ISPA
INSTITUTO UNIVERSITÁRIO
CIÊNCIAS PSICOLÓGICAS, SOCIAIS E DA VIDA

THE STRUCTURING ROLE OF VALENCE IN THE RELATIONSHIP BETWEEN AND WITHIN MODELS OF FACE AND TRAIT IMPRESSIONS

Manuel José Barbosa de Oliveira

Thesis supervised by

Teresa Garcia-Marques, Ph.D.
ISPA – Instituto Universitário, William James Center for Research

Co-supervised by

Leonel Garcia-Marques, Ph.D.
Faculdade de Psicologia, Universidade de Lisboa

Ron Dotsch, Ph.D.
Utrecht University

Tese submetida como requisito parcial para obtenção do grau de

Doutoramento em Psicologia

Área de especialidade..... Psicologia Social

2019

Tese apresentada para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Psicologia na área de especialização de Psicologia Social, realizada sob a orientação de Teresa Garcia-Marques e co-orientação de Leonel Garcia-Marques e Ron Dotsch, apresentada no ISPA - Instituto Universitário, no ano de 2019.

O presente trabalho foi financiado pelo Programa Doutoral Lisbon PhD in Social Psychology (LiSP) da Fundação para a Ciência e a Tecnologia (PD/BD/113471/2015).

To my family

you belong to the group of giants whose shoulders I am standing on

Acknowledgments

Thank you Teresa and Leonel for all your guidance throughout my academic journey, for continuously inspiring me with your contagious passion for science, and for showing me how to become an independent thinker. I feel lucky and grateful for having been trained by such brilliant minds.

Thank you Ron for accepting to tag along this journey, for all the valuable skills you taught me, and for inspiring me with your insatiable thirst for knowledge and innovation. I was very well-received in Utrecht and provided with a great environment where I matured both as a researcher and as a person.

Thank you to all the founding members of the LiSP doctoral programme for investing in me and making this project a reality.

Thank you to all my colleagues at ISPA, Alexandre, Cris, Filipe, Joana, Nuno, Pedro, Rita, and Samuel, for their comradery and turning the PhD into a more convivial experience. Thank you to all the LiSPers out there for everything I learned from you in our inter-institutional lab meetings, for your good company during the social meetings I was able to attend, and for all your efforts to hold the bridge between institutions.

Thank you to all my colleagues from the RC lab meetings, Loek, Mike, Jelmer, Irene, and Emma, for creating such a stimulating and collaborative learning environment. Thank you to Henk Aarts for approving my extended stay at UU and integrating me in the GOALLAB. Thank you also to all the GOALLAB members for their valuable input and teaching me so much about topics outside my own. Thank you Sammy and Myrthel for receiving me so well in your office, and for all the interesting conversations about stats and programming over coffee. I hope our office plant Pearson remains as gorgeous as ever. I look forward to meeting you again someday. Thank you to all my colleagues and staff at the SHOP department in Utrecht University for making me feel at home and creating such a pleasant and collaborative work environment. Special thanks to Ruud Custers, Hans Marien, all my officemates from G1.11 and G1.15, Anouk van der Weiden, and Catherine Evers for their warm goodbye events during my last week in Utrecht.

Thank you to all my friends in Utrecht and Aveiro for their understanding, support, and reminding me once in a while that there is a life beyond the PhD.

Special thanks to Anneloes for all your love, support, understanding, and positivity throughout the best and hardest moments of this journey. Let’s start the next chapter.

Thank you to my sister, who was always there to encourage me and remind me of my capabilities in moments of self-doubt.

Finally, words are not enough to express how indebted and grateful I am to my mother and to my father, for their unconditional love, support, and teaching me the value of hard-work and kindness.

Aveiro, October 18th, 2019

Palavras-chave:
Formação de Impressões; Perceção de Faces; Perceção de Pessoas; Valência

Key words:
Impression Formation; Face Perception; Person Perception; Valence

Categorias de Classificação da tese

PsycINFO Classification Categories and Codes:

2340 Cognitive Processes

3000 Social Psychology
    3040 Social Perception & Cognition

RESUMO

Na investigação em perceção social de faces, confiabilidade e dominância emergiram como as principais dimensões subjacentes às impressões de personalidade baseadas em aparência facial. Estas dimensões assemelham-se bastante às encontradas no domínio paralelo de perceção interpessoal, como as dimensões de simpatia e competência, ou *communion* e agência, que subjazem impressões de personalidade baseadas em descrições verbais de pessoas (e.g., baseadas em traços). Dada a coocorrência de ambos os tipos de informação na maioria de interações sociais reais, e o seu impacto na tomada de decisões sociais, torna-se crucial compreender como as impressões derivadas de ambas as fontes de informação se inter-relacionam. Porém, o grau de sobreposição entre as dimensões da perceção social de faces e da perceção interpessoal, assim como a natureza e direção das inter-relações entre essas dimensões, têm sido questões amplamente negligenciadas na literatura. O principal objetivo desta tese foi preencher essa lacuna na literatura e dar os primeiros passos em direção à integração dos modelos de impressões de personalidade associados à perceção social de faces e à perceção interpessoal.

No primeiro artigo, métodos de *reverse correlation* foram implementados para avaliar em que medida as dimensões da perceção social da face se sobrepõem às dimensões da perceção interpessoal na perspetiva dos próprios percipientes. Os resultados mostraram que as dimensões exibindo relações opostas com valência, como dominância e competência, foram percebidas como menos semelhantes entre si comparativamente às dimensões partilhando a mesma relação (positiva) com a valência, como confiabilidade, simpatia, e competência. Estes achados esclareceram que as dimensões de impressões faciais e da perceção interpessoal nem sempre são percebidas como redundantes, e destacam o papel da valência na estruturação das relações das dimensões entre os dois domínios de investigação.

No segundo artigo, desenvolveu-se um paradigma para avaliar diretamente a natureza da relação que a valência estabelece com as dimensões centrais das impressões faciais e da perceção interpessoal. Os resultados revelaram que as dimensões de traço relacionadas com habilidade, como competência e dominância, exibiram maior variabilidade na natureza e direção da sua relação com valência, comparativamente às dimensões relacionadas com moralidade e simpatia. Estes achados enfatizaram ainda que a sobreposição, ou dissociação, entre as dimensões é amplamente promovida pelas características das relações destas com valência.

O terceiro artigo focou-se exclusivamente na perceção social de faces, e utilizou métodos de reverse correlation para investigar como a confiabilidade e a dominância são naturalmente integradas em impressões unitárias de aparência facial. Os resultados mostraram que a dimensão mais fortemente relacionada com valência—confiabilidade—teve maior peso que dominância nas impressões faciais. Estes achados destacam o papel primário da valência na integração de dimensões em impressões faciais.

No geral, estes achados destacam o papel primário da valência na estruturação das relações entre dimensões de julgamento social, não apenas entre os modelos de impressões faciais e de perceção interpessoal, mas também dentro de cada modelo. Além disso, oferecem uma visão mais clara sobre a relação e integração destes modelos, e traçam direções claras para futuras investigações no domínio geral da perceção social.

ABSTRACT

In social face perception research, trustworthiness and dominance were found as the core dimensions underlying personality impressions based on facial appearance. These dimensions bear a striking resemblance to dimensions found in the parallel domain of person perception research such as the warmth and competence or communion and agency dimensions of personality impressions based on verbal person descriptions (e.g., trait-based descriptions). Given that both types of social information often co-occur in real social interactions and guide social decision making, it becomes crucial to understand how impressions derived from both sources are interrelated. Yet, so far, questions regarding the extent to which the dimensions of social face perception overlap with the dimensions of person perception, and regarding the nature and direction of the interrelationships between these dimensions, have been largely overlooked in the literature. The main goal of this thesis was to fill this gap in the literature and make initial steps towards the integration of social face perception and person perception models of personality impressions.

In the first paper, a reverse correlation methodology was used to assess the extent to which dimensions of social face perception were perceived to overlap with dimensions of person perception by perceivers themselves. The results showed that dimensions establishing opposite relationships with valence, such as dominance and competence, were perceived as less similar than dimensions establishing a common positive relationship with valence, such as trustworthiness, warmth, and competence. These findings clarified that the dimensions of facial impressions and of person perception are not always perceived as redundant, and further highlighted the role of valence in shaping the relationship between dimensions across domains.

The second paper employed a paradigm designed to directly assess the nature of the relationship that valence establishes with the core dimensions of social face perception and person perception. The results revealed that ability-related trait dimensions such as competence and dominance exhibited more variability in the nature and direction of their relationship with valence, comparatively to dimensions related with morality and warmth. These findings further emphasized that the overlap or dissociation between core dimensions of social judgment is largely driven by the features of the relationship they establish with valence.

The third paper focused exclusively on social face perception and used a reverse correlation methodology to investigate how trustworthiness and dominance are naturally integrated into unitary impressions of facial appearance. The results showed that the dimension more strongly related with valence—trustworthiness—outweighed dominance in the resulting impressions of facial appearance. These findings highlight the primary role of valence information in shaping how dimensions are integrated within social face perception.

Overall, these findings highlight the primary role of valence in structuring the relationship between dimensions of social judgment, not only across models of person perception and social face perception, but also within each model. Moreover, they offer a clearer picture on the relationship and integration of models of social face perception and person perception, and lay out clear new directions for future research on social perception in general.

Index

Section I :
Face Perception versus Person Perception Models of Social Perception

Section II: Empirical Section

Figures Index

Section I:
Face Perception versus Person Perception Models of Social Perception

Section II: Empirical Section

CHAPTER VI. Dominance and competence face to face: Dissociations obtained with a reverse correlation approach

CHAPTER VII. Good to bad or bad to bad? What is the relationship between valence and the trait content of the Big Two?

CHAPTER VIII. Combining traits into a face: A reverse correlation approach

Section III: General Discussion

Table Index

Section II: Empirical Section

CHAPTER VI. Dominance and competence face to face: Dissociations obtained with a reverse correlation approach

CHAPTER VII. Good to bad or bad to bad? What is the relationship between valence and the trait content of the Big Two?

CHAPTER VIII. Combining traits into a face: A reverse correlation approach

Section I

Face Perception versus Person Perception Models of Social Perception

Introduction

> *We look at a person and immediately a certain impression of his character forms itself in us. A glance, a few spoken words are sufficient to tell us a story about a highly complex matter.* (Asch, 1946, p. 48)

Solomon Asch's words elegantly abridge the vital role that impression formation plays in our social lives. This thesis falls within the broader domain of impression formation, and aims to contribute to the vast literature that was set in motion by Asch's work.

