competence and warmth are positively related (i.e., a halo effect) at both the judgmental and representational level (Imhoff et al., 2013; Oliveira, Garcia-Marques, Dotsch, & Garcia-Marques, 2019; Sutherland et al., 2016). However, these studies did not aim to test compensation *per se*, nor did they create an appropriate setting (e.g., comparison context; see Yzerbyt, 2018). The emergence of a visual correlate of the compensation effect would therefore constitute convincing evidence, especially in the light of the likely emergence of a halo effect already documented.

In Experiments 1 and 2, we tested if the compensation effect would be present in the facial representation of two social targets. To do so, we presented participants with two fictitious groups that differed on one of the two fundamental dimensions (competence in Experiment 1; warmth in Experiment 2) but were equal on the other (warmth in Experiment 1; competence in Experiment 2). The use of new groups as targets provides a more controlled setting and precludes the possibility of a perceptual bias caused by previous knowledge. We then relied on the reverse correlation procedure to capture participants' *near spontaneous* visual representation of the two targets. Independent judges then rated these visual renderings on the two fundamental dimensions of social perception. In both experiments, we predicted compensatory ratings of the prototypical faces. Because past findings have pointed out some divergence between the two models of social and face perception, in Experiment 3, we asked a new sample of judges to rate the previously produced faces on trustworthiness and dominance. We also predicted a negative relationship between these ratings. In Experiment 4, we sought to replicate and further validate the results from Experiments 1–3 by manipulating the design used to collect the data (i.e., within- vs. between-participants).

2. Power considerations

In our experiments, we relied on the standard implementation of the reverse correlation paradigm (Dotsch et al., 2008). Given that there exist no reliable method (see Brinkman et al., 2017) to estimate the number of participants (hereafter referred to as “producers”) required for a reverse correlation task, we followed previous studies recommendation (Dotsch & Todorov, 2012; see also Oliveira, Garcia-Marques, Dotsch et al., 2019) of at least twenty producers per condition. Likewise, previous studies have relied on a wide variety of samples sizes ranging between 31 and 400 participants (hereafter referred to as “judges”) to rate the visual outcomes generated from the RC task (e.g., Brinkman, Dotsch et al., 2019; Jackson, Hester, & Gray, 2018; Oliveira, Garcia-Marques, Dotsch, & Garcia-Marques, 2019). The

final sample sizes of our experiments, therefore, fulfilled the minimal number of producers and judges widely recommended and we increased these numbers in light of available financial resources and participants' pool availability.

3. Experiment 1

Experiment 1 sought to test whether the compensation effect would emerge at a visual level when confronted with two targets that initially differed on one of the two fundamental dimensions (i.e., competence). On this basis, participants formed an impression of two fictitious groups (i.e., the blue and green groups) and we captured their visual representation through a reverse correlation technique. Social targets were presented in such a way that one group was superior on competence (i.e., competence manipulated) but equal on warmth compared to the other group. We expected the visual prototypes of these groups to be rated in a compensatory way by external judges.

3.1 Method

Participants and design. Data were collected in two samples. The first sample of producers comprised 136 participants from the United States ($M_{\text{age}}=37.69$, $SD_{\text{age}}=14.00$, 79 females, 57 males) recruited online with Prolific Academic. The second sample of judges had 59 participants from a French-speaking university¹. The experiment involved a 2 (target group: high- vs. low-competence group) $\times$ 2 (judged dimension: competence vs. warmth) repeated measures design.

Materials. Six men's faces with similar shapes were selected from the Radboud Face Database (Langner et al., 2010) and split into two sets. We generated four noisy faces from each face by superimposing a random noise pattern with the default settings from the *rcirc* package (Dotsch, 2017). We used the resulting noisy faces in the impression formation phase as faces exemplar of group members (similar to Dotsch et al., 2013). Next, we morphed the six men's faces previously selected into a single base image using Psychomorph (Tiddeman, Stirrat, & Perrett, 2005). We then generated three hundred and fifty pairs of noisy faces with the *rcirc* R package (default settings) by adding or subtracting a noise pattern from the base image (see Dotsch & Todorov, 2012; Mangini & Biederman, 2004) as illustrated in Figure 1. The stimuli materials had a 512×512 pixels size.

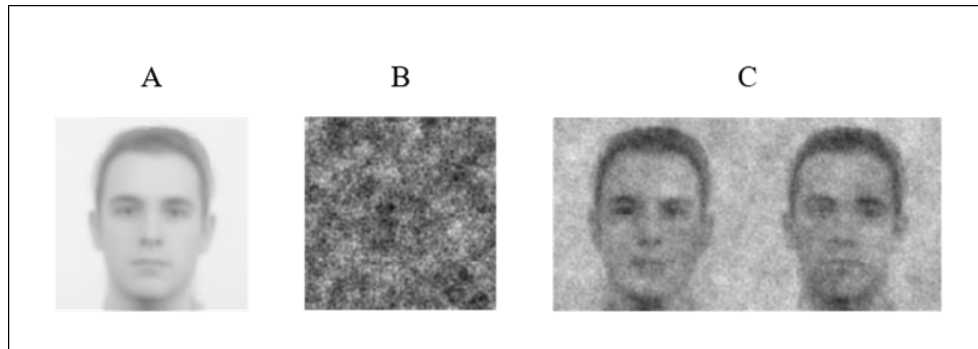


Figure 1. Base image (A), example of a random noise pattern (B), and an example of a pair of noisy faces produced by adding (left) or subtracting (right) the noise pattern from the base image (C).

Measures. Participants’ (producers) face representation of target groups, also known as classification images (CIs), were captured with a reverse correlation (RC) variant, namely the two images forced classification task (Dotsch et al., 2008). The literature shows that previous RC studies relied on 300–500 trials, which is large enough to generate reliable group-wise but not individual CIs (Brinkman et al., 2019). Since our focus was on the group-wise CIs, we opted for 350 trials. In each trial, participants had to “choose the face that looks most like a green [blue] group member” between two adjacent noisy faces. The resulting CI of a given target group was obtained by averaging all the noise patterns selected by all the participants and then superimposing this noise pattern to the base image.

We evaluated producers’ explicit impressions about the two groups through traits rated on a 7-point scale (0=strongly disagree; 6=strongly agree). We used competent, efficient, capable, and intelligent to assess the competence dimension, and sociable, nice, warm, and caring for the warmth dimension (all $\alpha \geq .90$).

We measured judges’ explicit impressions about the two CIs by means of traits rated on a 7-point scale (0=strongly disagree; 6=strongly agree). We used efficient, hardworking, capable, skilled, competent, and organized for the competence dimension, whereas kind, warm, nice, friendly, sociable, and caring tapped the warmth dimension (all α ’s $\geq .90$).

Procedure. The experiment comprised two phases. In the first phase, the producers underwent an impression formation task, a RC task, and a rating task. In the second phase, a new sample of independent judges was asked to rate the average visual representation (CIs) from the two target groups. The impression formation and the RC task were implemented in Psytoolkit (Stoet, 2010, 2017), and the rating tasks were implemented in Qualtrics.

The impression formation task was built upon Judd et al.’s (2005) procedure in which participants form an impression by reading a series of behaviors allegedly performed by members of two fictitious groups (i.e., the “Green group” and the “Blue group”). We adapted the twenty-four behaviors diagnostic of the two fundamental dimensions from previous studies (Judd et al., 2005; Kervyn et al., 2009; Yzerbyt et al., 2008). The high (low) group

was ascribed six positive (negative) and two negative (positive) competence related behaviors, as well as two negative and two positive warmth related behaviors. Therefore, the high group was portrayed as more competent (i.e., the manipulated dimension) than the low group, but both were equally warmth (i.e., the unmanipulated dimension). Group identity (i.e., “Blue” or “Green”) was randomized across participants. Behaviors were paced one at a time for 6000ms with an inter-stimulus delay of 400ms. Each behavior was anonymized and presented with a colored background (i.e., blue or green) to signal group membership. As in Dotsch et al. (2013) and in order to facilitate the forthcoming RC task, we randomly coupled each behavior with a noisy face of what we presented as a group member. On average, exemplar faces did not differ between target groups across participants. To solidify their impressions, participants went through the list of behaviors, once in random order, and once sorted by group identity. They then reported in a few lines their global impression about the two groups.

Next, participants performed a 350 trials-RC-task with a single target group (i.e., “Blue” or “Green” group) being randomized between participants. Trials appeared in random order, and the position of the noisy faces within a pair (i.e., left vs. right) was randomized at each trial. Participants then rated the two groups on a series of traits encompassing the warmth and competence dimension. Traits appeared in random order, with the identity of the first group being randomized across participants. Finally, participants provided some demographic information and were debriefed.

In the second phase of the experiment, a new sample of independent judges rated the two average CIs (i.e., one for the high group, and one for the low group) presented next to each other on a series of traits related to warmth and competence. CIs’ position and traits order were randomized. Finally, judges provided some demographic information and were debriefed.

3.2 Results

We submitted producers’ ratings of the two target groups to a 2 (target group: high- vs. low-competence group) $\times$ 2 (judged dimension: competence vs. warmth) repeated measures ANOVA. There was a main effect of target group, such that the high group received higher ratings than the low group ($M_{\text{high}}=3.91$; $SD_{\text{high}}=1.29$; $M_{\text{low}}=3.46$; $SD_{\text{low}}=1.37$), $F(1,132)=13.50$, $p<.001$, $\eta_p^2=.09$. There was also a main effect of judged dimension, such that the competence dimension was rated higher than the warmth dimension ($M_{\text{competence}}=3.78$; $SD_{\text{competence}}=1.38$; $M_{\text{warmth}}=3.59$; $SD_{\text{warmth}}=1.31$), $F(1,132)=13.88$, $p<.001$, $\eta_p^2=.09$. Importantly, the target group $\times$ judged dimension interaction was significant, $F(1,132)=52.07$, $p<.001$, $\eta_p^2=.28$. Confirming the impact of our manipulation, the high group came across as more competent ($M_{\text{high}}=4.40$; $SD_{\text{high}}=1.12$; $M_{\text{low}}=3.15$; $SD_{\text{low}}=1.34$), $F(1,132)=62.73$, $p<.001$, $\eta_p^2=.32$. Crucially, supporting compensation, the high group also appeared less warm than the low group ($M_{\text{high}}=3.42$; $SD_{\text{high}}=1.26$; $M_{\text{low}}=3.77$; $SD_{\text{low}}=1.33$), $F(1,132)=4.18$, $p=.043$, $\eta_p^2=.03$.

We then assessed the producers' visual representation of the two social targets. To compute the classification image (CI) of the high and low-competence group, we averaged every noise pattern selected by the producers and superimposed it on the base-face (see Figure 2).



Figure 2. High-competence group (left) and low-competence group (right) classification images (CIs). CIs were generated using the scaling constant method ($c = .004$).

We submitted judges' ratings of the two CIs captured from the producers to a 2 (target group: high- vs. low-competence group) $\times$ 2 (judged dimension: competence vs. warmth) two-way repeated measure ANOVA. There was a main effect of target group, such that the high-competence group CI was rated lower than the low-competence group CI ($M_{\text{high}}=4.01$; $SD_{\text{high}}=1.49$; $M_{\text{low}}=4.83$; $SD_{\text{low}}=1.25$), $F(1,54)=36.57$, $p<.001$, $\eta_p^2=.39$. There was also a main effect of judged dimension, such that the competence dimension was rated higher than the warmth dimension ($M_{\text{competence}}=4.60$; $SD_{\text{competence}}=1.28$; $M_{\text{warmth}}=4.24$; $SD_{\text{warmth}}=1.55$), $F(1,54)=13.17$, $p<.001$, $\eta_p^2=.19$. Importantly, the target group $\times$ judged dimension interaction proved significant, $F(1,54)=72.15$, $p<.001$, $\eta_p^2=.55$. Specifically, judges rated the high-competence CI as more competent ($M_{\text{high}}=4.77$; $SD_{\text{high}}=1.33$; $M_{\text{low}}=4.43$; $SD_{\text{low}}=1.22$), although this difference was marginal, $F(1,54)=3.41$, $p=.070$, $\eta_p^2=.06$. Crucially, and in line with the compensation effect, judges rated the high-competence CI less warm than the low-competence CI ($M_{\text{high}}=3.24$; $SD_{\text{high}}=1.22$; $M_{\text{low}}=5.24$; $SD_{\text{low}}=1.15$), $F(1,54)=95.51$, $p<.001$, $\eta_p^2=.62$.

3.3 Discussion

Experiment 1 aimed to test the visual correlate of the compensation effect. To do so, we captured both participants' (producers) verbal (i.e., traits ratings) and visual (i.e., CI) impressions of two fictitious groups that initially differed on competence but were equal on warmth. We predicted a compensation pattern at both levels.

In line with previous findings (Cambon, Yzerbyt, & Yakimova, 2015; Judd et al., 2005; Schmitz & Yzerbyt, in press; Yzerbyt & Cambon, 2017), our results reveal the presence of a clear-cut compensation pattern among producers' verbal ratings. Not surprisingly, the high-competence group was credited with more competence than the low-competence group. Also,

and in line with compensation, the high-competence group was credited with less warmth than the low-competence group. More importantly, we expected the same pattern to emerge on the visual facial prototypes of both groups as evaluated by independent judges. Indeed, the high-competence group CI was clearly rated as less warm but more competent than the low-competence CI. At the same time, however, the difference in competence in favor of the high group remained marginal—although producers unambiguously acknowledged such differences in their ratings. This pattern raises the possibility that the sample of traits encompassing the competence dimension used by the judges did not adequately capture this dimension, given that it only encompassed one of its two facets (i.e., ability). Moreover, the question remains whether this visual compensation pattern would hold when the initial difference between the groups resides on warmth rather than competence.

4. Experiment 2

Experiment 1 provided initial, and indeed encouraging, evidence for the visual manifestation of the dimensional compensation effect. In Experiment 2, we sought to extend these findings by manipulating the dimension of warmth instead of competence. It is noteworthy that this manipulation constitutes a more conservative test because compensation is reportedly weaker under such circumstances (Judd et al., 2005; Kervyn et al., 2011; Yzerbyt, 2016). Moreover, we enlarged the spectrum of traits used to rate the visual presentations to the four facets of the two fundamental dimensions (i.e., friendliness and morality for the warmth dimension, and assertiveness and ability for the competence dimension). As in Experiment 1, we hypothesized that a compensation effect would emerge at both the verbal (i.e., traits ratings) and visual level (i.e., CIs).

4.1 Method

Participants and design. Data were collected on two samples. The first sample of producers comprised 66 Belgians from a French-speaking university ($M_{\text{age}}=21.97$, $SD_{\text{age}}=2.47$; 51 females, 15 males). The second sample of judges involved 105 participants from the United Kingdom sampled through the online crowdsourcing platform Prolific Academic ($M_{\text{age}}=35.86$, $SD_{\text{age}}=11.84$; 68 female, 36 males, one other). The experiment involved a 2 (target group: high- vs. low-competence group) $\times$ 2 (judged dimension: competence vs. warmth) repeated measures design.

Measures. The same 350-trials-RC-task from Experiment 1 was used to capture participants' (producers) visual representation of the target group.

We evaluated producers' explicit impressions about the two groups by means of traits rated on a 7-point scale (0=strongly disagree; 6=strongly agree). We used competent, efficient, capable, and hardworking for the competence dimension, and friendly, agreeable, warm, and sociable for the warmth dimension (all α 's $\geq .80$).

We assessed judges' explicit impressions about the two CIs through traits rated on a 7-point scale (0=strongly disagree; 6=strongly agree). We used traits related to ability

(competent, efficient, and hardworking) and assertiveness (ambitious, determined, and self-confident) for the competence dimension, and friendliness (warm, sociable, and friendly) and morality (i.e., trustworthy, moral, honest) for warmth dimension (Abele et al., 2008, 2016). The competence and warmth dimensions had adequate levels of reliability (all α 's $\geq .77$).

Procedure. We adopted the same procedure as in Experiment 1, with three exceptions. First, we manipulated the warmth dimension (instead of competence) in the impression formation; that is, we presented behaviors in such a way that the high-warmth group was warmer but equally competent than the low group. Second, we did not provide exemplar faces of group members during the impression formation phase given that previous studies have shown that participants could form a visual representation of minimal groups even without seeing exemplar faces (Ratner et al., 2014). Third, we expanded our array of traits during the CIs' rating phase to tap the four facets of the two fundamental dimensions.

4.2 Results

We submitted producers' ratings of the two target groups to a 2 (target group: high- vs. low-competence group) $\times$ 2 (judged dimension: competence vs. warmth) repeated measures ANOVA. There was a main effect of target group, such that the high group received higher ratings than the low group ($M_{\text{high}}=4.07$; $SD_{\text{high}}=1.24$; $M_{\text{low}}=3.28$; $SD_{\text{low}}=1.43$), $F(1,62)=32.23$, $p<.001$, $\eta_p^2=.33$. There was also a main effect of judged dimension, such that the competence dimension was rated higher than the warmth dimension ($M_{\text{competence}}=3.90$; $SD_{\text{competence}}=1.11$; $M_{\text{warmth}}=3.45$; $SD_{\text{warmth}}=1.60$), $F(1,62)=32.36$, $p<.001$, $\eta_p^2=.33$. The target group $\times$ judged dimension interaction was significant, $F(1,62)=65.12$, $p<.001$, $\eta_p^2=.50$. In line with our manipulation, the high-warmth group came across as warmer ($M_{\text{high}}=4.57$; $SD_{\text{high}}=1.15$; $M_{\text{low}}=2.32$; $SD_{\text{low}}=1.13$), $F(1,62)=89.33$, $p<.001$, $\eta_p^2=.58$. Importantly, the high group appeared less competent than the low-warmth group ($M_{\text{high}}=3.56$; $SD_{\text{high}}=1.13$; $M_{\text{low}}=4.23$; $SD_{\text{low}}=0.99$), $F(1,62)=9.35$, $p<.01$, $\eta_p^2=.13$, thus supporting compensation.

Next, we assessed the producers' visual representation of the two social targets. To compute the classification image (CI) of the high and low-warmth group, we averaged every noise pattern selected by all the producers and superimposed it on the base image (see Figure 3).



Figure 3. High-warmth group (left) and low-warmth group (right) classification images (CIs). CIs were generated using the scaling constant method ($c = .006$).

We submitted judges' ratings of the two CIs generated by the producers to a 2 (target group: high- vs. low-competence group) $\times$ 2 (judged dimension: competence vs. warmth) repeated measures ANOVA. There was a main effect of target group, such that the high group CI was rated higher than the low group CI ($M_{\text{high}}=4.65$; $SD_{\text{high}}=1.08$; $M_{\text{low}}=3.58$; $SD_{\text{low}}=1.13$), $F(1,101)=80.90$, $p<.001$, $\eta_p^2=.44$. There was also a main effect of judged dimension, such that the competence dimension was rated higher than the warmth dimension ($M_{\text{competence}}=4.21$; $SD_{\text{competence}}=0.96$; $M_{\text{warmth}}=4.02$; $SD_{\text{warmth}}=1.44$), $F(1,101)=23.62$, $p<.001$, $\eta_p^2=.19$. Importantly, the target group $\times$ judged dimension interaction proved significant, $F(1,101)=149.77$, $p<.001$, $\eta_p^2=.59$. Follow-up tests revealed that the high-warmth CI came across as warmer ($M_{\text{high}}=5.02$; $SD_{\text{high}}=1.07$; $M_{\text{low}}=3.01$; $SD_{\text{low}}=1.00$), $F(1,101)=177.60$, $p<.001$, $\eta_p^2=.63$, but not significantly more competent than the low-warmth CI ($M_{\text{high}}=4.27$; $SD_{\text{high}}=0.96$; $M_{\text{low}}=4.15$; $SD_{\text{low}}=0.96$), $F(1,101)=0.80$, $p=.373$, $\eta_p^2=.01$.

4.3 Discussion

Experiment 2 sought to extend the results from Experiment 1 by enlarging the spectrum of traits used by independent judges to assess the visual outcomes to the four facets of social judgments (i.e., ability and assertiveness for the competence dimension; friendliness and morality for the warmth dimension). Moreover, we tested whether the compensation pattern would hold when the primary difference between groups resided on warmth. A compensatory pattern was predicted for traits ratings from producers as well as their visual renderings (CIs) of the two groups.

Consistent with our first experiment, Experiment 2 revealed that producers compensated on their rating of the two targets, such that the high-warmth group was depicted as warmer but also less competent than the low-warmth group. Contrary to our predictions, this pattern was not fully translated to the visual prototypes of the two groups. Indeed, the high-warmth group CI came across—as assessed by independent judges—as warmer than the low-warmth group CI. However, ratings did not differ on competence despite the inclusion of both facets (i.e., assertiveness and ability) in the sample of traits used by judges to appraise this

dimension. This finding is intriguing considering the fact that, in both experiments, producers' ratings clearly acknowledged a competence gap between the two social targets.

One possible account is that the marginal and non-significant differences on the manipulated dimension in Experiment 1 and on the non-manipulated dimension in Experiment 2, respectively, stem from the difficulty to encode or read competence content from faces. Along similar lines, the individual variability in the representations or in the inferences of this trait on faces may also explain the obtained pattern (Hehman, Sutherland, Flake, & Slepian, 2017; Oliveira, Garcia-Marques, Dotsch, & Garcia-Marques, 2019). Going one step further, and in light of the literature, it is possible that the producers visually encoded competence as dominance rather than competence (Sutherland et al., 2016).

5. Experiment 3

Experiments 1 and 2 were not conclusive with respect to the emergence of facial differences in terms of the social dimension of competence. In particular, it is possible that the perceived difference in competence between the two groups was facially translated as dominance instead of competence. We designed Experiment 3 to examine this hypothesis. Specifically, we asked a new sample of independent judges to rate the visual templates from the previous two experiments on two dimensions of the face perception model (Oosterhof & Todorov, 2008; Todorov et al., 2008), namely trustworthiness and dominance. In all likelihood, we would expect to find a marked compensation pattern using these facial dimensions.

5.1 Method

Participants and Design. We recruited a sample of 201 participants (judges) from the UK through Prolific Academic ($M_{\text{age}}=36.18$, $SD_{\text{age}}=12.94$; 138 females, 63 males). The experiment involved a 2 (manipulated dimension: competence vs. warmth) $\times$ 2 (target group: high- vs. low group on the manipulated dimension) $\times$ 2 (judged dimension: dominance vs. trustworthiness) design, with the first factor varying between participants and the last two within them.

Measures. We assessed judges' explicit impressions about the two CIs by means of two traits, dominance and trustworthiness, rated on a 7-point scale (0=strongly disagree; 6=strongly agree).

Procedure. Judges were randomly assigned to one of two conditions, that is, they either rated the CIs produced in Experiment 1 (i.e., competence manipulated) or those in Experiment 2 (i.e., warmth manipulated). In each condition, both high and low CIs were adjacent, and judges rated them on dominance and trustworthiness. CIs' position and traits order were randomized. After the ratings, judges provided some demographic information and were debriefed.

5.2 Results

We submitted judges' ratings of the CIs to a 2 (manipulated dimension: competence vs. warmth) $\times$ 2 (target group: high- vs. low group on the manipulated dimension) $\times$ 2 (judged dimension: dominance vs. trustworthiness) mixed model analysis, with the first factor varying between participants and the last two within them. There was no main effect of manipulated dimension, such that the ratings did not significantly differ when the manipulated dimension was competence or warmth ($M_{\text{competence}}=4.15$; $SD_{\text{competence}}=1.42$; $M_{\text{warmth}}=4.01$; $SD_{\text{warmth}}=1.62$), $F(1,194)=2.18$, $p=.141$, $\eta_p^2=.01$. There was no main effect of target group, such that the high group ratings were not significantly different from the low group ($M_{\text{high}}=4.05$; $SD_{\text{high}}=1.47$; $M_{\text{low}}=4.11$; $SD_{\text{low}}=1.57$), $F(1,194)=0.84$, $p=.360$, $\eta_p^2<.01$. There was a main effect of judged dimension, such that the dominance dimension was rated higher than the trustworthiness dimension ($M_{\text{trustworthiness}}=3.89$; $SD_{\text{trustworthiness}}=1.48$; $M_{\text{dominance}}=4.27$; $SD_{\text{dominance}}=1.55$), $F(1,194)=26.80$, $p<.001$, $\eta_p^2=.12$. The manipulated dimension $\times$ target group was significant, $F(1,194)=11.92$, $p<.001$, $\eta_p^2=.06$, as was the target group $\times$ judged dimension, $F(1,194)=7.71$, $p<.01$, $\eta_p^2=.04$. The manipulated dimension $\times$ judged dimension did not reach significance, $F(1,194)=0.10$, $p=.757$, $\eta_p^2<.01$.

Importantly, the manipulated dimension $\times$ target group $\times$ judged dimension interaction proved significant, $F(1,194)=314.56$, $p<.001$, $\eta_p^2=.61$. Specifically, the target group $\times$ judged dimension was significant when competence was the manipulated dimension, $F(1,194)=112.45$, $p<.001$, $\eta_p^2=.36$. Follow-up analyses confirmed that the high-competence CI came across as more dominant ($M_{\text{high}}=5.19$; $SD_{\text{high}}=1.17$; $M_{\text{low}}=3.52$; $SD_{\text{low}}=1.25$), $F(1,194)=97.05$, $p<.001$, $\eta_p^2=.33$, but also, in line with compensation, less trustworthy than the low-competence CI, ($M_{\text{high}}=3.29$; $SD_{\text{high}}=1.12$; $M_{\text{low}}=4.61$; $SD_{\text{low}}=1.22$), $F(1,194)=61.16$, $p<.001$, $\eta_p^2=.24$. Also, the target group $\times$ judged dimension was significant when warmth was the manipulated dimension, $F(1,194)=209.34$, $p<.001$, $\eta_p^2=.51$, such that judges rated the high-warmth CI as less dominant ($M_{\text{high}}=3.02$; $SD_{\text{high}}=0.97$; $M_{\text{low}}=5.36$; $SD_{\text{low}}=1.24$), $F(1,99)=190.17$, $p<.001$, $\eta_p^2=.49$, and, supporting compensation, more trustworthy than the low-warmth CI ($M_{\text{high}}=4.71$; $SD_{\text{high}}=1.34$; $M_{\text{low}}=2.95$; $SD_{\text{low}}=1.35$), $F(1,194)=106.56$, $p<.001$, $\eta_p^2=.35$.

5.3 Discussion

Experiment 3 tested whether a compensation pattern would emerge when the CIs from previous experiments were evaluated on the facial dimensions of dominance and trustworthiness. In line with our predictions, a compensation effect was present whether judges rated CIs from Experiment 1 (i.e., competence manipulated) or those from Experiment 2 (i.e., warmth manipulated). One limitation concerns whether facial judgments of competence and dominance can be distinguished. Indeed, previous research has shown that whether two ratings are measured at the same time or separately (i.e., within- vs. between-participants) may affect the results (Erlebacher, 1977; Yzerbyt & Judd, 2020).

6. Experiment 4

As a set, Experiments 1–3 reveal that a compensation pattern is more difficult to obtain when judges evaluate CIs on competence (i.e., social perception model; Experiments 1–2) even though producers manifest clear compensatory judgments on their verbal judgments. In sharp contrast, dimensional compensation takes place when judges evaluate CIs in terms of dominance (i.e., face perception model; Experiment 3). In Experiment 4, we sought to replicate and further ascertain these results by manipulating the experimental design (i.e., within-participants vs. between-participants), and tested whether the social dimension of competence could be sharply differentiated from the facial dimension of dominance even when measured at the same time. Specifically, participants either rated the CIs on the social dimensions (i.e., competence and warmth), the facial dimensions (i.e., dominance and trustworthiness), or both. We expected that a compensation pattern would emerge on the dominance and trustworthiness as well as on warmth but *not* on competence, independently of the design type.

6.1 Method

Participants and Design. We recruited 302 participants (judges) from the UK via Prolific Academic ($M_{age}=35.42$, $SD_{age}=11.46$; 233 females, 69 males). The experiment involved a 2 (design type: between-participants vs. within-participant) $\times$ 2 (manipulated dimension: competence vs. warmth) $\times$ 2 (target group: high- vs. low group on the manipulated dimension) $\times$ 2 (judged dimension: vertical vs. horizontal) $\times$ 2 (dimensional model: social vs. facial) design, with the first and the second factor varying between participants, the third and the fourth varying within them and the dimensional model factor varied both within and between them (Erlebacher, 1977; Yzerbyt & Judd, 2020).

Measures. We assessed judges' explicit impressions about the CIs by means of four traits (i.e., competence, warmth, dominance, and trustworthiness) rated on a 7-point scale (0=strongly disagree; 6=strongly agree).

Procedure. Judges were randomly assigned to one of six conditions, that is, they either saw the CIs produced in Experiment 1 (i.e., competence manipulated) or those produced in Experiment 2 (i.e., warmth manipulated) and rated them either on the social dimensions (i.e., competence and warmth), the facial dimensions (i.e., dominance and trustworthiness), or both. In each condition, both high and low CIs were adjacent, and judges rated them on two or four traits: competence, warmth, dominance, and trustworthiness. CIs position and traits order were randomized. After the ratings, judges provided some demographic information and were debriefed.

6.2 Results

We submitted judges' ratings of the CIs to a 2 (design type: between-participants vs. within-participant) $\times$ 2 (manipulated dimension: competence vs. warmth) $\times$ 2 (target group: high- vs. low group on the manipulated dimension) $\times$ 2 (judged dimension: vertical vs.

horizontal) $\times$ 2 (dimensional model: social vs. facial) mixed model analysis². We conducted the analysis with linear mixed models (Judd, Westfall, & Kenny, 2017; Yzerbyt & Judd, 2020). As we hoped, the five-way interaction was far from significant, $F(1,1296.99)=0.01$, $p=.926$, confirming that the specific design had no impact on the obtained pattern of judgments. In sharp contrast, and consistent with findings from Experiments 1–3, the manipulated dimension $\times$ target group $\times$ judged dimension $\times$ dimensional model proved significant, $F(1,1296.99)=27.11$, $p<.001$.

Consistent with those found in Experiments 1 and 2, follow-up analyses revealed that the manipulated dimension $\times$ target group $\times$ judged dimension was significant when the dimensional model was the social model, $F(1,1296.99)=132.53$, $p<.001$. As expected, the target group $\times$ judged dimension was significant when competence was the manipulated dimension (i.e., Experiment 1), $F(1,1296.99)=37.90$, $p<.001$. Specifically, there was no significant difference on competence between the high- and low-competence CIs ($M_{\text{high}}=4.34$; $SD_{\text{high}}=1.12$; $M_{\text{low}}=4.26$; $SD_{\text{low}}=1.03$), $F(1,1296.99)=0.23$, $p=.633$, whereas the latter was perceived as warmer ($M_{\text{high}}=3.01$; $SD_{\text{high}}=1.27$; $M_{\text{low}}=4.39$; $SD_{\text{low}}=1.48$), $F(1,1296.99)=67.1$, $p<.001$. Conversely, the target group $\times$ judged dimension was significant when warmth was the manipulated dimension (i.e., Experiment 2), $F(1,1296.99)=102.30$, $p<.001$. This time, the high-warmth CI was perceived as warmer than the low-warmth CI ($M_{\text{high}}=4.90$; $SD_{\text{high}}=1.45$; $M_{\text{low}}=2.38$; $SD_{\text{low}}=1.16$), $F(1,1296.99)=222.72$, $p<.001$, but there was no significant difference on competence between the two CIs ($M_{\text{high}}=4.11$; $SD_{\text{high}}=1.20$; $M_{\text{low}}=4.01$; $SD_{\text{low}}=1.24$), $F(1,1296.99)=0.38$, $p=.535$.

Turning to the facial dimensional model, the manipulated dimension $\times$ target group $\times$ judged dimension proved significant, $F(1,1296.99)=362.01$, $p<.001$. As expected, the target group $\times$ judged dimension was significant when competence was the manipulated dimension (i.e., Experiment 1), $F(1,1296.99)=127.38$, $p<.001$, such that the high-competence CI was perceived as more dominant ($M_{\text{high}}=5.27$; $SD_{\text{high}}=1.10$; $M_{\text{low}}=3.49$; $SD_{\text{low}}=1.21$), $F(1,1296.99)=114.10$, $p<.001$, but less trustworthy than the low-competence CI ($M_{\text{high}}=3.43$; $SD_{\text{high}}=1.16$; $M_{\text{low}}=4.31$; $SD_{\text{low}}=1.38$), $F(1,1296.99)=27.87$, $p<.001$. Conversely, the target group $\times$ judged dimension was significant when warmth was the manipulated dimension (i.e., Experiment 2), $F(1,1296.99)=244.36$, $p<.001$, such that the high-warmth CI was perceived as less dominant ($M_{\text{high}}=3.11$; $SD_{\text{high}}=1.28$; $M_{\text{low}}=5.36$; $SD_{\text{low}}=1.33$), $F(1,1296.99)=185.23$, $p<.001$, but more trustworthy than the low-warmth CI ($M_{\text{high}}=4.32$; $SD_{\text{high}}=1.46$; $M_{\text{low}}=2.91$; $SD_{\text{low}}=1.01$), $F(1,1296.99)=72.19$, $p<.001$.

6.3 Discussion

Experiment 4 sought to replicate and extend the results from Experiment 3 by manipulating the type of design in order to test whether the CIs from Experiments 1 and 2 could be clearly distinguished in terms of competence and dominance even when they were assessed simultaneously. In line with our predictions, facial compensation only took place when taking into account the facial dimensions of dominance and trustworthiness, independently of the design type (i.e., between-participants versus within-participant). Moreover, these results generally replicate those of Experiments 1 and 2 in which we

observed a significant difference on warmth but not on competence (although the difference on competence was marginal in Experiment 1), and with Experiment 3 in which a compensation pattern emerged when CIs were judged in terms of dominance and trustworthiness.

7. General discussion

Over more than a decade of research on the compensation effect has shown that the trade-off between the ascription of warmth and competence when forming an impression of two social targets is a robust phenomenon (for reviews, see Yzerbyt, 2016, 2018). However, this body of evidence has exclusively been built on verbal outputs such as written descriptions or reactions to group labels. Furthermore, the designs and measures used in previous studies restrained participants' responses to a limited set of options determined *a priori* by researchers (e.g., Cambon & Yzerbyt, 2018; Judd et al., 2005; Kervyn et al., 2011; Schmitz & Yzerbyt, in press). These two aspects are not trivial limitations when considering that everyday life social impressions are formed freely and comprise multi-modal components (e.g., auditory stereotypes; Fasoli, Hegarty, Maass, & Antonio, 2018; visual stereotypes; Dotsch, Wigboldus, & van Knippenberg, 2013). The current effort addresses these limitations by testing the compensation effect on a new sensory modality, namely the visual one, and in an unconstrained manner. In line with findings showing that top-down beliefs and stereotypes can affect the way we picture others in our mind (Brinkman et al., 2017; Imhoff et al., 2013; Todorov et al., 2015), our results indicate that compensation does also bias the visual template of social targets. Interestingly, however, this occurs on the horizontal dimensions of either trustworthiness or warmth and also on the vertical dimension of dominance but not on the vertical dimension of competence (Todorov & Oosthorov, 2008).