When we are getting to know someone we are faced with the task of gathering information that is relevant for interacting with the person. Throughout this process, we seem to be primarily inclined to gather information that signals others' social- and competence-related features such as warmth, trustworthiness, intelligence or dominance (Fiske, Cuddy, & Glick, 2007; Oosterhof & Todorov, 2008; Rosenberg, Nelson, & Vivekananthan, 1968; Wojciszke, 2005b). How benevolent are the intentions of this person? How capable is the person to implement her intentions? How approachable is the person? Can I trust this person, or should I be careful? Or ultimately, how threatening might this person be? These are all questions that we are trying to answer while forming an impression about someone's personality. As social perceivers, we find our answers in the traits we infer based on all the information we gather about a person, and especially from their behavior and facial appearance (e.g., Bodenhausen & Macrae, 2006; Hassin & Trope, 2000; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). And importantly, those trait inferences reflect the two primary dimensions of psychological content that have been consistently found to underlie how people subjectively perceive each other in terms of personality (i.e., person perception) (Abele & Wojciszke, 2014; Fiske, 2018; Fiske et al., 2007; Rosenberg & Sedlak, 1972).

In this thesis we will focus on two particular domains of social perception wherein two-dimensional models emerged as theoretical frameworks that subsume the core dimensions underlying personality inferences of unfamiliar others. The first domain is person perception, where the focus is placed on how people subjectively perceive each other's personality based on behavioral or trait information, (e.g., verbal descriptions of a target person) (e.g., Abele, Cuddy, Judd, & Yzerbyt, 2008; Rosenberg & Sedlak, 1972; Uleman & Kressel, 2013). The second domain is social face perception, which in turn is concerned with

how people infer personality traits based on facial appearance alone (e.g., Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015; Zebrowitz, 2006).

Despite of focusing on different modalities of social stimuli (i.e., visual vs. verbal), these two research domains converge in one aspect: Two primary and relatively independent dimensions were found to underlie personality impressions within each of these domains. In person perception, the two dimensions that subsume the content of personality impressions are warmth (also known as communion, morality, or social good-bad) and competence (also known as agency, or intellectual good-bad) (Abele & Wojciszke, 2014; Fiske et al., 2007; Rosenberg et al., 1968). In social face perception, however, it is the *trustworthiness* (or valence) and *dominance* (or power) dimensions that account for most of the variability in personality impressions derived solely from facial appearance (Oosterhof & Todorov, 2008; Todorov et al., 2015). The convergence of outcomes of the research developed in these two separate domains should not be surprising, as they both ultimately tap into the cognitive structure underpinning how the human mind evaluates human agents in a social environment. However, the literature has given little attention to *how* the two domains converge. In this thesis, we argue that a more thorough examination of how these two domains converge is necessary for a more complete understanding of how we perceive and mentally represent other people based on behavioral and facial information.

The resemblance between the two two-dimensional models becomes apparent from a functional perspective: First, both models include a dimension related with the evaluation of intentions (e.g., is the person a friend or a foe?). Within the domain of social face perception, the evaluation of intentions is reflected in the dimension of trustworthiness, sometimes described as valence (Jones et al., 2018; Oosterhof & Todorov, 2008). Within the domain of person perception, the dimension is known as warmth (Fiske, 2018), communion (Abele & Wojciszke, 2014), morality (Wojciszke, 2005b), or social good-bad (Rosenberg et al., 1968). Second, both models include a dimension related with the evaluation of ability (e.g., how able is the person to implement her intentions?). Within the domain of social face perception, the evaluation of ability reflects an evaluation of dominance, sometimes described as power (Oosterhof & Todorov, 2008). Within the domain of person perception, the ability dimension is known as competence (Fiske, 2018; Wojciszke, 2005b), agency (Abele & Wojciszke, 2014), or intellectual good-bad (Rosenberg et al., 1968). Thus, despite being best represented by different trait adjectives in the two different contexts of research, a parsimonious interpretation would be that only two dimensions underlie impressions of personality

regardless of the type of information about a target person: one encompassing information about intentions (trustworthiness or warmth), and another encompassing information about ability (dominance or competence).

A related aspect that is important to consider when examining the differences between the trustworthiness-by-dominance and warmth-by-competence (or communion-by-agency) models is that the most representative traits within each domain are related with the specific context in which personality inferences are made. Whereas trustworthiness and dominance seem to be the most relevant traits in a context where physical features are evaluated, warmth and competence are the most relevant traits in a context where behavioral and conceptual trait information are evaluated. At this point, it is important to clarify that both the trustworthiness-by-dominance of facial impressions and the seminal social-by-intellectual model of person perception (preceding the warmth-by-competence model) derived their respective dimensions using an inductive approach (i.e., data-driven), and thus, reached different solutions in function of the modality of their target stimulus. The comparison of these models is fruitful in that it generates questions whose answers will increase our understanding of the impression formation process. For instance, some questions we may ask at this point include: How differently perceived are the intention- and ability-related traits across these domains? And, do these trait inferences have a complementary or overlapping (confirmatory) informative role, within and across domains?

In the following chapters we will review and borrow support from previous literature to argue that, although these dimensions may ultimately amount to an appraisal of intentionality and ability, the traits that best represent intention and ability in each of the two contexts of personality perception under focus exhibit noteworthy differences at the conceptual and evaluative levels that call for a more nuanced examination of the apparent overlap of these dimensions across domains. Although both dominance and competence pertain to an ability, the ways in which this ability is expressed in terms of expected facial features or behaviors may be quite different, which in turn may lead the perceiver to expect different outcomes in a social interaction. Similarly, although trustworthiness and warmth may inform the perceiver about someone's intentions, each of these dimensions may be expressed differently in terms of facial features and behaviors. And, perhaps most importantly, the extent to which the trait concepts representing the dimensions overlap across the two models should be dependent upon how these traits are perceived in terms of valence (i.e., their positivity).

This introductory section (Section I) is composed of five chapters. All chapters, excluding Chapter V, include abridged reviews of the literature that provide the basis for a more complete comprehension of the fields of research, topics, and methodologies underpinning our research questions. In the last chapter (Chapter V, more details below) of this introductory section, we present the main research questions of this thesis, by articulating the literature reviewed in the four previous chapters.

In Chapter I we start with a focus on person perception and review the most influential two-dimensional models of personality impressions based on conceptual knowledge about behavior and trait categories. These models include the social-by-intellectual good-bad model, the warmth-by-competence model, the communion-by-agency (or Big Two) model, and the most recent Big Two facets model. Because some of these models encompass social perception at the interpersonal, intrapersonal, and intergroup levels, we must clarify that our focus will be placed on the interpersonal level. This focus is justified by our aim of comparing the dimensions of these models with the dimensions of facial impressions which occur at the interpersonal level.

In Chapter II we turn our focus to the literature of social face perception and highlight research on the basic dimensions of personality inferences from faces, and on the relationship between personality inferences and facial features resembling the expression of emotions. Within this chapter we will additionally review state-of-the-art methodologies used in social face perception research such as psychophysical reverse correlation, which we employed in some of our studies described in the empirical section of this thesis.

In Chapter III we shift our focus to the relationships between the warmth and competence dimensions of person perception and the trustworthiness and dominance dimensions of social face perception. First, we will focus on literature that sheds light on the extent to which we should expect a conceptual and evaluative overlap between the intention- and ability-related dimensions across domains, by comparing the trustworthiness and warmth dimensions, and the dominance and competence dimensions. We will discuss how previous research has been showing that, although warmth and trustworthiness share a consistent positive correlation with valence (e.g., the more warm or trustworthy a person is, the more positive is our final impression of her), the same does not occur for dominance and competence, which exhibit an opposite and inconsistent relationship with valence across the extant literature. Furthermore, this relationship with valence has been shown to influence the direction of the relationship established between the two dimensions within models of social

perception (e.g., Suitner & Maass, 2008). For this reason, we will conclude the chapter with a focus on how the relationship between dimensions within each model is shaped by the relationship that each of these dimensions establishes with valence.

In Chapter IV we move beyond the topic of the relationships between dimensions and turn our focus to the questions of whether perceivers attribute higher relevance to any of the dimensions within a model, and how these dimensions are integrated into unitary impressions of personality. We will briefly review the literature that provided insight into the interplay between these two aspects of impression formation and the relationship between dimensions (e.g., impact of processing one dimension before the other), while comparing how these questions have been investigated in the domains of person perception and social face perception.

Drawing on the literature reviewed throughout Chapters I to IV, we selected a set of specific questions which we empirically addressed in the three different papers presented in the empirical section (Section II) of this thesis. In sum, the research questions that we raise in this thesis are predominantly focused on clarifying and describing the interrelationships between core dimensions of social face perception and person perception. Importantly, based on previous research showing how valence plays a role in shaping the relationship between the basic dimensions of person perception, we placed a special focus on examining the role played by valence in structuring the relationships between dimensions not only across, but also within, the two domains of social perception under focus.

In Chapter V, we provide an overview of our research aims and methodological approaches used in each of the three papers presented in the empirical section (Section II), while at the same time clarifying how these papers are interrelated and provide clues to answering the questions raised in this thesis.

Chapter I. The Fundamental Dimensions of Person Perception

One of the main goals of this thesis is to examine the relationship between the dimensions of models of person perception and models of social face perception. To better understand how they are related it is important to first clarify their origins, similarities and distinctions. In this chapter we review the most influential two-dimensional models that emerged in the domain of person perception. These include the seminal social-by-intellectual model of person perception (Rosenberg et al., 1968), or the more contemporary warmth-by-competence (Stereotype Content Model; Fiske, Cuddy, Glick, & Xu, 2002) and communion-by-agency (Abele & Wojciszke, 2007, 2014) models. In light of the most recent developments in social perception research, we also review the Big Two facets model (Abele et al., 2016; see also Fiske, 2018), which is able to integrate a considerable body of research supporting a more granular view on the fundamental dimensions of person perception. Our aim is to provide insight into the origins of these models and how their dimensions have been interpreted, before we move on to explicitly comparing the dimensions of person perception with the dimensions of facial impressions in Chapter III. As it will become apparent throughout this chapter, the main distinction between these dimensional models and that of social face perception is that they emerged from paradigms using verbal person descriptors (e.g., behavioral descriptions, trait-based descriptions, or group labels) as the targets of judgments, instead of face stimuli.

The Social-by-Intellectual Model

Ever since Solomon Asch (1946) set out to systematically uncover the underpinnings of how people form impressions of personality about each other, a great number of milestones have been achieved in the topic of person perception. One of them was the finding by Seymour Rosenberg and colleagues (Rosenberg et al., 1968) that two relatively independent dimensions parsimoniously describe the content of subjective personality impressions of single individuals. Rosenberg and colleagues (Rosenberg et al., 1968) were interested in investigating the multidimensional structure underlying people's *lay theories of personality* (or implicit theories of personality), which reflect personal beliefs about how personality traits should co-occur in the same individual (Bruner & Tagiuri, 1954; Schneider, 1973). These lay theories are thought to reflect a semantic network of trait interrelationships in memory that

perceivers rely upon when they go beyond the information that is more accessible to them (i.e., "jumping into conclusions") and infer additional personality attributes about someone in the context of first impressions (Bruner & Tagiuri, 1954; Hays, 1958). By examining the fundamental structure underlying these lay theories, they would be able to identify the most relevant trait dimensions in impression formation.

In their seminal work, Rosenberg and colleagues (Rosenberg et al., 1968) asked each participant to sort personality trait cards into 10 piles, each representing a given familiar person (e.g., friends, family members, or public figures that the participant could think of), according to how the participant thought those "traits tend to go together in the same individual" (Rosenberg et al., 1968, p. 285). These data were then submitted to a Multidimensional Scaling analysis (MDS), a dimensionality reduction technique—to a certain extent similar to a PCA, or Exploratory Factor Analysis—that allowed them to identify how many dimensions sufficiently accounted for the variability in the data (for more details on MDS see Chapter VI). Their analyses indicated that two dimensions appropriately described the structure underlying the trait interrelationships in their data. To interpret these dimensions they submitted the traits used in their trait sorting task to ratings on several dimensions, previously suggested to be relevant in person perception (e.g., Hays, 1958; Osgood, Suci, & Tannenbaum, 1957). They found that traits differentiated along the first dimension (e.g., honest, warm, dishonest, cold) were best interpreted as ranging from good to bad for social activities, and traits differentiated along the second dimension (e.g., intelligent, skillful, unintelligent, clumsy) were best interpreted as ranging from good to bad for intellectual activities. As a result, they labelled the two dimensions as social good-bad and intellectual good-bad (see Figure 1, extracted from Rosenberg et al., 1968).

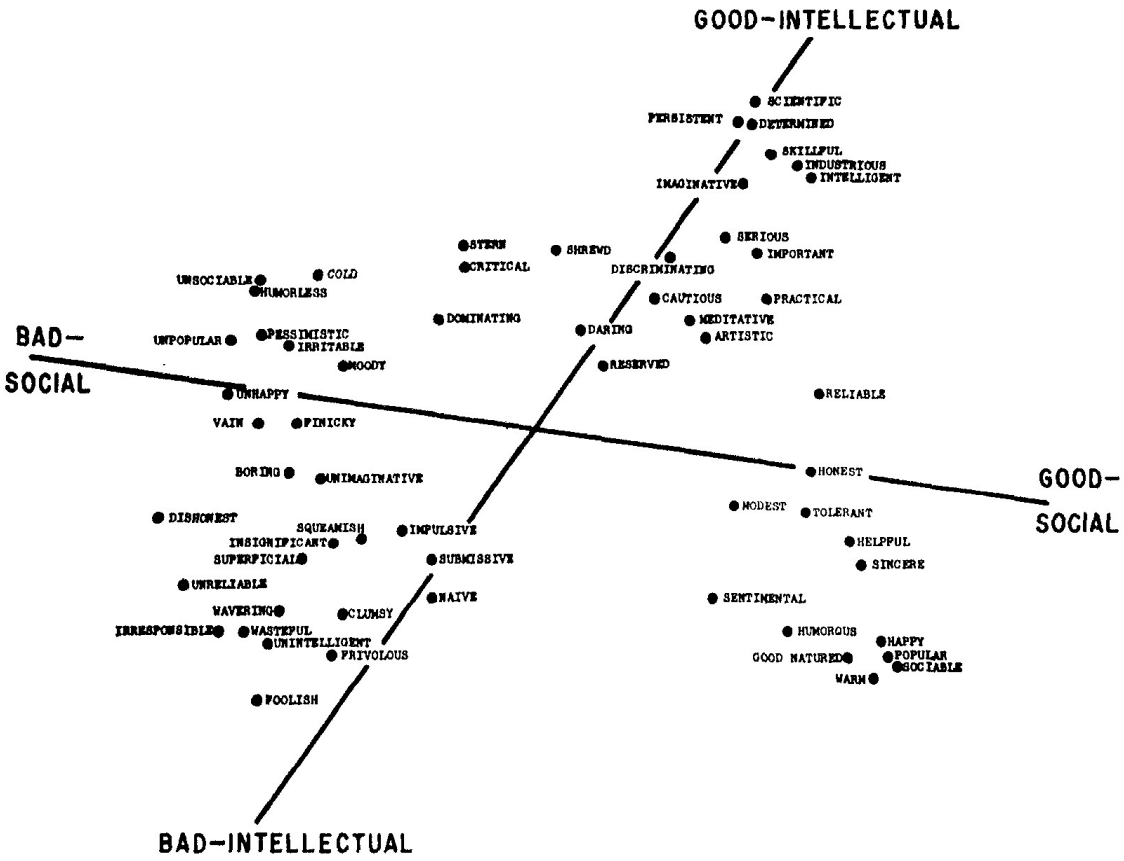

Figure 1. Illustration of the two-dimensional space (conceptual trait space) defined by the social good-bad and intellectual good-bad dimensions found by Rosenberg et al. (1968). Image extracted from Rosenberg et al. (1968, p. 290).

In sum, this model describes the two-dimensional structure underlying the content of our conceptual knowledge about traits and how they are interrelated. This new theoretical framework promoted a deeper understanding of how people form personality impressions about each other by shedding light on the dimensions that future research on impression formation should focus on. Since then, the social-by-intellectual model became a reference in most of the dimensional approaches to social perception that followed, including the more recent and influential warmth-by-competence model of group perception and stereotype content (Cuddy, Fiske, & Glick, 2008; Fiske et al., 2007, 2002), which we describe next.

The Warmth-by-Competence Model

Although the social and intellectual dimensions emerged in the context of abstract person perception—in the sense that participants retrieved the targets of their evaluations from memory—they were also found to parsimoniously describe the content of group stereotypes by Fiske and colleagues (Fiske et al., 2007, 2002; Fiske, Xu, Cuddy, & Glick, 1999), who aimed to empirically test the generalizability of these dimensions across the interpersonal and intergroup levels of social perception (see Fiske et al., 1999). However, these authors labeled these dimensions as warmth (cf. social) and competence (cf. intellectual), and called the resulting two-dimensional model the Stereotype Content Model (Fiske et al., 2002), which we will henceforth refer to as warmth-by-competence model. Thus, and importantly, the warmth and competence dimensions were largely derived, in a theory-driven manner from the social-by-intellectual model of person perception (e.g., see Koch, Imhoff, Dotsch, Unkelbach, & Alves, 2016 who contrasted their data-driven methodology with Fiske's approach). The labels of these dimensions were largely influenced by the warm-cold and competence-related (i.e., intelligence) traits used in Asch's (1946) classical studies on impression formation, and selected on the basis of the results of a survey conducted by Fiske et al. (1999). In this survey, participants were asked to rate 17 diverse societal groups on 27 trait adjectives related with communality (warmth) and agency (competence), selected from previous work on gender stereotypes (Conway, Pizzamiglio, & Mount, 1996). These ratings were then analyzed in separate factor analyses, one for each target societal group. The results indicated that the two most stable factors emerging across groups were best represented by the traits *competent* (competence factor), and *likeable* or *warm* (warmth factor). Other traits encapsulated by these factors were "*intelligent*, *confident*, *competitive*, and *independent*" (Fiske et al., 1999, p. 478) for the competence factor, and *sincere*, *good-natured*, and *tolerant* for the warmth/likeability factor. These findings served as a basis for subsequent measurements of warmth and competence perceptions, ultimately defining the trait content expected to best capture each dimension. Whereas warmth scales included traits such as "good-natured, trustworthy, tolerant, friendly, and sincere" (Cuddy et al., 2008, p. 65), competence scales included traits such as "capable, skillful, intelligent, and confident" (Cuddy et al., 2008, p. 65). Thus, similarly to the social-by-intellectual model, the warmth-by-competence model defines warmth in terms of traits that are relevant for social activities, and competence in terms of traits that are relevant for intellectual activities. For instance, warmth scales usually include traits such as "good-natured, trustworthy, tolerant, friendly, and

sincere" (Cuddy et al., 2008, p. 65), whereas competence scales usually include traits such as "capable, skillful, intelligent, and confident" (Cuddy et al., 2008, p. 65).

From a functional standpoint, Fiske and colleagues interpreted these dimensions as reflecting a primary focus on identifying others' intentions and capabilities, that ultimately allows perceivers to evaluate the degree to which someone represents a threat to their self-interests (Fiske et al., 2007). According to this view, the degree of threat is contingent on the intentions of the target person. A competent ill-intentioned target represents a higher threat to the perceivers' goals than an incompetent ill-intentioned target. On the other hand, targets with good intentions are likely to be perceived as less threatening, regardless of their capability to implement intentions. Importantly, these authors added to their interpretation that warmth-related information was given priority over competence-related information in social perception, drawing on previous findings showing that warmth-related information is processed faster (Ybarra, Chan, & Park, 2001) and given more weight in global impressions of others (Wojciszke, Bazinska, & Jaworski, 1998) comparatively to competence information. Put simply, upon meeting a person for the first time, we are first interested in knowing the person's intentions, before focusing on the person's capabilities to act upon them.

To conclude, we must at this point emphasize that, despite of emerging in the domain of group stereotypes, the warmth-by-competence model is considered to apply to interpersonal perception (Cuddy et al., 2008), not least because it was largely influenced by classical findings in person perception (see Fiske et al., 2007, 1999). Recognizing the accumulating evidence demonstrating that these two dimensions consistently emerged across different lines of research (e.g., person perception, group perception, self-perception) within the social perception literature (e.g., Abele, 2003; Asch, 1946; Peeters, 2002; Rosenberg et al., 1968; Wojciszke, 1994), Fiske and colleagues (Fiske et al., 2007) proposed referring to warmth and competence as the fundamental dimensions of social perception (Cuddy et al., 2008; Fiske et al., 2007). Moreover, the hypothetical universality of these dimensions received some support in subsequent research showing that warmth and competence reliably account for the variability in stereotypical content across several cultures (Cuddy et al., 2009). Thus far, the warmth-by-competence model has been one of the most influential of its class in the social psychological literature, and its status in the literature is only paralleled by its twin model of communion and agency, reviewed next.