Experiments 1 and 2 had participants (producers) read about two fictitious groups that differed on one of the two fundamental dimensions of social perception (competence in Experiment 1, and warmth in Experiment 2). Importantly, we probed for their expected facial content (i.e., CI) of the two targets in a near spontaneous and unconstrained way through a data-driven technique (i.e., the reverse correlation) that is agnostic to researchers' hypothesis (Brinkman et al., 2017; Dotsch et al., 2008; Jack & Schyns, 2017). Although participants displayed strong dimensional compensation in their verbal ratings, their facial prototypes of the two groups, at least as gauged by external judges, contrasted strongly in terms of warmth but only marginally (in Experiment 1, despite the fact that this was the manipulated dimension) or not at all (Experiment 2) in terms of competence. Building on existing literature (Sutherland et al., 2016) suggesting that cues of dominance (e.g., physical strength, aggressivity, or masculinity) may be more readily encoded into facial features than cues of competence (e.g., intelligence, status, or skills), we conjectured that the perceived gap in competence expressed in the verbal ratings could have been better translated in the face space as dominance.

In line with this hypothesis, Experiments 3 revealed a clear-cut compensation pattern when re-evaluating the CIs on the facial dimensions of dominance and trustworthiness.

Moreover, Experiment 4 showed that the pattern upheld whether the vertical dimension of dominance was gauged along with or independently of the vertical dimension of competence. Again, competence failed to reveal a compensatory pattern in judges' ratings of the target faces. This finding adds to the growing body of literature indicating that the two vertical dimensions seem to work in rather distinct ways (e.g., Carrier et al., 2014; Oliveira, Garcia-Marques, Garcia-Marques et al., 2019; Oliveira, Garcia-Marques, Dotsch, & Garcia-Marques, 2019; Imhoff & Koch, 2017; Sutherland et al., 2016).

Although our work focused on the evaluations made by human judges, future investigations may want to test for the presence of a compensatory pattern in the produced CIs by examining the amount of physical information associated with each dimension (warmth, competence, trustworthiness, and dominance) present in these visual renderings. That is, by examining correlation of the pixels luminance (see Brinkman et al., 2017; e.g., Dotsch & Todorov, 2012; Hinzman, & Maddox, 2017; Oliveira, Garcia-Marques, Garcia-Marques et al., 2019; Oliveira, Garcia-Marques, & Dotsch, 2019) found in the CIs generated in a compensatory setting and in those independently generated CIs diagnostic of each dimension (e.g., competence).

A limitation of the present research is that the CIs were computed and analyzed at the group level (i.e., averaged visual representation from all the participants within a given experimental condition). Indeed, average CIs do not take into account the inter-individual variability and do not allow for more fine-grained analyses (Brinkman et al., 2017; Ratner et al., 2014). For instance, it may be interesting for future research to test the degree to which producers' verbal compensatory ratings correlate (or not) with the evaluation by external judges of their idiosyncratic CIs. Another caveat is that we did not assess producers' verbal ratings of the two target groups on the facial dimensions, as it was not initially the scope of the current research. Although our manipulation relied on behaviors diagnostic of competence and warmth (see Judd et al., 2005; Kervyn et al., 2009; Yzerbyt, Kervyn, & Judd, 2008), they may also be informative in terms of dominance and trustworthiness, and thus account for the presence of dominance and trustworthiness in the CIs.

In conclusion, the present research is the first to provide direct experimental evidence for the visual correlate of the compensation effect, thus highlighting its pervasiveness beyond verbal measures. Clearly, our work opens new avenues to investigate compensation in a way that is more germane to really life social interactions, that is, under near spontaneous and unconstrained conditions.

8. Transparency statement

The data and R scripts to carry out the analyses are publicly available on Open Science Framework: https://osf.io/hk8av/?view_only=e4a9e9974b2d43b2a7cfa9351b1944d2

9. Endnotes

¹ Demographic data from this sample was lost due to a programming error.

² Because the regression analysis involved 32 terms, for the sake of clarity and conciseness, we only report here the statistical tests of the regression terms required to test our hypothesis. The complete statistical analyses are available on Open Science Framework

10. References

- Abele, A. E., Cuddy, A. J., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, 38(7), 1063–1065. <http://dx.doi.org/10.1002/ejsp.574>
- Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. Y. (2019). *Navigating the social world: Shared vertical and horizontal dimensions for evaluating self, individuals, and groups*. Manuscript submitted for publication.
- Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness: Communion with warmth and morality. *Frontiers in Psychology*, 7, 1–17. <http://dx.doi.org/10.3389/fpsyg.2016.01810>
- Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, 93(5), 751–63. <http://dx.doi.org/10.1037/0022-3514.93.5.751>
- Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. In *Advances in experimental social psychology* (Vol. 50, pp. 195–255). Academic Press.
- Brinkman, L., Goffin, S., van de Schoot, R., van Haren, N. E., Dotsch, R., & Aarts, H. (2019). Quantifying the informational value of classification images. *Behavior research methods*, 51(5), 2059–2073. <https://doi.org/10.3758/s13428-019-01232-2>
- Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, 28(1), 333–361. <https://doi.org/10.1080/10463283.2017.1381469>
- Brooks, J. A., Stoler, R. M., & Freeman, J. B. (2018). Stereotypes bias visual prototypes for sex and emotion categories. *Social Cognition*, 36(5), 481–493. <http://dx.doi.org/10.1521/soco.2018.36.5.481>
- Brown-Iannuzzi, J. L., Dotsch, R., Cooley, E., & Payne, B. K. (2017). The relationship between mental representations of welfare recipients and attitudes toward welfare. *Psychological science*, 28(1), 92–103. <https://doi.org/10.1177/0956797616674999>
- Bruce, V., & Young, A. W. (2012). *Face perception*. London, England: Psychology Press.
- Cambon, L., Yzerbyt, V., & Yakimova, S. (2015). Compensation in intergroup relations: An investigation of its structural and strategic foundations. *British Journal of Social Psychology*, 54(1), 140–158. <https://doi.org/10.1111/bjso.12067>
- Cambon, L., & Yzerbyt, V. Y. (2017). Compensation is for real: Evidence from existing groups in the context of actual relations. *Group Processes & Intergroup Relations*, 20(6), 745–756. <https://doi.org/10.1177/1368430215625782>

- Cambon, L., & Yzerbyt, V. Y. (2018). Two routes toward compensation: An investigation into the mechanisms of compensation for high-and low-status groups. *Journal of Experimental Social Psychology*, 77, 24-35. <https://doi.org/10.1016/j.jesp.2018.04.002>
- Carrier, A., Louvet, E., Chauvin, B., & Rohmer, O. (2014). The primacy of agency over competence in status perception. *Social Psychology*, 45, 347–356. <http://dx.doi.org/10.1027/1864-9335/a000176>.
- Chapais, B. (2015). Competence and the evolutionary origins of status and power in humans. *Human Nature*, 26(2), 83–161. <http://dx.doi.org/10.1007/s12110-015-9227-6>
- Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology*, 104(1), 103–125. <http://dx.doi.org/10.1037/a0030398>.
- Dotsch, R. (2017). rcicr: Reverse-Correlation Image-Classification Toolbox (R package version 0.4.1). Retrieved from <https://rdrr.io/cran/rcicr/>
- Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, 3(5), 562–571. <https://doi.org/10.1177/1948550611430272>
- 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. <https://doi.org/10.1111/j.1467-9280.2008.02186.x>
- Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of personality and social psychology*, 100(6), 999–1014. <https://doi.org/10.1037/a0023026>
- Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2013). Behavioral information biases the expected facial appearance of members of novel groups. *European Journal of Social Psychology*, 43(1), 116–125. <https://doi.org/10.1002/ejsp.1928>
- Eberhardt, J. L., Dasgupta, N., & Banaszynski, T. L. (2003). Believing is seeing: The effects of racial labels and implicit beliefs on face perception. *Personality and Social Psychology Bulletin*, 29(3), 360–370. <https://doi.org/10.1177/0146167202250215>
- Erlebacher, A. (1977). Design and analysis of experiments contrasting the within- and between-subjects manipulation of the independent variable. *Psychological Bulletin*, 84(2), 212–219. <http://dx.doi.org/10.1037/0033-2909.84.2.212>
- Fasoli, F., Hegarty, P., Maass, A., & Antonio, R. (2018). *Who wants to sound straight? Sexual majority and minority stereotypes, beliefs and desires about auditory gaydar*. *Personality and Individual Differences*, 130, 59–64. doi:10.1016/j.paid.2018.03.046
- 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. <https://doi.org/10.1037/0022-3514.82.6.878>

- Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, 56(2), 109–118. <http://dx.doi.org/10.1037/0003-066X.56.2.109>
- Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of personality and social psychology*, 78(5), 837–52. <https://doi.org/10.1037//0022-3514.78.5.837>
- Hehman, E., Flake, J. K., & Freeman, J. B. (2015). Static and dynamic facial cues differentially affect the consistency of social evaluations. *Personality and Social Psychology Bulletin*, 41(8), 1123–1134. <https://doi.org/10.1177/0146167215591495>
- Hehman, E., Stolier, R. M., Freeman, J. B., Flake, J. K., & Xie, S. Y. (2019). Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Social and Personality Psychology Compass*, 13(2), e12431. <https://doi.org/10.1111/spc3.12431>
- Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, 113, 513–529. <https://doi.org/10.1037/pspa0000090>
- Hilton, J. L., & Von Hippel, W. (1996). Stereotypes. *Annual review of psychology*, 47(1), 237–271. <https://doi.org/10.1146/annurev.psych.47.1.237>
- Hinzman, L., & Maddox, K. B. (2017). Conceptual and visual representations of racial categories: Distinguishing subtypes from subgroups. *Journal of Experimental Social Psychology*, 70, 95–109. <https://doi.org/10.1016/j.jesp.2016.12.012>
- Imhoff, R., & Dotsch, R. (2013). Do we look like me or like us? Visual projection as self-or ingroup-projection. *Social Cognition*, 31(6), 806–816. <http://dx.doi.org/10.1521/soco.2013.31.6.806>
- Imhoff, R., Dotsch, R., Bianchi, M., Banse, R., & Wigboldus, D. H. J. (2011). Facing Europe: Visualizing spontaneous in-group projection. *Psychological Science*, 22(12), 1583–1590. <https://doi.org/10.1177/0956797611419675>
- Imhoff, R., & Koch, A. (2017). How orthogonal are the Big Two of social perception? On the curvilinear relation between agency and communion. *Perspectives on Psychological Science*, 12(1), 122–137. <https://doi.org/10.1177/1745691616657334>
- Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in psychology*, 4, Article 386. <https://doi.org/10.3389/fpsyg.2013.00386>
- Jack, R. E., & Schyns, P. G. (2017). Toward a social psychophysics of face communication. *Annual review of psychology*, 68, 269–297. <https://doi.org/10.1146/annurev-psych-010416-044242>
- Jackson, J. C., Hester, N., & Gray, K. (2018). The faces of God in America: Revealing religious diversity across people and politics. *PloS one*, 13(6). <https://doi.org/10.1371/journal.pone.0198745>
- Jost, J. T., Kivetz, Y., Rubini, M., Guermendi, G., & Mosso, C. (2005). System-justifying functions of complementary regional and ethnic stereotypes: Cross-national

- evidence. *Social justice research*, 18(3), 305–333. <http://dx.doi.org/10.1007/s11211-005-6827-z>
- Judd, C. M., Garcia-Marques, T., & Yzerbyt, V. Y. (2019). The complexity of relations between dimensions of social perception: Decomposing bivariate associations with crossed random factors. *Journal of Experimental Social Psychology*, 82, 200–207. <https://doi.org/10.1016/j.jesp.2019.01.008>
- Judd, C. M., James-Hawkins, L., Yzerbyt, V. Y., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, 89(6), 899–913. <https://doi.org/10.1037/0022-3514.89.6.899>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual review of psychology*, 68, 17.1–17.25. <http://dx.doi.org/601-625>. 10.1146/annurev-psych-122414-033702
- Kervyn, N., Yzerbyt, V. Y., Demoulin, S., & Judd, C. M. (2008). Competence and warmth in context: The compensatory nature of stereotypic views of national groups. *European Journal of Social Psychology*, 38(7), 1175–1183. <http://dx.doi.org/10.1002/ejsp.526>
- Kervyn, N., Yzerbyt, V. Y., & Judd, C. M. (2011). When compensation guides inferences: Indirect and implicit measures of the compensation effect. *European Journal of Social Psychology*, 41(2), 144–150. <https://doi.org/10.1002/ejsp.748>
- Kervyn, N., Yzerbyt, V. Y., Judd, C. M., & Nunes, A. (2009). A question of compensation: the social life of the fundamental dimensions of social perception. *Journal of Personality and Social Psychology*, 96(4), 828–842. <https://doi.org/10.1037/a0013320>
- Knutson, B. J. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, 20(3), 165–182. <https://doi.org/10.1007/BF02281954>
- Krosch, A. R., & Amodio, D. M. (2014). Economic scarcity alters the perception of race. *Proceedings of the National Academy of Sciences*, 111(25), 9079–9084. <https://doi.org/10.1073/pnas.1404448111>
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, 24(8), 1377–1388. <https://doi.org/10.1080/02699930903485076>
- Lin, C., Keles, U., & Adolphs, R. (2019, October 2). Comprehensive trait attributions show that face impressions are organized in four dimensions. <https://doi.org/10.31234/osf.io/87nex>
- Lippman, W. (1922). *Public opinion*. New York: Harcourt.
- Mangini, M., & Biederman, I. (2004). Making the ineffable explicit: estimating the information employed for face classifications. *Cognitive Science*, 28(2), 209–226. <https://doi.org/10.1016/j.cogsci.2003.11.004>
- Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal behavior*, 27(4), 237–254. <https://doi.org/10.1023/A:1027332800296>

- Oldmeadow, J. A., Sutherland, C. A. M., & Young, A. W. (2012). Facial stereotype visualization through image averaging. *Social Psychology and Personality Science*, 00(0), 1–10. <https://doi.org/10.1177/1948550612469820>
- Oliveira, M., Garcia-Marques, T., & Dotsch, R. (2019). Combining Traits Into a Face: A Reverse Correlation Approach. *Social Cognition*, 37(5), 516–545. <https://doi.org/10.1521/soco.2019.37.5.516>
- Oliveira, M., Garcia-Marques, T., Dotsch, R., & Garcia-Marques, L. (2019). Dominance and competence face to face: Dissociations obtained with a reverse correlation approach. *European Journal of Social Psychology*, 00, 1–15. <https://doi.org/10.1002/ejsp.2569>
- Oliveira, M., Garcia-Marques, T., Garcia-Marques, L., & Dotsch, R. (2019). Good to Bad or Bad to Bad? What is the relationship between valence and the trait content of the Big Two?. *European Journal of Social Psychology*, 00, 1–21. <https://doi.org/10.1002/ejsp.2618>
- Olivola, C. Y., Tingley, D., & Todorov, A. (2018). Republican voters prefer candidates who have conservative-looking faces: new evidence from exit polls. *Political Psychology*, 39(5), 1157–1171. <https://doi.org/10.1111/pops.12489>
- Olivola, C. Y., & Todorov, A. J. (2010). Elected in 100 milliseconds: Appearance-Based Trait Inferences and Voting. *Journal of Nonverbal Behavior*, 34(2), 83–110. <https://doi.org/10.1007/s10919-009-0082-1>
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087–11092. <https://doi.org/10.1073/pnas.0805664105>
- Peabody, D. (1967). Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology*, 7(4), 1–18. <https://doi.org/10.1037/h0025230>
- Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law*, 16(6), 477–491. <https://doi.org/10.1080/10683160902926141>
- Ratner, K. G., Dotsch, R., Wigboldus, D. H., van Knippenberg, A., & Amodio, D. M. (2014). Visualizing minimal ingroup and outgroup faces: Implications for impressions, attitudes, and behavior. *Journal of personality and social psychology*, 106(6), 897–911. <https://doi.org/10.1037/a0036498>
- Schmitz, M., & Yzerbyt, V. Y. (in press). Direct and indirect compensation: Is there a difference between outsiders and insiders?. *European Journal of Social Psychology*.
- Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096–1104. <https://doi.org/10.3758/BRM.42.4.1096>
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24–31. <https://doi.org/10.1177/0098628316677643>

- Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour*, 1–11. <https://doi.org/10.1038/s41562-019-0800-6>
- Stolier, R. M., Hehman, E., Keller, M. D., Walker, M., & Freeman, J. B. (2018). The conceptual structure of face impressions. *Proceedings of the National Academy of Sciences*, 115(37), 9210–9215. <http://dx.doi.org/10.1073/pnas.1807222115>
- Sutherland, C. A. M., Liu, X., Chu, Y., Zhang, L., Oldmeadow, J., & Young, A. W. (2015). Chinese perceivers' facial first impressions. *Journal of Vision*, 15(12), 1218. <http://dx.doi.org/10.1167/15.12.1218>
- Sutherland, C. A. M., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Personality and Social Psychology Bulletin*, 44(4), 521–537. <http://dx.doi.org/10.1177/0146167217744194>
- Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, 127(1), 105–118. <http://dx.doi.org/10.1016/j.cognition.2012.12.001>
- Sutherland, C. A. M., Oldmeadow, J. A., & Young, A. W. (2016). Integrating social and facial models of person perception: Converging and diverging dimensions. *Cognition*, 157, 257–267. <https://doi.org/10.1016/j.cognition.2016.09.006>
- Tiddeman, B. P., Stirrat, M. R., & Perrett, D. I. (2005). Towards realism in facial image transformation: Results of a Wavelet MRF method. *Computer Graphics Forum*, 24, 449–456. <https://doi.org/10.1111/j.1467-8659.2005.00870.x>
- Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, 1124(1), 208–224. <https://doi.org/10.1196/annals.1440.012>
- Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: a model based approach. *Social cognitive and affective neuroscience*, 3(2), 119–127. <https://doi.org/10.1093/scan/nsn009>
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308(5728), 1623–1626. <https://doi.org/10.1126/science.1110589>
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*, 66, 519–45. <https://doi.org/10.1146/annurev-psych-113011-143831>
- Toscano, H., Schubert, T. W., Dotsch, R., Falvello, V., & Todorov, A. (2016). Physical strength as a cue to dominance: A data-driven approach. *Personality and Social Psychology Bulletin*, 42(12), 1603–1616. <https://doi.org/10.1177/0146167216666266>
- Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *PNAS*, 111(32), E3353–E3361. <http://dx.doi.org/10.1073/pnas.1409860111>

- Yzerbyt, V. Y. (2016). Intergroup stereotyping. *Current Opinion in Psychology*, 11, 90–95. <https://doi.org/10.1016/j.copsyc.2016.06.009>
- Yzerbyt, V. Y. (2018). The dimensional compensation model: Reality and strategic constraints on warmth and competence in intergroup perceptions. In *Agency and communion in social psychology* (pp. 126–141). Routledge. <https://doi.org/10.4324/9780203703663>
- Yzerbyt, V. Y., & Cambon, L. (2017). The dynamics of compensation: When ingroup favoritism paves the way for outgroup praise. *Personality and Social Psychology Bulletin*, 43(5), 587–600. <https://doi.org/10.1177/0146167216689066>
- Yzerbyt, V. Y., & Judd, C. M. (2020). *Comparing effect sizes between within-participant and between-participants designs*. Manuscript in preparation, University of Louvain.
- Yzerbyt, V. Y., Kervyn, N., & Judd, C. M. (2008). Compensation versus halo: The unique relations between the fundamental dimensions of social judgment. *Personality and Social Psychology Bulletin*, 34(8), 1110–1123. <https://doi.org/10.1177/0146167208318602>
- Yzerbyt, V. Y., Provost, V., & Corneille, O. (2005). Not Competent but Warm... Really? Compensatory stereotypes in the French-speaking world. *Group Processes & Intergroup Relations*, 8(3), 291–308. <https://doi.org/10.1177/1368430205053944>
- Zebrowitz, L. A., Fellous, J. M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review*, 7(3), 194–215. https://doi.org/10.1207/S15327957PSPR0703_01
- Zebrowitz, L. A., & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Social and personality psychology compass*, 2(3), 1497. <https://doi.org/10.1111/j.1751-9004.2008.00109.x>

Chapter 6

Introducing the Brief Reverse Correlation

An improved tool to assess visual representations

The reverse correlation (RC) is an innovative method to capture visual mental representations (i.e., classification images, CIs) of social targets that has become increasingly popular in social psychology. Because CIs of high quality are difficult to obtain without a large number of trials, the majority of past research relied on CIs extracted from samples of participants (average CIs). This strategy, however, leads to inflated false positivity rates. Using the representation from each participant (individual CIs) offers one solution to this problem. Still, this approach requires large numbers of trials and is thus economically costly, time demanding, demotivating for the participants, or simply impractical. We introduce a new version of the reverse correlation method, namely the Brief-RC. The Brief-RC increases the quality of individual (and average) CIs and reduces the overall task length by increasing the number of stimuli (i.e., noisy faces) presented at each trial. In two experiments, assessments by external judges confirm that the new method delivers equally good (Experiment 1) or higher-quality (Experiment 2) outcomes than the traditional method for the same number of trials, time length, and number of stimuli. The Brief-RC may thus facilitate the production of higher-quality individual CIs and alleviate the risk of false positivity rate.

Keywords: reverse correlation, Brief-RC, classification images, rcirc package, infoVal

Reference

Schmitz, M., Rougier, M., & Yzerbyt, V. (under review). Introducing the Brief Reverse Correlation: An improved tool to assess visual representations. Manuscript under review in *Journal of Experimental Social Psychology*.

Introducing the Brief Reverse Correlation

An improved tool to assess visual representations

1. Introduction

The two-image forced choice reverse correlation (Brinkman, et al., 2017; Dotsch & Todorov, 2011; Dotsch et al., 2008; Mangini & Biederman, 2004) is a method that provides visual proxies of mental representations (mostly faces). In recent years, the method has become increasingly popular, especially in social psychology (for a review, see Brinkman et al., 2017). This approach allows capturing facial representations at a group-level (i.e., from a sample of participants in a given condition) as well as at an individual-level (i.e., from a single participant). Although facial renderings produced at a group-level are of better quality than those at an individual-level, the evaluation of the former is prone to inflated Type I errors (Cone et al., 2020). Additionally, only the individual-level allows more fine-grained analyses (e.g., correlating judgments of the visual rendering with another individual-level variable). However, whereas individual-level visual outcomes are preferable, a large number of trials is usually required to achieve high-quality outcomes, which entails other issues (i.e., economically costly, time demanding, decreased participants' motivation to complete the task in a conscientious manner; Brinkman, Goffin et al., 2019; Todorov et al., 2011). Researchers have therefore noted the need to improve the method in order to “generate higher quality outcomes or reduce the number of trials” (Todorov et al., 2011, p. 787). With these concerns in mind, we propose an improved version of the method, namely the Brief Reverse Correlation (Brief-RC). This new method aims to address the issues just mentioned by increasing the number of stimuli at each trial, thereby reducing the overall number of trials and task length while improving the outcome quality.

1.1 The Reverse Correlation Paradigm

The reverse correlation (RC) is a data-driven method which provides access to the visual representation (generally facial) that people have of a given target (e.g., how a criminal person's face looks like). Over the years, the procedure has become widespread in social psychology and has proven to be particularly useful to identify the diagnostic features that drive social perception or to examine how top-down processes can bias mental images of social targets (see Brinkman et al., 2017; Jack & Schyns, 2017; Todorov et al., 2011; Todorov et al., 2013). For instance, it contributed to uncover facial diagnostic components of social categories such as ethnicity and race (e.g., Dotsch et al., 2008; Dotsch et al., 2011; Hinzman & Maddox, 2017; Krosch & Amodio, 2014; Kunst, Dovidio et al., 2017), gender (e.g., Brooks et al., 2018; Degner et al., 2019; Gundersen & Kunst, 2018), country of origin (e.g., Imhoff & Dotsch, 2013; Imhoff et al., 2011), profession and occupation (e.g., Degner et al., 2019; Lloyd et al., 2020), age (e.g., Albohn & Adams, 2020), religion (e.g., Brown-Iannuzzi et al., 2018) but also personality traits (e.g., Dotsch & Todorov, 2011; Imhoff et al., 2013; Lin et al.,

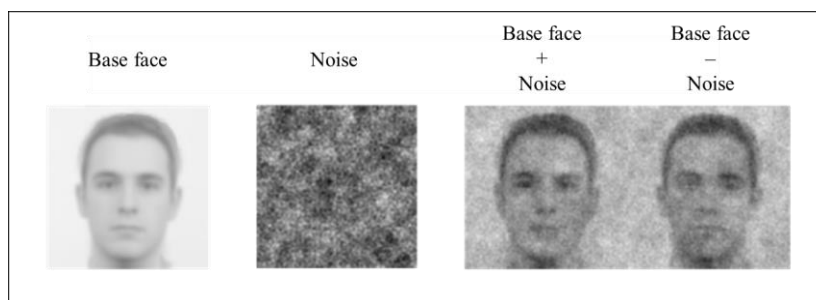
2018; Oliveira et al., 2019), and emotions (e.g., Albohn & Adams, 2020; Brooks et al., 2018). Furthermore, the RC method provides insights on how these mental templates could be distorted by a priori preferences, attitudes, beliefs, and knowledge, such as political ideology (e.g., Jackson et al., 2018; Young et al., 2013), love and attraction (e.g., Gunaydin & DeLong, 2015; Karremans et al., 2011), stereotypes and prejudice (e.g., Brown-Iannuzzi et al., 2016; Brown-Iannuzzi et al., 2018; Dotsch et al., 2008; Hinzman & Maddox, 2017), group membership (e.g., Hong & Ratner, 2020; Ratner et al., 2014), dehumanization (Kunst, Kteily et al., 2017; Petsko et al., in press), or behavioral information about novel groups (Dotsch et al., 2013).

The RC is a method rooted in signal-detection theory (Ahumada & Lovell, 1971; Ahumada et al., 1975) that aims to identify the information that underlies perception. Essentially, the method estimates the diagnostic information (i.e., the signal) that drives perception in random variations of the stimulus (Jack & Schyns, 2017). In other words, this technique tries to capture people's expectations about a given target (i.e., the signal; e.g., a happy face in visual perception) by presenting them with noisy stimulus (e.g., a face with random noise added) and retaining those that, by pure coincidence, happen to match their expectations. For instance, a random noise would likely be selected as a happy face if it slightly distorts the mouth so that it appears smiling (e.g., Kontsevich & Tyler, 2004).

Although several implementations of this paradigm exist in the domain of visual perception (see Jack & Schyns, 2017; Todorov et al., 2011; Todorov et al., 2015), the two-image forced choice noised-based reverse correlation (hereafter Traditional-RC) developed by Dotsch et al. (2008; see also Dotsch & Todorov, 2011) is by far the most prevalent. A typical procedure (e.g., Dotsch et al., 2008) involves two steps. In the first step, participants have to choose, across many trials, between two random variations of the same base face (i.e., noisy faces) the one that best matches their mental representation of the category of interest (e.g., "Choose the most Moroccan-looking face"). The random instantiations consist in superimposing (by either adding or subtracting) a noise pattern (for the generation of the noise pattern see Dotsch & Todorov, 2011; Mangini & Biederman, 2004) onto a base image (usually a morph of several faces) as illustrated in Figure 1.

Figure 1

Base face (from Experiment 1), example of a random noise pattern, and example of a pair of stimuli (i.e., noisy faces) produced by adding (left) or subtracting (right) the noise pattern from the base face



Each noisy face in a pair is maximally different from the other of the same pair since one is the mathematical opposite of the other (e.g., the luminance value of a given pixel in one image will be the mathematical opposite on the other). In essence, both faces are equidistant to the base face and any difference in classifications can only stem from the noise pattern (Dotsch & Todorov, 2011). To build the representation of a participant or a group of participants, one adds the average of all selected noise patterns to the base face. This allows obtaining the so-called *classification image* (CI), i.e., the visual proxy of the mental representation of the target by a single participant (i.e. participant-level or individual CI) or by a group of participants (i.e., group-level or average CI) (for the interpretation of CIs, see Brinkman, et al., 2017). In the second step, a new sample of participants (i.e., the judges) rates the CIs on the variables of interest, allowing a test of researchers' hypotheses. In one illustrative research, Dotsch et al. (2008) captured the average facial template of Moroccans from low, moderate, and highly prejudiced individuals (as measured with an Implicit Association Task). The authors then asked a sample of independent judges to rate these representations on criminality and trustworthiness. Results revealed that the more prejudiced the individual, the more negative their representation of this group.

This procedure offers several advantages to study psychological effects (see Brinkman et al., 2017; Dotsch & Todorov, 2011; Jack & Schyns, 2017). First, it makes no prior assumption about what people's internal representations may look like. In other words, it allows sampling a vast set of hypotheses (that the researcher may not even have thought of or theoretically anticipated) instead of a single one and doing so agnostically. Second, participants' responses are not primed in any direction, as with pre-selected response labels which could bias their responses (e.g., "How aggressive are Moroccans?" from 1 = not at all to 7 = totally). The RC thus reveals "near spontaneous use of information" (Brinkman et al., 2017, p. 336) because participants may adopt any criteria they want for the classification task (i.e., the method is said to be unconstrained). Actually, it may even allow probing representations that are ineffable for the participant (Mangini & Biederman, 2004).

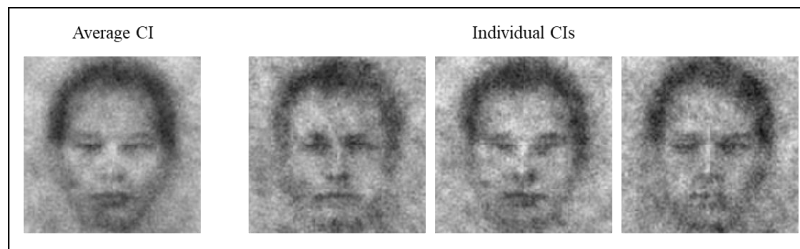
1.2 Current Limitations of the Traditional-RC

The majority of studies based on the Traditional-RC have relied solely on the average CIs instead of individual CIs. The reason is that the former are usually less noisy than the latter (see Figure 2; Cone et al., 2020; Imhoff et al., 2011; Imhoff et al., 2013). However, average CIs present several shortcomings. They prevent carrying out idiosyncratic analyses, for instance, by analyzing the relation between racial bias and visual representation bias at the participant-level (Dotsch et al., 2008). In addition, they are only valid to the extent that participants actually share a common representation of the target category, which is not necessarily the case (Brinkman et al., 2017; Ratner et al., 2014). Also, judges may detect significant differences between group-wise CIs that may not be materialize if the inter-individual variability were to be considered, that is, if judges had to rate the individual CIs. Indeed, Cone et al. (2020) conducted computer simulations and showed that using average CIs in typical two-phase reverse correlations procedures (the first phase corresponding to the production of the CIs and the second to their rating by external judges) may increase

considerably Type I error rates because they do not take into account the inter-individual variability of the underlying CIs. These authors estimated that about 66% of past research might suffer from inflated false positivity rates. Building on this result, they recommend using individual CIs (and other methods that are currently being developed) to address this problem.

Figure 2

Example of an average CI and some individual CIs issued from the Traditional-RC after 526 trials (from Experiment 2)



Although relying on individual CIs would address the abovementioned concerns (i.e., they reflect the inter-individual variability and thus reduce inflated Type I error rates; they grant access to fine-grained analyses), they require a very large number of trials to achieve an acceptable level of quality. This is because they are composed of much fewer trials than the average CIs (Brinkman et al., 2017; Cone et al., 2020; Jack & Schyns, 2017; Ratner et al., 2014; Todorov et al., 2011). Indeed, RC procedures usually comprise 300–1,000 trials per individual-level CI (and thus 300–1,000 trials multiplied by the number of participants within a condition for average-level CIs; Brinkman et al., 2017; Brinkman, Goffin et al., 2019; Cone et al., 2020; Dotsch & Todorov, 2011). Although adding more trials should in principle enhance the quality of individual CIs, in practice, however, it does not always bear fruit. In fact, doing so may even be detrimental because this hinders participants' motivation to mindfully complete the task (Brinkman, Goffin et al., 2019; Todorov et al., 2011; see also Lick et al., 2013). Furthermore, increasing the length of the task can be time-consuming, economically costly, or simply impractical. Importantly, if the task is too short to properly cancel out the noise or if participants' responses become erratic (e.g., because of their lack of motivation), the risk is to interpret a noisy CI as being meaningful. This would increase the chances of false positives when judging low quality individual CIs as well as the resulting average CIs. Indeed, the RC method will always yield a CI whatever the number of trials or the meaningfulness of responses (Brinkman, Goffin et al., 2019). To address this issue, Brinkman, Goffin et al. (2019) developed infoVal, an information value metric that assesses the degree to which an individual CI derives from signal rather than noise. However, although this metric allows identifying noisy individual CIs, it does not preclude obtaining poor quality renderings.

A different approach to improving individual CIs would be to take into account the response confidence by providing multiple response alternatives (e.g., “probably happy”,

“possibly happy”, “possibly unhappy”, or “probably unhappy”) and only built the CIs from trials with high confidence responses (e.g., Brinkman, Dotsch, et al., 2019; Mangini & Biederman, 2004; see also Dai & Michéyl, 2010). However, this technique implies a contrast between target categories (e.g., happy vs. unhappy) that is not necessarily desired as researchers may want to tap into a single target category (e.g., happy) without contrasting it with another one (Dotsch & Todorov, 2011). Furthermore, this strategy fails to shorten task length and may require even more trials if the rate of high-confidence responses is low (Brinkman et al., 2017).

In sum, and as of today, there is no practical and reliable solution to improve the quality of individual CIs and thus reduce the inflated Type I errors rates stemming from the use of average CIs or poor-quality individual CIs. This matter becomes even more crucial in light of the replicability crisis in psychological science where strong and reliable science should be encouraged (Open Science Collaboration, 2015; Simmons et al., 2011).

1.3 The Brief Reverse Correlation

Because no strategy currently allows reducing the task length while simultaneously improving the outcome quality, only a limited number of RC-based studies conducted analyses of individual CIs (e.g., Brooks et al., 2018; Degner et al., 2019; Dotsch et al., 2008, 2013; Imhoff et al., 2013). In short, researchers have urged to “generate higher quality outcomes or reduce the number of trials” (Todorov et al., 2011, p. 787). In light of these concerns, we propose an improved version of the paradigm, the Brief-RC. The key difference between the Brief-RC and the Traditional-RC is that the former present participants with a greater number of noisy faces to select from at each trial.