The Communion-by-Agency Model

The communion and agency dimensions, also referred to as the *Big Two* of social perception, are often considered to be akin to the warmth/social and competence/intellectual dimensions (Abele & Wojciszke, 2007, 2013; Cuddy et al., 2008). There are, however, some distinctions worth mentioning regarding how these dimensions are conceptually defined.

The communion and agency constructs were initially proposed by David Bakan (1966) in a philosophical discussion within the context of personality psychology. Bakan (1966) argued that there were two modalities characterizing human existence: "agency for the existence of an organism as an individual" which manifested "itself in the urge to master" (Bakan, 1966, pp. 14–15), and "communion for the participation of the individual in some larger organism of which the individual is part", manifesting itself in "non-contractual cooperation" (Bakan, 1966, pp. 14–15). This idea was later adopted by several social psychologists who built upon it to conceptually define the two dimensions of content underlying social judgments that consistently emerged across the social perception literature (for reviews see Abele & Wojciszke, 2014; Cuddy et al., 2008). In more recent literature, the communion dimension has been defined as emerging from "strivings to integrate the self in a larger social unit through caring for others" (Abele & Wojciszke, 2007, p. 751) and is associated with traits promoting cooperation, emotional expression, and well-being (in both others and ourselves). In turn, agency has been defined as arising from "strivings to individuate and expand the self" (Abele & Wojciszke, 2007, p. 751) and encapsulates traits related with "ambition, dominance, competence, and efficiency in goal attainment" (Abele & Wojciszke, 2007, p. 751). The communion-by-agency model itself was empirically supported by the studies conducted by Abele and Wojciszke (2007). These authors asked participants to rate 300 trait words on unidimensional scales of competence, morality, individualism, collectivism, agency, communion, masculinity and femininity. Because some of the traits, such as agency and communion, are scientifically devised constructs that are not spontaneously used by people, the participants were presented with a definition of each construct before they could rate it. The content of their trait set was selected on the basis of previous research and included traits considered to be representative of agency and communion, traits associated with gender categories, traits from the social-by-intellectual model, warmth and competence, morality- and competence-related traits, traits perceived as more profitable for oneself or more profitable for others, traits associated with individualism and collectivism, and traits associated with the Big Five factors of personality (for more

details see Abele & Wojciszke, 2007, p. 754). (In this way, Abele and Wojciszke's (2007) work also strove to integrate the trait content of multiple previous models of social perception, encompassing person perception, self-perception and group perception, coming from multiple different theoretical frameworks.) The trait ratings were then submitted to a Factor Analysis, which yielded a two-factorial structure (i.e., two dimensions). The first factor, interpreted as communion, accounted for the highest amount of variability in the items (around two thirds), and was mostly associated with (from highest to lowest factor loading) collectivism, communion, morality, and femininity. The second factor, interpreted as agency, accounted for a lower amount of item variability (around one third), and was mostly associated with (from highest to lowest factor loading) agency, individualism, competence, and masculinity. These findings showed that the communion and agency dimensions were able to capture the trait content of the dimensions defining other models (e.g., social-by-intellectual and warmth-by-competence). This is one of the main reasons that led Abele and Wojciszke to later refer to them as the "Big Two" of social perception (Abele & Wojciszke, 2013). In addition, Abele and Wojciszke (2007) asked participants to rate the same traits on the degree to which a trait was beneficial to the self versus to others (i.e., self-interest vs. other-interest), and found that whereas communion was more strongly and positively related with other-interest, agency was more strongly and positively related with self-interest. Put simply, from the perspective of the trait possessor, communion-related traits are perceived as more beneficial to others than to oneself; and agency-related traits are perceived as more beneficial to oneself than to others.

At this point, we should realize that one key difference between communion and agency and the warmth and competence dimensions is that the former are scientifically devised constructs capable of capturing a broader range of trait content comparatively to the trait-based labels that characterize the warmth and competence dimensions (Abele & Wojciszke, 2014). However, some disagreement emerged regarding the competence and agency constructs regarding their conceptual breadth. Although Fiske and colleagues considered their warmth and competence dimensions to overlap with the communion and agency dimensions, they also clarified that they understood the overlap to be greater between warmth and communion, than between competence and agency. In their understanding, the competence construct captures both traits related with potential action (e.g., intelligence, competence) and with taking action (e.g., assertive, efficient, dominant), while the agency construct seems to place a higher focus on traits related with taking action (Cuddy et al., 2008, p. 65), thus only partially capturing the content of this "ability-related" dimension. Adding to

the resulting definitional debate, evidence started to accrue showing the multifaceted nature of the Big Two dimensions, which led to the more recent proposal of the Big two facets model (Abele et al., 2016), and an update of how these dimensions are defined (e.g., Fiske, 2018).

Facets of the Big Two

Over the past two decades, several findings in the social perception literature began to accumulate and suggest that people differentiated between sociability- and morality-related information within the communion/warmth dimension (Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Goodwin, Piazza, & Rozin, 2014; Leach, Ellemers, & Barreto, 2007), and between assertiveness- and competence-related information within the agency/competence dimension (Carrier, Louvet, Chauvin, & Rohmer, 2014).

In the case of communion/warmth, Goodwin and colleagues (Goodwin et al., 2014) argued that traits varied in the extent to which they were morally relevant. For instance, traits such as trustworthy or honest are morally relevant in that they are informative about someone's intentions. On the other hand, traits such as friendly or sociable are less able to inform about someone's moral character, and are more related with the perception of someone's ability to navigate the social world or recruit support for their intentions. With this distinction in mind, Goodwin and colleagues (Goodwin et al., 2014) differentiated between morality-related traits and social warmth-related traits, while at the same time clarifying that some traits could be relevant to both the morality and sociability/social warmth dimensions (e.g., cooperative, humble).

Regarding the agency/competence dimension, Carrier and colleagues (Carrier et al., 2014) made the distinction between agentic traits that implied an underlying motivation to advance the self (e.g., dominance, ambition, competitiveness), and competence-related traits that merely implied the possession of resources to implement intentions (e.g., intelligence, competence), further showing that agentic traits were more predictive of status perceptions than competence-related traits (see also Louvet, Cambon, Milhabet, & Rohmer, 2019; Mollaret & Miraucourt, 2016 for additional support for a distinction between competence and agentic traits).

These findings were taken into account in the recently proposed facet model of the Big Two by Abele and colleagues (Abele et al., 2016), which branches the communion dimension into the morality and sociability facets, and the agency dimension into the assertiveness and

competence facets (for an illustration see Figure 2). In a series of studies, Abele and colleagues (Abele et al., 2016) assessed self-perceptions on communion- and agency-related traits across six different countries/cultures (Australia, China, France, Germany, Poland, and USA). Their trait set included the most frequently used trait-adjectives in prior research guided by the communion-by-agency model. At the same time, the trait content was counterbalanced in regard to their hypothesized membership to the facets of the Big Two, based on the previous findings in person and group perception suggesting the branching of the Big Two (see above). Regarding the communion dimension, some of the traits used to represent the morality facet included "trustworthy", "honest", or "just"; whereas traits used to represent the warmth facet included "warm", "caring", or "friendly" (Abele et al., 2016, p. 5). Regarding the agency dimension, some of the traits used to represent the assertiveness facet included "assertive", "ambitious", and "feel superior"; whereas traits used to represent the competence facet included "competent", "intelligent", and "capable" (Abele et al., 2016, p. 5). The ratings were then first submitted to Exploratory Factor Analyses (EFAs), separately by country, and solutions for different dimensionalities were inspected. For every country, Abele and colleagues (Abele et al., 2016) found evidence for a two-factor solution, differentiating between the Big Two, and for the four facets of the Big Two in higher dimensionality solutions. Next, they submitted the data to Confirmatory Factor Analyses, and observed that, although both the two-factor and four-factor models exhibited a good fit to the data, the four-factor facets model showed a better fit comparatively to the two-factor communion-by-agency model, in all of the countries. These results suggest cross-cultural stability for both the Big Two model and the Big Two facets model, and supported Abele and colleagues' (Abele et al., 2016, p. 3) expectation that the different labels used for the dimensions in previous literature (e.g., warmth vs. morality, Fiske et al., 2002; Wojciszke, 1994) were not entirely spurious, but reflected a meaningful distinction hidden within the trait content used to represent each Big Two dimension.

Although Abele and colleagues (Abele et al., 2016) only provided empirical evidence for the Big Two facets model at the self-perception level, new evidence emerged in the meantime supporting the extension of this model to the domain of stereotypes (Hentschel, Heilman, & Peus, 2019). Moreover, there is a substantial body of evidence accruing in the literature lending support to the multi-faceted nature of the Big Two in domains such as person perception, interpersonal status perception, and group perception (see above). Only further research will be able to ascertain the boundaries of the applicability of the Big Two

facets model across multiple sub-domains of social perception. As it stands, however, the evidence favoring the existence of the Big Two facets is sufficiently compelling to draw the attention of contemporary research on social perception. This can be seen, for instance, in a more recent definition of the dimensions of the warmth-by-competence model that takes into account their multifaceted nature (Fiske, 2018). At the expense of simplicity, the facet model constitutes a more accurate description of the trait content of the Big Two, and promises to amend for the definitional confusion regarding these dimensions in previous literature.

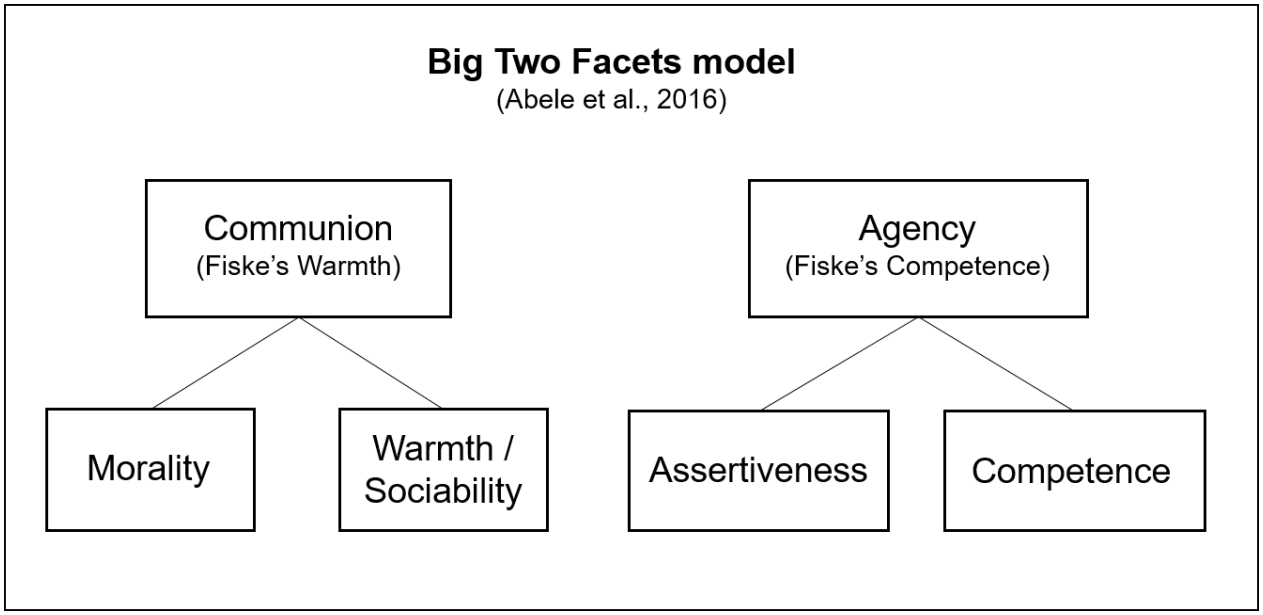

Figure 2. Diagram illustrating the structure of the Big Two facets model proposed by Abele et al. (2016).

This section concludes our review of the most influential models describing the fundamental dimensions of psychological content underlying person perception. In the next chapter we shift the focus to the domain of social face perception, and review some of the literature concerned with the question of how personality is inferred from facial appearance alone.

Chapter II. Inferring Personality from Faces

In this chapter, we start with a clarification of the role that facial appearance plays in impression formation. Subsequently, we review the work underlying the two-dimensional model of personality inferences from faces, as well as the work that provided the basis for the interpretation of this model's dimensions. In the last two sections of the chapter, we additionally provide a short review of one of the methodological approaches (i.e., reverse correlation) we employed to examine our questions throughout the thesis, after a brief clarification of the theoretical assumptions underlying it. These reviews are deemed important for a better understanding of the terminology and concepts associated with our research questions and methodologies.

The Role of Facial Appearance in Personality Impressions

In our social environment, few person features can rival the face as the targets of our attention. Not only the human face is ubiquitous across most social interactions, but it is also a highly efficient medium for social communication in that it rapidly communicates relevant information about a person. A mere glimpse of a face allows us to identify the person, ascribe the person to multiple social categories (e.g., gender, age, ethnicity; Bodenhausen & Macrae, 2006; Mason, Cloutier, & Macrae, 2006; Zebrowitz, 2006), or learn about the person's emotional state (Darwin, 1872; Ekman & Oster, 1979; Horstmann, 2003). But perhaps the most fascinating aspect of face inferences is how perceivers rely on facial features to infer unobservable dispositional qualities such as personality traits (Secord, 1958; Todorov et al., 2015; Zebrowitz, 2017) in the absence of any other information about the target, and regardless of accuracy in doing so (e.g., Olivola & Todorov, 2010b).

The idea that the face provides clues about someone's character is an old one. It can be traced back to classical Greek scholars such as Aristotle, who argued in his treatise *Physiognomica* that any physical resemblance between humans and animals reflected the character associated with those animals (Todorov, 2017; Zebrowitz, 1997). This belief that the face can act as a "window to the soul"—as Leslie Zebrowitz puts it in her influential book on social face perception (Zebrowitz, 1997)—achieved its peak of popularity by the end of the 18th century with the emergence of physiognomy—a pseudoscience that aimed to establish direct relationships between facial appearance and human character and ability, by assuming

that facial appearance is just as diagnostic of personality as behavior is (for a review see Todorov et al., 2015; Todorov, 2017). The most influential proponent of physiognomy was the pastor and physician Johann Kaspar Lavater, who developed a formal set of rules specifying associations between facial features and character traits. Physiognomists claimed that with enough training, people were able to predict real character traits based on facial appearance alone. The widespread acceptance of physiognomy within the intellectual elites of the 19th century led to real-world consequences. Cesare Lombroso, the founder of criminal anthropology, relied on physiognomy's ideas to testify in court during criminal trials. The impact of these ideas was such that Charles Darwin was almost refused to depart on his highly influential Beagle voyage, merely due to the skepticism that his facial features raised on the Beagle's captain regarding Darwin's determination to endure the trip (Todorov, 2017). Fortunately, as we know, Darwin proved the captain wrong.

Although physiognomy is nowadays discredited as a science, its main premise that facial appearance is predictive of underlying personality features reflects, nevertheless, how social perceivers rely on facial appearance to subjectively make sense of each other. In support of this assertion, a brief survey conducted by Hassin and Trope (2000) using a representative sample of Israelis, suggested that approximately 75% of that population believed that it is possible "to know an individual's true personality traits from looking at his or her face" (Hassin & Trope, 2000, p. 837). Such a reliance on facial cues in inferences of personality occurs regardless of any actual accuracy or reliability (e.g., see Olivola, Eubanks, & Lovelace, 2014; Olivola & Todorov, 2010b). Moreover, the reliance on judgments of facial appearance during social decision making appears to be more strongly driven by the ease of accessibility and processing of facial cues than by beliefs in the validity of such judgments (Jaeger, Evans, Stel, & van Beest, 2019).

At this point, we must clarify that the focus of the present thesis is on subjective perceptions of personality regardless of their accuracy. However accurate people may be, what the evidence in the literature clearly shows is that people regularly engage in personality assessments based on facial appearance alone, and that these judgments dynamically interact with people's behavior, thereby leading to real-world consequences in a variety of situations (e.g., Ballew & Todorov, 2007; Hassin & Trope, 2000; Jaeger et al., 2019; Livingston & Pearce, 2009; Olivola & Todorov, 2010a; Zebrowitz, Andreoletti, Collins, Lee, & Blumenthal, 1998).

Given the impact that these judgments can have on our lives, it becomes important to understand whether some personality traits are more relevant than others on how we form impressions of personality based on facial appearance. What sort of information are perceivers primarily looking for in a face? Next, we review the research that aimed to identify the traits lying at the core of personality inferences from faces.

The Trustworthiness-by-Dominance Model of Social Face Perception

There are many traits that one can infer from the face alone. Simply due to the large number of possibilities offered by a language's lexicon, different people may use different trait-adjectives to describe their personality impression based on someone's face. But, despite of the variability in the communication of one's own impressions, could there be any more fundamental information underlying our impressions of personality based on facial appearance?

Oosterhof and Todorov (2008) systematically examined this question in their seminal work on the basic dimensions of social face perception. In a number of studies, these authors aimed not only to identify the basic dimensions underlying personality attributions based on facial appearance, but also to identify the facial features that predicted inferences along the continuum of each of those dimensions. Importantly, unlike the theory-driven approach taken by the authors who proposed the warmth-by-competence, communion-by-agency, and Big Two facets models (i.e., who selected traits on the basis of previous models and preferred theoretical framework), Oosterhof and Todorov (2008) opted for an inductive, or data-driven, approach. In their approach, these authors started by collecting a large amount (above 1100) of unconstrained person descriptions of a set of male and female emotionally neutral face photographs with direct eye-gaze—from the Karolisnka face database (Lundqvist, Flykt, & Ohman, 1998). The content of those descriptions was then classified into a smaller number of 14 trait categories by three independent judges. Next, they asked another sample of participants to rate the exact same set of faces on the resulting 14 trait categories and on dominance—a trait added by the authors after they acknowledged its relevance in prior research on interpersonal perception (Oosterhof & Todorov, 2008, p. 11091). These ratings were subsequently submitted to a principal component analysis (PCA): a dimensionality-reduction statistical procedure that allowed the authors to convert a large set of potentially highly correlated (i.e., redundant) trait ratings into a smaller set of orthogonal variables known as principal components that are still able to account for most of the variability in the

original data. By applying this technique, the authors were able to identify the minimum amount of dimensions (i.e., principal components) sufficient to parsimoniously account for most of the variability in trait ratings of facial appearance.

Their results indicated that two principal components were sufficient to account for slightly more than 80% of the variability in their ratings' data. The first principal component differentiated between positive (e.g., caring, sociable) and negative traits (e.g., aggressive, mean), and accounted for the highest portion of variance (i.e., 63.3%) in their data. For this reason, they interpreted this principal component as a general valence dimension. The traits that exhibited the highest correlations with the second principal component were dominance, confidence, and aggressiveness, which led the authors to interpret this principal component as a dimension of dominance evaluation (Oosterhof & Todorov, 2008, see Table S3 of this article's Supporting Information). Importantly, this second dimension accounted for a smaller portion of the variance (18.3%) in their data, compared to the first dimension of general valence. (This apparent imbalance will be one of our main topics of discussion in Chapter VIII.) Finally, the authors found that the judgments more highly correlated with the first and second principal components were trustworthiness and dominance, respectively (Oosterhof & Todorov, 2008, see Figure S2 of this article's Supporting Information). The close overlap of these judgments with the two dimensions (or called principal components in the output of a PCA) led researchers to refer to them as trustworthiness and dominance, although Oosterhof and Todorov (2008) and other researchers (e.g., Jones et al., 2018) also refer to them as valence and dominance, or as goodness and power (Todorov, 2017). Put simply, what the seminal work by Oosterhof and Todorov (2008) is suggesting is that when perceivers engage in personality inferences from the facial appearance of unfamiliar others, they are essentially evaluating how trustworthy and how dominant someone appears to be, regardless of how accurately these perceptions reflect actual behavior. This two-dimensional model of social face perception has since become a cornerstone in social face perception research.

These dimensions were subsequently replicated by Sutherland and colleagues (Sutherland et al., 2013) who used a larger set of highly variable and more naturalistic face stimuli known as *ambient images* (i.e., images intended to be highly representative of faces as they occur in naturalistic daily life settings; see Jenkins, White, Van Montfort, & Mike Burton, 2011). In doing so, Sutherland and colleagues (Sutherland et al., 2013) offered stronger support for Oosterhof and Todorov's (2008) model, while additionally adding to its external validity. In addition, it is noteworthy to mention that Sutherland and colleagues

(Sutherland et al., 2013) found a third additional dimension which captured the variance of age and attractiveness judgments (i.e., youthfulness-attractiveness dimension). Although age and attractiveness are undoubtedly dimensions that carry a great weight in facial first impressions, they do not reflect inferences of unobservable personality features to the extent that the trustworthiness and dominance dimensions do. Instead, this third dimension reflects more accurate inferences based on more tangible links between physical features and social categories. For this reason, and given the main goal of the present thesis of examining and describing the relationships between dimensions that reflect the inference of unobservable personality features, our focus will be placed on Oosterhof and Todorov's (2008) two-dimensional trustworthiness-by-dominance model.