In the Brief-RC, the likelihood to find a stimulus that carries more diagnostic information of the expected signal, i.e., with a higher signal-to-noise ratio (SNR), at each trial should be higher because the set of options is larger. For instance, if we are searching for a happy-looking face, the higher the number of noisy faces at each trial, the higher the probability that one of them will match (to some degree and by chance alone) our expectation of a happy-looking face. Conversely, the fewer the noisy faces, the more likely it is that we end up picking one purely at random because none happens to prove close enough to our mental template of the target. A greater panel of faces should therefore improve the SNR of the selected noisy face at each trial and consequently accelerate the convergence toward a higher quality and more robust CI, both at the individual and at the average levels. To be sure, the improvement in SNR resulting from the increase of the stimulus set at each trial should reach a plateau because the visual processing capacity of the human brain is limited (Marois & Ivanoff, 2005; Palmer, 1990). Still, a reasonable increase in the selection presented to the participants should deliver clearer CIs for a given number of trials. This means that, in general, one could obtain high quality CIs with fewer trials and a shorter time of completion. Said otherwise, the same quality-level CI issued from a Traditional-RC might be achieved with a Brief-RC after fewer trials and, thus, a shorter amount of time. In sum, we argue that in comparison to the Traditional-RC, the Brief-RC should improve the quality of the individual CIs without hampering participants’ motivation with additional trials.

1.4 The Present Research

In two experiments, we compared the quality of individual and average CIs produced from the Traditional-RC with 2 stimuli (i.e., noisy faces) per trial to the Brief-RC with more than two stimuli per trial. Because factors such as limitations to process visual information may constrain the benefits of the increase in the number of trials, we tested this assumption empirically with two versions of the Brief-RC—namely the Brief-RC12, with 12 stimuli per trial, and the Brief-RC20, with 20 stimuli per trial. We assessed the quality of the visual renderings by means of a subjective (ratings of the CIs from independent judges) and an objective measure (infoVal). In order to minimize the variability of both the to-be-measured facial representation and the ratings from judges, we opted for a social group—Chinese people—for which there is a relatively clear and homogeneous visual representation among our participants—i.e., Americans (e.g., other-race effect; Ge et al., 2009; Kelly et al., 2007; out-group homogeneity effect; Judd & Park, 1988; Lee & Ottati, 1995).

In Experiment 1, we fixed the overall number of stimuli. Because of the improved SNR, we expected the two variants of the Brief-RC to perform at the same level or better than the Traditional-RC. The decision to rely on the same number of stimuli overall can be seen as rather conservative because the number of trials in the Traditional-RC will be much larger than the Brief-RC for the same number of stimuli presented. In Experiment 2, we sought to replicate findings from Experiment 1. Furthermore, we increased the length of the task and compared the RC variants across different criteria (task length, number of trials, and number of stimuli). Indeed, all these features may have important practical implications in terms of improving research practices, namely, to achieve better quality outcomes, to reduce the task length (number of trials and time), to sustain participants' motivation, and to reduce the researcher's financial cost and time to carry out the task. We expected the Brief-RC (Brief-RC12 and Brief-RC20) to outperform the Traditional-RC when relying on the same task length and number of trials, while we predicted at least as good a performance when holding constant the number of stimuli (as in Experiment 1). We specifically geared our predictions toward individual CIs because they are usually noisier than average CIs and because enhancing individual CIs quality will contribute to a more replicable social psychology science (Cone et al., 2020; Open Science Collaboration, 2015; Simmons et al., 2011).

2. Overview and Analytical Strategy

The experiments reported in the present research followed the standard implementation of the reverse correlation paradigm (Dotsch et al., 2008). Each experiment comprised two parts involving two distinct samples of participants. In the first part, a sample of participants called “producers” completed one of the three version of the RC task (i.e., Traditional-RC, Brief-RC12, or Brief-RC20) in order to capture their visual representation of a Chinese-looking face. In the second part, another sample of participants called “judges” evaluated these CIs on how Chinese they look. All participants took part via the online crowdsourcing platform Prolific Academic (www.prolific.ac) in exchange for a monetary compensation (5£/hour). Participants were from the United States, they did not participate in any of our

previous similar studies, and they had an approval rate of at least 95% to ensure data quality (Peer et al., 2013). The present research was conducted in a manner consistent with the APA's Ethical Principles.

Whenever possible, we conducted the analyses using linear mixed models (LMM; Judd et al., 2012; Judd et al., 2017; Westfall et al., 2014). Usually more conservative, a mixed model approach also allows generalizing the results across both “judges” (judges' unique ids) and “producers” (producers' unique ids or equivalently CIs' unique ids), our two random factors. Maximal LMM (i.e., the ones that fit the full variance-covariance structure of random effects) usually comes with a significant loss of power and may fail to converge. Therefore, we only report the most parsimonious LMM (i.e., the ones that maximize power while trying to minimize the Type I error rate), following Bates et al. guidelines (Bates et al., 2015; see also Matuschek et al., 2017). The R scripts (available in the Open Science Framework link, see below) detail the model selection process. We report effect sizes for LMM computed with the *r2glmm* package (version 0.1.2, Jaeger, 2017). Of note, LLM effect sizes tend to be considerably smaller than those for by-judges or by-faces analyses because averaging across responses (either across judges or faces) drastically reduces the standard errors (Brysbaert & Stevens, 2018).

We created a set of a priori orthogonal contrast codes to compare the performance of the different RC variants. The first contrast, C_1 , compares the Traditional-RC (Traditional-RC coded $-2/3$) to the Brief-RCs versions (Brief-RC12 and Brief-RC20 both coded $+1/3$), whereas the second contrast, C_2 compares the two Brief-RC variants to each other, that is Brief-RC12 (coded -0.5) and Brief-RC20 (coded $+0.5$; with Traditional-RC coded 0).

On some occasions, we predicted an absence of difference between conditions. To better gauge the evidence in favor of the null hypothesis, we report the Bayes factors¹ (BF_{01}) associated with the specific predictor when the OLS or LMM did not yield a significant result (Dienes, 2014). We interpret the Bayes factor according to Jeffreys (1961) guidelines in which a BF_{01} is considered as anecdotal (1–3), substantial (3–10), strong (10–30), very strong (30–100), or decisive (>100) evidence in support of null hypothesis (H_0) as opposed to the alternative hypothesis (H_1).

To assess the quality of the average and individual CIs, we relied on infoVal, an innovative informational value metric developed by Brinkman, Goffin et al. (2019; see also Schmitz, Rougier, & Yzerbyt, 2020; Schmitz, Rougier, Yzerbyt, Brinkman et al., 2020) that quantifies the degree to which a CI contains signal rather than noise. The infoVal metric comes as a new standard of good practice because it addresses the issue of erroneously interpreting the CI as meaningful when it is not. To improve the accuracy of the metric, we applied an oval-shaped mask to the CIs in order to extract the face region (e.g., Oliveira et al., 2019; Ratner et al., 2014) and compute the infoVal score on this region.

¹ Bayes Factor (BF) estimations were derived from the Bayesian Information Criterion (BIC) with a “unit information prior” following the guidelines from Wagenmakers (2007).

We preregistered all experiments on Open Science Framework (OSF; including a priori theoretical reasoning, hypotheses, power estimations, procedures, and statistical analyses). We report any significant deviations from the initial pre-registrations in the core manuscript. We report all measures, manipulations, and exclusions in these studies. Our pre-registrations, data, data analysis R scripts for all experiments, and JavaScript scripts to run both Traditional-RC and the Brief-RC variants are available on the following link: https://osf.io/ps9wu/?view_only=028597d60e7342bfaeb6881051cb6bca.

3. Experiment 1

In Experiment 1, we compared the performance of two variants of the Brief-RC (Brief-RC12 and Brief-RC20) to the Traditional-RC (Traditional-RC). We kept constant the overall number of stimuli (720 noisy faces) across conditions, while the number of trials, stimuli per trial, and completion time varied accordingly. We expected the Brief-RC to perform as well or possibly better than the Traditional-RC.

3.1 Method

Participants. Although there are no specific good practices regarding how to design an RC experiment (Brinkman et al., 2017), we relied on a sample size that was similar to previous studies (e.g., $N = 28$ for a three-level between-participants design; Dotsch et al., 2008, Experiment 1). In the first part, we implemented a similar design. To increase statistical power, we recruited a sample of 67 producers (and thus 67 individual CIs). In line with pre-registration, we excluded one participant (from the Traditional-RC condition) due to abnormally high percentage (48%) of fast reaction times ($< 200\text{ms}$). The final sample comprised 66 “producers” ($n = 22$ per condition; $M_{\text{age}} = 31.86$, $SD_{\text{age}} = 11.52$; 36 males). In the second part, we recruited 70 independent judges ($M_{\text{age}} = 29.79$, $SD_{\text{age}} = 8.52$; 33 males and 2 others—not identifying as either male or female). We based the judges’ sample size on previous research that usually relies on 30–90 raters (e.g., Dotsch et al., 2008; Dotsch et al., 2013; Imhoff et al., 2013). Sample size was determined before any data analysis. A post-hoc sensitivity analysis revealed that the current configuration had an 80% power to detect an effect size $\eta_p^2 \geq 0.044$ for the C_1 contrast (opposing the Traditional-RC to the Brief-RC12 and Brief-RC20).

Experimental design and procedure. We first asked producers to complete one of the RC variants and then invited judges to rate the produced CIs.

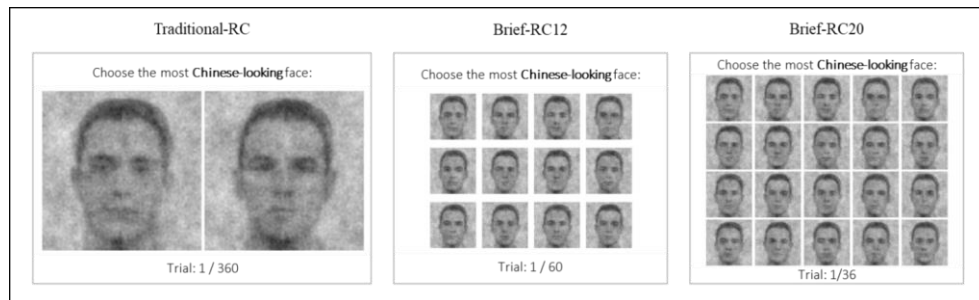
Part 1: Reverse correlation task. We randomly assigned producers to one of the three RC tasks (Traditional-RC, Brief-RC12, or Brief-RC20). We selected six Caucasian male faces from the Radboud Face Database (Langner et al., 2010), merged and blurred them (using a low spatial frequency filter) into a single black and white base image using Psychomorph (Tiddeman et al., 2005). With the base image, we then generated 360 pairs of

stimuli (i.e., noisy faces) using the *rcicr* R package (development version²) with the default settings. Pairs consisted in either adding (“oriented”) or subtracting (“inverted”) a sinusoidal noise pattern from the base image (see Dotsch & Todorov, 2011; Mangini & Biederman, 2004). All stimuli initially had a 512×512 pixels size and were rescaled to 150×150 pixels size in the case of the Brief-RC12 and Brief-RC20 so they could fit into the window screen, while they kept their original size in the Traditional-RC. We built the reverse correlation tasks with a JavaScript library called jsPsych (www.jspsych.org; de Leeuw, 2014; version 6.0.3).

All participants received the following instruction: “*In this task, you will be presented with a series of noisy faces. Your task will be to ‘Choose the most Chinese-looking face’ in each trial. Use your mouse to make a choice. Note that the faces may look similar to each other, yet they are different. Try to rely on your intuition to make the best choice. Please remain fully concentrated*”. The Traditional-RC task comprised 360 trials with two adjacent noisy faces (one oriented and one inverted). The Brief-RC12 comprised 60 trials with 12 noisy faces per trial (6 oriented and 6 inverted) arranged in a 3×4 grid. Finally, the Brief-RC20 comprised 36 trials with 20 noisy faces per trial (10 oriented and 10 inverted) arranged in a 4×5 grid (see Figure 3). In each condition, we randomly generated a trial order and a stimuli position within each trial and kept it constant across all the participants. At the end, participants provided demographic information through a Qualtrics survey before being debriefed.

Figure 3

Illustration of a single trial from the Traditional-RC, Brief-RC12 and Brief-RC20 tasks. The instruction is displayed at the top of the screen and reads: “Choose the most Chinese-looking face”. The current and total number of trials are displayed at the bottom of the screen



We computed each participant’s individual CI by averaging all the noise patterns extracted from the selected noisy faces and superimposing it on the base image³. The CIs were created with a constant scaling that was different as a function of the type of CI

² We used the development version in order to obtain the noise matrices, which ease the computation of the CIs and infoVal. The development version is available at <https://github.com/rdotsch/rcicr/tree/development>

³ We adapted the function from the *rcicr* package (Dotsch, 2017) to compute the CIs in our experiments in order to better handle large noise matrices.

(individual vs. average) and the type of condition (Traditional-RC vs. Brief-RC12 vs. Brief-RC20). We opted for a constant scaling because the noise range differed between conditions and levels of aggregation (average vs. individual). The selected constants⁴ were those minimizing the perceptual difference (based on visual inspection) in noise level in the CIs between conditions. Finally, we computed the *infoVal* metric for the individual CIs.

Part 2: CIs rating task. A new sample of participants (judges) rated the 66 individual CIs and the 3 average CIs via a Qualtrics survey. Judges were told that they were about to rate several blurred faces by the means of a 7-point scale based on how Chinese-looking the faces were (from 0 = *neutral* to 6 = *extremely Chinese looking*)⁵. Before the rating task, all individual CIs were presented at once (through a matrix grid) for one minute to allow the judges to gauge the similarities and differences between them. Then, judges rated in a first block the randomly ordered individual CIs one by one. In a second block, they rated the three average CIs presented simultaneously in a matrix with one row per CI (CIs position was randomized between participants). Next, judges provided demographic information, had the option to make a comment about the study, and were debriefed.

3.2 Results

Task completion time. We submitted the task completion time (in minutes) to our two contrasts of interest C_1 (opposing Traditional-RC to Brief-RC12 and Brief-RC20) and C_2 (comparing Brief-RC12 to Brief-RC20) in an OLS analysis. As expected, the task completion time was significantly longer—almost twice as long—for the Traditional-RC ($M = 10.81$, $Mdn = 10.74$, $SD = 4.55$) than for the Brief-RC12 and Brief-RC20 ($M = 5.72$, $Mdn = 5.45$, $SD = 3.49$), $F(1, 67) = 25.22$, $p < .001$, $\eta_p^2 = .286$. Interestingly, the completion time did not significantly differ between the Brief-RC12 ($M = 6.20$, $Mdn = 5.38$, $SD = 3.70$) and Brief-RC20 ($M = 5.23$, $Mdn = 5.45$, $SD = 3.28$), $F(1, 67) = 0.69$, $p = .410$, $\eta_p^2 = .011$, $BF_{01} = 5.67$.

Judges' ratings. We submitted judges' ratings on the Chinese-looking scale of the individual CIs to C_1 and C_2 in a LMM analysis, using judges and producers as random factors. In line with expectations, the individual CIs ratings did not differ significantly between the Traditional-RC ($M = 3.13$, $SD = 1.83$) and the Brief-RC12 and Brief-RC20 ($M = 3.39$, $SD = 1.82$), $F(1, 65.81) = 1.13$, $p = .292$, $\eta_p^2 = .004$, $BF_{01} = 38.86$, nor between the Brief-RC12 ($M = 3.46$, $SD = 1.78$) and the Brief-RC20 ($M = 3.32$, $SD = 1.86$), $F(1, 65.81) = 0.25$, $p = .616$, $\eta_p^2 = .001$, $BF_{01} = 59.89$.

Next, we submitted judges' ratings of the average CIs to C_1 and C_2 in an OLS analysis. The average CI ratings did not significantly differ between the Traditional-RC ($M = 4.77$, $SD = 1.08$) and the Brief-RC12 and Brief-RC20 ($M = 4.69$, $SD = 1.49$), $F(1, 67) = 0.16$, $p = .692$,

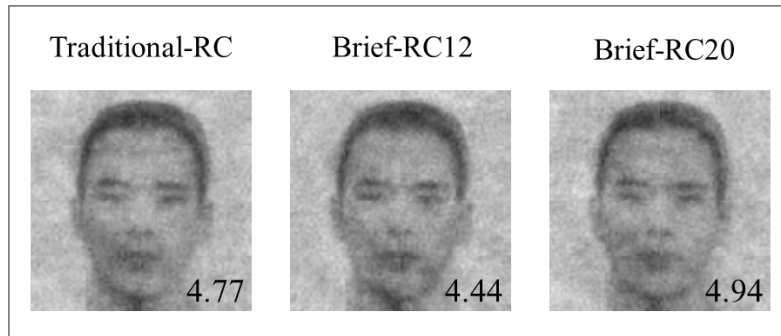
⁴ The following constants were used to scale the noise patterns for the average CIs: Traditional-RC = .0055, Brief-RC12 = .0130, Brief-RC20 = .0150; and for the individual CIs: Traditional-RC = .0190, Brief-RC12 = .0450, Brief-RC20 = .0600.

⁵ We slightly deviated from the pre-register scale values (-3 = *not at all Chinese-looking* to +3 = *very Chinese-looking*) to make the scale more meaningful to participants as in Brinkman, Goffin et al. (2019; 1 = “not masculine”, 9 = “very masculine”).

$\eta_p^2 = .001$, $BF_{01} = 13.38$. Interestingly, the average CI ratings of the Brief-RC12 ($M = 4.44$, $SD = 1.53$) were significantly lower than those of the Brief-RC20 ($M = 4.94$, $SD = 1.41$), $F(1, 67) = 4.77$, $p = .030$, $\eta_p^2 = .023$ (see Figure 4).

Figure 4

Average CIs ratings by condition. Average ratings are displayed in the lower-right corner of the CIs



InfoVal. We submitted the infoVal⁶ scores of the individual CIs to C_1 and C_2 in an OLS analysis. There were no significant differences between the Traditional-RC ($M = 1.54$, $SD = 2.08$) and the Brief-RC12 and Brief-RC20 ($M = 1.83$, $SD = 1.80$), $F(1, 67) = 0.34$, $p = .560$, $\eta_p^2 = .005$, $BF_{01} = 6.79$, nor between the Brief-RC12 ($M = 1.93$, $SD = 1.74$) and the Brief-RC20 ($M = 1.74$, $SD = 1.90$), $F(1, 67) = 0.10$, $p = .748$, $\eta_p^2 = .002$, $BF_{01} = 7.69$.

3.3 Discussion

In Experiment 1, we compared the visual representation of a Chinese-looking face as captured by the Brief-RC (Brief-RC12 and Brief-RC20) versus the Traditional-RC (Traditional-RC) after participants had been exposed to the same number of stimuli (720 noisy faces). Neither the individual nor the average CIs ratings from judges (on how Chinese they looked-like) differed significantly between the various methods. The same pattern emerged when relying on a more objective metric, namely the infoVal score of individual CIs. Overall, frequentist as well as Bayesian analyses of our data confirm that the Brief-RC and the Traditional-RC delivered similar outcomes. A crucial difference, however, is that fixing the number of stimuli as we did here resulted in the fact that the Brief-RC took half the time than the Traditional-RC and only required as little as 60 or 36 trials. A researcher with limited resources would thus have obvious interest in opting for the Brief-RC rather than for the Traditional-RC.

⁶ We did not declare a priori hypothesis concerning infoVal for Experiment 1 in the preregister since the reference for this metric (Brinkman, Goffin et al., 2019) was not available at that time. Nonetheless, we decided to include these analyses to better gauge the differences between the Brief-RC and Traditional-RC.

One caveat of Experiment 1 concerns the selected number of stimuli (720 noisy faces). Indeed, RC-based studies generally rely on more than 360 trials. This means that one would expect to see more stimuli presented to the participants in the first part of the experiment. Moreover, one could argue that the comparison between methods should also rest on such criteria as completion time and number of trials rather than number of stimuli alone because these factors affect participants' motivation and attention, as well as the time and economic resources from the researchers. Experiment 2 increased the length of the task and relied on completion time, number of trials, and number of stimuli as comparison criteria.

4. Experiment 2

In Experiment 2, we sought to replicate and extend the findings from Experiment 1 by comparing the performance of the Brief-RC (Brief-RC12 and Brief-RC20) versus the Traditional-RC (Traditional-RC) not only after the same number of stimuli (and thus for different points in time and numbers of trials) but also after the same time (and thus for different numbers of trials and stimuli) and after the same number of trials (and thus for different points in time and numbers of stimuli). We expected the Brief-RC variants to outperform the Traditional-RC when comparing ratings after the same time and number of trials. Regarding the number of stimuli, we expected to replicate the findings from Experiment 1. To increase statistical power, we increased the sample sizes of both producers, who deliver the individual CIs, and judges.

4.1 Method

Participants. Because our design involved linear mixed models, with producers (i.e., individual CIs) and judges as random factors, we relied on Judd, Westfall, and Kenny (2012; Judd et al., 2017) and Westfall et al. (2014) to select the number of producers and judges. Accordingly, we needed about 60 producers per condition and 150 judges to achieve a power of .80 to detect an effect of $d = 0.20$. To maximize power, we opted for 100 producers (and thus 100 individual CIs) per condition and 250 judges. We thus recruited a sample of 300 producers ($n = 100$ per condition; $M_{\text{age}} = 31.30$, $SD_{\text{age}} = 10.35$; 169 males, 4 others) and another sample of 253 independent judges. We excluded one judge who gave the same rating to all faces⁷. The final sample thus comprised 252 judges ($M_{\text{age}} = 30.50$, $SD_{\text{age}} = 10.04$; 126 females, 3 others). Sample size was determined before any data analysis. A post-hoc sensitivity analysis revealed that the current configuration had an 80% power to detect an effect size $\eta_p^2 \geq 0.013$ for the C_1 contrast (opposing the Traditional-RC to the Brief-RC12 and Brief-RC20).

⁷ It was established in the preregister that judges with less than 5% variation on their ratings or/and with a median rating time of less than 1 second per stimulus would be excluded. However, the median rating time per stimulus was 0.5 second, and 94.44% of judges had rating durations shorter than 1 second/stimulus. Given that this exclusion criterion was underestimated, we omitted it.

Experimental design and procedure. The overall procedure was similar to Experiment 1 with a few differences detailed below.

Part 1: Reverse correlation task. We estimated the number of trials for each RC task from Experiment 1 such that the median completion time length for the RC task in all three conditions would be approximately 10 minutes. We created two thousand pairs of noisy faces (stimuli) from a different base image and a different seed than Experiment 1 for generalizability purposes. Specifically, we relied on the same base image as in Dotsch et al. (2008). The Traditional-RC comprised 550 trials (2 noisy faces per trial; 1100 noisy faces in total), the Brief-RC12 comprised 250 trials (12 noisy faces per trial; 3000 noisy faces in total), and the Brief-RC20 comprised 200 trials (20 noisy faces per trial; 4000 in total). Because the total number of stimuli varied from one condition to another, we presented the noisy faces in the same order and kept the order of the trials constant between conditions. For instance, the two noisy faces in the first trial of the Traditional-RC were also part of the first trial of the Brief-RC12 and Brief-RC20.

Next, participants answered two questions concerning their perceived threat (2 items) and attitude (1 item) regarding Chinese people. We do not report about these measures here as they belonged to a separate study. Producers then gave their feedback about how boring (slider from 0 = *not boring at all* to 100 = *very boring*) and difficult (slider from 0 = *very easy* to 100 = *very difficult*) they found the task. Finally, they provided the same demographic information, had the option to make a comment about the study, and were debriefed.

We computed the individual and average CIs within each task at different points according to the criteria of comparison, namely: time (approximately 5 and 10 minutes), number of trials (90 and 167), and number of stimuli presented (approximately 1050). Table 1 presents the 11 distinct points of comparison⁸.

⁸ We deviated slightly from the points of comparison established in our preregister to minimize the number of CIs and judgments required. To do so, we selected points that could be used to compare more than one criterion (see Table 1). Importantly, these points were determined *before* computing the visual renderings.

Table 1

Points of comparison by Time, Number of trials, and Number of stimuli as a function of condition (Traditional-RC, Brief-RC12, and Brief-RC20)

Condition	Time	Number of trials	Number of stimuli	Points of comparison		
				By time	By trials	By stimuli
Traditional-RC	2.34	90	180		90 trials	
	4.19	167	334		167 trials	
	5.00	203	406	~5 mins		
	9.99	526	1052	~10 mins		~1050 stimuli
Brief-RC12	5.01	90	1080	~5 mins	90 trials	~1050 stimuli
	7.73	167	2004		167 trials	
	9.99	240	2880	~10 mins		
Brief-RC20	4.42	51	1020			~1050 stimuli
	5.02	62	1240	~5 mins		
	6.56	90	1800		90 trials	
	10.05	167	3340	~10 mins	167 trials	

We created the individual and average CIs with the auto-scale method that matches the noise pattern to the range of pixel of the base image. Although this method is suboptimal, it is common in RC experiments because it prevents from selecting an arbitrary constant (called “constant scaling”) that may introduce a bias (see Brinkman et al., 2017; and the documentation from the *rcicr* package at <https://rdrr.io/cran/rcicr/man/generateCI.html>). We computed the infoVal as in Experiment 1.

Part 2: CIs rating task. Judges first saw the same 55 randomly sampled individual CIs (5 CIs randomly sampled from each of the 11 distinct points of comparison; see Table 1) as to better gauge the differences between CIs. These illustrative CIs appeared as a grid for one minute on the same page in a random fixed order. Judges then rated 88 individual CIs (8 individual CIs randomly sampled from each point of comparison; see Table 1). We presented individual CIs in a matrix (each row corresponding to one CI) with the row order randomized across participants. Next, judges saw the 11 average CIs corresponding to the 11 points of comparison for twenty seconds before rating each of them. In the rating task, average CIs appeared one by one and in random order across participants. Next, judges provided demographic information, had the possibility to make a comment about the study, and were debriefed.

4.2 Results

Judges' ratings as a function of time. We submitted judges' ratings of the individual CIs to a LMM analysis with C_1 (Traditional-RC versus Brief-RC12 and Brief-RC20), C_2 (Brief-RC12 versus Brief-RC20), time (5 versus 10 minutes) and their interaction as predictors, using judges and producers as random factors. As predicted, C_1 was significant such that the individual CIs from the Traditional-RC ($M = 2.28$, $SD = 1.95$) looked less Chinese than those from the Brief-RC12 and Brief-RC20 ($M = 2.81$, $SD = 2.06$), $F(1, 319.34) = 13.71$, $p < .001$, $\eta_p^2 = .010$. As expected, there was a main effect of time such that the individual CIs after 5 minutes ($M = 2.56$, $SD = 1.99$) looked significantly less Chinese than those after 10 minutes ($M = 2.72$, $SD = 2.08$), $F(1, 247.84) = 19.14$, $p < .001$, $\eta_p^2 = .001$. Moreover, the $C_1 \times$ time interaction was significant, $F(1, 247.44) = 5.73$, $p = .017$, $\eta_p^2 < .001$. The individual CIs from the Traditional-RC ($M = 2.25$, $SD = 1.90$) looked less Chinese than those from the Brief-RC12 and Brief-RC20 ($M = 2.71$, $SD = 2.02$) after 5 minutes, $F(1, 341.57) = 9.27$, $p = .003$, $\eta_p^2 = .005$, this pattern being even more pronounced after 10 minutes ($M_{\text{traditional-RC}} = 2.31$, $SD_{\text{traditional-RC}} = 2.00$; $M_{\text{Brief-RC12\&Brief-RC20}} = 2.92$, $SD_{\text{Brief-RC12\&Brief-RC20}} = 2.10$), $F(1, 336.07) = 17.99$, $p < .001$, $\eta_p^2 = .010$. There was no significant effect of C_2 nor of $C_2 \times$ time (all p 's $> .10$, BF_{01} 's > 10).

Next, we submitted judges' ratings of the average CIs to an OLS analysis with C_1 , C_2 , time, and their interaction as predictors. As expected, there was a significant main effect of C_1 such that the average CIs looked less Chinese for the Traditional-RC ($M = 4.55$, $SD = 1.38$) than for the Brief-RC12 and Brief-RC20 ($M = 4.85$, $SD = 1.21$), $F(1, 246) = 19.74$, $p < .001$, $\eta_p^2 = .013$. There was no significant effect of C_2 , time, $C_1 \times$ time, nor $C_2 \times$ time (all p 's $> .10$, BF_{01} 's > 10).

Figure 5 (panel A) shows the results from judges' ratings of the individual and average ratings of CIs as a function of time. Figure 6 illustrates the average CIs as a function of time and their ratings.

Figure 5

Relation between individual CIs ratings (averaged) and (A) the time (median cumulative time in minutes), (B) the number of trials, and (C) the number of stimuli, as a function of condition (different colored lines: Traditional-RC, Brief-RC12, and Brief-RC20), and CI type (solid lines represent average CIs and the dashed lines individual CIs). The dots represent average ratings at given points of comparison (cf. Table 1). The grey frames highlight the points of comparison, i.e., (A) at approximately 5 and 10 minutes, (B) at 90 and 167 trials, and (C) at approximately 1050 stimuli.

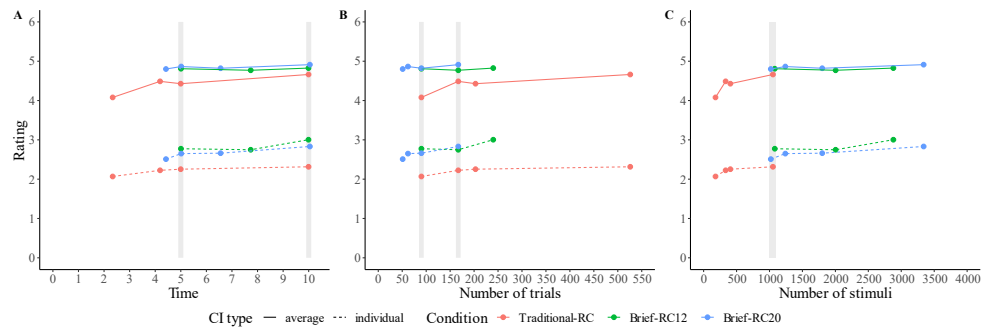
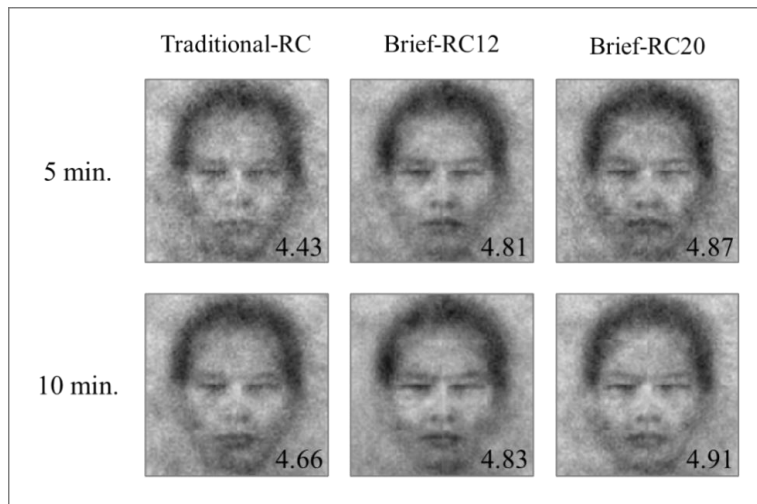


Figure 6

Average CIs by condition and time (5 and 10 minutes). Average ratings are displayed in the lower-right corner of the CIs



Judges' ratings as a function of number of trials. We submitted judges' ratings of the individual CIs to a LMM analysis with C_1 , C_2 , trials (90 versus 167 trials), and their interaction as predictors, using judges and producers as random factors. As predicted, there was a main effect of C_1 such that the individual CIs from the Traditional-RC ($M = 2.15$, SD

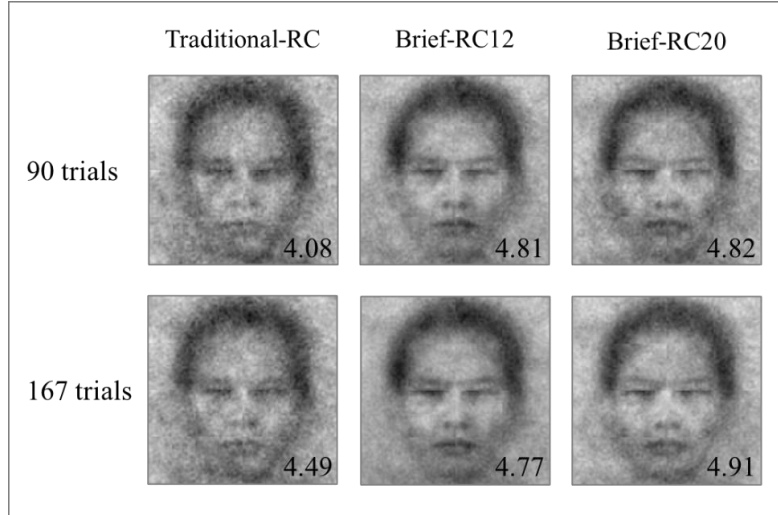
= 1.87) looked significantly less Chinese than those from the Brief-RC12 and Brief-RC20 ($M = 2.75$, $SD = 2.05$), $F(1, 339.58) = 20.19$, $p < .001$, $\eta_p^2 = .020$. Unsurprisingly, the individual CIs after 90 trials ($M = 2.50$, $SD = 1.98$) looked significantly less Chinese than those after 167 trials ($M = 2.60$, $SD = 2.04$), $F(1, 246.43) = 8.12$, $p = .005$, $\eta_p^2 = .001$. Moreover, there was a significant $C_2 \times \text{trials}$ interaction $F(1, 245.72) = 4.67$, $p = .032$, $\eta_p^2 < .001$. Although the simple effects were not significant, the individual CIs from the Brief-RC12 ($M = 2.77$, $SD = 2.01$) looked descriptively more Chinese than those from the Brief-RC20 ($M = 2.66$, $SD = 2.04$) at 90 trials, $F(1, 339.22) = 0.42$, $p = .519$, $\eta_p^2 < .001$, whereas the individual CIs from the Brief-RC12 ($M = 2.75$, $SD = 2.01$) looked less Chinese than those from the Brief-RC20 ($M = 2.83$, $SD = 2.14$) at 167 trials, $F(1, 338.13) = 0.34$, $p = .558$, $\eta_p^2 < .001$. There were no significant effects of C_2 nor of $C_1 \times \text{trials}$ (all p 's $> .10$, BF_{01} 's > 10).

Next, we submitted judges' ratings of the average CIs to an OLS analysis with C_1 , C_2 , trials, and their interaction as predictors. The predicted main effect of C_1 was significant such that the average CIs looked less Chinese for Traditional-RC ($M = 4.28$, $SD = 1.44$) than for Brief-RC12 and Brief-RC20 ($M = 4.83$, $SD = 1.20$), $F(1, 246) = 60.38$, $p < .001$, $\eta_p^2 = .039$. There was also a main effect of trials such that the average CIs looked significantly less Chinese after 90 trials ($M = 4.57$, $SD = 1.34$) than after 167 trials ($M = 4.72$, $SD = 1.28$), $F(1, 246) = 5.39$, $p = .020$, $\eta_p^2 = .004$. Moreover, there was a significant $C_1 \times \text{trials}$ interaction $F(1, 246) = 7.46$, $p = .007$, $\eta_p^2 = .005$. Specifically, the average CIs from the Traditional-RC ($M = 4.08$, $SD = 1.50$) looked significantly less Chinese than the Brief-RC12 and Brief-RC20 ($M = 4.82$, $SD = 1.19$) after 90 trials, $F(1, 246) = 55.15$, $p < .001$, $\eta_p^2 = .035$, this in a more pronounced manner than at 167 trials ($M_{\text{traditional-RC}} = 4.49$, $SD_{\text{traditional-RC}} = 1.36$; $M_{\text{brief-RC12\&Brief-RC20}} = 4.84$, $SD_{\text{brief-RC12\&Brief-RC20}} = 1.22$), $F(1, 246) = 12.69$, $p < .001$, $\eta_p^2 = .008$. There were no significant effects of C_2 , nor $C_2 \times \text{trials}$ (all p 's $> .10$, BF_{01} 's > 10).

Figure 5 (panel B) shows the results from judges' ratings of the individual and average ratings of CIs as a function of number of trials. Figure 7 illustrates the average CIs as a function of trials.

Figure 7

Average CIs by condition and trial (90 and 167 trials). Average ratings are displayed in the lower-right corner of the CIs



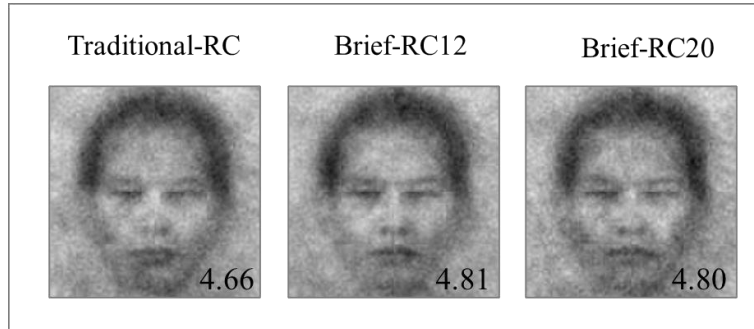
Judges' ratings as a function of number of stimuli. We submitted judges' ratings of the individual CIs produced after producers had been exposed to approximately 1050 stimuli (noisy faces) to a LMM analysis with C_1 and C_2 as predictors, using judges and producers as random factors. There was a main effect of C_1 such that the individual CIs from the Traditional-RC ($M = 2.31$, $SD = 2.00$) looked significantly less Chinese than those of the Brief-RC12 and Brief-RC20 ($M = 2.64$, $SD = 2.02$), $F(1, 331.36) = 4.78$, $p = .029$, $\eta_p^2 = .005$. There was no significant effect of C_2 , $F(1, 326.76) = 2.52$, $p = .113$, $\eta_p^2 = .003$, $BF_{01} = 22.12$.

Next, we submitted judges' ratings of the average CIs to an OLS analysis with C_1 and C_2 as predictors. The significant main effect of C_1 confirmed the average CIs looked less Chinese for Traditional-RC ($M = 4.66$, $SD = 1.29$) than for the Brief-RC12 and Brief-RC20 ($M = 4.81$, $SD = 1.18$), $F(1, 249) = 59.96$, $p < .001$, $\eta_p^2 = .038$. There was no significant effect of C_2 , $F(1, 249) = 0.91$, $p = .341$, $\eta_p^2 = .001$, $BF_{01} = 24.68$.

Figure 5 (panel C) shows the results from judges' ratings of the individual and average ratings of CIs as a function of number of stimuli. Figure 8 illustrates the average CIs and their ratings. Means and standard deviations of the ratings of individual and average CIs for each of the experimental cells are available in the Supplementary Materials (Table S1).

Figure 8

Average CIs produced after producers were exposed to approximately 1050 stimuli (noisy faces). Average ratings are displayed in the lower-right corner of the CIs



InfoVal as a function of time. We submitted the infoVal scores of the individual CIs to an OLS analysis with C_1 , C_2 , time (5 vs. 10 minutes), and their interactions as predictors in an OLS analysis. As a reminder, the higher infoVal, the more likely the individual CI was generated from meaningful (versus random) responses. As for C_1 , there was no difference between the infoVal scores from the Traditional-RC ($M = 0.98$, $SD = 1.66$) and those from the Brief-RC12 and Brief-RC20 ($M = 1.09$, $SD = 1.79$), $F(1, 294) = 0.61$, $p = .434$, $\eta_p^2 = .001$, $BF_{01} = 17.99$. There was a significant time effect, such that the infoVal scores after 5 minutes ($M = 0.77$, $SD = 1.43$) were lower than those after 10 minutes ($M = 1.35$, $SD = 1.98$), $F(1, 294) = 16.89$, $p < .001$, $\eta_p^2 = .028$. There was no significant effect of C_2 , $C_1 \times$ time, or $C_2 \times$ time (all p 's $> .10$, BF_{01} 's > 10). Figure 9 (panel A) shows the averaged infoVal scores of individual CIs as a function of time.

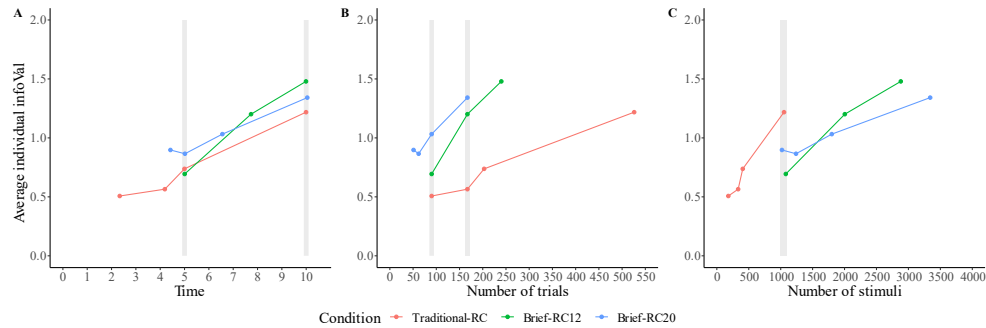
InfoVal as a function of number of trials. We submitted the infoVal scores of the individual CIs to an OLS analysis with C_1 , C_2 , trials (90 vs. 167 trials), and their interactions. C_1 was significant such that infoVal scores from the Traditional-RC ($M = 0.54$, $SD = 1.24$) were lower than those from the Brief-RC12 and Brief-RC20 ($M = 1.07$, $SD = 1.70$), $F(1, 294) = 15.60$, $p < .001$, $\eta_p^2 = .026$. We also found a trials main effect in that infoVal scores after 90 trials ($M = 0.74$, $SD = 1.47$) were significantly lower than those after 167 trials ($M = 1.04$, $SD = 1.67$), $F(1, 294) = 5.29$, $p = .022$, $\eta_p^2 = .009$. C_2 was not significant such that infoVal scores from the Brief-RC12 ($M = 0.95$, $SD = 1.64$) did not significantly differ from those of the Brief-RC20 ($M = 1.19$, $SD = 1.75$), $F(1, 294) = 2.38$, $p = .124$, $\eta_p^2 = .004$. There was no significant effect of $C_1 \times$ trials, nor $C_2 \times$ trials (all p 's $> .10$, BF_{01} 's > 10). Figure 9 (panel B) shows the averaged infoVal scores of individual CIs as a function of number of trials.

InfoVal as a function of number of stimuli. We submitted the infoVal scores of the individual CIs produced after producers had seen approximately 1050 stimuli (noisy faces) to an OLS analysis using C_1 and C_2 as predictors. C_1 was significant such that the infoVal scores from the Traditional-RC ($M = 1.22$, $SD = 1.89$) was higher than those for the Brief-RC12 and Brief-RC20 ($M = 0.80$, $SD = 1.41$), $F(1, 297) = 4.73$, $p = .030$, $\eta_p^2 = .016$. There were no significant differences between the infoVal scores of the Brief-RC12 ($M = 0.69$, SD

= 1.49) and Brief-RC20 ($M = 0.90$ $SD = 1.32$), $F(1, 297) = 0.82$, $p = .364$, $\eta_p^2 = .003$, $BF_{01} = 11.43$. Figure 9 (panel C) shows the averaged infoVal scores of individual CIs as a function of number of stimuli. Means and standard deviations of the infoVal scores of individual CIs for each of the experimental cells are available in the Supplementary Materials (Table S2).

Figure 9

Relation between the average infoVal (information value) of individual CIs and (A) the time (median cumulative time in minutes), (B) the number of trials, and (C) the number of stimuli, as a function of condition (different colored lines: Traditional-RC, Brief-RC12, and Brief-RC20). The dots represent the averaged infoVal at given points of comparison (cf. Table 1). The grey frames highlight the points of comparison, i.e., (A) at approximately 5 and 10 minutes, (B) at 90 and 167 trials, and (C) at approximately 1050 stimuli.



4.3 Discussion

As a reminder, Experiment 1 revealed that when participants saw the same number of stimuli, the Brief-RC performed as well as the Traditional-RC. Experiment 2 sought to replicate these findings while taking into account additional criteria that prove relevant when designing RC-based studies, namely the time of completion and the number of trials. Moreover, we increased the overall length of task and increased the statistical power to maximize the reliability of comparisons.

In line with our expectations, subjective ratings from independent judges revealed that the Brief-RC outperformed the Traditional-RC on all three criteria. With respect to time, judges evaluated individual CIs as more prototypical of the social target (i.e., a Chinese-looking face) after only 5 minutes and even more so after 10 minutes, suggesting that this trend may further improve over time. As for the number of trials, the ratings were also more satisfactory with the new as opposed to the traditional method and, this, whether participants had completed 90 or 167 trials. In contrast with Experiment 1, Experiment 2 revealed that the Brief-RC was not as good but did in fact surpass the Traditional-RC after participant had seen to the same number of stimuli (i.e., noisy faces). To be sure, the latter comparison entailed a large number of stimuli in Experiment 2 than in Experiment 1. Finally, the average CIs also came across as more prototypical on all the three criteria when we relied on the Brief-RC as compared to the Traditional-RC.

We also explored the differences between methods using an objective metric. The message proves somewhat less clear than the one emerging from the subjective ratings. On the one hand, infoVal scores of individual CIs revealed that the two methods performed at a similar level after the same time, while the Brief-RC outperformed the Traditional-RC after the same number of trials. The opposite pattern emerged when holding constant the number of stimuli.

Taken together, these findings clearly suggest that the Brief-RC produces superior individual and average CIs at least as assessed on a subjective scale.

5. General Discussion

The reverse correlation (RC) is an innovative data-driven method designed to capture visual representations (or so-called classification images; CIs) of social targets. The procedure has proven very useful to uncover various important phenomena such as the top-down distortions of facial representation of social groups (e.g., Brown-Iannuzzi et al., 2016; Brown-Iannuzzi et al., 2018; Dotsch et al., 2008; Hinzman & Maddox, 2017). One issue, however, is that past research relied almost solely on ratings from average CIs, a strategy that may drastically increase Type I errors (Cone et al., 2020). This state of affairs poses a serious threat in terms of the replicability of psychological science (e.g., see the Open Science Collaboration, 2015; Simmons et al., 2011). An obvious solution is the use of individual CIs (Cone et al., 2020). This is even more tempting that individual CIs also pave the way to more fine-grained analyses (e.g., analyzing the relation between racial bias and visual representation bias at the participant-level; Dotsch et al., 2008).

However, producing robust and high-quality individual CIs is far from being an easy task as it requires a very large number of trials. This is not only economically costly and time consuming but also challenging in terms of participants' motivation (Brinkman, Goffin et al., 2019; Todorov et al., 2011). For these reasons, researchers called for an improvement of the method to produce higher quality outcomes, while reducing the number of trials (Todorov et al., 2011). With exactly this goal in mind, we conducted two pre-registered experiments that examine a new method, namely the Brief-RC. This new version aims to reduce the number of trials and to improve the outcome quality by increasing the number of stimuli at each trial.

In Experiment 1, we compared two variants of the Brief-RC (the Brief-RC12 with 12 stimuli per trial and Brief-RC20 with 20 stimuli per trial) to the Traditional-RC (Traditional-RC with 2 stimuli per trial) while fixing the overall number of stimuli (i.e., noisy faces) presented to the participants across all trials. Our findings reveal that individual CIs produced from both methods performed similarly, as assessed by judges (i.e., on how prototypically Chinese the faces looked) and by the infoVal scores—which determines the degree to which a CI stems from meaningful responses (i.e., signal) rather than random one (i.e., noise) (Brinkman, Goffin et al., 2019). Moreover, judges' ratings of average CIs between the two methods did not differ. In itself, this is a noteworthy achievement because the Brief-RC allows presenting the same amount of information (i.e., number of stimuli) in almost half the time and only a fraction of trials compared to the traditional method.

Experiment 2 extended these findings by also comparing the end-products of the different RC methods on additional criteria. Moreover, we increased the task length to produce CIs that are more robust. Finally, we also increased statistical power. This time, subjective ratings from independent judges of the individual CIs revealed that the Brief-RC outperformed the Traditional-RC after participants had seen the same number of stimuli. One reasonable account for the difference between the two experiments may reside in the fact that, compared to the point of comparison in Experiment 2 (i.e., approximately 1050 stimuli), the point of comparison in Experiment 1 (i.e., 720 stimuli) may have been set too early to spot any divergence. It should also be noted that Experiment 2 had more statistical power than Experiment 1.

Importantly, the Brief-RC variants also outperformed the Traditional-RC after the same time length (after 5 or 10 minutes) and after the same number of trials (after 90 or 167 trials). Moreover, these results replicated when considering the average CIs. As for infoVal, although more exploratory, the two methods performed on similar levels when considering the same time length, whereas the Brief-RC scored better when comparing across number of trials, but worse when comparing across number of stimuli. This divergence between the results on the subjective and objective measures suggests that the relation between the two is not as straightforward as it may seem and needs further investigations. One possible explanation may be that the signal detected by infoVal did not (or did only partly) correspond with the one assessed by subjective ratings. For instance, it may be the case that most of the signal identified from infoVal came from the jaw, whereas subjective ratings of external judges regarding how Chinese-looking were the faces were mainly driven by the eye's region. In other words, some slight qualitative changes—that are not quantitatively noticeable enough for the infoVal metric—could be highly informative for judges. All of this remains very tentative at this stage and requires additional research.

An interesting finding from Experiment 2 stems from the visual inspection of Figure 5. Indeed, the gap between judgments of individual CIs from the Brief-RC variants versus the Traditional-RC widens as the task progresses. Specifically, individual CIs' ratings from the Traditional-RC seem to reach an horizontal asymptote not long after the beginning of the task (around 4 minutes, or 200 trials) whereas for the Brief-RC (and particularly for the Brief-RC12) they appear to further improve, even beyond our fixed comparison points. This suggests that, on average, individual CIs may not substantially get better after a few minutes (or trials) when relying on the traditional procedure, whereas there seems to be room for improvement with our new method. At the same time, a different pattern seems to take place when considering ratings of average CIs. That is, the Brief-RC offers the advantage that it rapidly converges towards a stable average CI, whereas the Traditional-RC may take longer. In both cases, these improvements could be attributed to the enhanced signal-to-noise ratio of the Brief-RC, and perhaps also to the lack of motivation when choosing from a set of stimuli that are less likely to carry signal (as in the Traditional-RC). We also note that ratings of average CIs are largely superior to those of individual CIs. This is hardly surprising given that they build on thousands of trials instead of a few hundreds (Cone et al., 2020; Imhoff et al., 2011; Imhoff et al., 2013).

Overall, the present findings strongly suggest that the Brief-RC offers a substantial improvement over the Traditional-RC on both individual and average CIs, at least in terms of subjective judgments. This is an important accomplishment because it should allow researchers to run RC-based studies in a much more efficient manner. A notable dividend is that the Brief-RC may grant an access to higher quality individual CIs thus paving the way for more fine-grained analyses with other variables of interest (e.g., inter-individual differences). Clearly, the advantage of increasing the number of stimuli per trial should also benefit other variants of the RC paradigm, as in methods that rely on three- and four-dimensional (instead of two-dimensional) stimuli (see Jack & Schyns, 2017; Walker & Keller, 2019; Walker & Vetter, 2016).

As anticipated, the differences between the two variants of the Brief-RC (i.e., Brief-RC12 and Brief-RC20) were minimal. This is an interesting outcome in and of itself. Indeed, one might conjecture that an ideal observer should perform better as the number of stimuli per trial increases from 12 to 20 per trials. This was not the case here with our participants. A possible reason may stem from the limitations of the human brain when it comes to processing visual information (Marois & Ivanoff, 2005; Palmer, 1990). In light of the present findings, the Brief-RC12 should be preferred over the Brief-RC20 as it requires less stimuli to achieve a similar result. However, our research was limited to two version of the Brief-RC (Brief-RC12 and Brief-RC20) and further work should investigate the optimal configuration in terms of number of stimuli per trial (e.g., 4, 6, 8, 10).

Although not directly tested in the present work, another remarkable advantage of the Brief-RC is that it likely increases the external validity of the outcomes. CIs are essentially a linear combination of the stimuli (noisy faces). Increasing the number of trials and, even more so, the number of stimuli per trial substantially enlarges the unique set of stimulus (e.g., noisy faces) and thus the CI's potential face space—the universe of potential CIs that can be built from all possible combinations of stimuli from a given RC task (Jack & Schyns, 2017; Todorov et al., 2011). In particular, the visual renderings from the RC-Brief should gain in external validity because they derive from an exponentially larger sample of stimuli⁹ and therefore have more chances to match the underlying psychological representations. Overall, this also brings the Brief-RC a step closer to adequately sampling larger stimulus space spanned by additional dimensions (e.g., color, depth, texture, or higher dimensional noise patterns; Brinkman et al., 2017; Jack & Schyns, 2017) and to capturing more complex structures unlikely to emerge with fewer stimuli (e.g., a wrinkle on the forehead when selecting older faces; Jack & Schyns, 2017). Further research should provide empirical evidence for the increased ecological validity of Brief-RC's outcomes.

⁹ A reverse correlation task with m stimuli per trial and n trials can generate m^n unique individual CIs (assuming that each stimulus is unique). Therefore, a Brief-RC with m stimuli can generate $(m/2)^n - 1$ times more unique individual CIs than a Traditional-RC with the same number of trials. For instance, after $n = 3$ trials, the Brief-RC12 with $m = 12$ stimuli per trial can produce $12^3 = 1728$ unique individual CIs, whereas a Traditional-RC can only generate $2^3 = 8$. That is, the Brief-RC12 can already generate $(12/2)^3 - 1 = 215$ times more unique individual CIs after only 3 trials than the Traditional-RC ($8 + 8 \times 215 = 1728$).

In conclusion, the current work introduces an improved reverse correlation method, namely the Brief-RC. This new technique enhances the quality of average CIs and even more so of individual CIs by improving the signal-to-noise ratio through the presentation of a greater number of stimuli at every trial. The access to more robust and better-quality individual CIs is key to address inflated Type I errors rates prevalent in traditional two-phase reverse correlations procedures. Moreover, the Brief-RC significantly reduces the overall task length, which offers clear-cut benefits in terms of maintaining participant's motivation, but also in terms of economic resources and time constraints for the researcher.

6. Open Practices

The data, R code, and materials are publicly available on Open Science Framework (https://osf.io/ps9wu/?view_only=028597d60e7342bfaeb6881051cb6bca).

7. References

- Ahumada, A., & Lovell, J. (1971). Stimulus Features in Signal Detection. *The Journal of the Acoustical Society of America*, 49(6B), 1751–1756. <https://doi.org/10.1121/1.1912577>
- Ahumada, A., Marken, R., & Sandusky, A. (1975). Time and frequency analyses of auditory signal detection. *The Journal of the Acoustical Society of America*, 57(2), 385–390. <https://doi.org/10.1121/1.380453>
- Albohn, D. N., & Adams, R. B. (2020). Everyday Beliefs About Emotion Perceptually Derived From Neutral Facial Appearance. *Frontiers in Psychology*, 11, 264. <https://doi.org/10.3389/fpsyg.2020.00264>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint*. arXiv:1506.04967.
- Brinkman, L., Dotsch, R., Zondergeld, J., Koevoets, M. G. J. C., Aarts, H., & van Haren, N. E. M. (2019). Visualizing mental representations in schizophrenia patients: A reverse correlation approach. *Schizophrenia Research: Cognition*, 17, 100138. <https://doi.org/10.1016/j.scog.2019.100138>
- Brinkman, L., Goffin, S., van de Schoot, R., van Haren, N. E. M., Dotsch, R., & Aarts, H. (2019). Quantifying the informational value of classification images. *Behavior Research Methods*, 51(5), 2059–2073. <https://doi.org/10.3758/s13428-019-01232-2>
- Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, 28(1), 333–361. <https://doi.org/10.1080/10463283.2017.1381469>
- Brooks, J. A., Stoler, R. M., & Freeman, J. B. (2018). Stereotypes Bias Visual Prototypes for Sex and Emotion Categories. *Social Cognition*, 36(5), 481–493. <https://doi.org/10.1521/soco.2018.36.5.481>
- Brown-Iannuzzi, J. L., Dotsch, R., Cooley, E., & Payne, B. K. (2016). The Relationship Between Mental Representations of Welfare Recipients and Attitudes Toward Welfare. *Psychological Science*, 28(1), 92–103. <https://doi.org/10.1177/0956797616674999>

- Brown-Iannuzzi, J. L., McKee, S., & Gervais, W. M. (2018). Atheist horns and religious halos: Mental representations of atheists and theists. *Journal of Experimental Psychology: General*, 147(2), 292–297. <https://doi.org/10.1037/xge0000376>
- Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, 1(1), 1–20. <https://doi.org/10.5334/joc.10>
- Cone, J., Brown-Iannuzzi, J. L., Lei, R., & Dotsch, R. (2020). Type I Error Is Inflated in the Two-Phase Reverse Correlation Procedure. *Social Psychological and Personality Science*, 194855062093861. <https://doi.org/10.1177/1948550620938616>
- Dai, H., & Micheyl, C. (2010). Psychophysical reverse correlation with multiple response alternatives. *Journal of Experimental Psychology: Human Perception and Performance*, 36(4), 976–993. <https://doi.org/10.1037/a0017171>
- Degner, J., Mangels, J., & Zander, L. (2019). Visualizing Gendered Representations of Male and Female Teachers Using a Reverse Correlation Paradigm. *Social Psychology*, 50(4), 233–251. <https://doi.org/10.1027/1864-9335/a000382>
- de Leeuw, J. R. (2014). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y>
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781. <https://doi.org/10.3389/fpsyg.2014.00781>
- Dotsch, R., & Todorov, A. (2011). Reverse Correlating Social Face Perception. *Social Psychological and Personality Science*, 3(5), 562–571. <https://doi.org/10.1177/1948550611430272>
- 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. <https://doi.org/10.1111/j.1467-9280.2008.02186.x>
- Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of Personality and Social Psychology*, 100(6), 999–1014. <https://doi.org/10.1037/a0023026>
- Dotsch, R., Wigboldus, D. H. J., & Van Knippenberg, A. (2013). Behavioral information biases the expected facial appearance of members of novel groups. *European Journal of Social Psychology*, 43(1), 116–125. <https://doi.org/10.1002/ejsp.1928>
- Ge, L., Zhang, H., Wang, Z., Quinn, P. C., Pascalis, O., Kelly, D., Slater, A., Tian, J., & Lee, K. (2009). Two Faces of the Other-Race Effect: Recognition and Categorisation of Caucasian and Chinese Faces. *Perception*, 38(8), 1199–1210. <https://doi.org/10.1068/p6136>
- Gunaydin, G., & DeLong, J. E. (2015). Reverse Correlating Love: Highly Passionate Women Idealize Their Partner's Facial Appearance. *PLOS ONE*, 10(3), e0121094. <https://doi.org/10.1371/journal.pone.0121094>
- Gundersen, A. B., & Kunst, J. R. (2018). Feminist ≠ Feminine? Feminist Women Are Visually Masculinized Whereas Feminist Men Are Feminized. *Sex Roles*, 80(5–6), 291–309. <https://doi.org/10.1007/s11199-018-0931-7>

- Hinzman, L., & Maddox, K. B. (2017). Conceptual and visual representations of racial categories: Distinguishing subtypes from subgroups. *Journal of Experimental Social Psychology*, 70, 95–109. <https://doi.org/10.1016/j.jesp.2016.12.012>
- Hong, Y., & Ratner, K. G. (2020). Minimal but not meaningless: Seemingly arbitrary category labels can imply more than group membership. *Journal of Personality and Social Psychology*, 1–25. <https://doi.org/10.1037/pspa0000255>
- Imhoff, R., & Dotsch, R. (2013). Do We Look Like Me or Like Us? Visual Projection as Self- or Ingroup-Projection. *Social Cognition*, 31(6), 806–816. <https://doi.org/10.1521/soco.2013.31.6.806>
- Imhoff, R., Dotsch, R., Bianchi, M., Banse, R., & Wigboldus, D. H. J. (2011). Facing Europe. *Psychological Science*, 22(12), 1583–1590. <https://doi.org/10.1177/0956797611419675>
- Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in Psychology*, 4, 1–8. <https://doi.org/10.3389/fpsyg.2013.00386>
- Jack, R. E., & Schyns, P. G. (2017). Toward a Social Psychophysics of Face Communication. *Annual Review of Psychology*, 68(1), 269–297. <https://doi.org/10.1146/annurev-psych-010416-044242>
- Jackson, J. C., Hester, N., & Gray, K. (2018). The faces of God in America: Revealing religious diversity across people and politics. *PLOS ONE*, 13(6), e0198745. <https://doi.org/10.1371/journal.pone.0198745>
- Jaeger, B. (2017). r2glmm: Computes R Squared for Mixed (Multilevel) Models (R package version 0.1.2). Retrieved from <https://CRAN.R-project.org/package=r2glmm>
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.
- Judd, C. M., & Park, B. (1988). Out-group homogeneity: Judgments of variability at the individual and group levels. *Journal of Personality and Social Psychology*, 54(5), 778–788. <https://doi.org/10.1037/0022-3514.54.5.778>
- 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. <https://doi.org/10.1037/a0028347>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with More Than One Random Factor: Designs, Analytic Models, and Statistical Power. *Annual Review of Psychology*, 68(1), 601–625. <https://doi.org/10.1146/annurev-psych-122414-033702>
- Karremans, J. C., Dotsch, R., & Corneille, O. (2011). Romantic relationship status biases memory of faces of attractive opposite-sex others: Evidence from a reverse-correlation paradigm. *Cognition*, 121(3), 422–426. <https://doi.org/10.1016/j.cognition.2011.07.008>
- Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The Other-Race Effect Develops During Infancy. *Psychological Science*, 18(12), 1084–1089. <https://doi.org/10.1111/j.1467-9280.2007.02029.x>
- Kontsevich, L. L., & Tyler, C. W. (2004). What makes Mona Lisa smile? *Vision Research*, 44(13), 1493–1498. <https://doi.org/10.1016/j.visres.2003.11.027>

- Krosch, A. R., & Amodio, D. M. (2014). Economic scarcity alters the perception of race. *Proceedings of the National Academy of Sciences*, 111(25), 9079–9084. <https://doi.org/10.1073/pnas.1404448111>
- Kunst, J. R., Dovidio, J. F., & Dotsch, R. (2017). White Look-Alikes: Mainstream Culture Adoption Makes Immigrants “Look” Phenotypically White. *Personality and Social Psychology Bulletin*, 44(2), 265–282. <https://doi.org/10.1177/0146167217739279>
- Kunst, J. R., Kteily, N., & Thomsen, L. (2017). “You Little Creep”: Evidence of Blatant Dehumanization of Short Groups. *Social Psychological and Personality Science*, 10(2), 160–171. <https://doi.org/10.1177/1948550617740613>
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, 24(8), 1377–1388. <https://doi.org/10.1080/02699930903485076>
- Lee, Y.-T., & Ottati, V. (1995). Perceived In-Group Homogeneity as a Function of Group Membership Salience and Stereotype Threat. *Personality and Social Psychology Bulletin*, 21(6), 610–619. <https://doi.org/10.1177/0146167295216007>
- Lick, D. J., Carpinella, C. M., Preciado, M. A., Spunt, R. P., & Johnson, K. L. (2013). Reverse-correlating mental representations of sex-typed bodies: the effect of number of trials on image quality. *Frontiers in Psychology*, 4, 1–9. <https://doi.org/10.3389/fpsyg.2013.00476>
- Lin, C., Adolphs, R., & Alvarez, R. M. (2018). Inferring Whether Officials Are Corruptible From Looking at Their Faces. *Psychological Science*, 29(11), 1807–1823. <https://doi.org/10.1177/0956797618788882>
- Lloyd, E. P., Sim, M., Smalley, E., Bernstein, M. J., & Hugenberg, K. (2020). Good Cop, Bad Cop: Race-Based Differences in Mental Representations of Police. *Personality and Social Psychology Bulletin*, 46(8), 1205–1218. <https://doi.org/10.1177/0146167219898562>
- Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: estimating the information employed for face classifications. *Cognitive Science*, 28(2), 209–226. https://doi.org/10.1207/s15516709cog2802_4
- Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, 9(6), 296–305. <https://doi.org/10.1016/j.tics.2005.04.010>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- Oliveira, M., Garcia-Marques, T., Dotsch, R., & Garcia-Marques, L. (2019). Dominance and competence face to face: Dissociations obtained with a reverse correlation approach. *European Journal of Social Psychology*, 49(5), 888–902. <https://doi.org/10.1002/ejsp.2569>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>

- Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, 16(2), 332–350. <https://doi.org/10.1037/0096-1523.16.2.332>
- Peer, E., Vosgerau, J., & Acquisti, A. (2013). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, 46(4), 1023–1031. <https://doi.org/10.3758/s13428-013-0434-y>
- Petsko, C. D., Lei, R. F., Kunst, J. R., Bruneau, E., & Kteily, N. (in press). Blatant dehumanization in the mind’s eye: Prevalent even among those who explicitly reject it? *Journal of Experimental Psychology: General*. Retrieved from <https://psyarxiv.com/g7w4b/>
- Ratner, K. G., Dotsch, R., Wigboldus, D. H. J., van Knippenberg, A., & Amodio, D. M. (2014). Visualizing minimal ingroup and outgroup faces: Implications for impressions, attitudes, and behavior. *Journal of Personality and Social Psychology*, 106(6), 897–911. <https://doi.org/10.1037/a0036498>
- Schmitz, M., Rougier, M., & Yzerbyt, V. (2020). Comment on “Quantifying the informational value of classification images”: A miscomputation of the infoVal metric. *Behavior Research Methods*, 52(3), 1383–1386. <https://doi.org/10.3758/s13428-019-01295-1>
- Schmitz, M., Rougier, M., Yzerbyt, V., Brinkman, L., & Dotsch, R. (2020). Erratum to: Comment on “Quantifying the informational value of classification images”: Miscomputation of infoVal metric was a minor issue and is now corrected. *Behavior Research Methods*, 52(4), 1800–1801. <https://doi.org/10.3758/s13428-020-01367-7>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Tiddeman, B. P., Stirrat, M. R., & Perrett, D. I. (2005). Towards Realism in Facial Image Transformation: Results of a Wavelet MRF Method. *Computer Graphics Forum*, 24(3), 449–456. <https://doi.org/10.1111/j.1467-8659.2005.00870.x>
- Todorov, A., Dotsch, R., Wigboldus, D. H. J., & Said, C. P. (2011). Data-driven Methods for Modeling Social Perception. *Social and Personality Psychology Compass*, 5(10), 775–791. <https://doi.org/10.1111/j.1751-9004.2011.00389.x>
- Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current Opinion in Neurobiology*, 23(3), 373–380. <https://doi.org/10.1016/j.conb.2012.12.010>
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*, 66(1), 519–545. <https://doi.org/10.1146/annurev-psych-113011-143831>
- Walker, M., & Keller, M. (2019). Beyond attractiveness: A multimethod approach to study enhancement in self-recognition on the Big Two personality dimensions. *Journal of Personality and Social Psychology*, 117(3), 483–499. <https://doi.org/10.1037/pspa0000157>

- Walker, M., & Vetter, T. (2016). Changing the personality of a face: Perceived Big Two and Big Five personality factors modeled in real photographs. *Journal of Personality and Social Psychology*, 110(4), 609–624. <https://doi.org/10.1037/pspp0000064>
- Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143(5), 2020–2045. <https://doi.org/10.1037/xge0000014>
- Young, A. I., Ratner, K. G., & Fazio, R. H. (2013). Political Attitudes Bias the Mental Representation of a Presidential Candidate's Face. *Psychological Science*, 25(2), 503–510. <https://doi.org/10.1177/0956797613510717>

Chapter 7

Facial compensation in existing groups

Compensation characterizes the trade-off between the fundamental dimensions of competence and warmth when judging two social targets. The present work aimed to test, with existing and meaningful social groups, the external validity of the visual correlate of the compensation effect. In line with compensation, the facial representation of the allegedly low-status group (i.e., Walloons) came across as warmer and more trustworthy but less competent and dominant than the high-status group (i.e., Flemings). These findings increase the external validity of the facial compensation effect and contribute to the current understanding about the Belgian conflict between the two linguistic communities.

Keywords: compensation effect, reverse correlation, Walloons and Flemings, warmth and competence, dominance and trustworthiness

Reference

Schmitz, M., Terache, J., & Yzerbyt, V. (in preparation). Facial compensation in existing groups. *Manuscript in preparation*.

Facial compensation in existing groups

1. Introduction

The compensation effect refers to the negative relation between the two fundamental dimensions of warmth and competence when judging two social targets (Yzerbyt et al., 2005). Although the effect has proven widespread and robust in measures such as verbal reports, mostly in the form of rating scales but also via causal attribution or descriptions, only one study has directly examined if compensation could extend beyond the conceptual level and bias the facial representation of social targets (Schmitz, Vanbeneden et al., under review). Albeit quite informative, these authors' findings relied on fictitious groups, leaving open the question of external validity. The present work addresses this important limitation by making use of existing and relevant social group (i.e., Walloons and Flemings) with long-lasting intergroup relations.

1.1 Two fundamental dimensions of social and facial perception

Several social evaluation models suggest that two fundamental dimensions map the content of social perception (for a review, see Abele et al., in press). According to the Stereotype Content Model (SCM; Fiske et al., 2002), for instance, *Warmth* derives from the target's inferred intention ("is it a friend or a foe?"), whereas *competence* relates to the capacity of the target to carry out these intentions. The first dimension is associated with traits such as sociable, friendly, warm, whereas the second dimension is characterized by traits such as competent, efficient, hardworking.

A similar bi-dimensional configuration has also emerged in the domain of face perception (Oosterhof & Todorov, 2008; Todorov et al., 2008). *Trustworthiness* is characterized by traits such as trustworthy, sociable, or caring, and is strongly related to the emotional expression of happiness that may signal the intentions of the target (i.e., help vs. harm) and triggers responses in terms of approach and avoidance (Oosterhof & Todorov, 2008; Said et al., 2009; Todorov, 2008; Todorov et al., 2008). In contrast, *dominance* encompasses traits such as dominant, aggressive, or confident, and stems from cues of physical strength, facial maturity and masculinity, which may indicate the capability of the target to enact his or her intentions (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Toscano et al., 2016).