A logical follow-up question is then *why* would trustworthiness and dominance in particular occupy a central role in social face perception? From a functional and evolutionary perspective, these dimensions are thought to reflect a primary concern with assessing someone's intentions (i.e., trustworthiness) and someone's ability to implement their intentions (i.e. dominance) (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Todorov et al., 2015). Together these pieces of information ultimately allow the perceiver to assess the degree of threat someone might represent (e.g., "how much harm can this person potentially inflict in a social interaction?"), and ultimately drive approach-avoidance behavior (Oosterhof & Todorov, 2008). In other words, these dimensions may be reflecting an underlying threat detection mechanism, presumably shaped throughout our evolutionary history. This functional perspective on the dimensions of facial impressions overlaps with the functional interpretation of the warmth and competence dimensions by Fiske and colleagues (e.g., Fiske et al., 2007), and constitutes one of the arguments supporting an overlap between the dimensions of person perception and the dimensions of facial impressions. Whether or not the dimensions of person perception overlap with the dimensions of social face perception, is one of the main questions that we address in this thesis (see Chapter VI).

Resemblance with Person Perception Models

The most noteworthy aspect of the finding of a two-dimensional structure underlying face impressions is its remarkable resemblance to previous well-established models of interpersonal and intergroup perception, including the social-by-intellectual model of person perception (Rosenberg et al., 1968), or the more influential warmth-by-competence

(Stereotype Content Model; Fiske et al., 2002) and communion-by-agency (Abele & Wojciszke, 2007) models.

It is important, at this point, to make a distinction between these models in regard to the methodology employed to support them. Models like the warmth-by-competence, communion-by-agency, and the Big Two facets models, were all supported by trait sets selected on the basis of previous models and preferred theoretical frameworks of the authors. For instance, the warmth-by-competence model (Fiske et al., 2002) was largely derived from the social-by-intellectual model (Rosenberg et al., 1968); in a similar vein, the communion-by-agency model (Abele & Wojciszke, 2007) was derived from the trait content of the warmth-by-competence and social-by-intellectual models, although also from additional traits guided by theory; and finally, The Big Two facets model (Abele et al., 2016) was largely derived from studies using the communion-by-agency framework. In this sense, all of these models followed a theory-driven approach, in that they all predicted a two-factor solution to emerge in their analyses, and all used traits that can be traced back to the seminal work supporting the social-by-intellectual model. As a result, most of these models tend to overlap in their trait content, and their differences (e.g., dimension definitions and interpretations) tend to result from the different theoretical backgrounds associated with them. By contrast, both the social-by-intellectual model of person perception and the trustworthiness-by-dominance model of facial impressions followed a (predominantly, see below) data-driven approach. Although the social-by-intellectual model relied on traits used on previous research (e.g., Asch, 1946), the methodology may be understood as data-driven in the sense that Rosenberg and colleagues (Rosenberg et al., 1968) were faced with the task of identifying the optimal number of dimensions—unknown in the literature at the time—to describe their results, and conducted additional studies to interpret these dimensions. Moreover, in subsequent studies by Rosenberg and Sedlak (1972), the social-by-intellectual model received further support from a more clearly data-driven methodology, where the authors started by extracting their trait set from spontaneous person descriptions, similarly to the first step of the approach taken by Oosterhof and Todorov (2008) for facial impressions. Of relevance to the present work, we should note that the models obtained via a data-driven approach were the ones arriving at different conclusions regarding the traits that best represent the dimensions. Given that a similar approach underlies the social-by-intellectual and the trustworthiness-by-dominance models, their seemingly divergent dimensions are likely to be related with the modality of the target stimulus (i.e., verbal person descriptors vs. faces) focused by each

model. Nevertheless, the traits representing the correspondent dimensions between these two data-driven models (i.e., social vs. trustworthiness; intellectual vs. dominance) still exhibit a remarkable resemblance in terms of their functional interpretation, despite resulting from different classes of stimuli. That is, in both models the traits are relevant for intentionality and ability appraisal. What remains to be clarified, however, is the extent to which these traits are perceived as redundant by perceivers themselves, or instead, if any distinctions in their perception play an important role in guiding social behavior.

A second clarification should additionally be made regarding the approach taken by Oosterhof and Todorov (2008). Although these authors took a data-driven approach, they were at some point also guided by two-dimensional models of person perception in their search for the dimensions of facial impressions. In this sense, their approach can be said to have been *predominantly* data-driven. For instance, the trait dominance did not emerge in the content of the spontaneous descriptions of faces obtained in the first step of their methodology. Instead, the decision to include dominance in the set of traits that was ultimately submitted to the PCA that supports the trustworthiness-by-dominance model, was based on the dominance-by-affiliation circumplex model of interpersonal perception previously found by Wiggins and colleagues (Wiggins, 1979; Wiggins, Trapnell, & Phillips, 1988)—considered to be akin to the warmth-by-competence or communion-by-agency models in the literature (see Abele & Wojciszke, 2014). In this sense, dominance can be said to have been obtained in a less data-driven fashion comparatively to the other traits that made it into the final set of candidate dimensions.

This is not to say that dominance is not a central dimension in facial impressions. Both the higher correlation of dominance with the second principal component in Oosterhof and Todorov's (2008) results, and the conceptual replication of the trustworthiness-by-dominance model by Sutherland and colleagues (Sutherland et al., 2013), do suggest that dominance is a core dimension in facial impressions. We must note, however, that Sutherland and colleagues' (Sutherland et al., 2013) conceptual replication was theory-driven (vs. data-driven), and relied heavily on the traits selected by Oosterhof and Todorov (2008). Thus, one could wonder whether only valence/trustworthiness would have been found as the core dimension of facial impressions, or whether any other dimension would emerge as the second (e.g., confidence or aggressiveness), if dominance would not have been included in the trait set. Nevertheless, and more importantly to our goals (see Chapter VI), the fact is that competence did not emerge as the second dimension of facial impressions, as it did in the person perception domain.

The Trustworthiness-by-Dominance Face Space

After identifying the basic dimensions of social face perception, Oosterhof and Todorov (2008) moved on to identify the configurations of diagnostic facial features for each of the two dimensions, with the aim of revealing how these personality dimensions are physically represented. To do so, they relied on a data-driven methodology built upon the idea of a multidimensional face-space model. The idea was introduced by Valentine (1991; see also Valentine, Lewis, & Hills, 2016). In a face-space each point corresponds to a face, and the distance between each face reflects their degree of similarity, with higher spatial proximity representing higher similarity. A face-space can have any number of dimensions, each representing a given variation on a face, which can be either a single parameter (e.g., eye size, jaw width) or global face properties (e.g., age, masculinity). It is also assumed that faces are normally distributed along each dimension. A face reflecting the average on all dimensions (i.e., the most typical-looking face on that space) is located at the point of origin of the *n*-dimensional face-space where all dimensions converge. The further away a face is from the origin the more exaggerated it becomes on a given dimension (e.g., smaller eyes at -3 *SD*s away from average face, bigger eyes at +3 *SD*s away from average face; or younger at -3 *SD*s and older at +3 *SD*s), and thus more distinctive (see Figure 3 for an example of a face space).

Oosterhof and Todorov (2008) tailored this methodology to the context of social face perception to identify the changes in physical features of the face that maximally predict changes in trustworthiness and dominance judgments. To do so, they asked participants to judge the trustworthiness and dominance of a large set of computer-generated faces created with the FaceGen face modelling software (Singular Inversions, 2005), which in turn is based on a statistical face model with more than 50 parameters of facial shape and reflectance. By statistically deriving the relationships between the trait ratings and the facial shape attributes of the target faces, they were able to obtain statistical face models for the trustworthiness and dominance dimensions. These models allow to visualize how changes in facial appearance track changes in judgments on each of the two dimensions.

Subsequently, both models were used to randomly generate 300 novel emotionally neutral faces for which judgments of trustworthiness and dominance were obtained. Because a small negative correlation was found between trustworthiness and dominance judgments, Oosterhof and Todorov (2008) orthogonalized the dimensions by rotating the dominance axis in relation to the trustworthiness axis (so they became related by a 90° angle), for the sake of

theoretical utility. This resulted in a two-dimensional face-space defined by the orthogonal axes of trustworthiness and dominance. Additionally, they derived morphable facial models for each dimension, which enabled them to manipulate the extent to which newly generated faces conveyed each dimension (e.g., highly untrustworthy face at -4.5 *SDs* away from average face, and highly trustworthy face at +4.5 *SD*s away from average face, see Figure 3) (Oosterhof & Todorov, 2008). The two statistical face models were then validated by trustworthiness and dominance judgments of novel faces generated with them.