Comparisons between the two models (see Oliveira, Garcia-Marques, Dotsch et al., 2019; Sutherland et al., 2016) reveals a marked overlap between the two horizontal dimensions of warmth and trustworthiness as they both strongly correlate with valence (i.e., good vs. bad) and presage the target's intentions (Fiske et al., 2002; Kervyn et al., 2013; Oosterhof & Todorov, 2008). At the same time, there is a divergence between the two vertical dimensions of competence and dominance, such that the former is positively related to valence whereas the latter exhibits a negative relation with valence (Chen et al., 2014; Oliveira, Garcia-

Marques, Dotsch et al., 2019; Oliveira, Garcia-Marques, Garcia-Marques et al., 2019). A plausible explanation for this pattern is that competence and dominance represent distinct strategies to attain status: Competence is potentially beneficial to others (e.g., by sharing and displaying expertise) whereas dominance is likely detrimental to others (e.g., by inducing coercion or intimidation) (Chapais, 2015; Cheng et al., 2013).

1.2 The compensation effect

Although the bidimensional space proposed by social evaluation models materializes the fact that the two dimensions are generally independent from each other, there is also indication that the so-called vertical (competence, agency, etc.) and horizontal (warmth, communion) dimensions may be negatively related, at least when it comes to the perception of social groups (Fiske, 2015). One account for the fact that most stereotypes are mixed or ambivalent (i.e., groups are high on one dimension but low on the other; e.g., stereotypes about rich and poor people, young and elderly people, men and women) can be found in the so-called compensation effect (Yzerbyt et al., 2005). This effect characterizes the negative relation that exists between warmth and competence when judging two social targets. For instance, Italians are portrayed as warmer but less competent than Germans (Kervyn et al., 2008).

Fifteen years of research has shown that dimensional compensation is quite robust and widespread (for a recent review see Yzerbyt, 2018). Interestingly, this pattern arises when evaluating existing groups in real-life settings (Cambon & Yzerbyt, 2016; Kervyn et al., 2008; Yzerbyt & Cambon, 2017; Yzerbyt et al., 2005) as well as fictitious groups in experimental settings (e.g., Cambon et al., 2015; Judd et al., 2005; Kervyn et al., 2016), and whether perceivers are themselves members of one of the evaluated groups or external observers (e.g., Cambon & Yzerbyt, 2018; Judd et al., 2005; Schmitz & Yzerbyt, under review; Yzerbyt et al., 2005).

As of now, the evidence in favor of the dimensional compensation rests on verbal reports, be it under the form of rating scales, causal attribution judgments with hardly any attempt to examine the phenomenon on other, less conceptual, aspects. Although not directly testing compensation, one study by Imhoff et al. (Imhoff et al., 2013) showed evidence that can be interpreted as facial compensation in social groups with ambivalent stereotypes (e.g., nursery teachers and managers). In a more recent effort that specifically addressed dimensional compensation at the facial level in an experimental setting, Schmitz et al. (Schmitz, Vanbeneden et al., under review) found that the facial representation of two fictitious groups that were initially asymmetrical on one dimension (i.e., competence) showed a compensatory pattern on the other dimension.

1.3 Relations Between Walloons and Flemings

Despite being a small country, Belgium is composed of three territorial regions. Flanders is in the northern part of the country and the official language of Flemings is Dutch. Wallonia is in the southern part of Belgium and the official language of Walloons is French. In the

center, the Brussels-capital region is a bilingual region although the majority of inhabitants speak French. In total, the Dutch-speaking and the French-speaking Belgians account for approximatively 60% and 40% of the population respectively¹.

Since Belgium's independence in 1830, there has been long-lasting, non-violent, rivalries between the two communities (see Klein et al., 2011; Van der Linden & Roets, 2017). Until the early 1960s, the state was dominated by the French-speaking elites and French was considered a sign of status and power. From the early twentieth century on, a Flemish movement started to claim for more political autonomy and the recognition of the Flemish language and culture. Little by little, the movement developed into a sub-nationalism characterized in ethnic terms. The 1960s brought about a reversal in the relationship. Linguistic borders were established and there was an important shift in economic power, such that Flanders' exportation-based economy took over the declining mining and steel industry of Wallonia. Over the following decades, the regions acquired more power and autonomy. These structural changes were deemed mostly sufficient for the Walloons' political parties but insufficient for the Flemings ones (especially the nationalist parties; e.g., N-VA) (Van der Linden & Roets, 2017). Tensions between the two communities are still quite obvious today. As a case in point, the divergent demands of the leaders from the Flemish and Walloon (and more generally French-speaking) political leaders make it extremely difficult to form coalition governments at the federal level. As a matter of fact, some 500 days have passed without a federal government after the May 2019 elections.

According to the so-called Waffle model (Klein et al., 2011; see also Klein et al., 2017), the Walloon-Flemish conflict can be structured around two dimensions. Regarding the language and territoriality dimension, the Flemings (and especially supporters of the N-VA nationalist political party) champion the principle that local administration language should be restrained to the language of the territory (e.g., Dutch in Flanders) while Walloons defend the freedom to use the mother tongue to deal with bureaucratic matters. As for the autonomy and economic dimension, the equity principle is favored by the Flemings, whereas the solidarity principle is preferred by their French-speaking counterpart.

Although both dimensions are thought to map the content of collective memories and stereotypes, the economic/autonomy dimension has acquired more importance over the years (Klein et al., 2011; Meuleman et al., 2017; see also Meeusen et al., 2017; Rimé et al., 2015). In this context, the few existing survey reports suggest that the Francophones now constitute the low-status group. Indeed, Walloons are often depicted as the "lazy" and "profiteers" (i.e., less competent) in contrast with the "hard-working" and "autonomous" Flemings, but also as more "open-minded", "tolerant" and "generous" (i.e., more sociable/moral) (Klein et al., 2011; Leyens & Yzerbyt, 1992). This pattern thus suggests that compensation might be at work (Klein et al., 2011; Yzerbyt et al., 2005).

¹ There is a third, German-speaking, community that accounts for less than 1% of the population and is located in Wallonia, near the German border.

1.4 The Present Research

A noteworthy limitation from the compensation literature is that outcomes were almost exclusively investigated at a verbal level (e.g., traits ratings, reaction times to labels, behavior descriptions) while overlooking the impact they may have on visual components such as faces which are crucial for everyday-life social interactions (Jack & Schyns, 2017; Todorov et al., 2015). Although quite informative, Imhoff et al.'s (2013) work did not aim to test compensation *per se*, whereas Schmitz, Vanbeneden et al.'s (under review) work is silent about the external validity. In light of this state of affairs, the present work examines whether the visual correlate of compensation can be extended to real-life scenarios. Specifically, we turned to a sample of Walloons participants to test whether, as we would predict, the conceptual and facial representation of Walloons and Flemings are biased in a compensatory way such that Walloons would be seen as lower than Flemings on the vertical dimension but higher on the horizontal dimension. Based on the review, we expected a compensation pattern to emerge at both the conceptual (e.g., ratings) and facial level. To examine the latter aspect, we relied on an improved version of the reverse correlation method (RC; Dotsch et al., 2008), namely the Brief-RC (Schmitz, Rougier, & Yzerbyt, under review) to capture participants' visual representation of the social targets. One notable asset of this data-driven method is that it makes no prior assumption about what those representation may look-like (see Brinkman et al., 2017).

2. Method

Participants and design. Data collection proceeded in two-steps. The first sample comprised 72 participants (hereafter producers) from a Francophone university in Wallonia ($M_{\text{age}} = 20.21$, $SD_{\text{age}} = 4.28$, 62 females). Producers took part in the study (in French) in exchange for university credits. Only Walloon participants were retained² and sample size was based on the availability of participants. The second sample initially comprised 298 US participants (hereafter judges) from Prolific Academic. Based on *a priori* criteria, 16 judges were excluded because they had an invariant response styles for two or more faces. The final sample comprised 282 judges ($M_{\text{age}} = 24.40$, $SD_{\text{age}} = 10.13$, 149 males).

Materials. The base face for the RC was the same as in Schmitz, Rougier et al. (under review) and Schmitz, Vanbeneden et al. (under review). Specifically, the base face was composed of a morph from six real men faces from the Radboud Face Database (Langner et al., 2010). We created a total of 1800 noisy faces (i.e., stimuli) for the RC-task by adding ("oriented") or subtracting ("inverted") a random noise pattern from the base face (for the noise algorithm see Mangini & Biederman, 2004; see Figure 1) with the *rcicr* R package

² A filter question read as follows: "Which community do you feel you belong to the most?" with "Flanders", "Wallonia", "Brussels" and "I'm not Belgian" as possible responses. We retained only those who answered "Wallonia".

(development version³). We rescaled stimuli to a 150×150 pixels so 12 of them could simultaneously fit the screen.

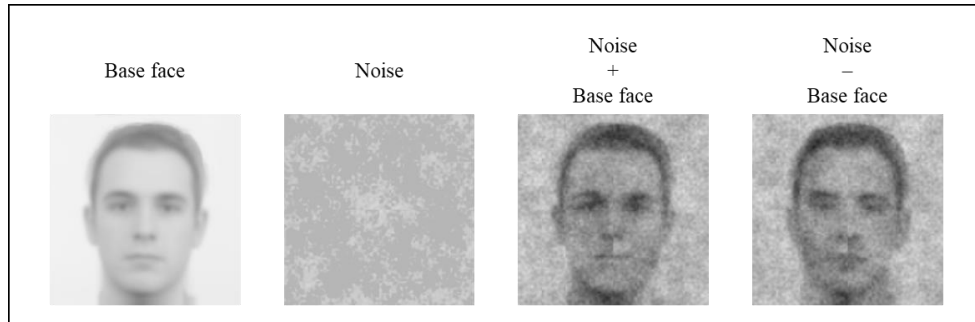


Figure 1. Base face, example of noise pattern, and example of two noisy faces created by adding or subtracting the noise from the base face.

Measures

Classification images. We collected producers’ facial representations of the social targets, also known as classification images (CIs; for the interpretation see Brinkman, et al., 2017) by means of the Brief-RC12 task. In this task, participants are presented with 12 randomly chosen⁴ noisy faces at each of the 150 trials and are asked to “Choose the face that most resembles a Walloon [Fleming]”. We computed individual CIs (i.e., participant-level CIs) by averaging all the selected stimuli’s noise layer and superimposing the average to the base face. We obtained the average CIs (i.e., group-level CIs) by averaging all the noise layers selected by a group of participants. In this case, we obtained the Walloon (Fleming) average CI by averaging all the noise layers selected by all the participants when the RC-target was “Walloon” (“Fleming”). We built the Brief-RC12 task with a JavaScript library called jsPsych (www.jspsych.org; de Leeuw, 2014; version 6.0.3).

Ratings of the social targets. We assessed producers’ impressions about the Walloons and Flemings with a series of traits encompassing the two fundamental dimensions of social perception (from 1 = “Totally agree” to 7 = “Totally disagree”). Specifically, we used warm, sociable, and friendly for the Warmth dimension and competent, efficient, and hard-working for the Competence dimension (all α ’s $\geq .79$).

Ratings of CIs. We assessed judges’ evaluation of average and individual CIs on traits encompassing the two dimensions of the social and facial perception model (from 1 = “Totally agree” to 7 = “Totally disagree”). Specifically, we used the same traits as above for the social dimensions of warmth and competence, and we used dominant, strong, and

³ We relied on the development version in order to obtain the noise matrices, which ease the computation of the CIs. The development version is available at <https://github.com/rdotsch/rcicr/tree/development>

⁴ Noisy faces were randomly selected in pairs, such that each pair of “oriented” and “inverted” noisy face were in the same trial to try to maximize the information presented at each trial.

masculine for the Dominance facial dimension, and trustworthy, reliable, and honest for the Trustworthiness facial dimension (all α 's $\geq .80$).

Demographic information. Both producers and judges provided their age and gender.

Procedure

Step one: Collecting producers' CIs. The first step took place in a laboratory room in the university. Producers learned that they had to take two Brief-RC12, one with the Walloon target and the other with the Flemish target. Each RC-task had 150 trials and the order of the tasks was randomized. Producers were instructed verbally that they could take small breaks whenever they felt during the study. After completing the two RCs, they were to provide their impression about the two groups on traits encompassing the warmth and competence dimensions. Targets (Walloon vs. Fleming) and traits appeared in a random order. Finally, producers completed questions about demographic s and were debriefed.

Step two: Collecting judges' ratings of the CIs. The second step took place online. From the previous step, we generated 72 individual Walloons CIs and 72 individual Fleming CIs, as well as the average Walloon CI and the average Fleming CI. Each judge had to rate 10 randomly sampled individual Walloon CIs, 10 randomly sample individual Fleming CIs, and the two average CIs. All CIs appeared in random other, one at a time. Judges were rated each CI on the traits encompassing the two social and two facial dimensions. Traits appeared in random order for each judge. Finally, judges completed questions about demographics and were debriefed.

3. Results

Producers' ratings of social targets. We submitted producers' ratings of the social targets to a 2 (Target Group: Walloon vs. Fleming) $\times$ 2 (Judged Dimension: Competence vs. Warmth) repeated measures ANOVA. There was a main effect of Target Group, such that producers gave Walloons ($M = 5.22$; $SD = 1.23$) higher scores than Flemings ($M = 4.47$; $SD = 1.61$), $F(1, 68) = 51.22$, $p < .001$, $\eta_p^2 = .419$. There was also a main effect of Judged Dimension, such that producers gave higher ratings for the Competence dimension ($M = 5.07$; $SD = 1.06$) than for the Warmth dimension ($M = 4.62$; $SD = 1.78$), $F(1, 68) = 29.49$, $p < .001$, $\eta_p^2 = .293$. Moreover, the Target Group $\times$ Judged Dimension interaction was significant, $F(1, 68) = 227.55$, $p < .001$, $\eta_p^2 = .762$. Specifically, and in line with the compensation effect, producers rated Walloons ($M = 5.99$; $SD = 0.93$) as warmer than Flemings ($M = 3.25$; $SD = 1.31$), $F(1, 68) = 201.06$, $p < .001$, $\eta_p^2 = .739$, whereas they rated Flemings ($M = 5.69$; $SD = 0.72$) as more competent than Walloons ($M = 4.46$; $SD = 0.99$), $F(1, 68) = 76.71$, $p < .001$, $\eta_p^2 = .519$. Figure 2 shows the average ratings for each dimension.

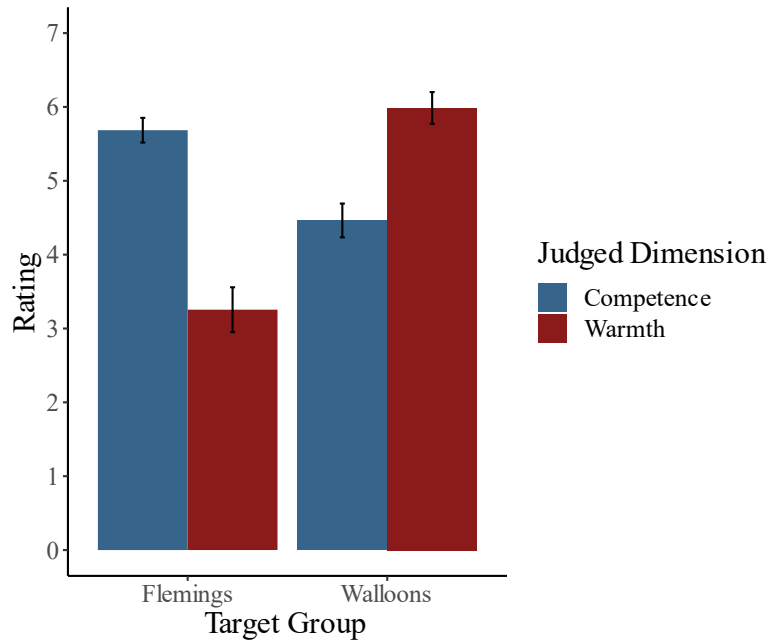


Figure 2. Producers' ratings as a function of Target Group (Flemings vs. Walloons) and Judged Dimension (warmth vs. competence). Error bars represent 95% confidence intervals.

Judges ratings of individual CIs. We submitted judges' ratings of the individual CIs to a 2 (Target Group: Walloons vs. Flemings) $\times$ 2 (Judged Dimension: Vertical vs. Horizontal) $\times$ 2 (Dimensional Model: Social vs. Facial) to a linear mixed model⁵ with all factors varying within participants and with judges and producers' faces as random factors. The Target Group $\times$ Judged Dimension $\times$ Dimensional Model interaction was not significant, $F(1, 67249.50) = 0.19$, $p = .784$, $\eta_p^2 < .001$, indicating that the Dimensional Model (i.e., Social vs. Facial model) failed to moderate the predicted Target Group $\times$ Judged Dimension interaction.

As predicted, the Target Group $\times$ Judged Dimension proved significant, $F(1, 67249.50) = 2262.40$, $p < .001$, $\eta_p^2 = .010$. In line with the compensation effect, judges gave Walloons ($M = 4.54$; $SD = 1.40$) higher scores on the Horizontal dimension (i.e., Warmth and Trustworthiness) than they gave to Flemings ($M = 3.76$; $SD = 1.44$), $F(1, 154.19) = 305.93$, $p < .001$, $\eta_p^2 = .040$, whereas they rated Flemings ($M = 4.66$; $SD = 1.32$) higher on the Vertical dimension (i.e., Competence and Dominance) than they rated Walloons ($M = 4.59$; $SD =$

⁵ Linear mixed models (LMM) take into account the variability from more than one random factor (i.e., judges and producers' faces) (Judd, Westfall, & Kenny, 2012). Therefore, results obtained from these models can be generalized to all random factors. For instance, our results are expected to hold if we were to use other producers and judges (from the same population). It is also for this reason that LMM are usually more conservative than linear models. We computed effect sizes for LMM with the *r2glmm* package (version 0.1.2, Jaeger, 2017). For the sake of clarity and conciseness, we only report the statistical test that are relevant for our hypotheses. The complete analyses are available on Open Science Framework.

1.29), $F(1, 154.19) = 3.09$, $p = .081$, $\eta_p^2 = .030$, although this effect was marginal. Figure 3 shows the average ratings for each CI and dimension.

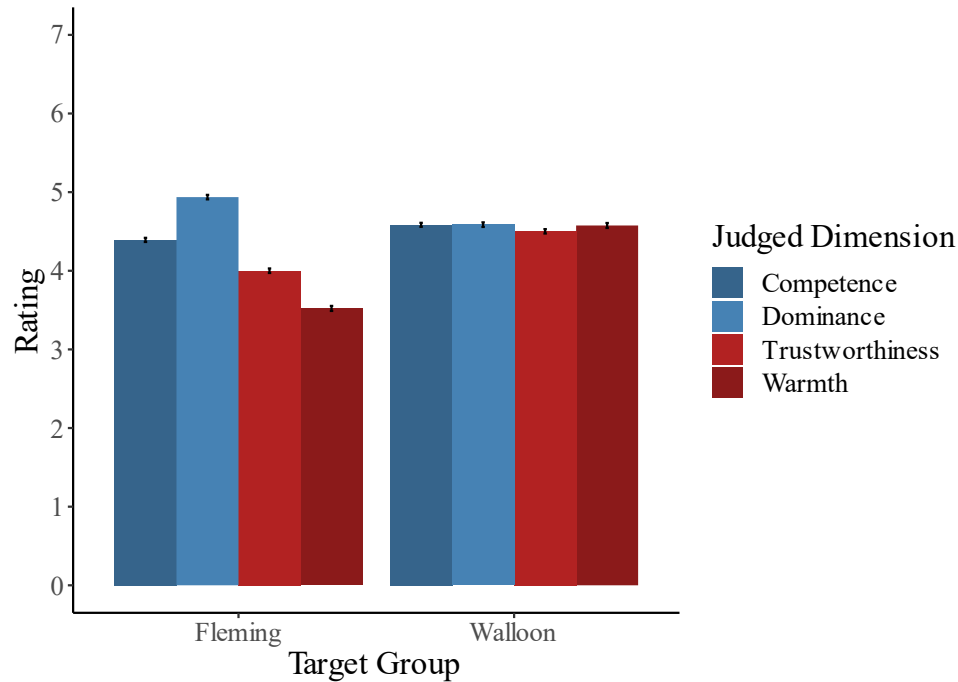


Figure 3. Judges' ratings of the individual CIs as a function of Target Group (Flemings vs. Walloons) and Judged Dimension (horizontal: warmth and trustworthiness vs. vertical: competence and dominance). Error bars represent 95% confidence intervals.

Judges ratings of average CIs. We submitted judges' ratings of the average CIs to a 2 (Target Group: Walloons vs. Flemings) $\times$ 2 (Judged Dimension: Vertical vs. Horizontal) $\times$ 2 (Dimensional Model: Social vs. Facial) ANOVA with all factors varying within participants. The Target Group $\times$ Judged Dimension $\times$ Dimensional Model interaction was not significant, $F(1, 274) = 0.21$, $p = .644$, $\eta_p^2 = .001$, indicating that the Dimensional Model (i.e., Social vs. Facial model) failed to moderate the Target Group $\times$ Judged Dimension interaction.

The Target Group $\times$ Judged Dimension proved significant, $F(1, 274) = 682.97$, $p < .001$, $\eta_p^2 = .708$. In line with the compensation effect, judges gave Walloons ($M = 5.32$; $SD = 1.06$) higher scores on the Horizontal dimension (i.e., warmth and trustworthiness) than they gave Flemings ($M = 3.30$; $SD = 1.33$), $F(1, 274) = 585.48$, $p < .001$, $\eta_p^2 = .676$, whereas they rated Flemings ($M = 4.50$; $SD = 1.07$) higher on the Vertical dimension (i.e., competence and dominance) than they rated Walloons ($M = 4.95$; $SD = 1.14$), $F(1, 274) = 63.18$, $p < .001$, $\eta_p^2 = .184$. Average CIs are presented in Figure 4. Average ratings for each average CI and dimension are presented in Figure 5.

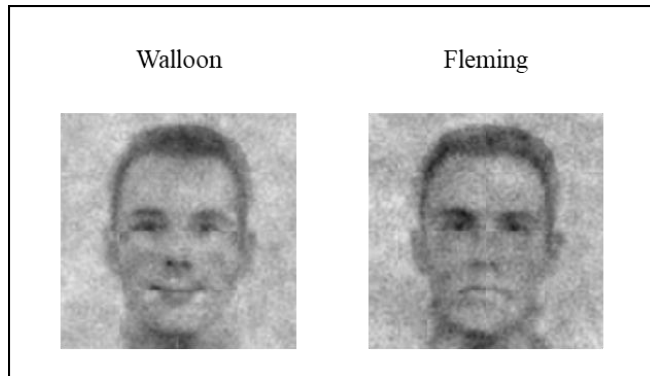


Figure 4. Average classification image (CI) of the Walloon (left) and the Fleming (right).

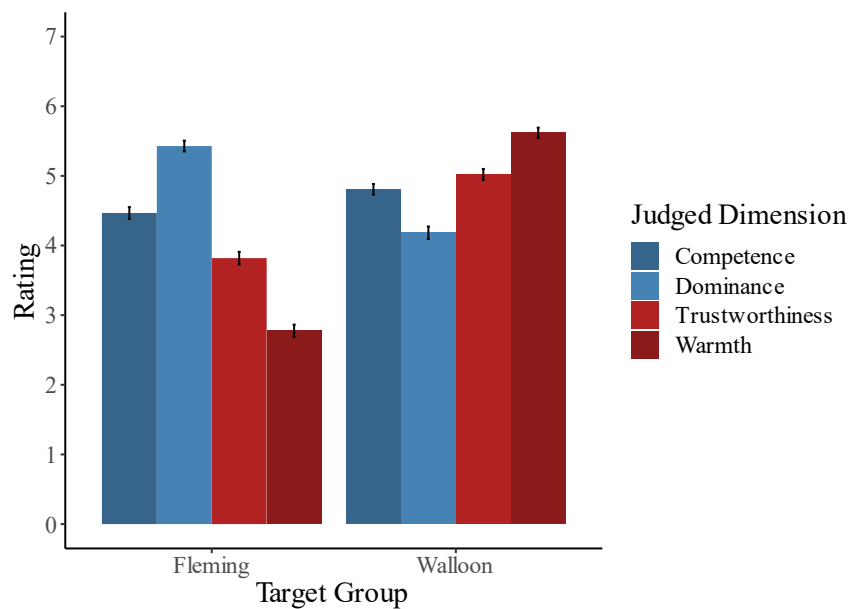


Figure 5. Judges' ratings of the averages CIs as a function of Target Group (Flemings vs. Walloons) and Judged Dimension (horizontal: warmth and trustworthiness vs. vertical: competence and dominance). Error bars represent 95% confidence intervals.

4. General Discussion

Compensation takes place when one target (e.g., a group) is perceived as being higher on one of the fundamental dimensions of social perception (e.g., warmth) but lower on the other dimension (e.g., competence) in comparison to another target (Yzerbyt et al., 2005). Although a large body of evidence points to the robustness of this effect, it is only recently that researchers have started to investigate whether visual representations of social targets

were also distorted in a compensatory way (Schmitz, Vanbeneden et al., under review; see also Imhoff et al., 2013). Unfortunately, the scarce evidence available is limited to fictitious groups and may thus not generalize to real-life scenarios. The aim of the present work was therefore to contribute to the current literature by testing if the compensation effect at a facial level could also take place with existing groups, thus increasing its external validity. To do so, we turned to two large social groups with long-lasting rivalries, namely the Walloons and Flemings and examined the conceptual and facial representations of these two groups in a sample of Walloon citizens. We did so because earlier work (ref) suggests that Walloons perceive these status differences (stemming from economic disparities) as legitimate and stable and the level of conflict as moderate (Klein et al., 2011; Meuleman et al., 2017; see also Rimé et al., 2015), a series of perceptual and structural conditions that facilitate the emergence of a compensation pattern (for a review, see Yzerbyt, 2018).

In the first part of the study, we asked Walloon participants to report their impressions of Walloons and Flemings on a series of traits. In line with our expectations and earlier studies (Klein et al., 2011; Leyens & Yzerbyt, 1992), a clear-cut compensation pattern emerged. Specifically, Walloons participants depicted their ingroup as warmer but less competent than the Flemings. The presence of compensatory judgments among Walloon participants nicely dovetails with previous findings suggesting that this response may serve self-affirmation concerns (Yzerbyt et al., 2005; Cambon & Yzerbyt, 2018; Schmitz & Yzerbyt, under review). Indeed, a self-serving judgment on the Warmth dimension seems to be especially likely among group members (e.g., Walloons) who have their social identity threatened because of their low status (e.g., “lazy Walloons”). This is because compensation allows shifting the comparison focus from a dimension on which one comes across less favorably than the outgroup (i.e., competence) to a dimension on which one can shine (e.g., warmth), thus redeeming some esteem. This process has been labelled social creativity (Lemaine, 1974; Mummendey & Schreiber, 1983; Tajfel & Turner, 1979). Moreover, acknowledging the outgroup superiority on one dimension (e.g., “they are competent”) while achieving positive distinctiveness on the other (e.g., “we are sociable”) may help to promote positive and peaceful intergroup relationship (Mummendey & Schreiber, 1983; Tajfel & Turner, 1979; Yzerbyt et al., 2005). This possibility echoes the observation that younger generations are adopting less regionalist and more integrative positions in comparison to their predecessors (Rimé et al., 2015).

In the second part of the study, an independent sample of judges assessed the visual representations (or classification images; CIs) of the two groups. Consistent with our compensation hypothesis, judges deemed the Walloons facial renderings higher on the horizontal dimension (warmth and trustworthiness) than the Flemings faces, while the reverse was true for the vertical dimension (competence and dominance). Furthermore, the pattern proved stronger for the average CIs than the individual CIs. This is likely because the former takes into account inter-individual variability (i.e., different representation of the target for each producer), whereas the latter ignores it. Moreover, individual CIs are usually

noisier than average CIs because they build on far less trials⁶ (Brinkman et al., 2019; Cone et al., 2020; see also Schmitz, Rougier et al., under review).

Interestingly, we did not observe any difference in pattern as a function of the type of model used to assess the CIs, namely the competence-by-warmth social model and the dominance-by-trustworthiness facial model. This is somewhat surprising in the case of the vertical dimension given that its constituents do not usually convey the same information. For instance, the two sub-dimensions differ in their relation to valence, i.e., negative for dominance, positive for competence (see Oliveira, Garcia-Marques, Dotsch et al., 2019; Schmitz, Vanbeneden, et al., under review; Sutherland et al., 2016). However, visual inspection of Figures 3 and 5 suggest that we had insufficient statistical power to detect this specific interaction. In fact, differences in competence between the two targets were parallel to those on the horizontal dimensions of warmth and trustworthiness (i.e., valence); whereas it was the opposite for dominance. At a descriptive level, these results are thus in line with the divergence between dominance and competence found in Schmitz, Vanbeneden et al. (under review).

Taken together, the present findings support the idea that compensation is a pervasive phenomenon that goes beyond the conceptualization of groups in terms of abstract traits. Specifically, we obtained evidence that compensation can distort the way people represent the faces of social targets. Importantly, our results nicely replicate and increase the external validity of the findings reported in earlier efforts by showing that facial compensation also takes place with real and meaningful groups that share a long-lasting relationship.

One caveat of the current study is that the sample was limited to Walloon participants and thus did not include Flemings participants. It would therefore be interesting for future research to assess if the compensatory pattern observed among Walloons is mirrored among Flemings, or on the contrary, if Flemings would turn out to be reluctant to concede any advantage to the Walloons as a result of their superior status. Furthermore, the present sample mostly comprised young participants in their early twenties. Future investigations could examine if compensatory judgments and visual representations are also present in older generation. Given that past generations have higher regionalist aspirations and are less favorable to integration than younger ones (Rimé et al., 2015), they may prove less willing to display compensation.

In conclusion, the present endeavor increases the external validity of the visual correlate of the compensation effect by relying on meaningful and existing social groups. Moreover, it enriches the literature about stereotypes regarding Walloons and Flemings, and therefore contribute to the better understanding of the Belgian intergroup conflict.

⁶ In the present research, each individual CI was built on 150 trials, whereas each average CI was built on $150 \times 72 = 10,800$ trials.

5. Open Practices

The data, R code, and material are publicly available on Open Science Framework (https://osf.io/rj3xk/?view_only=c7064782e1524a48992a401722675ea8).

6. References

- Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (in press). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*.
- Brinkman, L., Goffin, S., van de Schoot, R., van Haren, N. E. M., Dotsch, R., & Aarts, H. (2019). Quantifying the informational value of classification images. *Behavior Research Methods*, 51(5), 2059–2073. <https://doi.org/10.3758/s13428-019-01232-2>
- Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, 28(1), 333–361. <https://doi.org/10.1080/10463283.2017.1381469>
- Cambon, L., & Yzerbyt, V. Y. (2016). Compensation is for real: Evidence from existing groups in the context of actual relations. *Group Processes & Intergroup Relations*, 20(6), 745–756. <https://doi.org/10.1177/1368430215625782>
- Cambon, L., & Yzerbyt, V. Y. (2018). Two routes toward compensation: An investigation into the mechanisms of compensation for high-and low-status groups. *Journal of Experimental Social Psychology*, 77, 24–35. <https://doi.org/10.1016/j.jesp.2018.04.002>
- Cambon, L., Yzerbyt, V. Y., & Yakimova, S. (2015). Compensation in intergroup relations: An investigation of its structural and strategic foundations. *British Journal of Social Psychology*, 54(1), 140–158. <https://doi.org/10.1111/bjso.12067>
- Chapais, B. (2015). Competence and the evolutionary origins of status and power in humans. *Human Nature*, 26(2), 83–161. <http://dx.doi.org/10.1007/s12110-015-9227-6>
- Chen, F. F., Jing, Y., & Lee, J. M. (2014). The looks of a leader: Competent and trustworthy, but not dominant. *Journal of Experimental Social Psychology*, 51, 27–33. <https://doi.org/10.1016/j.jesp.2013.10.008>
- Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology*, 104(1), 103–125. <http://dx.doi.org/10.1037/a0030398>.
- Cone, J., Brown-Iannuzzi, J. L., Lei, R., & Dotsch, R. (2020). Type I Error Is Inflated in the Two-Phase Reverse Correlation Procedure. *Social Psychological and Personality Science*, 194855062093861. <https://doi.org/10.1177/1948550620938616>
- de Leeuw, J. R. (2014). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y>

- 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. <https://doi.org/10.1111/j.1467-9280.2008.02186.x>
- Fiske, S. T. (2015). Intergroup biases: a focus on stereotype content. *Current Opinion in Behavioral Sciences*, 3, 45–50. <https://doi.org/10.1016/j.cobeha.2015.01.010>
- 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. <https://doi.org/10.1037/0022-3514.82.6.878>
- Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in psychology*, 4, Article 386. <https://doi.org/10.3389/fpsyg.2013.00386>
- Jack, R. E., & Schyns, P. G. (2017). Toward a Social Psychophysics of Face Communication. *Annual Review of Psychology*, 68(1), 269–297. <https://doi.org/10.1146/annurev-psych-010416-044242>
- Jaeger, B. (2017). r2glmm: Computes R Squared for Mixed (Multilevel) Models (R package version 0.1.2). Retrieved from <https://CRAN.R-project.org/package=r2glmm>
- Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, 89(6), 899–913. <https://doi.org/10.1037/0022-3514.89.6.899>
- 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. <https://doi.org/10.1037/a0028347>
- Kervyn, N., Bergsieker, H. B., Grignard, F., & Yzerbyt, V. (2016). An advantage of appearing mean or lazy: Amplified impressions of competence or warmth after mixed descriptions. *Journal of Experimental Social Psychology*, 62, 17–23. <https://doi.org/10.1016/j.jesp.2015.09.004>
- Kervyn, N., Fiske, S. T., & Yzerbyt, V. (2013). Integrating the stereotype content model (warmth and competence) and the Osgood semantic differential (evaluation, potency, and activity). *European Journal of Social Psychology*, 43(7), 673–681. <https://doi.org/10.1002/ejsp.1978>
- Kervyn, N., Yzerbyt, V., Demoulin, S., & Judd, C. M. (2008). Competence and warmth in context: The compensatory nature of stereotypic views of national groups. *European Journal of Social Psychology*, 38(7), 1175–1183. <http://dx.doi.org/10.1002/ejsp.526>
- Klein, O., Bouchat, P., Azzi, A., & Luminet, O. (2017). Principled Disagreements: Adhesion to Intergroup Justice Standards in the Context of the Belgian Linguistic Conflict. *Psychologica Belgica*, 57(3), 13. <https://doi.org/10.5334/pb.345>
- Klein, O., Licata, L., Van der Linden, N., Mercy, A., & Luminet, O. (2011). A waffle-shaped model for how realistic dimensions of the Belgian conflict structure collective memories and stereotypes. *Memory Studies*, 5(1), 16–31. <https://doi.org/10.1177/1750698011424028>

- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, 24(8), 1377–1388. <https://doi.org/10.1080/02699930903485076>
- Lemaine, G. (1974). Social differentiation and social originality. *European Journal of Social Psychology*, 4(1), 17–52. <https://doi.org/10.1002/ejsp.2420040103>
- Leyens, J.-P., & Yzerbyt, V. Y. (1992). The ingroup overexclusion effect: Impact of valence and confirmation on stereotypical information search. *European Journal of Social Psychology*, 22(6), 549–569. <https://doi.org/10.1002/ejsp.2420220604>
- Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: estimating the information employed for face classifications. *Cognitive Science*, 28(2), 209–226. https://doi.org/10.1207/s15516709cog2802_4
- Meeusen, C., Boonen, J., & Dassonneville, R. (2017). The Structure of Prejudice and Its Relation to Party Preferences in Belgium: Flanders and Wallonia Compared. *Psychologica Belgica*, 57(3), 52. <https://doi.org/10.5334/pb.335>
- Meuleman, B., Abts, K., & Meeusen, C. (2017). Walloons as General or Specific Others? A Comparison of anti-Walloon and anti-immigrant Attitudes in Flanders. *Psychologica Belgica*, 57(3), 75. <https://doi.org/10.5334/pb.336>
- Mummendey, A., & Schreiber, H.-J. (1983). Better or just different? Positive social identity by discrimination against, or by differentiation from outgroups. *European Journal of Social Psychology*, 13(4), 389–397. <https://doi.org/10.1002/ejsp.2420130406>
- Oliveira, M., Garcia-Marques, T., Dotsch, R., & Garcia-Marques, L. (2019). Dominance and competence face to face: Dissociations obtained with a reverse correlation approach. *European Journal of Social Psychology*, 00, 1–15. <https://doi.org/10.1002/ejsp.2569>
- Oliveira, M., Garcia-Marques, T., Garcia-Marques, L., & Dotsch, R. (2019). Good to Bad or Bad to Bad? What is the relationship between valence and the trait content of the Big Two?. *European Journal of Social Psychology*, 00, 1–21. <https://doi.org/10.1002/ejsp.2618>
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087–11092. <https://doi.org/10.1073/pnas.0805664105>
- Rimé, B., Bouchat, P., Klein, O., & Licata, L. (2015). When collective memories of victimhood fade: Generational evolution of intergroup attitudes and political aspirations in Belgium. *European Journal of Social Psychology*, 45(4), 515–532. <https://doi.org/10.1002/ejsp.2104>
- Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, 9(2), 260–264. <https://doi.org/10.1037/a0014681>
- Schmitz, M., Rougier, M., & Yzerbyt, V. Y. (under review). Introducing the Brief Reverse Correlation: An improved tool to assess visual representations.
- Schmitz, M., Vanbeneden, A., & Yzerbyt, V. Y. (under review). When you look dominant but cold: Taking compensation at the face value.

- Schmitz, M., & Yzerbyt, V. Y. (in press). Direct and indirect dimensional compensation: Is there a difference between observers and group members? *Group Processes & Intergroup Relations*. <https://doi.org/10.1177/1368430220963176>
- Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, 127(1), 105–118. <http://dx.doi.org/10.1016/j.cognition.2012.12.001>.
- Sutherland, C. A. M., Oldmeadow, J. A., & Young, A. W. (2016). Integrating social and facial models of person perception: Converging and diverging dimensions. *Cognition*, 157, 257–267. <https://doi.org/10.1016/j.cognition.2016.09.006>
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup relations. In W. G. Austin & S. Worchel (Eds.), *The psychology of intergroup relations*. Monterey, CA: Brooks-Cole.
- Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, 1(1124), 208–224. <http://doi.org/10.1196/annals.1440.012>
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*, 66(1), 519–545. <https://doi.org/10.1146/annurev-psych-113011-143831>
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in cognitive sciences*, 12(12), 455–460. <https://doi.org/10.1016/j.tics.2008.10.001>
- Toscano, H., Schubert, T. W., Dotsch, R., Falvello, V., & Todorov, A. (2016). *Physical strength as a cue to dominance: A data-driven approach*. *Personality and Social Psychology Bulletin*, 42(12), 1603–1616. <https://doi.org/10.1177/0146167216666266>
- Van der Linden, N., & Roets, A. (2017). Insights into the Belgian Linguistic Conflict from a (Social) Psychological Perspective: Introduction to the Special Issue. *Psychologica Belgica*, 57(3), 1–12. <https://doi.org/10.5334/pb.418>
- Yzerbyt, V. (2018). The dimensional compensation model: Reality and strategic constraints on warmth and competence in intergroup perceptions. In *Agency and communion in social psychology* (pp. 126–141). Routledge. <https://doi.org/10.4324/9780203703663>
- Yzerbyt, V. Y., & Cambon, L. (2017). The dynamics of compensation: When ingroup favoritism paves the way for outgroup praise. *Personality and Social Psychology Bulletin*, 43(5), 587–600. <https://doi.org/10.1177/0146167216689066>
- Yzerbyt, V. Y., Provost, V., & Corneille, O. (2005). Not Competent but Warm... Really? Compensatory stereotypes in the French-speaking world. *Group Processes & Intergroup Relations*, 8(3), 291–308. <https://doi.org/10.1177/1368430205053944>

General discussion

General discussion

The present doctoral dissertation builds on a series of empirical articles. In each of them, we carefully discussed the specific findings. For this reason, our General Discussion aims to offer a comprehensive picture of the different lines of research and to extract conclusions from those findings that emerged across the research program. Finally, we highlight some of the strengths and limitations of our work and offer new avenues of research.

1. Objective of the thesis

More than half a century after Asch's (1946, 1952) seminal experiments on impression formation, research has convincingly shown that two fundamental dimensions map the content of social perception. In the domain of self-other perception, they have traditionally been called communion and agency (Abele, Cuddy, Judd, & Yzerbyt, 2008; Abele & Wojciszke, 2007), whereas in the domain of group perception, they often come under the labels of warmth and competence (Fiske, Cuddy, Glick, & Xu, 2002). A similar bi-dimensional configuration also emerged in the domain of face perception with the dimensions of trustworthiness and dominance (Oosterhof & Todorov, 2008). More recently, and as a means to unify these various efforts, these two dimensions have been presented as the horizontal (communion, warmth, and trustworthiness) and vertical (agency, competence, dominance) dimensions (Abele et al., in press). Although different combinations of the two dimensions can give rise to a wide variety of impressions, it so happens that stereotypes of social groups are often mixed or ambivalent (Fiske et al., 2002). This pattern of assigning a higher standing on different attributes (i.e., warmth vs. competence) to each one of the targets refers to the so-called compensation effect (Yzerbyt et al., 2005).

Research on compensation has shown that the phenomenon is quite robust and rather widespread (for a review, see Yzerbyt, 2018). In fact, it emerges when evaluating fictitious groups in experimental settings or real groups in more ecological settings, whether the social targets are groups or individual, and whether perceivers belong to one of the evaluated groups or if they are external observers. Investigations have also identified a series of boundary conditions. Particularly, compensation requires a comparison context, a difference in status between targets that is perceived as legitimate and stable, and a low or moderate level of conflict.

Despite this large body of empirical research, little is known about the motivational antecedents that may drive compensation. In our view, this is mainly due to the bulk of studies relying on the so-called direct measures. Indeed, the emergence of compensatory judgments on Likert-type scales may stem from different motivations that cannot be properly assessed and disentangled by simply asking people to report their impressions. The sole use of direct measures is not only open to demand characteristics but also to social desirability issues (Fazio & Olson, 2003; Greenwald & Banaji, 1995). This state of affairs is all the more problematic that there is some evidence suggesting that compensation is prone to self-

presentation concerns (Cambon & Yzerbyt, 2018; Yzerbyt & Cambon, 2017). In this context, indirect measures may come as a particularly valuable tool to circumvent demand characteristics and social desirability issues (Fazio & Olson, 2003; Greenwald & Banaji, 1995). More importantly, discrepancies (and consistencies) between directly and indirectly measured judgments may help to better understand the motivations that foster compensation.

On a different note, we observe that there is essentially no research on the consequences of compensation beyond the conceptual level. Indeed, compensation research generally assessed participants' representations by means of verbal reports such as ratings scales, open descriptions, or, in some instances, casual attribution judgments (e.g., Judd et al., 2005; Kervyn, Yzerbyt, & Judd, 2011). This is unfortunate when considering the crucial role of non-verbal and non-conceptual components such as faces on everyday life social interactions (Jack & Schyns, 2017; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015) as well as the consequences of face-based social attributions on important social outcomes like voting intention, mating preferences, or sentencing decisions (e.g., Karremans, Dotsch, & Corneille, 2011; Olivola, Tingley, & Todorov, 2018; Porter, ten Brinke, & Gustaw, 2010). As a result, we would argue that compensation should probably influence a larger array of aspects. In particular, we pose here that the visual representation of the social targets should likely reveal the impact of dimensional compensation.

Building on this analysis, the general objective of the present dissertation was to shed light on some of the antecedents and consequences of dimensional compensation on social evaluation by making use of direct and indirect measures. To do so, we relied on two distinct indirect measures, namely the Brief Implicit Association Test (BIAT; Sriram & Greenwald, 2009) and the (Brief) Reverse Correlation (Dotsch et al., 2008; Chapter 6, Schmitz, Rougier et al., under review).

2. Summary of the empirical findings

In Chapter 4 (Schmitz & Yzerbyt, in press), we tested whether compensation could emerge on both direct and indirect measures depending on whether individuals were in a position of external observers or group members. In Experiment 1, participants were external observers and learned about two fictitious groups that differed on the manipulated dimension (either competence or warmth; e.g., competence) but were similarly ambiguous on the non-manipulated dimension (e.g., warmth). After the impression formation task, they reported their impressions about the two targets on Likert-type scales (i.e., direct measure) tapping the two fundamental dimensions of social judgment. Participants then completed a BIAT (i.e., indirect measures) to gauge how strongly they associated each group with each dimension. As expected, external observers compensated on both the direct and indirect level. Experiment 2 went a step further by assigning participants to either the low- or high-competence group. In line with the standard compensation pattern, the low-competence group members perceived themselves as less competent but warmer than the high-status group on both types of measures. In sharp contrast, high-competence group members were reluctant to grant superiority on warmth to the low-competence group as revealed by

reaction-time-based measures, although they did display a compensation pattern on traits ratings.

In Chapter 5 (Schmitz, Vabeneden et al., under review) we relied on an innovative, data-driven, and allegedly indirect (see Chapter 3) technique, namely the RC method (Dotsch et al., 2008) to investigate if compensation could extend beyond the semantic level onto the visual perception of social targets. In Experiment 1, we relied on this method to capture participants' (producers) facial prototypes of two fictitious group that differed on competence but were equal on warmth. We then had another sample of participants (judges) evaluate the average classification images (CIs) on warmth- and competence-related traits. Although producers displayed a clear-cut compensatory pattern when simply asked to rate the groups, their average CIs' evaluation by external judges contrasted on warmth but only marginally on competence, thus providing some evidence for facial compensation. Experiment 2 sought to extend these initial findings by using warmth (instead of competence) as the manipulated dimension. Results were globally similar to those of Experiment 1, with the exception that the gap between CIs in terms of competence did not reach significance. The lack of dissimilarity in terms of competence came as a surprise and led us to think that the perceived difference on this dimension between the two groups may have been facially encoded as dominance. Therefore, in Experiment 3, we re-assessed the CIs from the previous two experiments with a new sample of judges, but this time on the facial dimensions of trustworthiness and dominance. As expected, a marked compensation pattern emerged for both pairs of average CIs. Finally, in Experiment 4, we sought to replicate and further ascertain these results by manipulating the experimental design (i.e., within-participants vs. between-participants) and test if competence could be distinguished from dominance, even when measured simultaneously. Results showed that the pattern from the previous experiment replicated whether the two vertical dimensions of dominance and competence were gauged together or separately.

One possible limitation from the previous findings is that the facial templates (i.e., classification images or CIs) of the two groups were examined at the average level. Indeed, shortly after our submission, an article by Cone et al. (2020) conducted computational simulations to show that evaluations of average CIs by external judges are prone to inflated Type I errors because they ignore the (potential) inter-individual variability. Although the evaluation of individual CIs offers a straightforward solution, they are usually very noisy and require an overly large number of trials to achieve an adequate quality level. Doing so turns out to be time-consuming, economically costly, or simply impractical. In light of these obstacles, we decided to develop an enhanced method (see Chapter 6, Schmitz, Rougier et al., under review), the Brief-RC, that aimed to increase the quality of individual (and average) CIs by increasing the number of stimuli (i.e., noisy faces) presented at each trial and thus improving the signal-to-noise ratio. In Experiment 1, we showed that our new method produced equally good individual and average CIs than the traditional one in half the time. In Experiment 2, we replicated these findings and showed that the Brief-RC outperformed the Traditional-RC when compared across the same number of trials, time length, and number of stimuli.

Finally, because our previous experiments (Chapters 4 and 5) were based exclusively on fictitious groups, our Chapter 7 (Schmitz, Terache et al., in preparation) allowed us to use this enhanced method to test the external validity of the visual correlate of the compensation effect. To do so, we turned to two groups that proved to be real and meaningful with long-lasting intergroup relations, namely the Walloons and Flemings. In this study, French speaking Walloon participants provided their impressions on a series of traits encompassing the warmth and competence dimensions, before undergoing two Brief-RC that captured their facial representation of each of the two groups. Blind judges then rated the two average CIs as well as the individual CIs on the social (i.e., warmth and competence) and facial (i.e., dominance and trustworthiness) dimensions. In line with our hypotheses, Walloons came across as less competent but warmer than Flemings at the conceptual level (i.e., traits rating by producers). This pattern was replicated at the facial level, although it proved stronger for average than individual CIs. Moreover, visual inspection of the results suggested a potential divergence between the vertical dimensions of competence and dominance comparable to the findings from Chapter 5.

3. Discussion

A first important and consistent finding from our research is that compensation does show up when assessed with indirect measures. That is, compensatory judgments emerge when the individual is not informed of what is being measured nor requested to self-report (Greenwald & Banaji, 1995). Moreover, we identified a moderating factor to the relation between direct and indirect measures, namely the vantage points of the perceiver. In line with earlier efforts (e.g., Imhoff et al., 2013; Kervyn, Yzerbyt, & Judd, 2011), we showed in three experiments (Schmitz, Vanbeneden et al., under review, Exp. 3–4; Schmitz & Yzerbyt, in press, Exp. 1) that when participants adopted the position of external observers, they compensated at both the direct and indirect level, whether the evaluated target groups initially differed in terms of competence or warmth (i.e., the manipulated dimension). Additionally, we showed in two studies (Schmitz, Terache et al., in preparation; Schmitz & Yzerbyt, in press, Exp. 2) that a similar pattern emerged when participants were members of low-competence groups, whether these groups were fictitious or real (i.e., Walloons). In stark contrast, we provided empirical evidence (Schmitz & Yzerbyt, in press, Exp. 1) that high-competence group members compensated at the direct level but not at the indirect level.

Taken together, the aforementioned findings help us to better understand and to further disentangle in a systematic manner the motivational antecedents of compensation. Specifically, the diverging pattern between self-reports and less controllable measures observed in privileged group members contrasts with the converging pattern displayed by external observers and disadvantaged group members and suggests, at the very least, that distinct driving forces may underlie compensatory judgment for these participants. Based on previous empirical work and theoretical analyses (see Chapter 2 and 3), we reasoned that the main incentive behind the public expression of compensation would be, respectively, self-presentation concerns (i.e., to appear non-discriminatory) for high-competence group members, self-affirmation (coupled with structural constraints; i.e., the objectivity of

competence vs. malleability of warmth) for low-competence group members, and justice concerns for external observers.

A second noteworthy and systematic result is that the consequences of compensation can transcend abstract group perception and show up in the visual representation of social targets. Across three studies (Chapter 7, Schmitz, Terache et al., in preparation; Chapter 5, Schmitz et al., under review, Exp. 3–4), we showed that the classification images (CIs) of fictitious and real groups could be biased in a compensatory way on the horizontal and vertical dimensions. An interesting additional outcome is that the visual bias appears to operate more markedly on the facial dimension of dominance rather than on the social dimension of competence. These slight discrepancies between the two vertical dimensions add to the thriving body of literature suggesting that they operate in somewhat distinct manners (i.e., dominance tends to be negatively associated with valence, whereas competence is usually positively associated with valence) (e.g., Carrier, Louvet, & Rohmer, 2014; Imhoff & Koch, 2017; Oliveira, Garcia-Marques, Garcia-Marques et al., 2019; Sutherland et al., 2016).

A third interesting and consistent finding is that compensation can emerge in a very spontaneous manner. Indeed, previous research on the topic restrained participants' responses to a limited set of options determined *a priori* by the researchers. For instance, when conveying their impressions, participants could only do so on a handful of pre-selected traits encompassing the two dimensions, and not on any trait or dimension they may have freely chosen or that may have been pertinent if presented with. This shortcoming extends to indirect measure that make use of handpicked traits and group labels. These constraints mean that only a limited set of specific pre-established hypotheses can be tested, instead of any set of hypotheses (which the researcher may not even be aware of or that were not theoretically anticipated). Our use of the reverse correlation method offered an opportunity to overcome these limitations (Brinkman et al., 2017) by allowing participants to spontaneously rely on whatever criteria they wanted in order to make their choices (i.e., to select the target face). Because the set of possible outcomes (i.e., the CIs) is immensely large, this allows researchers to test a vast sample of hypotheses instead of just a few. In this sense, this data-driven technique makes no prior assumption about what these end-products may look like. It is under these unconstrained conditions that we showed across several studies the spontaneous appearance of facial compensation (Schmitz, Terache et al., in preparation; Schmitz, Vanbeneden et al., under review).

In light of these various messages, we think that the present endeavor, by making use of direct and indirect measures, greatly contributes to the literature on social evaluation in general and compensation in particular by uncovering some of the motivational antecedents of compensation. It also constitutes the first attempt to do so by incorporating a crucial factor, namely, the social position of individuals as observers or group members. Moreover, we provide the first systematic evidence that compensation can have concrete consequences not only on how we think about social groups, but also how we visually picture them in our minds, thus emphasizing the pervasiveness of this phenomenon. Likewise, we collected the first empirical and systematic evidence showing that compensation can take place spontaneously, that is, without constraining people's responses. Also, our findings have

practical implications as they widen the current knowledge about the long-lasting conflict between the two Belgian linguistic communities (i.e., Walloons and Flemings). Along the way, we also developed an enhanced reverse correlation method that greatly facilitates the access to higher quality CIs and will contribute to the reproducibility of research by substantially reducing the risk of false positives.

Before examining the limitations of this dissertation, we deem it important to highlight some of its strengths. For instance, our work made use of fictitious groups with tightly controlled descriptions as well as real and meaningful groups, thus respectively increasing the internal and external validity of our findings (Brewer, 2000). In a similar vein, we relied on cutting-edge statistical techniques, namely linear mixed models, to reduce Type I errors and further generalize our results to more than one random factor (e.g., to other samples of producers and judges) (Judd, Westfall, & Kenny, 2017), and Bayesian analyses to better gauge the evidence in favor of our null hypotheses (Dienes, 2014). Furthermore, we developed and made use of an improved version of the reverse correlation (i.e., the Brief-RC; Schmitz, Rougier et al., under review) that efficiently delivers higher quality individual (and average) CIs thus minimizing the risk of false positives (Cone et al., 2020). Additionally, we increased the external validity of our findings by collecting samples from demographically diverse and more representative populations than simply W.E.I.R.D. (Western, Educated, Industrialized, Rich, and Democratic) populations via online crowdsourcing platforms (i.e., Prolific Academic; Gleibs, 2017; Peer, Brandimarte, Samat, & Acquisti, 2017). Last but not least, we contributed to foster a more transparent and replicable social psychology science by pre-registering (some of) our research and by making our data, analyses, and materials publicly accessible to researchers (Open Science Collaboration, 2015; Simmons et al., 2011).

4. Limitations and future research

Our research program examined several motivations that potentially fuel compensation, namely the need for harmonious intergroup relations, justice concerns, self-affirmation concerns, and self-presentation concerns (see Chapter 3). The divergences (convergences) observed between direct and indirect measures as a function of participants' social position (as external observers or group members) provided valuable insights about those motivations that might be at work as a function of these contextual factors. However, because more than one motive may produce a given pattern, more research is required to further disentangle each specific motive. For instance, the prevalence of justice concerns could be assessed by the means of a lexical decision task (LDT), such that people who are more sensitive to justice concerns would be faster at recognizing justice-related words than neutral or control words (see Kay & Jost, 2003; see also Hafer, 2000). Our expectation would be that the impact of justice concerns (as measured by the LDT) on compensatory tendencies would be larger among external observers than with advantaged or disadvantaged group members, who are presumably more concerned about self-presentation and self-affirmation, respectively.

Another caveat is that we only tested direct and indirect compensation in real groups when people were members of the low-status group (i.e., Walloons). The question thus remains whether the pattern of results observed among high-status group members and external observers in our controlled experiments generalizes to real-life scenarios. For instance, it would be very interesting to examine if Walloons' facial compensation about themselves and Flemings also holds for Flemings (i.e., the high-status group). Based on the results from Schmitz and Yzerbyt (in press; Chapter 4), we would expect Flemings to show compensation on traits ratings (with perhaps a slightly more pronounced pattern on competence and a less pronounced pattern on warmth than the one found among Walloons, see Yzerbyt et al., 2005). In contrast, the evaluation of their CIs would signal that Flemings less readily concede an advantage on warmth to the Walloons. Likewise, new studies could rely on a third, admittedly neutral, community or national group (e.g., German-speaking Belgian or German participants) as external observers to contrast their direct and indirect representation of the two linguistic groups. In all likelihood, we would predict compensation to take place in such a case.

Two additional shortcomings are related to the way we gauged the reverse correlation outputs. First, and as in most RC-based studies, we tested our hypotheses by relying on subjective human judgments of the CIs (i.e., traits ratings), which may be biased and, in some instances, quite variable (Hehman et al., 2019; Hehman et al., 2017). An alternative strategy is to rely on pixel luminance correlations (see Brinkman et al., 2017; Dotsch & Todorov, 2012; for an illustrative example see Imhoff et al., 2013) which assesses the objective and physical (dis)similarity between the target CI (e.g., the Walloon CI) and a CI that is diagnostic of the dimension of interest (e.g., the CI of a warmth-looking person). Second, even though the CIs are produced in a spontaneous (i.e., unconstrained) manner, we conducted their evaluation by judges on a limited and predefined set of traits. To the best of our knowledge, no solution has been proposed so far to overcome this undiscussed issue. In the case of judges, we believe a solution might be to ask them to describe the CIs on a free basis and to use text mining techniques (see Chen & Wojcik, 2016) to extract the traits and their respective weights used in the descriptions. This specific method could also be employed to examine the spontaneous conceptual representation of target groups (e.g., “describe in a few lines your impression about the Walloons”).

A final noteworthy limitation is that in Chapter 5 (Schmitz, Vanbeneden et al., under review) we did not directly test what we would call the “encoding hypothesis”: The fact that the perceived conceptual difference in competence between the two groups (as reported by producers' traits ratings) is presumably facially encoded as differences in dominance instead of competence (as determined by judges). Beside the fact that testing this hypothesis was outside the scope of the present endeavor, it was not possible to probe this assumption because the behavioral statements that we used to describe the fictitious groups had been pre-selected on the basis of their unique diagnosticity in terms of warmth and competence (see Judd et al., 2005), and not dominance. To examine this conjecture, future research could, as an initial step, make use of CIs generated with two-trait combinations (see Oliveira, Garcia-Marques, & Dotsch, 2019). For instance, by generating the CIs of a “competent but *not*

dominant” and “dominant but *not* competent” target person or group. The objective and/or subjective similarities (dissimilarities) between the CIs could be indicative of whether our hypothesis holds (or not).

Conclusion

Conclusion

The objective of the present dissertation was to shed light on some of the antecedents and consequences of the compensation effect on social evaluation by making use of direct and indirect measures. Overall, discrepancies and consistencies between the two types of measurement suggest that distinct compensatory motivations are likely at play, depending notably on the vantage point of the perceiver. Our results suggest, at the very least, that the underlying motivations from high-standing group members differ from those of the low-standing group members and external observers. We go further by positing that disadvantaged group members compensate primarily because of self-affirmation concerns, advantaged group members do so because of self-presentation concerns, and external observers mainly as a way to satisfy their justice concerns. More research should help to further disentangle and nail down these various motivations and conditions that actualize each of them.

Clearly, our findings also stress the pervasiveness of dimensional compensation by showing that its consequences can transcend the abstract conceptualization of social targets and spontaneously alter the way they are represented visually in people's mind. Because most of our research program relied on experimental settings with fictitious groups as social targets, future work should definitely seek to generalize our findings to real groups and real-life scenarios. Our research agenda will definitely include this as a goal for the future.

Bibliographical references

Bibliographical references

- Abele, A. E. (2003). The dynamics of masculine-agentic and feminine-communal traits: Findings from a prospective study. *Journal of Personality and Social Psychology*, 85(4), 768–776. <https://doi.org/10.1037/0022-3514.85.4.768>
- Abele, A. E., & Brack, S. (2013). Preference for other persons' traits is dependent on the kind of social relationship. *Social Psychology*, 44(2), 84–94. <https://doi.org/10.1027/1864-9335/a000138>
- Abele, A. E., & Bruckmüller, S. (2011). The bigger one of the “Big Two”? Preferential processing of communal information. *Journal of Experimental Social Psychology*, 47(5), 935–948. <https://doi.org/10.1016/j.jesp.2011.03.028>
- Abele, A. E., Cuddy, A. J., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, 38(7), 1063–1065. <http://dx.doi.org/10.1002/ejsp.574>
- Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. Y. (in press). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*.
- Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness—Communion with warmth and morality. *Frontiers in psychology*, 7, 1810. <https://doi.org/10.3389/fpsyg.2016.01810>
- Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, 93(5), 751–763. <https://doi.org/10.1037/0022-3514.93.5.751>
- Abele, A. E., & Wojciszke, B. (2013). The Big Two in social judgment and behavior [Editorial]. *Social Psychology*, 44(2), 61–62. <https://doi.org/10.1027/1864-9335/a000137>
- Abele, A. E., & Wojciszke, B. (2014). Communal and Agentic Content in Social Cognition. *Advances in Experimental Social Psychology*, 195–255. <https://doi.org/10.1016/b978-0-12-800284-1.00004-7>
- Ahumada, A., & Lovell, J. (1971). Stimulus Features in Signal Detection. *The Journal of the Acoustical Society of America*, 49(6B), 1751–1756. <https://doi.org/10.1121/1.1912577>
- Ahumada, A., Marken, R., & Sandusky, A. (1975). Time and frequency analyses of auditory signal detection. *The Journal of the Acoustical Society of America*, 57(2), 385–390. <https://doi.org/10.1121/1.380453>
- Albohn, D. N., & Adams, R. B. (2020). Everyday Beliefs About Emotion Perceptually Derived From Neutral Facial Appearance. *Frontiers in Psychology*, 11, 264. <https://doi.org/10.3389/fpsyg.2020.00264>
- Ames, D., Lee, A., & Wazlawek, A. (2017). Interpersonal assertiveness: Inside the balancing act. *Social and Personality Psychology Compass*, 11(6), Article e12317. <https://doi.org/10.1111/spc3.12317>

- Anderson, N. H., & Barrios, A. A. (1961). Primacy effects in personality impression formation. *The Journal of Abnormal and Social Psychology*, 63(2), 346–350. <https://doi.org/10.1037/h0046719>
- Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, 41(3), 258–290. <https://doi.org/10.1037/h0055756>
- Asch, S. E. (1952). *Social Psychology*. Englewood Cliffs, New Jersey: Prentice Hall.
- Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion*. Rand McNally.
- Banise, R., Seise, J., & Zerbis, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie*, 48(2), 145–160. <https://doi.org/10.1026/0949-3946.48.2.145>
- Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, 6(2), 269–278. <https://doi.org/10.1037/1528-3542.6.2.269>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint*. arXiv:1506.04967.
- Bergsieker, H. B., Leslie, L. M., Constantine, V. S., & Fiske, S. T. (2012). Stereotyping by omission: Eliminate the negative, accentuate the positive. *Journal of Personality and Social Psychology*, 102(6), 1214–1238. <http://dx.doi.org/10.1037/a0027717>
- Berry, D. S. (1990). Taking people at face value: Evidence for the kernel of truth hypothesis. *Social Cognition*, 8(4), 343–361. <https://doi.org/10.1521/soco.1990.8.4.343>
- Blake, R. R., & Mouton, J. S. (1964). *The Managerial Grid*. Gulf Publishing Company.
- Brambilla, M., Sacchi, S., Pagliaro, S., & Ellemers, N. (2013). Morality and intergroup relations: Threats to safety and group image predict the desire to interact with outgroup and ingroup members. *Journal of Experimental Social Psychology*, 49(5), 811–821. <https://doi.org/10.1016/j.jesp.2013.04.005>
- Brewer, M. B. (2000). Research design and issues of validity. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (p. 3–16). Cambridge University Press.
- Brinkman, L., Dotsch, R., Zondergeld, J., Koevoets, M. G. J. C., Aarts, H., & van Haren, N. E. M. (2019). Visualizing mental representations in schizophrenia patients: A reverse correlation approach. *Schizophrenia Research: Cognition*, 17, 100138. <https://doi.org/10.1016/j.scog.2019.100138>
- Brinkman, L., Goffin, S., van de Schoot, R., van Haren, N. E. M., Dotsch, R., & Aarts, H. (2019). Quantifying the informational value of classification images. *Behavior Research Methods*, 51(5), 2059–2073. <https://doi.org/10.3758/s13428-019-01232-2>
- Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, 28(1), 333–361. <https://doi.org/10.1080/10463283.2017.1381469>
- Brooks, J. A., Stoller, R. M., & Freeman, J. B. (2018). Stereotypes Bias Visual Prototypes for Sex and Emotion Categories. *Social Cognition*, 36(5), 481–493. <https://doi.org/10.1521/soco.2018.36.5.481>

- Brown-Iannuzzi, J. L., Dotsch, R., Cooley, E., & Payne, B. K. (2017). The relationship between mental representations of welfare recipients and attitudes toward welfare. *Psychological Science*, 28(1), 92–103. <https://doi.org/10.1177/0956797616674999>
- Brown-Iannuzzi, J. L., McKee, S., & Gervais, W. M. (2018). Atheist horns and religious halos: Mental representations of atheists and theists. *Journal of Experimental Psychology: General*, 147(2), 292–297. <https://doi.org/10.1037/xge0000376>
- Bruce, V., & Young, A. W. (2012). *Face perception*. London, England: Psychology Press.
- Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, 1(1), 1–20. <https://doi.org/10.5334/joc.10>
- Cambon, L., & Yzerbyt, V. Y. (2016). Compensation is for real: Evidence from existing groups in the context of actual relations. *Group Processes & Intergroup Relations*, 20(6), 745–756. <https://doi.org/10.1177/1368430215625782>
- Cambon, L., & Yzerbyt, V. Y. (2018). Two routes toward compensation: An investigation into the mechanisms of compensation for high-and low-status groups. *Journal of Experimental Social Psychology*, 77, 24–35. <https://doi.org/10.1016/j.jesp.2018.04.002>
- Cambon, L., Yzerbyt, V. Y., & Yakimova, S. (2015). Compensation in intergroup relations: An investigation of its structural and strategic foundations. *British Journal of Social Psychology*, 54(1), 140–158. <http://dx.doi.org/10.1111/bjso.12067>
- Carlsson, R., & Björklund, F. (2010). Implicit stereotype content: Mixed stereotypes can be measured with the implicit association test. *Social Psychology*, 41(4), 213–222. <http://dx.doi.org/10.1027/1864-9335/a000029>
- Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). *Facial Structure Is a Reliable Cue of Aggressive Behavior*. *Psychological Science*, 20(10), 1194–1198. <https://doi.org/10.1111/j.1467-9280.2009.02423.x>
- Carrier, A., Dompnier, B., & Yzerbyt, V. Y. (2019). Of nice and mean: The personal relevance of others' competence drives perceptions of warmth. *Personality and Social Psychology Bulletin*, 45(11), 1549–1562. <https://doi.org/10.1177/0146167219835213>
- Carrier, A., Louvet, E., Chauvin, B., & Rohmer, O. (2014). The primacy of agency over competence in status perception. *Social Psychology*, 45(5), 347–356. <https://doi.org/10.1027/1864-9335/a000176>
- Carrier, A., Louvet, E., & Rohmer, O. (2014). Compétence et agentisme dans le jugement social [Competence and agency in social judgment]. *Revue Internationale de Psychologie Sociale*, 27(1), 95–125.
- Chapais, B. (2015). Competence and the evolutionary origins of status and power in humans. *Human Nature*, 26(2), 83–161. <http://dx.doi.org/10.1007/s12110-015-9227-6>
- Chen, E. E., & Wojcik, S. P. (2016). A practical guide to big data research in psychology. *Psychological Methods*, 21(4), 458–474. <http://dx.doi.org/10.1037/met0000111>
- Chen, F. F., Jing, Y., & Lee, J. M. (2014). The looks of a leader: Competent and trustworthy, but not dominant. *Journal of Experimental Social Psychology*, 51, 27–33. <https://doi.org/10.1016/j.jesp.2013.10.008>

- Cheng, J. T., & Tracy, J. L. (2014). Toward a unified science of hierarchy: Dominance and prestige are two fundamental pathways to human social rank. In J. T. Cheng, J. L. Tracy, & C. Anderson (Eds.), *The psychology of social status* (p. 3–27). Springer Science + Business Media. https://doi.org/10.1007/978-1-4939-0867-7_1
- Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology*, 104(1), 103–125. <http://dx.doi.org/10.1037/a0030398>.
- Chiao, J. Y., Bowman, N. E., & Gill, H. (2008). The political gender gap: Gender bias in facial inferences that predict voting behavior. *PloS one*, 3(10), e3666. <https://doi.org/10.1371/journal.pone.0003666>
- Cichocka, A., Winiewski, M., Bilewicz, M., Bukowski, M., & Jost, J. T. (2015). Complementary stereotyping of ethnic minorities predicts system justification in Poland. *Group Processes & Intergroup Relations*, 18(6), 788–800. <https://doi.org/10.1177/1368430214566891>
- Cislak, A., & Wojciszke, B. (2008). Agency and communion are inferred from actions serving interests of self or others. *European Journal of Social Psychology*, 38(7), 1103–1110. <https://doi.org/10.1002/ejsp.554>
- Clément-Guillotin, C., Rohmer, O., Forestier, C., Guillotin, P., Deshayes, M., & d'Arripe-Longueville, F. (2018). Implicit and explicit stereotype content associated with people with physical disability: Does sport change anything? *Psychology of Sport and Exercise*, 38, 192–201. <http://dx.doi.org/10.1016/j.psychsport.2018.06.014>
- Collange, J., Fiske, S. T., & Sanitioso, R. (2009). Maintaining a positive self-image by stereotyping others: Self-threat and the stereotype content model. *Social cognition*, 27(1), 138–149. <https://doi.org/10.1521/soco.2009.27.1.138>
- Cone, J., Brown-Iannuzzi, J. L., Lei, R., & Dotsch, R. (2020). Type I Error Is Inflated in the Two-Phase Reverse Correlation Procedure. *Social Psychological and Personality Science*, 194855062093861. <https://doi.org/10.1177/1948550620938616>
- 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*, 1088868320911325. <https://doi.org/10.1177/1088868320911325>
- Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, 82(3), 359–378. <http://dx.doi.org/10.1037/0022-3514.82.3.359>
- Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2004). When Professionals Become Mothers, Warmth Doesn't Cut the Ice. *Journal of Social Issues*, 60(4), 701–718. <https://doi.org/10.1111/j.0022-4537.2004.00381.x>
- 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. <https://doi.org/10.1037/0022-3514.92.4.631>
- Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. In M. P. Zanna (Ed.), *Advances in experimental social psychology: Vol. 40. Advances in*