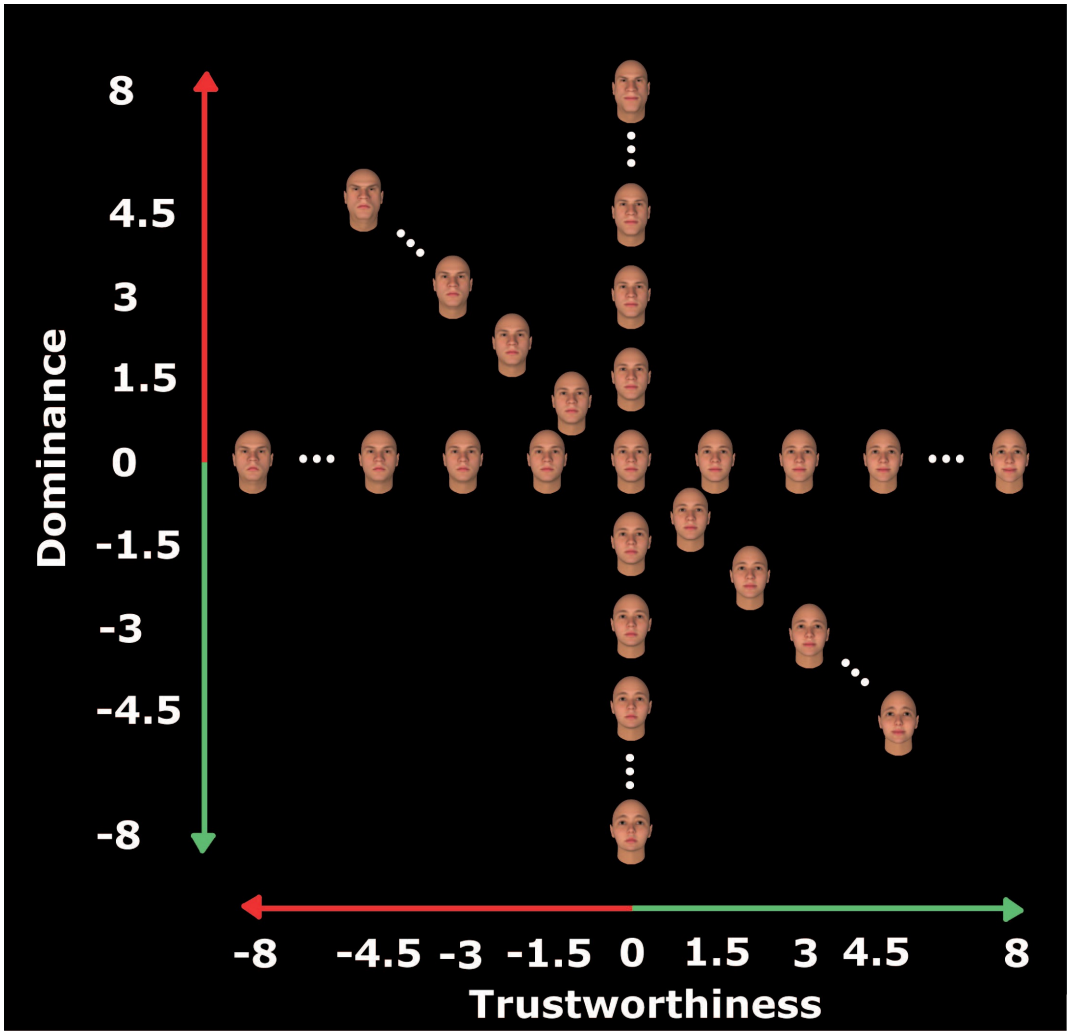

Figure 3. Example of the two-dimensional face space defined by the orthogonal axes of trustworthiness and dominance. The synthetic faces along these axes vary along the trustworthiness and dominance dimensions. The threat dimension is shown in the diagonal of the face space. Values represent the distance of a given face from the average face at the point of origin (in standard deviation units). Image extracted from Oosterhof and Todorov (2008; Figure S9 in their Supporting Information).

In subsequent studies, Oosterhof and Todorov (2008; see also Todorov, Said, Engell, & Oosterhof, 2008) aimed to interpret the diagonals of the trustworthiness-by-dominance two-dimensional face-space. To do so, they generated faces for the four quadrants of the face-space by averaging the trustworthiness and dominance dimensions. After submitting these new faces to several judgments, they found that threat judgments were the most highly correlated with changes in the faces representing the diagonal running from the "high dominance-low trustworthiness" quadrant to the "low dominance-high trustworthiness" quadrant (see Figure 3). Specifically, the less trustworthy and the more dominant the face looked, the more threatening it was perceived to be. In turn, judgments of attractiveness, likeability, and competence tracked variations in the faces representing the remaining diagonal, such that the more trustworthy and dominant the faces looked, the higher they were rated in those three dimensions (Todorov et al., 2008). However, as we will demonstrate in Chapter VIII, this approach assumes a linear and equal weighing of the two dimensions—since the synthetic faces of the face-space diagonals were averages—that may not fully capture how perceivers naturally integrate the two dimensions, especially when we take into account that the trustworthiness dimension accounts for more variance in face judgments. Nevertheless, Oosterhof and Todorov's (2008) finding of a threat dimension is highly informative regarding the functional role of the two dimensions.

Next, we turn our focus to the meaning of the facial content identified by this methodology.

Facial Cues of Trustworthiness and Dominance and their Meaning

The identification of the facial content that is predictive of judgments on the core traits of face impressions is relevant in that it provides clues to interpret the meaning of the two dimensions. In additional studies, Oosterhof and Todorov (2008) found that trustworthiness judgments tracked variations in structural resemblance to emotional expressions. Extremely untrustworthy faces were perceived as angry, whereas extremely trustworthy faces were perceived as happy. In turn, they found that dominance judgments tracked variations in emotional expression to a lesser extent. Extremely submissive faces were perceived as fearful. And although they found that the classification of extremely dominant as angry was not above chance level, Oosterhof and Todorov (2008) suspected this resulted from their (forced) orthogonalization of the two dimensions, given past research demonstrating an association between high dominance and anger perceptions (e.g., Knutson, 1996). However, they did find

that dominance judgments tracked judgments of masculinity and facial maturity (or *babyfacedness*), and interpreted this as suggesting that dominance judgments are related with perceptions of physical strength and power (Oosterhof & Todorov, 2008). The relationship between perceived dominance and perceived physical strength from faces gained empirical support in subsequent studies conducted by Toscano, Schubert, and Sell (2014) who found a positive correlation between dominance and physical strength judgments of computer-generated and naturalistic images of male faces. In addition, these authors identified some of the facial features that were predictive of both dominance and physical strength judgments from male faces, namely: brow height, chin length, eye length, mouth width and nose width (Toscano et al., 2014).

Regarding the association between these dimensions and emotional expressions, it must be noted that all the target faces in these studies were manipulated to be emotionally neutral, in the sense that they were expressionless. However, the structural properties of an expressionless face can still resemble the pattern of a given emotion and be perceived as if it is expressing it. Zebrowitz and colleagues (Montepare & Dobish, 2003; Zebrowitz, 1997; Zebrowitz, Kikuchi, & Fellous, 2010; Zebrowitz & Montepare, 2008) hypothesized that this reflects an *emotion overgeneralization* mechanism, through which emotional states are perceived as a result of our preparedness to adaptively respond to emotional expressions, at the expense of accuracy. Such a lowered threshold to detect emotions in others is adaptive in the sense that failing to detect emotions in others (i.e., false negatives) may be more costly to the perceiver than inaccurate inferences of emotional states (i.e., false positives).

From the perspective of ecological theory (McArthur & Baron, 1983; Zebrowitz & Montepare, 2008), facial expressions of emotion not only communicate a person's internal affective state, but are also informative about her potential behavior (e.g., Darwin, 1872; Knutson, 1996). For instance, emotional displays can signal intentions of approach, avoidance or attack. When someone smiles at us, that person may be signaling an intention to cooperate with us (Frank, 1988). Likewise, if someone's face expresses anger towards us, the person may be signaling an intention to harm us. In other words, emotional expressions have a social predictive function in that they allow perceivers to predict each other's behaviors during present and future social interactions. Although emotional states are temporary, perceivers show a tendency to extrapolate transient expressions of emotions to more stable personality traits of a given target person, regardless of accuracy—a process that Secord (1958) referred to as *temporal extension*. In sum, the emotion overgeneralization hypothesis and the process

of temporal extension provide a potential explanation for how people make personality attributions based on facial appearance alone. At the same time, they highlight the role of emotional expressions in personality impressions. These hypothesized mechanisms make it less surprising that variations in how faces convey trustworthiness track variations on their perceptual similarity to emotional expressions.

Of importance to the present thesis is that, despite of being often described as independent dimensions for theoretical convenience, both dominance and trustworthiness share a relationship with valence. Face evaluation always seems to entail a judgment about the positivity or negativity of the stimulus, which appears to be informed by emotional cues. This pervasiveness of valence in facial impressions further reflects Osgood and colleagues' (Osgood et al., 1957) findings showing that evaluation (i.e., valence) is one of the most fundamental dimensions of semantic meaning, since the perceived valence of an object (e.g., social stimulus such as a face) has the ability to guide behavior (e.g., positivity fosters approach, while negativity fosters avoidance; Chen & Bargh, 1999; but see also Rotteveel et al., 2015).

The literature reviewed so far suggests that dominance and trustworthiness establish distinct relationships with valence, that differ both in direction and strength. Trustworthiness judgments are more strongly related with general valence and account for a higher proportion of the variability in trait inferences from faces. And while dominant-looking faces resemble negative emotions, trustworthy-looking faces resemble positive emotions. In turn, this may be contributing to their small but, nevertheless, negative correlation (see Oosterhof & Todorov, 2008). Despite of this, little attention has been placed into a closer examination of the features of the relationship of valence with the two dimensions of social face perception, or how the relationship established between valence and these dimensions impacts how they are integrated into unitary impressions of facial appearance. In this thesis, we aimed to fill this gap by directly examining the nature of the relationship between valence and each of the basic dimensions of social face perception (see Chapter VII), and clarifying how the strength of the relationship between a trait dimension and valence impacts how dominance and trustworthiness are integrated into unitary face impressions (see Chapter VIII).

Link between Facial Appearance and Conceptual Knowledge

The literature reviewed so far suggests that tangible features of the face (e.g., emotional expressions or mere resemblance to them, or gender and age) have a clear influence on how perceivers infer latent person features such as personality, regardless of how accurate these inferences may be. Tangible configurations of facial features resembling emotional signal seem to be relied upon by the perceiver to predict others' potential behavior, and ultimately influence personality trait attributions (e.g., through a process of temporal extension; Secord, 1958). This seems to suggest that perceivers develop associations between low-level physiognomic information (i.e., facial features) and high-level conceptual knowledge about personality traits (see Over & Cook, 2018). Through these associations, certain patterns or configurations of facial features would lead to particular personality judgments. This leads to the question of whether the same facial information leads to similar personality judgments across different perceivers. In other words, to what extent do different perceivers infer the same personality traits from the same target face?

The identification of facial features that lead to certain trait judgments hinges on earlier findings showing that perceivers tend to highly agree in their impressions based on the same target face (Secord & Bevan, 1956; Todorov et al., 2015). Together with findings showing that this consensus extends across different cultures (Rule et al., 2010; Zebrowitz et al., 2012), and that both male and female perceivers use the same facial information in face-driven personality judgments (Oh, Dotsch, Porter, & Todorov, 2019), this suggests that personality judgments from faces may be emerging from a mental process shared by all social perceivers (see also Jones et al., 2018, for ongoing work on the universality of the trustworthiness-by-dominance model).

Another important question to consider, in light of the aims of the present thesis, relates with how low-level physiognomic information (e.g., facial features) is associated, and interacts, with high-level conceptual features (e.g., conceptual knowledge about traits). An answer to this question requires some level of integration of social face perception research with person perception research. However, bearing in mind that these domains developed in parallel throughout the second half of the last century, only more recently researchers began to look into how these domains are related (Bodenhausen & Macrae, 2006; Hassin & Trope, 2000; Over & Cook, 2018; Quinn & Macrae, 2011; Stolier, Hehman, Keller, Walker, & Freeman, 2018; Sutherland, Oldmeadow, & Young, 2016; Zebrowitz, 2006). As a result, there is still a relatively small, but steadily growing, understanding of how bottom-up facial

information interacts with top-down conceptual information to produce impressions of personality integrating both pieces of information. One important step in that direction was taken in the work by Hassin and Trope (2000) showing that conceptual knowledge about personality and facial information influence each other. These authors made a distinction between reading personality "from faces" (RFF) and reading personality "into faces" (RIF). Whereas RFF reflects the act of inferring personality based on a target's facial appearance, RIF reflects the act of "mentally visualizing" facial appearance based on verbal personality information. Taking that distinction into account, Hassin and Trope (2000) demonstrated that, not only facial information influences how one perceives the target person's personality, but verbal information about the target person's personality (in the absence of a face) influences the expected facial appearance of that person. The fact that people are able to mentally visualize a face conveying some personality feature, suggests the existence of an associative network where facial information establishes connections with trait concepts in semantic memory. This top-down bottom-up interactivity is taken into account in more recent models of person construal such as the Dynamic Interactive Model (Freeman & Ambady, 2011). Simply put, this model describes how our perceptions of a person progressively emerges from a dynamic interaction between low-level information (e.g., facial, bodily, or vocal cues) and higher-level information (e.g., social categories, traits, stereotypes, high-level cognitive states). Of importance to our current aims, is that this model and the empirical work supporting it (see Freeman & Ambady, 2011) lay out predictions about how low-level facial cues are associated and interact with high-level conceptual trait knowledge. In line with this model, inadvertently or not, most studies involving personality inferences from faces tend to assume that, when perceivers are judging facial appearance on some target trait, they are matching the visual input (i.e., face stimulus) with some internal representation of a trait concept (e.g., Dotsch & Todorov, 2012; Imhoff, Woelki, Hanke, & Dotsch, 2013). For instance, most of us are able to imagine what a friendly (or unfriendly) face looks like, and it is very likely that we would agree on the features exhibited by such a face. Our ability to evoke such information highly suggests that our internal representation of the concept of friendliness is linked to specific visual information about facial appearance. It should, therefore, be possible to capture these "mental templates" about facial appearance by asking perceivers to describe them.

However, such a task would prove difficult for the perceiver, for at least two reasons. First, it is unclear what constitutes a facial feature (e.g., is it the whole nose, the tip of the

nose, the width of a nostril?) or whether language includes labels to communicate them. And second, perceivers themselves may not even be aware of which features, or combinations thereof, they are relying upon to make the judgment (see Brinkman, Todorov, & Dotsch, 2017; Todorov, Dotsch, Wigboldus, & Said, 2011). Moreover, even if researchers rely on a theory-driven approach to select and manipulate a finite set of facial features with the aim of observing of how they correlate with social judgments, they still cannot be entirely sure that those features are indeed the strongest predictors of the target social judgments. Another approach could be to greatly increase the sample of face stimuli: As the sampling space increases to infinite, the more it represents the variability found under natural settings, thus allowing researchers to capture all sources of variability in faces that predict social judgments. However, the high number of features of the human face rapidly leads to a combinatorial explosion that poses an intractable challenge to researchers, since experimental time and resources are finite (see Jack & Schyns, 2017). But there is an alternative approach that allow researchers to circumvent the aforementioned methodological challenges: reverse correlation methods.

Reverse Correlation Methodology

Reverse correlation (RC) is a data-driven method developed within the domain of auditory cognition (Ahumada & Lovell, 1971) that was later extended to other domains such as vision, neurophysiology, and more recently to social face perception (for reviews see Brinkman et al., 2017; Jack & Schyns, 2017). The main feature of RC paradigms is that they *reverse* the statistical relationship between the stimulus and the response. In more conventional paradigms, researchers place the focus on how responses vary as a function of theory-driven manipulations of a given target stimulus' features, and subsequently analyze how the responses correlate with those fixed stimulus' features. In RC paradigms, on the other hand, the variation of the stimulus' features is random, while the response variable is fixed. In this sense, RC paradigms are data-driven, since researchers make no *a priori* assumptions about which stimulus' features are associated with changes in the response variable. Instead, researchers let each perceiver "specify" the signal in the stimulus (i.e., facial features) that elicits the target judgment. By doing so, researchers sidestep making decisions about the colossal amount of possible feature combinations that may be underlying the target judgment, while still being able to capture the stimulus' features that elicit the judgment, including those that might have not been known to be relevant.

Within the social face perception domain, RC has proven to be a useful tool for researchers interested in visualizing internal representations of facial content associated with trait concepts (Dotsch & Todorov, 2012; Oosterhof & Todorov, 2008), or in how these representations are modulated by top-down biases (e.g., Dotsch, Wigboldus, Langner, & Van Knippenberg, 2008; Dotsch, Wigboldus, & van Knippenberg, 2011; Imhoff, Dotsch, Bianchi, Banse, & Wigboldus, 2011; Ratner, Dotsch, Wigboldus, van Knippenberg, & Amodio, 2014). Most studies in social face perception have relied on two variants of RC methods (for a review see Todorov et al., 2011): a noise-based psychophysical RC method known as the classification image technique (e.g., Dotsch & Todorov, 2012; Gosselin & Schyns, 2003; Imhoff et al., 2013; Mangini & Biederman, 2004), and the face space based RC approach[1] (e.g., Oosterhof & Todorov, 2008; Toscano, Schubert, Dotsch, Falvello, & Todorov, 2016; Toscano et al., 2014; Walker & Vetter, 2009). Despite their differences (noise-based vs. face space based, see Todorov et al., 2011), both techniques were shown to converge on their output regarding systematic relationships between facial content and social judgments (see Dotsch & Todorov, 2012). Here, however, we will place our focus on the noise-based classification image technique since the empirical work presented in this thesis relied on it to address our research questions (in Chapters VI and VIII).

The implementation of the classification image technique starts with the construction of the stimulus set. First, researchers select a base face image. For instance, in their studies, Dotsch and Todorov (2012) used as their base image the average male face[2] of the Karolinska face database (Lundqvist et al., 1998). In the next step, random visual noise (e.g., sine-wave-based, Gabor noise, or white noise; for an illustration see Figure 3 in the article of Brinkman et al., 2017) is superimposed on the base image in order to create a large amount of images reflecting variations of the base image. These images will constitute the stimulus set. The noise superimposed on each image distorts the pixels of the base face image at random locations. In this sense, the stimulus set can be understood as a large amount of random variations of the same face. This stimulus set is then typically submitted to perceptual

[1] Although we did not present the face space based approach used by Oosterhof and Todorov (2008) as a reverse correlation method, for the sake of readability, it is considered as such (Todorov et al., 2011), given that the stimulus set used in that approach is randomly generated based on a statistical face space model, and subsequently submitted to trait judgments in order to reveal systematic relationships between facial content and trait concepts.

[2] One reason to rely on an averaged face is that it retains only the features shared across all the several faces with different identities from which it was derived, thus approximating the base face as close as possible to a prototypical face (in this case, a male face). Another advantage of using averaged face images as base faces is that these are typically blurred, which in turn facilitates their blending with the superimposed noise. Moreover,

classifications by participants in a two-image forced-choice task (2IFC) (for other variations of this task, see Brinkman et al., 2017). On each trial of the 2IFC task, participants are presented with a pair of face images (previously set to exhibit maximally distinct noise patterns) of the stimulus set, and are asked to select the one that most resembles a person possessing a given target personality trait (e.g., "Which face looks more Dominant?"). The target judgment is fixed throughout the whole task, and the number of trials typically ranges between 300 and 1000 in social face perception studies (e.g., Dotsch & Todorov, 2012; Imhoff et al., 2013). Once the data is collected, the noise patches of the images selected by participants as containing the target trait signal (e.g., dominant) are averaged at the group-level (although it can also be averaged at the individual level, depending on the research question). The averaged noise is then re-superimposed in the base face image, thus producing a classification image (CI): a face image whose features maximally predict the target trait judgment. When computed at the group-level, these CIs average out all individual differences regarding the facial content expected to be associated with the trait concept, thus retaining only the facial content on which most participants agree upon. CIs are usually understood as visual approximations to the visual component of the internal representations of a trait concept. Whatever mental template the perceiver may have used to select the face images throughout the 2IFC task, it should be captured by the CI to a certain extent. However, because researchers are not immune to biases (e.g., confirmation bias) when they interpret the CIs, the obtained CIs are usually validated by another sample of participants, independent from the one used to generate the CI. For instance, if a CI was generated for the trait "Dominant", it is then judged on dominance, and possibly other traits, by the new sample of judges, so the researchers can ascertain that the CI conveys the intended trait signal.

But beyond allowing researchers to gain insight into the visual content of internal representations of trait concepts, the classification image technique further allows to tackle questions regarding the similarity between the visual content of different internal representations. For instance, the correlation between the pixel luminance values of two CIs generated for two different traits quantifies the extent to which these two CIs are objectively similar in terms of visual information. In turn, this serves as a proxy measure of the degree of overlap between the visual content of any two internal trait representations (e.g., see analytical approaches by Dotsch & Todorov, 2012; Imhoff et al., 2013; Oliveira, Garcia-Marques, Dotsch, & Garcia-Marques, 2019). This possibility offered by RC methods was

---

their blurriness makes them less detailed, which in turn decreases the amount of trials required in a noise-based

instrumental in answering some of the main research questions raised in this thesis (see Chapter V).

Interim Summary

Throughout this chapter we aimed to provide the basis for a more complete understanding of the origins of the social face perception model, the concept of a face space, the relevance of facial emotional cues in social face perception, and the methodological approaches that have been pivotal in this domain of research. These pieces of information are essential to understand the context within which we raised some of the questions examined in this thesis. For instance, both the concept of a face space and knowledge about reverse correlation methods facilitate the comprehension of our analytical and methodological approaches in the empirical work presented in Chapters VI and VIII. Moreover, knowledge on the link between emotional information and personality inferences from faces is helpful to better understand the role that valence plays in social face perception, and to better follow some of the points we raise in regard to our findings in the General Discussion section.

In the next chapter, we review the literature providing insight into the similarities and differences between the models of different domains, taking into account our goal of examining their relationship. While doing so, we will consider not only how the dimensions might be related between models (e.g., dominance vs. competence; trustworthiness vs. warmth), but also how the relationship established between dimensions within a model (e.g., warmth and competence) differs from that