- experimental social psychology* (Vol. 40, p. 61–149). Elsevier Academic Press. [https://doi.org/10.1016/S0065-2601\(07\)00002-0](https://doi.org/10.1016/S0065-2601(07)00002-0)
- Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., Bond, M. H., Croizet, J.-C., Ellemers, N., Sleebos, E., Htun, T. T., Kim, H.-J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., ... Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, 48(1), 1–33. <https://doi.org/10.1348/014466608x314935>
- Cuddy, A. J. C., Glick, P., & Beninger, A. (2011). The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in organizational behavior*, 31, 73–98. <https://doi.org/10.1016/j.riob.2011.10.004>
- Cuddy, A. J. C., Norton, M., & Fiske, S. T. (2005). This old stereotype: The stubbornness and pervasiveness of the elderly stereotype. *Journal of Social Issues*, 61(2), 267–285.
- Dai, H., & Micheyl, C. (2010). Psychophysical reverse correlation with multiple response alternatives. *Journal of Experimental Psychology: Human Perception and Performance*, 36(4), 976–993. <https://doi.org/10.1037/a0017171>
- Dambrun, M., & Guimond, S. (2004). Implicit and explicit measures of prejudice and stereotyping: Do they assess the same underlying knowledge structure? *European Journal of Social Psychology*, 34(6), 663–676. <http://dx.doi.org/10.1002/ejsp.223>
- Dang, J., Liu, L., Liang, Y., & Ren, D. (2017). Expected hierarchical integration reduces perceptions of a low status group as less competent than a high status group while maintaining the same level of perception of warmth. *Frontiers in Psychology*, 7, Article 2068. <https://doi.org/10.3389/fpsyg.2016.02068>
- Degner, J., Mangels, J., & Zander, L. (2019). Visualizing Gendered Representations of Male and Female Teachers Using a Reverse Correlation Paradigm. *Social Psychology*, 50(4), 233–251. <https://doi.org/10.1027/1864-9335/a000382>
- De Houwer, J., Beckers, T., & Moors, A. (2007). Novel attitudes can be faked on the Implicit Association Test. *Journal of Experimental Social Psychology*, 43(6), 972–978. <https://doi.org/10.1016/j.jesp.2006.10.007>
- de Leeuw, J. R. (2014). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y>
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781. <https://doi.org/10.3389/fpsyg.2014.00781>
- Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, 73(6), 1246–1256. <https://doi.org/10.1037/0022-3514.73.6.1246>
- Dotsch, R. (2017). rcicr: Reverse-Correlation Image-Classification Toolbox (R package version 0.4.1). Retrieved from <https://rdr.io/cran/rcicr/>
- Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, 3(5), 562–571. <https://doi.org/10.1177/1948550611430272>

- 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. <https://doi.org/10.1111/j.1467-9280.2008.02186.x>
- Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of personality and social psychology*, 100(6), 999–1014. <https://doi.org/10.1037/a0023026>
- Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2013). Behavioral information biases the expected facial appearance of members of novel groups. *European Journal of Social Psychology*, 43(1), 116–125. <https://doi.org/10.1002/ejsp.1928>
- Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 36, p. 1–52). Elsevier Academic Press. [https://doi.org/10.1016/S0065-2601\(04\)36001-6](https://doi.org/10.1016/S0065-2601(04)36001-6)
- Dufner, M., Denissen, J., Sedikides, C., Van Zalk, M., Meeus, W. H. J., & Van Aken, M. (2013). Are actual and perceived intellectual self-enhancers evaluated differently by social perceivers? *European Journal of Personality*, 27(6), 621–633. <https://doi.org/10.1002/per.1934>
- Durante, F., Fiske, S. T., Gelfand, M. J., Crippa, F., Suttora, C., Stillwell, A., Asbrock, F., Aycan, Z., Bye, H. H., Carlsson, R., Björklund, F., Dagher, M., Geller, A., Larsen, C. A., Latif, A.-H. A., Mähönen, T. A., Jasinskaja-Lahti, I., & Teymoori, A. (2017). Ambivalent stereotypes link to peace, conflict, and inequality across 38 nations. *Proceedings of the National Academy of Sciences*, 114(4), 669–674. <https://doi.org/10.1073/pnas.1611874114>
- Durante, F., Fiske, S. T., Kervyn, N., Cuddy, A. J. C., Akande, A. D., Adetoun, B. E., Adewuyi, M. F., Tserere, M. M., Ramiah, A. A., Mastor, K. A., Barlow, F. K., Bonn, G., Tafarodi, R. W., Bosak, J., Cairns, E., Doherty, C., Capozza, D., Chandran, A., Chrysoschoou, X., ... Storari, C. C. (2013). Nations' income inequality predicts ambivalence in stereotype content: How societies mind the gap. *British Journal of Social Psychology*, 52(4), 726–746. <https://doi.org/10.1111/bjso.12005>
- Eagly, A. H. (1987). *John M. MacEachran memorial lecture series; 1985. Sex differences in social behavior: A social-role interpretation*. Lawrence Erlbaum Associates, Inc.
- Eberhardt, J. L., Dasgupta, N., & Banaszynski, T. L. (2003). Believing is seeing: The effects of racial labels and implicit beliefs on face perception. *Personality and Social Psychology Bulletin*, 29(3), 360–370. <https://doi.org/10.1177/0146167202250215>
- Ebert, I. D., Steffens, M. C., & Kroth, A. (2014). Warm, but maybe not so competent?—Contemporary implicit stereotypes of women and men in Germany. *Sex roles*, 70(9-10), 359–375. <http://dx.doi.org/10.1007/s11199-014-0369-5>
- Eckes, T. (2002). Paternalistic and envious gender stereotypes: Testing predictions from the stereotype content model. *Sex Roles: A Journal of Research*, 47, 99–114. <https://doi.org/10.1023/A:1021020920715>
- Ellemers, N. (2017). *Morality and the regulation of social behavior: Groups as moral anchors*. Psychology Press. <https://doi.org/10.4324/9781315661322>
- Ellemers, N., Pagliaro, S., Barreto, M., & Leach, C. W. (2008). Is it better to be moral than smart? The effects of morality and competence norms on the decision to work at group

- status improvement. *Journal of Personality and Social Psychology*, 95(6), 1397–1410. <https://doi.org/10.1037/a0012628>
- Erlebacher, A. (1977). Design and analysis of experiments contrasting the within- and between-subjects manipulation of the independent variable. *Psychological Bulletin*, 84(2), 212–219. <http://dx.doi.org/10.1037/0033-2909.84.2.212>
- Fasoli, F., Hegarty, P., Maass, A., & Antonio, R. (2018). *Who wants to sound straight? Sexual majority and minority stereotypes, beliefs and desires about auditory gaydar. Personality and Individual Differences*, 130, 59–64. <http://dx.doi.org/10.1016/j.paid.2018.03.046>
- Fasoli, F., Maass, A., Paladino, M. P., & Sulpizio, S. (2017). Gay-and lesbian-sounding auditory cues elicit stereotyping and discrimination. *Archives of sexual behavior*, 46(5), 1261–1277. <http://dx.doi.org/10.1007/s10508-017-0962-0>
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual review of psychology*, 54(1), 297–327. <https://doi.org/10.1146/annurev.psych.54.101601.145225>
- Felipe, A. I. (1970). Evaluative versus descriptive consistency in trait inferences. *Journal of Personality and Social Psychology*, 16(4), 627–638. <https://doi.org/10.1037/h0030090>
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, 58(2), 203–210. <https://doi.org/10.1037/h0041593>
- Fiedler, F. E. (1967). *A Theory of Leadership Effectiveness*. McGraw-Hill, New York.
- Fiske, S. T. (1992). Thinking is for doing: Portraits of social cognition from Daguerreotype to laserphoto. *Journal of Personality and Social Psychology*, 63(6), 877–889. <https://doi.org/10.1037/0022-3514.63.6.877>
- Fiske, S. T. (2015). Intergroup biases: a focus on stereotype content. *Current Opinion in Behavioral Sciences*, 3, 45–50. <https://doi.org/10.1016/j.cobeha.2015.01.010>
- Fiske, S. T. (2018a). Stereotype Content: Warmth and Competence Endure. *Current Directions in Psychological Science*, 27(2), 67–73. <https://doi.org/10.1177/0963721417738825>
- Fiske, S. T. (2018b). Warmth and competence are parallels to communion and agency: Stereotype Content Model. In *Agency and communion in social psychology* (pp. 39–51). Taylor and Francis.
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences*, 11(2), 77–83. <http://dx.doi.org/10.1016/j.tics.2006.11.005>
- 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. <https://doi.org/10.1037/0022-3514.82.6.878>
- Fiske, S. T., & Ruscher, J. B. (1993). *Negative interdependence and prejudice: Whence the affect?* In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (p. 239–268). Academic Press.

- Fiske, S. T., & Tablante, C. B. (2015). Stereotyping: Processes and content. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbooks in psychology®. APA handbook of personality and social psychology, Vol. 1. Attitudes and social cognition* (p. 457–507). American Psychological Association. <https://doi.org/10.1037/14341-015>
- Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis)respecting versus (dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues*, 55(3), 473–489. <https://doi.org/10.1111/0022-4537.00128>
- Flowe, H. D. (2012). Do characteristics of faces that convey trustworthiness and dominance underlie perceptions of criminality?. *PloS one*, 7(6), e37253. <https://doi.org/10.1371/journal.pone.0037253>
- 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* (p. 283–310). Cambridge University Press.
- Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us?: Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, 2(2), 181–193. <https://doi.org/10.1111/j.1745-6916.2007.00036.x>
- Ge, L., Zhang, H., Wang, Z., Quinn, P. C., Pascalis, O., Kelly, D., Slater, A., Tian, J., & Lee, K. (2009). Two Faces of the Other-Race Effect: Recognition and Categorisation of Caucasian and Chinese Faces. *Perception*, 38(8), 1199–1210. <https://doi.org/10.1068/p6136>
- Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (p. 89–150). McGraw-Hill.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117(1), 21–38. <https://doi.org/10.1037/0033-2909.117.1.21>
- Gleibs, I. H. (2017). Are all “research fields” equal? Rethinking practice for the use of data from crowdsourcing marketplaces. *Behavior Research Methods*, 49(4), 1333–1342. <https://doi.org/10.3758/s13428-016-0789-y>
- Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70(3), 491–512. <https://doi.org/10.1037/0022-3514.70.3.491>
- Glick, P., & Fiske, S. T. (2001a). Ambivalent stereotypes as legitimizing ideologies: Differentiating paternalistic and envious prejudice. In J. T. Jost & B. Major (Eds.), *The psychology of legitimacy: Emerging perspectives on ideology, justice, and intergroup relations* (p. 278–306). Cambridge University Press.
- Glick, P., & Fiske, S. T. (2001b). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, 56(2), 109–118. <http://dx.doi.org/10.1037/0003-066X.56.2.109>
- Glick, P., & Fiske, S. T. (2011). Ambivalent sexism revisited. *Psychology of Women Quarterly*, 35(3), 530–535. <https://doi.org/10.1177/0361684311414832>

- Grant, A. M., & Schwartz, B. (2011). Too much of a good thing: The challenge and opportunity of the inverted U. *Perspectives on Psychological Science*, 6(1), 61–76. <https://doi.org/10.1177/1745691610393523>
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4–27. <https://doi.org/10.1037/0033-295X.102.1.4>
- Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, 15(4), 553–561. <http://dx.doi.org/10.1037/pspa0000016>
- Greenwald, A. G., & Lai, C. K. (2020). Implicit Social Cognition. *Annual Review of Psychology*, 71(1), 1–25. <http://dx.doi.org/10.1146/annurev-psych-010419-050837>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6), 1464–1480. <http://dx.doi.org/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. <http://dx.doi.org/10.1037/0022-3514.85.2.197>
- Greenwald, A. G., Pickrell, J. E., & Farnham, S. D. (2002). Implicit partisanship: Taking sides for no reason. *Journal of Personality and Social Psychology*, 83(2), 367–379. <https://doi.org/10.1037/0022-3514.83.2.367>
- Grill-Spector, K., Weiner, K. S., Kay, K., & Gomez, J. (2017). The Functional Neuroanatomy of Human Face Perception. *Annual Review of Vision Science*, 3(1), 167–196. <https://doi.org/10.1146/annurev-vision-102016-061214>
- Gunaydin, G., & DeLong, J. E. (2015). Reverse Correlating Love: Highly Passionate Women Idealize Their Partner's Facial Appearance. *PLOS ONE*, 10(3), e0121094. <https://doi.org/10.1371/journal.pone.0121094>
- Gundersen, A. B., & Kunst, J. R. (2018). Feminist ≠ Feminine? Feminist Women Are Visually Masculinized Whereas Feminist Men Are Feminized. *Sex Roles*, 80(5–6), 291–309. <https://doi.org/10.1007/s11199-018-0931-7>
- Hafer, C. L. (2000). Do innocent victims threaten the belief in a just world? Evidence from a modified Stroop task. *Journal of Personality and Social Psychology*, 79(2), 165–173. <https://doi.org/10.1037/0022-3514.79.2.165>
- Hahn, A., Banchevsky, S., Park, B., & Judd, C. M. (2015). Measuring intergroup ideologies: Positive and negative aspects of emphasizing versus looking beyond group differences. *Personality and Social Psychology Bulletin*, 41(12), 1646–1664. <https://doi.org/10.1177/0146167215607351>
- Hall, J. A., Coats, E. J., & LeBeau, L. S. (2005). Nonverbal Behavior and the Vertical Dimension of Social Relations: A Meta-Analysis. *Psychological Bulletin*, 131(6), 898–924. <https://doi.org/10.1037/0033-2909.131.6.898>
- Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of personality and social psychology*, 78(5), 837–852. <https://doi.org/10.1037/0022-3514.78.5.837>

- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, 4(6), 223–233. [https://doi.org/10.1016/S1364-6613\(00\)01482-0](https://doi.org/10.1016/S1364-6613(00)01482-0)
- Hehman, E., Flake, J. K., & Freeman, J. B. (2015). Static and dynamic facial cues differentially affect the consistency of social evaluations. *Personality and Social Psychology Bulletin*, 41(8), 1123–1134. <https://doi.org/10.1177/0146167215591495>
- Hehman, E., Stoller, R. M., Freeman, J. B., Flake, J. K., & Xie, S. Y. (2019). Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Social and Personality Psychology Compass*, 13(2), e12431. <https://doi.org/10.1111/spc3.12431>
- Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, 113, 513–529. <https://doi.org/10.1037/pspa0000090>
- Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, 22(3), 165–196. [https://doi.org/10.1016/S1090-5138\(00\)00071-4](https://doi.org/10.1016/S1090-5138(00)00071-4)
- Hewstone, M., & Brown, R. (1986). *Contact is not enough: An intergroup perspective on the 'contact hypothesis.'* In M. Hewstone & R. Brown (Eds.), *Social psychology and society. Contact and conflict in intergroup encounters* (p. 1–44). Basil Blackwell.
- Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, 47, 237–271. <https://doi.org/10.1146/annurev.psych.47.1.237>
- Hinzman, L., & Maddox, K. B. (2017). Conceptual and visual representations of racial categories: Distinguishing subtypes from subgroups. *Journal of Experimental Social Psychology*, 70, 95–109. <https://doi.org/10.1016/j.jesp.2016.12.012>
- 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. <https://doi.org/10.1177/0146167205275613>
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24(1), 1–55. [https://doi.org/10.1016/0010-0285\(92\)90002-J](https://doi.org/10.1016/0010-0285(92)90002-J)
- Holoien, D. S., & Fiske, S. T. (2013). Downplaying positive impressions: Compensation between warmth and competence in impression management. *Journal of Experimental Social Psychology*, 49(1), 33–41. <https://doi.org/10.1016/j.jesp.2012.09.001>
- Hong, Y., & Ratner, K. G. (2020). Minimal but not meaningless: Seemingly arbitrary category labels can imply more than group membership. *Journal of Personality and Social Psychology*, 1–25. <https://doi.org/10.1037/pspa0000255>
- Hornsey, M. J., & Hogg, M. A. (2000). Subgroup relations: A comparison of mutual intergroup differentiation and common ingroup identity models of prejudice reduction. *Personality and Social Psychology Bulletin*, 26(2), 242–256. <https://doi.org/10.1177/0146167200264010>

- Imhoff, R., & Dotsch, R. (2013). Do we look like me or like us? Visual projection as self-or ingroup-projection. *Social Cognition*, 31(6), 806–816. <http://dx.doi.org/10.1521/soco.2013.31.6.806>
- Imhoff, R., Dotsch, R., Bianchi, M., Banse, R., & Wigboldus, D. H. J. (2011). Facing Europe: Visualizing spontaneous in-group projection. *Psychological Science*, 22(12), 1583–1590. <https://doi.org/10.1177/0956797611419675>
- Imhoff, R., & Koch, A. (2017). How orthogonal are the Big Two of social perception? On the curvilinear relation between agency and communion. *Perspectives on Psychological Science*, 12(1), 122–137. <https://doi.org/10.1177/1745691616657334>
- Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in psychology*, 4, 386. <https://doi.org/10.3389/fpsyg.2013.00386>
- Jack, R. E., & Schyns, P. G. (2017). Toward a Social Psychophysics of Face Communication. *Annual Review of Psychology*, 68(1), 269–297. <https://doi.org/10.1146/annurev-psych-010416-044242>
- Jackman, M. R. (1994). *The velvet glove: Paternalism and conflict in gender, class, and race relations*. University of California Press.
- Jackson, J. C., Hester, N., & Gray, K. (2018). The faces of God in America: Revealing religious diversity across people and politics. *PloS one*, 13(6). <https://doi.org/10.1371/journal.pone.0198745>
- Jackson, L. A., Sullivan, L. A., Harnish, R., & Hodge, C. N. (1996). Achieving positive social identity: Social mobility, social creativity, and permeability of group boundaries. *Journal of Personality and Social Psychology*, 70(2), 241–254. <https://doi.org/10.1037/0022-3514.70.2.241>
- Jaeger, B. (2017). r2glmm: Computes R Squared for Mixed (Multilevel) Models (R package version 0.1.2). Retrieved from <https://CRAN.R-project.org/package=r2glmm>
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.
- Jeffries, C. H., Hornsey, M. J., Sutton, R. M., Douglas, K. M., & Bain, P. G. (2012). The David and Goliath principle. Cultural, ideological, and attitudinal underpinnings of the normative protection of low-status groups from criticism. *Personality and Social Psychology Bulletin*, 38(8), 1053–1065. <http://dx.doi.org/10.1177/0146167212444454>
- Jones, B. C., DeBruine, L. M., Main, J. C., Little, A. C., Welling, L. L., Feinberg, D. R., & Tiddman, B. P. (2009). Facial cues of dominance modulate the short-term gaze-cuing effect in human observers. *Proceedings of the Royal Society B: Biological Sciences* 277(1681), 617–624. <http://dx.doi.org/10.1098/rspb.2009.1575>
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 220–266). Academic Press.
- 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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1111/j.1467-9221.2004.00402.x>
- Jost, J. T., Federico, C. M., & Napier, J. L. (2009). Political ideology: Its structure, functions, and elective affinities. *Annual Review of Psychology*, 60, 307–337. <https://doi.org/10.1146/annurev.psych.60.110707.163600>
- Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129(3), 339–375. <https://doi.org/10.1037/0033-2909.129.3.339>
- Jost, J. T., & Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: consequences for specific and diffuse forms of system justification. *Journal of Personality and Social Psychology*, 88(3), 498–509. <http://dx.doi.org/10.1037/0022-3514.88.3.498>
- Jost, J. T., Kivetz, Y., Rubini, M., Guermendi, G., & Mosso, C. (2005). System-justifying functions of complementary regional and ethnic stereotypes: Cross-national evidence. *Social justice research*, 18(3), 305–333. <http://dx.doi.org/10.1007/s11211-005-6827-z>
- Jost, J. T., Pelham, B. W., & Carvallo, M. R. (2002). Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups. *Journal of Experimental Social Psychology*, 38(6), 586–602. [https://doi.org/10.1016/S0022-1031\(02\)00505-X](https://doi.org/10.1016/S0022-1031(02)00505-X)
- Jost, J. T., & Thompson, E. P. (2000). Group-based dominance and opposition to equality as independent predictors of self-esteem, ethnocentrism, and social policy attitudes among African Americans and European Americans. *Journal of Experimental Social Psychology*, 36(3), 209–232. <https://doi.org/10.1006/jesp.1999.1403>
- Judd, C. M., Garcia-Marques, T., & Yzerbyt, V. Y. (2019). The complexity of relations between dimensions of social perception: Decomposing bivariate associations with crossed random factors. *Journal of Experimental Social Psychology*, 82, 200–207. <https://doi.org/10.1016/j.jesp.2019.01.008>
- Judd, C. M., James-Hawkins, L., Yzerbyt, V. Y., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, 89(6), 899–913. <https://doi.org/10.1037/0022-3514.89.6.899>
- Judd, C. M., & Park, B. (1988). Out-group homogeneity: Judgments of variability at the individual and group levels. *Journal of Personality and Social Psychology*, 54(5), 778–788. <https://doi.org/10.1037/0022-3514.54.5.778>
- 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. <https://doi.org/10.1037/a0028347>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with More Than One Random Factor: Designs, Analytic Models, and Statistical Power. *Annual Review of Psychology*, 68(1), 601–625. <https://doi.org/10.1146/annurev-psych-122414-033702>
- Karremans, J. C., Dotsch, R., & Corneille, O. (2011). Romantic relationship status biases memory of faces of attractive opposite-sex others: Evidence from a reverse-correlation paradigm. *Cognition*, 121(3), 422–426. <https://doi.org/10.1016/j.cognition.2011.07.008>
- Kay, A. C., & Jost, J. T. (2003). Complementary justice: effects of "poor but happy" and "poor but honest" stereotype exemplars on system justification and implicit activation of the justice motive. *Journal of Personality and Social Psychology*, 85(5), 823–837. <http://dx.doi.org/10.1037/0022-3514.85.5.823>
- Kay, A. C., Jost, J. T., Mandisodza, A. N., Sherman, S. J., Petrocelli, J. V., & Johnson, A. L. (2007). Panglossian ideology in the service of system justification: How complementary stereotypes help us to rationalize inequality. *Advances in experimental social psychology*, 39, 305–358. [http://dx.doi.org/10.1016/S0065-2601\(06\)39006-5](http://dx.doi.org/10.1016/S0065-2601(06)39006-5)
- Kelley, H. H. (1950). The warm-cold variable in first impressions of persons. *Journal of Personality*, 18, 431–439. <https://doi.org/10.1111/j.1467-6494.1950.tb01260.x>
- Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The Other-Race Effect Develops During Infancy. *Psychological Science*, 18(12), 1084–1089. <https://doi.org/10.1111/j.1467-9280.2007.02029.x>
- Kervyn, N., Bergsieker, H. B., Grignard, F., & Yzerbyt, V. Y. (2016). An advantage of appearing mean or lazy: Amplified impressions of competence or warmth after mixed descriptions. *Journal of Experimental Social Psychology*, 62, 17–23. <https://doi.org/10.1016/j.jesp.2015.09.004>
- Kervyn, N., Fiske, S. T., & Yzerbyt, V. Y. (2013). Integrating the stereotype content model (warmth and competence) and the Osgood semantic differential (evaluation, potency, and activity). *European Journal of Social Psychology*, 43(7), 673–681. <https://doi.org/10.1002/ejsp.1978>
- Kervyn, N., Fiske, S. T., & Yzerbyt, V. Y. (2015). Forecasting the primary dimension of social perception: Symbolic and realistic threats together predict warmth in the Stereotype Content Model. *Social Psychology*, 46(1), 36–45. <https://doi.org/10.1027/1864-9335/a000219>
- Kervyn, N., Judd, C. M., & Yzerbyt, V. Y. (2009). You want to appear competent? Be mean! You want to appear sociable? Be lazy! Group differentiation and the compensation effect. *Journal of Experimental Social Psychology*, 45(2), 363–367. <https://doi.org/10.1016/j.jesp.2008.08.006>
- Kervyn, N., Yzerbyt, V. Y., Demoulin, S., & Judd, C. M. (2008). Competence and warmth in context: The compensatory nature of stereotypic views of national groups. *European Journal of Social Psychology*, 38(7), 1175–1183. <https://doi.org/10.1002/ejsp.526>
- Kervyn, N., Yzerbyt, V. Y., & Judd, C. M. (2010). Compensation between warmth and competence: Antecedents and consequences of a negative relation between the two

- fundamental dimensions of social perception. *European Review of Social Psychology*, 21(1), 155–187. <https://doi.org/10.1080/13546805.2010.517997>
- Kervyn, N., Yzerbyt, V. Y., & Judd, C. M. (2011). When compensation guides inferences: Indirect and implicit measures of the compensation effect. *European Journal of Social Psychology*, 41(2), 144–150. <https://doi.org/10.1002/ejsp.748>
- Kervyn, N., Yzerbyt, V. Y., Judd, C. M., & Nunes, A. (2009). A question of compensation: the social life of the fundamental dimensions of social perception. *Journal of Personality and Social Psychology*, 96(4), 828–842. <https://doi.org/10.1037/a0013320>
- Kim, D.-Y. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly*, 66(1), 83–96. <https://doi.org/10.2307/3090143>
- Kite, M. E., & Whitley, B. E. (2016). *Psychology of prejudice and discrimination*. Psychology Press.
- Klein, O., Bouchat, P., Azzi, A., & Luminet, O. (2017). Principled Disagreements: Adhesion to Intergroup Justice Standards in the Context of the Belgian Linguistic Conflict. *Psychologica Belgica*, 57(3), 13. <https://doi.org/10.5334/pb.345>
- Klein, O., Licata, L., Van der Linden, N., Mercy, A., & Luminet, O. (2011). A waffle-shaped model for how realistic dimensions of the Belgian conflict structure collective memories and stereotypes. *Memory Studies*, 5(1), 16–31. <https://doi.org/10.1177/1750698011424028>
- Kleisner, K., Priplatova, L., Frost, P., & Flegel, J. (2013). Trustworthy-looking face meets brown eyes. *PloS one*, 8(1), e53285. <https://doi.org/10.1371/journal.pone.0053285>
- Knutson, B. J. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, 20(3), 165–182. <https://doi.org/10.1007/BF02281954>
- Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative–progressive beliefs, and communion. *Journal of Personality and Social Psychology*, 110(5), 675–709. <https://doi.org/10.1037/pspa0000046>
- Koch, A., Imhoff, R., Unkelbach, C., Nicolas, G., Fiske, S. T., Terache, J., Carrier, A., & Yzerbyt, V. Y. (2020). Groups' warmth is a personal matter: Understanding consensus on stereotype dimensions reconciles adversarial models of social evaluation. *Journal of Experimental Social Psychology*, 89, 103995. <https://doi.org/10.1016/j.jesp.2020.103995>
- Kontsevich, L. L., & Tyler, C. W. (2004). What makes Mona Lisa smile? *Vision Research*, 44(13), 1493–1498. <https://doi.org/10.1016/j.visres.2003.11.027>
- Krosch, A. R., & Amodio, D. M. (2014). Economic scarcity alters the perception of race. *Proceedings of the National Academy of Sciences*, 111(25), 9079–9084. <https://doi.org/10.1073/pnas.1404448111>
- Kunst, J. R., Dovidio, J. F., & Dotsch, R. (2017). White Look-Alikes: Mainstream Culture Adoption Makes Immigrants “Look” Phenotypically White. *Personality and Social Psychology Bulletin*, 44(2), 265–282. <https://doi.org/10.1177/0146167217739279>
- Kunst, J. R., Kteily, N., & Thomsen, L. (2017). “You Little Creep”: Evidence of Blatant Dehumanization of Short Groups. *Social Psychological and Personality Science*, 10(2), 160–171. <https://doi.org/10.1177/1948550617740613>

- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, 24(8), 1377–1388. <https://doi.org/10.1080/02699930903485076>
- Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, 93(2), 234–249. <https://doi.org/10.1037/0022-3514.93.2.234>
- Leary, M. R. (1995). *Social psychology series. Self-presentation: Impression management and interpersonal behavior*. Brown & Benchmark Publishers.
- Lee, Y.-T., & Ottati, V. (1995). Perceived In-Group Homogeneity as a Function of Group Membership Salience and Stereotype Threat. *Personality and Social Psychology Bulletin*, 21(6), 610–619. <https://doi.org/10.1177/0146167295216007>
- Lemaine, G. (1974). Social differentiation and social originality. *European Journal of Social Psychology*, 4(1), 17–52. <https://doi.org/10.1002/ejsp.2420040103>
- Lerner, M. (1980). *The belief in a just world: A fundamental delusion*. Plenum Press.
- Leyens, J.-P., & Yzerbyt, V. Y. (1992). The ingroup overexclusion effect: Impact of valence and confirmation on stereotypical information search. *European Journal of Social Psychology*, 22(6), 549–569. <https://doi.org/10.1002/ejsp.2420220604>
- Lick, D. J., Carpinella, C. M., Preciado, M. A., Spunt, R. P., & Johnson, K. L. (2013). Reverse-correlating mental representations of sex-typed bodies: the effect of number of trials on image quality. *Frontiers in Psychology*, 4, 1–9. <https://doi.org/10.3389/fpsyg.2013.00476>
- Lin, C., Adolphs, R., & Alvarez, R. M. (2018). Inferring Whether Officials Are Corruptible From Looking at Their Faces. *Psychological Science*, 29(11), 1807–1823. <https://doi.org/10.1177/0956797618788882>
- Lin, C., Keles, U., & Adolphs, R. (2019, October 2). Comprehensive trait attributions show that face impressions are organized in four dimensions. <https://doi.org/10.31234/osf.io/87nex>
- Lindholm, T., & Yzerbyt, V. Y. (2018). When being nice or being smart could bring you down: Compensatory dynamics in strategic self-presentation. *International Review of Social Psychology*, 31(1), Article 16. <https://doi.org/10.5334/irsp.136>
- Lippman, W. (1922). *Public opinion*. Harcourt.
- Little, A. C., Burt, D. M., & Perrett, D. I. (2006). What is good is beautiful: Face preference reflects desired personality. *Personality and Individual Differences*, 41, 1107–1118. <https://doi.org/10.1016/J.PAID.2006.04.015>
- Lloyd, E. P., Sim, M., Smalley, E., Bernstein, M. J., & Hugenberg, K. (2020). Good Cop, Bad Cop: Race-Based Differences in Mental Representations of Police. *Personality and Social Psychology Bulletin*, 46(8), 1205–1218. <https://doi.org/10.1177/0146167219898562>
- Locke, K. D. (2003). Status and solidarity in social comparison: Agentic and communal values and vertical and horizontal directions. *Journal of Personality and Social Psychology*, 84(3), 619–631. <https://doi.org/10.1037/0022-3514.84.3.619>

- Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: estimating the information employed for face classifications. *Cognitive Science*, 28(2), 209–226. https://doi.org/10.1207/s15516709cog2802_4
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224–253. <https://doi.org/10.1037/0033-295X.98.2.224>
- Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, 9(6), 296–305. <https://doi.org/10.1016/j.tics.2005.04.010>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- McGarty, C., Yzerbyt, V. Y., & Spears, R. (Eds.). (2002). *Stereotypes as explanations: The formation of meaningful beliefs about social groups*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511489877>
- Meeusen, C., Boonen, J. & Dassonneville, R. (2017). The Structure of Prejudice and Its Relation to Party Preferences in Belgium: Flanders and Wallonia Compared. *Psychologica Belgica*, 57(3), 52–74. <http://doi.org/10.5334/pb.335>
- Meuleman, B., Abts, K. and Meeusen, C. (2017). Walloons as General or Specific Others? A Comparison of anti-Walloon and anti-immigrant Attitudes in Flanders. *Psychologica Belgica*, 57(3), 75–97. <http://doi.org/10.5334/pb.336>
- Monteith, M. J., Deneen, N. E., & Tooman, G. D. (1996). The effect of social norm activation on the expression of opinions concerning gay men and Blacks. *Basic and Applied Social Psychology*, 18(3), 267–288. https://doi.org/10.1207/s15324834basp1803_2
- Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal behavior*, 27(4), 237–254. <https://doi.org/10.1023/A:1027332800296>
- Moon, K., Kim, S., Kim, J., Kim, H., & Ko, Y. (2020). The Mirror of Mind: Visualizing Mental Representations of Self Through Reverse Correlation. *Frontiers in Psychology*, 11, 1149. <https://doi.org/10.3389/fpsyg.2020.01149>
- Moore, F. R., Filippou, D., & Perrett, D. I. (2011). Intelligence and attractiveness in the face: Beyond the attractiveness halo effect. *Journal of Evolutionary Psychology*, 9(3), 205–217. <https://doi.org/10.1556/JEP.9.2011.3.2>
- Mueller, U., & Mazur, A. (1996). Facial dominance of West Point cadets as a predictor of later military rank. *Social forces*, 74(3), 823–850. <https://doi.org/10.1093/sf/74.3.823>
- Mummendey, A., & Schreiber, H.-J. (1983). Better or just different? Positive social identity by discrimination against, or by differentiation from outgroups. *European Journal of Social Psychology*, 13(4), 389–397. <https://doi.org/10.1002/ejsp.2420130406>
- Mussweiler, T. (2007). Assimilation and Contrast as Comparison Effects: A Selective Accessibility Model. In D. A. Stapel & J. Suls (Eds.), *Assimilation and contrast in social psychology* (p. 165–185). Psychology Press.

- Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, 134(4), 565–584. <http://dx.doi.org/10.1037/0096-3445.134.4.565>
- Nosek, B. A., Bar-Anan, Y., Sriram, N., Axt, J., & Greenwald, A. G. (2014). Understanding and using the brief implicit association test: Recommended scoring procedures. *PLoS One*, 9(12), e110938. <http://dx.doi.org/10.1371/journal.pone.0110938>
- Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2020). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General*, 149(2), 323–342. <https://doi.org/10.1037/xge0000638>
- Oldmeadow, J. A., & Fiske, S. T. (2007). System-justifying ideologies moderate status = competence stereotypes: Roles for belief in a just world and social dominance orientation. *European Journal of Social Psychology*, 37(6), 1135–1148. <http://dx.doi.org/10.1002/ejsp.428>
- Oldmeadow, J. A., & Fiske, S. T. (2010). Social status and the pursuit of positive social identity: Systematic domains of intergroup differentiation and discrimination for high- and low-status groups. *Group Processes & Intergroup Relations*, 13(4), 425–444. <https://doi.org/10.1177/1368430209355650>
- Oldmeadow, J. A., Sutherland, C. A. M., & Young, A. W. (2012). Facial stereotype visualization through image averaging. *Social Psychology and Personality Science*, 00(0), 1–10. <https://doi.org/10.1177/1948550612469820>
- Oliveira, M., Garcia-Marques, T., & Dotsch, R. (2019). Combining Traits Into a Face: A Reverse Correlation Approach. *Social Cognition*, 37(5), 516–545. <https://doi.org/10.1521/soco.2019.37.5.516>
- Oliveira, M., Garcia-Marques, T., Dotsch, R., & Garcia-Marques, L. (2019). Dominance and competence face to face: Dissociations obtained with a reverse correlation approach. *European Journal of Social Psychology*, 49(5), 888–902. <https://doi.org/10.1002/ejsp.2569>
- Oliveira, M., Garcia-Marques, T., Garcia-Marques, L., & Dotsch, R. (2019). Good to Bad or Bad to Bad? What is the relationship between valence and the trait content of the Big Two?. *European Journal of Social Psychology*, 00, 1–21. <https://doi.org/10.1002/ejsp.2618>
- Olivola, C. Y., Tingley, D., & Todorov, A. (2018). Republican voters prefer candidates who have conservative-looking faces: new evidence from exit polls. *Political Psychology*, 39(5), 1157–1171. <https://doi.org/10.1111/pops.12489>
- Olivola, C. Y., & Todorov, A. J. (2010). Elected in 100 milliseconds: Appearance-Based Trait Inferences and Voting. *Journal of Nonverbal Behavior*, 34(2), 83–110. <https://doi.org/10.1007/s10919-009-0082-1>
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087–11092. <https://doi.org/10.1073/pnas.0805664105>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>

- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Univer. Illinois Press.
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, 105(2), 171–192. <https://doi.org/10.1037/a0032734>
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology*, 108(4), 553–561. <http://dx.doi.org/10.1037/pspa0000016>
- Otten, S., & Moskowitz, G. B. (2000). Evidence for implicit evaluative in-group bias: Affect-biased spontaneous trait inference in a minimal group paradigm. *Journal of Experimental Social Psychology*, 36(1), 77–89. <https://doi.org/10.1006/jesp.1999.1399>
- Owuamalam, C. K., Wong, K. X., & Rubin, M. (2016). Chubby but cheerful? Investigating the compensatory judgments of high, medium, and low status weight groups in Malaysia. *Cogent Psychology*, 3(1), 1188441. <http://dx.doi.org/10.1080/23311908.2016.1188441>
- Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, 16(2), 332–350. <https://doi.org/10.1037/0096-1523.16.2.332>
- Paulhus, D. L., & Trapnell, P. D. (2008). Self-presentation of personality: An agency-communion framework. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (p. 492–517). The Guilford Press.
- Peabody, D. (1967). Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology*, 7(4), 1–18. <https://doi.org/10.1037/h0025230>
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163. <https://doi.org/10.1016/j.jesp.2017.01.006>
- Peer, E., Vosgerau, J., & Acquisti, A. (2013). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, 46(4), 1023–1031. <https://doi.org/10.3758/s13428-013-0434-y>
- Peeters, G. (1992). Evaluative meanings of adjectives in vitro and in context: Some theoretical implications and practical consequences of positive-negative asymmetry and behavioral-adaptive concepts of evaluation. *Psychologica Belgica*, 32(2), 211–231.
- Peeters, G. (2008). The evaluative face of a descriptive model: Communion and agency in Peabody's tetradic model of trait organization. *European Journal of Social Psychology*, 38(7), 1066–1072. <https://doi.org/10.1002/ejsp.524>
- Petsko, C. D., Lei, R. F., Kunst, J. R., Bruneau, E., & Kteily, N. (in press). Blatant dehumanization in the mind's eye: Prevalent even among those who explicitly reject it? *Journal of Experimental Psychology: General*. Retrieved from <https://psyarxiv.com/g7w4b/>
- Phalet, K., & Poppe, E. (1997). Competence and morality dimensions of national and ethnic stereotypes: A study in six eastern-European countries. *European Journal of Social*

- Psychology*, 27(6), 703–723. [https://doi.org/10.1002/\(SICI\)1099-0992\(199711/12\)27:6<703::AID-EJSP841>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1099-0992(199711/12)27:6<703::AID-EJSP841>3.0.CO;2-K)
- Pinter, B., & Greenwald, A. G. (2004). Exploring Implicit Partisanship: Enigmatic (But Genuine) Group Identification and Attraction. *Group Processes & Intergroup Relations*, 7(3), 283–296. <https://doi.org/10.1177/1368430204046112>
- Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law*, 16(6), 477–491. <https://doi.org/10.1080/10683160902926141>
- Ramos, T., Oliveira, M., Santos, A. S., Garcia-Marques, L., & Carneiro, P. (2016). Evaluating young and old faces on social dimensions: Trustworthiness and dominance. *Psicológica*, 37(2), 169–185.
- Ratner, K. G., Dotsch, R., Wigboldus, D. H. J., van Knippenberg, A., & Amodio, D. M. (2014). Visualizing minimal ingroup and outgroup faces: Implications for impressions, attitudes, and behavior. *Journal of Personality and Social Psychology*, 106(6), 897–911. <https://doi.org/10.1037/a0036498>
- Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PloS one*, 7(3), e34293. <https://doi.org/10.1371/journal.pone.0034293>
- Richards, L., Rollerson, B., & Phillips, J. (1991). Perceptions of submissiveness: Implications for victimization. *The Journal of Psychology: Interdisciplinary and Applied*, 125(4), 407–411. <https://doi.org/10.1080/00223980.1991.10543302>
- Richeson, J. A., & Ambady, N. (2001). Who's in charge? Effects of situational roles on automatic gender bias. *Sex Roles: A Journal of Research*, 44(9-10), 493–512. <https://doi.org/10.1023/A:1012242123824>
- Richetin, J., Durante, F., Mari, S., Perugini, M., & Volpato, C. (2012). Primacy of warmth versus competence: A motivated bias? *The Journal of Social Psychology*, 152(4), 417–435. <https://doi.org/10.1080/00224545.2011.623735>
- Rimé, B., Bouchat, P., Klein, O., & Licata, L. (2015). When collective memories of victimhood fade: Generational evolution of intergroup attitudes and political aspirations in Belgium. *European Journal of Social Psychology*, 45(4), 515–532. <https://doi.org/10.1002/ejsp.2104>
- Rohmer, O., & Louvet, E. (2012). Implicit measures of the stereotype content associated with disability. *British Journal of Social Psychology*, 51(4), 732–740. <http://dx.doi.org/10.1111/j.2044-8309.2011.02087.x>
- Rohmer, O., & Louvet, E. (2018). Implicit stereotyping against people with disability. *Group Processes & Intergroup Relations*, 21(1), 127–140. <http://dx.doi.org/10.1177/1368430216638536>
- Röhner, J., Schröder-Abé, M., & Schütz, A. (2013). What do fakers actually do to fake the IAT? An investigation of faking strategies under different faking conditions. *Journal of Research in Personality*, 47(4), 330–338. <https://doi.org/10.1016/j.jrp.2013.02.009>

- Rosenberg, S., & Jones, R. (1972). A method for investigating and representing a person's implicit theory of personality: Theodore Dreiser's view of people. *Journal of Personality and Social Psychology*, 22(3), 372–386. <https://doi.org/10.1037/h0032891>
- Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, 9(4), 283–294. <https://doi.org/10.1037/h0026086>
- Rosenberg, S., & Olshan, K. (1970). Evaluative and descriptive aspects in personality perception. *Journal of Personality and Social Psychology*, 16(4), 619–626. <https://doi.org/10.1037/h0030081>
- Rubin, Z., & Peplau, L. A. (1975). Who believes in a just world? *Journal of Social Issues*, 31(3), 65–89. <https://doi.org/10.1111/j.1540-4560.1975.tb00997.x>
- Rudman, L. A., & Fairchild, K. (2004). Reactions to Counterstereotypic Behavior: The Role of Backlash in Cultural Stereotype Maintenance. *Journal of Personality and Social Psychology*, 87(2), 157–176. <https://doi.org/10.1037/0022-3514.87.2.157>
- Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, 77(5), 1004–1010. <https://doi.org/10.1037/0022-3514.77.5.1004>
- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, 57(4), 743–762. <https://doi.org/10.1111/0022-4537.00239>
- Rule, N. O., & Ambady, N. (2009). She's got the look: Inferences from female chief executive officers' faces predict their success. *Sex Roles: A Journal of Research*, 61(9–10), 644–652. <https://doi.org/10.1007/s11199-009-9658-9>
- Russell, A. M. T., & Fiske, S. T. (2008). It's all relative: Competition and status drive interpersonal perception. *European Journal of Social Psychology*, 38(7), 1193–1201. <https://doi.org/10.1002/ejsp.539>
- Sachdev, I., & Bourhis, R. Y. (1985). Social categorization and power differentials in group relations. *European Journal of Social Psychology*, 15(4), 415–434. <https://doi.org/10.1002/ejsp.2420150405>
- Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, 9(2), 260–264. <https://doi.org/10.1037/a0014681>
- Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3), 501–518. <https://doi.org/10.1037/0278-7393.13.3.501>
- Schlenker, B. R., & Leary, M. R. (1982). Social anxiety and self-presentation: A conceptualization model. *Psychological Bulletin*, 92(3), 641–669. <https://doi.org/10.1037/0033-2909.92.3.641>
- Schmitz, M., Rougier, M., & Yzerbyt, V. Y. (2020). Comment on “Quantifying the informational value of classification images”: A miscomputation of the infoVal metric.

- Behavior Research Methods*, 52(3), 1383–1386. <https://doi.org/10.3758/s13428-019-01295-1>
- Schmitz, M., Rougier, M., & Yzerbyt, V. Y. (under review). Introducing the Brief Reverse Correlation: An improved tool to assess visual representations. Manuscript under review in *Journal of Experimental Social Psychology*.
- Schmitz, M., Rougier, M., Yzerbyt, V. Y., Brinkman, L., & Dotsch, R. (2020). Erratum to: Comment on “Quantifying the informational value of classification images”: Miscomputation of infoVal metric was a minor issue and is now corrected. *Behavior Research Methods*, 52(4), 1800–1801. <https://doi.org/10.3758/s13428-020-01367-7>
- Schmitz, M., Terache, J., & Yzerbyt, V. Y. (in preparation). Facial compensation in existing groups. *Manuscript in preparation*.
- Schmitz, M., Vanbeneden, A., & Yzerbyt, V. Y. (under review). When you look competent but cold: Taking compensation at the face value. Manuscript under review in *European Journal of Social Psychology*.
- Schmitz, M., & Yzerbyt, V. Y. (in press). Direct and indirect dimensional compensation: Is there a difference between observers and group members? *Group Processes & Intergroup Relations*. <https://doi.org/10.1177/1368430220963176>
- Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin*, 79(5), 294–309. <https://doi.org/10.1037/h0034496>
- Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, 54(4), 558–568. <https://doi.org/10.1037/0022-3514.54.4.558>
- Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139175043>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Son Hing, L. S., Bobocel, D. R., Zanna, M. P., Garcia, D. M., Gee, S. S., & Oraziatti, K. (2011). The merit of meritocracy. *Journal of Personality and Social Psychology*, 101(3), 433–450. <http://dx.doi.org/10.1037/a0024618>
- Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology*, 56(4), 283–294. <http://dx.doi.org/10.1027/1618-3169.56.4.283>
- Stangor, C. (2009). 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* (p. 1–22). Psychology Press.
- Stangor, C., & Leary, S. P. (2006). Intergroup beliefs: Investigations from the social side. In M. P. Zanna (Ed.), *Advances in experimental social psychology: Vol. 38. Advances in experimental social psychology, Vol. 38* (p. 243–281). Elsevier Academic Press. [https://doi.org/10.1016/S0065-2601\(06\)38005-7](https://doi.org/10.1016/S0065-2601(06)38005-7)
- Stephan, W. G., Ybarra, O., & Morrison, K. R. (2009). Intergroup threat theory. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (p. 43–59). Psychology Press.

- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24–31. <http://dx.doi.org/10.1177/0098628316677643>
- Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour*, 4, 361–371. <https://doi.org/10.1038/s41562-019-0800-6>
- Stolier, R. M., Hehman, E., Keller, M. D., Walker, M., & Freeman, J. B. (2018). The conceptual structure of face impressions. *Proceedings of the National Academy of Sciences*, 115(37), 9210–9215. <http://dx.doi.org/10.1073/pnas.1807222115>
- Sutherland, C. A. M., Liu, X., Chu, Y., Zhang, L., Oldmeadow, J. A., & Young, A. W. (2015). Chinese perceivers' facial first impressions. *Journal of Vision*, 15(12), 1218. <http://dx.doi.org/10.1167/15.12.1218>
- Sutherland, C. A. M., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Personality and Social Psychology Bulletin*, 44(4), 521–537. <http://dx.doi.org/10.1177/0146167217744194>
- Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, 127(1), 105–118. <http://dx.doi.org/10.1016/j.cognition.2012.12.001>
- Sutherland, C. A. M., Oldmeadow, J. A., & Young, A. W. (2016). Integrating social and facial models of person perception: Converging and diverging dimensions. *Cognition*, 157, 257–267. <https://doi.org/10.1016/j.cognition.2016.09.006>
- Sutherland, C. A. M., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology*, 106(2), 186–208. <https://doi.org/10.1111/bjop.12085>
- Swencionis, J. K., Dupree, C. H., & Fiske, S. T. (2017). Warmth-competence tradeoffs in impression management across race and social-class divides. *Journal of Social Issues*, 73(1), 175–191. <https://doi.org/10.1111/josi.12210>
- Swencionis, J. K., & Fiske, S. T. (2016). Promote up, ingratiate down: Status comparisons drive warmth-competence tradeoffs in impression management. *Journal of Experimental Social Psychology*, 64, 27–34. <https://doi.org/10.1016/j.jesp.2016.01.004>
- Tajfel, H. (1981). *Human groups and social categories*. Cambridge: Cambridge University Press.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup relations. In W. G. Austin & S. Worchel (Eds.), *The psychology of intergroup relations*. Brooks-Cole.
- Terache, J., Demoulin, S., & Yzerbyt, V. Y. (2020). Warmth and Competence in Interpersonal Comparisons: The Quiz Master Paradigm through the Lens of Compensation. *International Review of Social Psychology*, 33(1), 4. <http://doi.org/10.5334/irsp.275>
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1), 25–29. <https://doi.org/10.1037/h0071663>

- Tiddeman, B. P., Stirrat, M. R., & Perrett, D. I. (2005). Towards Realism in Facial Image Transformation: Results of a Wavelet MRF Method. *Computer Graphics Forum*, 24(3), 449–456. <https://doi.org/10.1111/j.1467-8659.2005.00870.x>
- Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, 1(1124), 208–224. <http://doi.org/10.1196/annals.1440.012>
- Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: a model based approach. *Social cognitive and affective neuroscience*, 3(2), 119–127. <https://doi.org/10.1093/scan/nsn009>
- Todorov, A., Dotsch, R., Wigboldus, D. H. J., & Said, C. P. (2011). Data-driven Methods for Modeling Social Perception. *Social and Personality Psychology Compass*, 5(10), 775–791. <https://doi.org/10.1111/j.1751-9004.2011.00389.x>
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308(5728), 1623–1626. <https://doi.org/10.1126/science.1110589>
- Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current Opinion in Neurobiology*, 23(3), 373–380. <https://doi.org/10.1016/j.conb.2012.12.010>
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*, 66(1), 519–545. <https://doi.org/10.1146/annurev-psych-113011-143831>
- Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, 27(6), 813–833. <https://doi.org/10.1521/soco.2009.27.6.813>
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in cognitive sciences*, 12(12), 455–460. <https://doi.org/10.1016/j.tics.2008.10.001>
- Toscano, H., Schubert, T. W., Dotsch, R., Falvello, V., & Todorov, A. (2016). *Physical strength as a cue to dominance: A data-driven approach. Personality and Social Psychology Bulletin*, 42(12), 1603–1616. <https://doi.org/10.1177/0146167216666266>
- Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, 93(3), 239–257. <https://doi.org/10.1037/0033-295X.93.3.239>
- Turner, J. C., & Brown, R. (1978). Social status, cognitive alternatives, and intergroup relations. In H. Tajfel (Ed.), *Differentiation between social groups: Studies in the social psychology of intergroup relations* (pp. 202–234). Academic Press.
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.
- Uleman, J., & Kressel, L. M. (2013). A brief history of theory and research on impression formation. In D. E. Carlston (Ed.), *Oxford handbook of social cognition* (pp. 53–73). Oxford University Press.

- Vanbeselaere, N., Boen, F., van Avermaet, E., & Buelens, H. (2006). The Janus face of power in intergroup contexts: A further exploration of the noblesse oblige effect. *The Journal of Social Psychology*, 146(6), 685–699. <http://dx.doi.org/10.3200/SOCP.146.6.685-699>
- Van der Linden, N., & Roets, A. (2017). Insights into the Belgian Linguistic Conflict from a (Social) Psychological Perspective: Introduction to the Special Issue. *Psychologica Belgica*, 57(3), 1–12. <https://doi.org/10.5334/pb.418>
- Van Dessel, P., Cummins, J., Hughes, S., Kasran, S., Cathelyn, F., & Moran, T. (in press). Reflecting on Twenty-Five Years of Research Using Implicit Measures: Recommendations for their Future Use. *Social Cognition*.
- Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *PNAS*, 111(32), E3353–E3361. <http://dx.doi.org/10.1073/pnas.1409860111>
- Wade, M. L., & Brewer, M. B. (2006). The structure of female subgroups: An exploration of ambivalent stereotypes. *Sex Roles*, 54(11-12), 753–765. <http://dx.doi.org/10.1007/s11199-006-9043-x>
- Walker, M., & Keller, M. (2019). Beyond attractiveness: A multimethod approach to study enhancement in self-recognition on the Big Two personality dimensions. *Journal of Personality and Social Psychology*, 117(3), 483–499. <https://doi.org/10.1037/pspa0000157>
- Walker, M., & Vetter, T. (2016). Changing the personality of a face: Perceived Big Two and Big Five personality factors modeled in real photographs. *Journal of Personality and Social Psychology*, 110(4), 609–624. <https://doi.org/10.1037/pspp0000064>
- Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143(5), 2020–2045. <https://doi.org/10.1037/xge0000014>
- Whitley, B. E., Jr. (1999). Right-wing authoritarianism, social dominance orientation, and prejudice. *Journal of Personality and Social Psychology*, 77(1), 126–134. <https://doi.org/10.1037/0022-3514.77.1.126>
- Wigboldus, D. H. J., Semin, G. R., & Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. *Journal of Personality and Social Psychology*, 78(1), 5–18. <http://dx.doi.org/10.1037/0022-3514.78.1.5>
- Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, 37(3), 395–412. <https://doi.org/10.1037/0022-3514.37.3.395>
- Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl, Vol. 1. Matters of public interest; Vol. 2. Personality and psychopathology* (p. 89–113). University of Minnesota Press.
- Willis, J., & Todorov, A. (2006). First Impressions: Making Up Your Mind After a 100-Ms Exposure to a Face. *Psychological Science*, 17(7), 592–598. <https://doi.org/10.1111/j.1467-9280.2006.01750.x>

- 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. <https://doi.org/10.1037/0022-3514.72.2.262>
- Wojciszke, B. (1994). Multiple meanings of behavior: Construing actions in terms of competence or morality. *Journal of Personality and Social Psychology*, 67(2), 222–232. <https://doi.org/10.1037/0022-3514.67.2.222>
- Wojciszke, B., & Abele, A. E. (2008). The primacy of communion over agency and its reversals in evaluations. *European Journal of Social Psychology*, 38(7), 1139–1147. <https://doi.org/10.1002/ejsp.549>
- Wojciszke, B., Baryla, W., Parzuchowski, M., Szymkow, A., & Abele, A. E. (2011). Self-esteem is dominated by agentic over communal information. *European Journal of Social Psychology*, 41(5), 617–627. <https://doi.org/10.1002/ejsp.791>
- Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, 24(12), 1251–1263. <https://doi.org/10.1177/01461672982412001>
- Wolsko, C., Park, B., Judd, C. M., & Wittenbrink, B. (2000). Framing interethnic ideology: Effects of multicultural and color-blind perspectives on judgments of groups and individuals. *Journal of Personality and Social Psychology*, 78(4), 635–654. <https://doi.org/10.1037/0022-3514.78.4.635>
- Yada, N., & Ikegami, T. (2014). Compensation versus halo effects in competitive or cooperative social settings: Mediation effects of social comparison-based emotions. *UrbanScope: e-journal of the Urban-Culture Research Center*, 5, 27–40.
- Yada, N., & Ikegami, T. (2017). Social Inequality and System-Justifying Function of Compensatory Judgments in Person Perception. *Psychology*, 8(12), 2031–2046. <https://doi.org/10.4236/psych.2017.812130>
- Yang, J., Shi, Y., Luo, Y. L. L., Shi, J., & Cai, H. (2014). The Brief Implicit Association Test is valid: Experimental evidence. *Social Cognition*, 32(5), 449–465. <http://dx.doi.org/10.1521/soco.2014.32.5.449>
- Ybarra, O., Chan, E., & Park, D. (2001). Young and old adults' concerns about morality. *Motivation and Emotion*, 25, 85–100. <http://dx.doi.org/10.1023/A:1010633908298>
- Young, A. I., Ratner, K. G., & Fazio, R. H. (2013). Political Attitudes Bias the Mental Representation of a Presidential Candidate's Face. *Psychological Science*, 25(2), 503–510. <https://doi.org/10.1177/0956797613510717>
- Yovel, G., & Duchaine, B. (2006). Specialized Face Perception Mechanisms Extract Both Part and Spacing Information: Evidence from Developmental Prosopagnosia. *Journal of Cognitive Neuroscience*, 18(4), 580–593. <https://doi.org/10.1162/jocn.2006.18.4.580>
- Yzerbyt, V. Y. (2016). Intergroup stereotyping. *Current Opinion in Psychology*, 11, 90–95. <http://dx.doi.org/10.1016/j.copsy.2016.06.009>
- Yzerbyt, V. Y. (2018). The dimensional compensation model: Reality and strategic constraints on warmth and competence in intergroup perceptions. In *Agency and communion in social psychology* (pp. 126–141). Routledge. <https://doi.org/10.4324/9780203703663>

- Yzerbyt, V. Y., & Cambon, L. (2017). The dynamics of compensation: When ingroup favoritism paves the way for outgroup praise. *Personality and Social Psychology Bulletin*, 43(5), 587–600. <https://doi.org/10.1177/0146167216689066>
- Yzerbyt, V. Y., & Demoulin, S. (2010). Intergroup relations. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (p. 1024–1083). John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470561119.socpsy002028>
- Yzerbyt, V. Y., & Judd, C. M. (2020). Comparing effect sizes between within-participant and between-participants designs. *Manuscript in preparation*.
- Yzerbyt, V. Y., Kervyn, N., & Judd, C. M. (2008). Compensation versus halo: The unique relations between the fundamental dimensions of social judgment. *Personality and Social Psychology Bulletin*, 34(8), 1110–1123. <https://doi.org/10.1177/0146167208318602>
- Yzerbyt, V. Y., Provost, V., & Corneille, O. (2005). Not Competent but Warm... Really? Compensatory stereotypes in the French-speaking world. *Group Processes & Intergroup Relations*, 8(3), 291–308. <https://doi.org/10.1177/1368430205053944>
- Zebrowitz, L. A., Fellous, J. M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review*, 7(3), 194–215. https://doi.org/10.1207/S15327957PSPR0703_01
- Zebrowitz, L. A., & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Social and personality psychology compass*, 2(3), 1497. <https://doi.org/10.1111/j.1751-9004.2008.00109.x>
- Zuckerman, M., & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior*, 13(2), 67–82. <https://doi.org/10.1007/BF00990791>

Appendixes

Appendixes

1. Bogus dot estimation task used in Chapter 4

The bogus dot estimation task was used in Chapter 4 (Schmitz & Yzerbyt, in press; Experiment 2) to randomly categorize participants into one of two fictitious groups (i.e., green or blue group).

Instructions:

TEST DE PERSONNALITÉ

Dans la tâche qui suit, nous allons vous demander de passer un test psychologique de personnalité basé sur la perception des stimuli.

Votre objectif est d'estimer du mieux que vous pouvez s'il y a plus de points **jaunes** ou si il y a plus de points **oranges** (tout en ignorant la taille des points). À chaque essai, vous aurez un temps **maximum** de **3 secondes**. Notez qu'il est souvent impossible de compter tous les points, c'est pourquoi nous vous demandons de faire un **choix rapide** en vous fiant à votre **intuition**. Aucun feedback ne vous sera donné pendant la tâche. Pour enregistrer vos réponses utilisez les touches < S > et < L > comme indiqué ci-dessous:

plus de points
jaunes

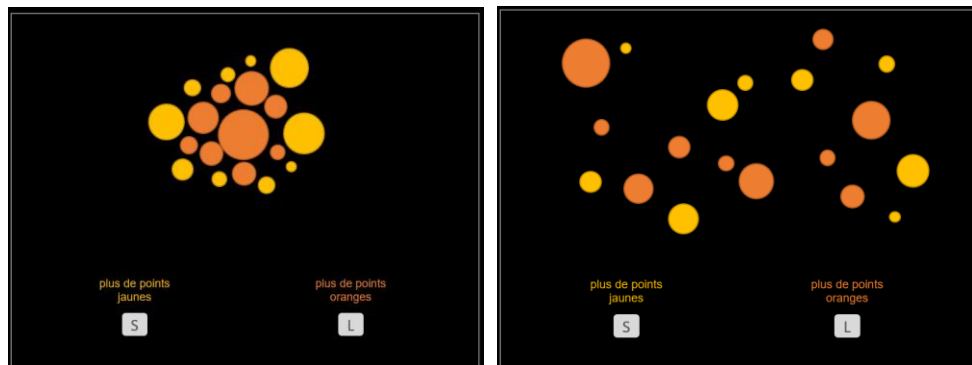
S

plus de points
oranges

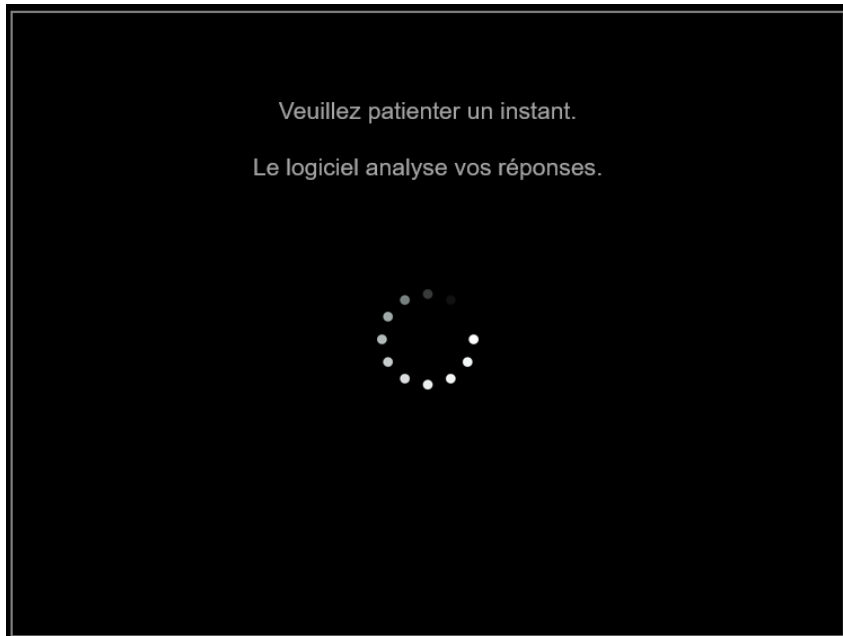
L

appuyez sur
<ESPACE>
pour commencer la tâche

Two examples of dot estimation trials (out of twelve trials):



Fake loading screen presented after the dot estimation task:



Bogus result of the dot estimation task used to randomly categorize participants into the green (as in the picture below) or blue group:



2. Behavioral statements used in Chapter 4 and 5

The behavioral statements were slightly adapted from previous studies (Judd et al., 2005; Kervyn et al., 2009; Yzerbyt et al., 2008). These behaviors were used in Experiments 1 (English) and 2 (translated into French) from Chapter 4 (Schmitz & Yzerbyt, in press) as well as for Experiments 1 (English) and 2 (English) from Chapter 5 (Schmitz, Vanbeneden et al., under review). Each behavior was either diagnostic of high competence, low competence, high warmth, or low warmth. Behaviors were presented inside a green or blue colored background rectangle to signal the actor's group membership.

High competence

- "X won the yearly award for the employee who contributes most to the company's profit."
- "X practiced the violin piece 20 times a day and after a month felt that he/she had it right."
- "X wrote a small computer program that solved a tough calculus integration problem."
- "X speaks fluently three different languages."
- "X presented his/her research at a chemistry congress."
- "X published a short story in a literary magazine while still in college."
- "X worked hard on the extra-credit assignment in linear algebra."
- "X is very careful when it comes to savings so that buying that first house will be possible."

Low competence

- "X went hiking in the mountains and had to be rescued because he/she was not paying attention to where he/she was going."
- "X was fired from his/her job because he/she rarely got things done on time."
- "X's bicycle was stolen several times because he/she forgot to set the lock."
- "X often gets lost when driving in the city, even though he/she has been living there for several years."
- "X's electricity was turned off because he/she did not pay the bill on time."
- "X had trouble finding work because he/she was always late for job interviews."
- "X took almost an hour to find his/her car after parking it in a shopping mall."
- "X did not pass his/her driving test for the third time and ended up abandoning it."

High warmth

- "X spent hours with a friend whose dog died."
- "X loves to be with other people."
- "X helped a blind woman cross the street."
- "X always greets friends with a big hug."
- "X always smiles at strangers on the street just to make their day better."
- "X loves to hold hands while walking."
- "X enjoys having long conversations with friends."
- "X gave up his/her seat on the crowded bus when an elderly woman got on."

Low warmth

- "X could not be bothered to give directions to a stranger."

“X decided that everyone at the party was pretty shallow and left early.”
“X prefers to go to the movie alone rather than with a friend.”
“X yelled at the driver who took the last parking space.”
“X rarely talked to the other people in the house that a bunch of them shared.”
“X did not want to congratulate the winner of the competition.”
“X did not go to his/her grandmother’s funeral because he/she was too busy with work.”
“X yelled at a little girl for coloring outside the lines.”

Instructions:

IMPRESSION FORMATION TASK

In this task we will present you a series of statements describing behaviors that have been performed by individuals belonging to one of two groups:

GREEN GROUP or **BLUE GROUP**

Each statement will be presented with a colored background, indicating the group membership of the individual who performed it.

Your task is to simply **read** each statement and to do your best to **form an impression** about these two groups. Note that this is not a memory task.

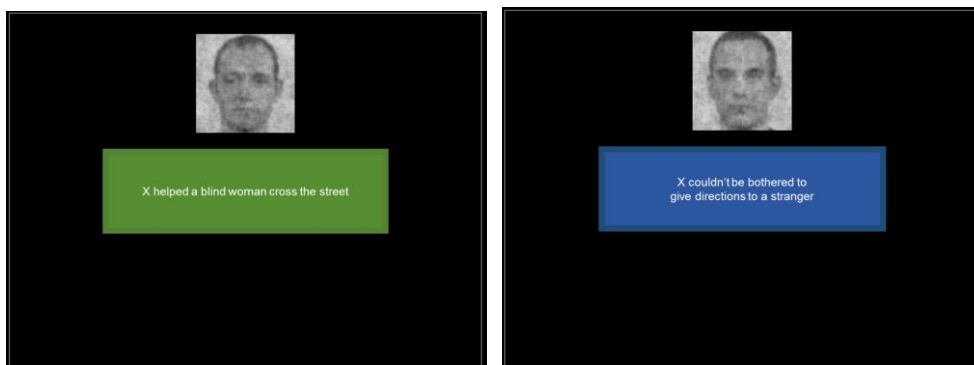
Please stay concentrated.

press
<SPACEBAR>
to begin the task

In Experiments 1 and 2 from Chapter 4 (Schmitz & Yzerbyt, in press), behaviors were presented together with four colored face shaped figures (green or blue). These face shapes were later used as stimuli for the BIAT. Below are displayed to examples:



In Experiment 1 from Chapter 5 (Schmitz, Vanbeneden et al., under review) behaviors were presented together with noisy face of what we presented as a group member (as in Dotsch et al., 2013) to facilitate the forthcoming reverse correlation task. Below are displayed two examples:



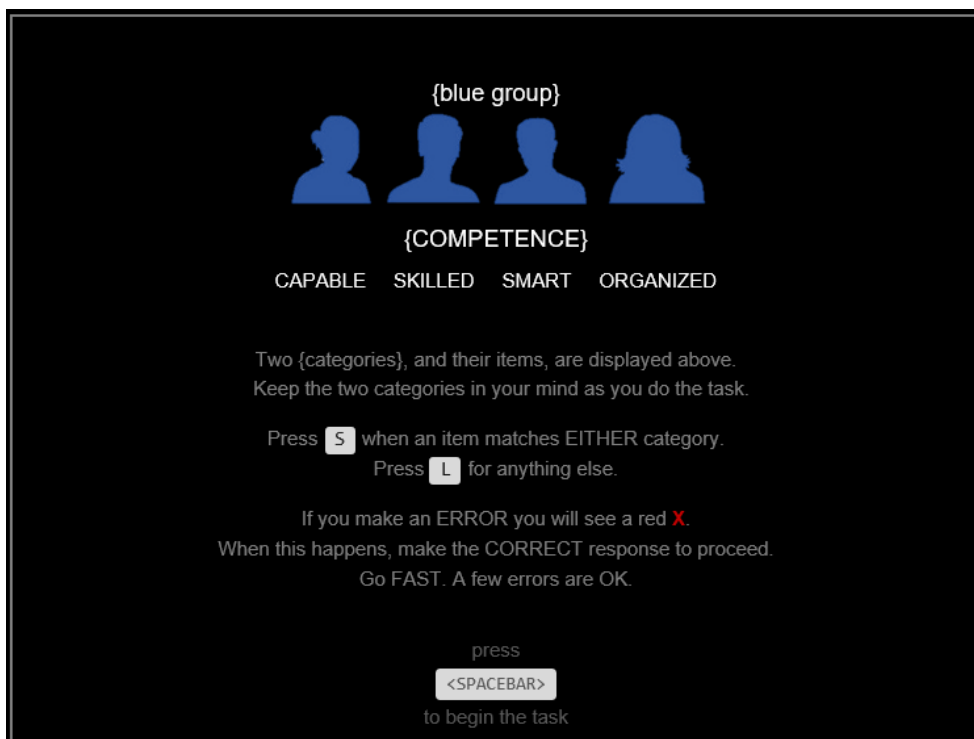
In Experiment 2 from Chapter 5 behaviors were presented without the exemplar noisy face. Below are displayed two examples:



3. Brief Implicit Association Test

In Experiments 1 (English) and 2 (translated to French) from Chapter 4 we used the Brief Implicit Association Test (BIAT; Sriram & Greenwald, 2009).

Below is an example of the BIAT instructions for the block with “competence” and “blue group” as the focal categories.



Below are four examples of trials with different stimuli from the same block:

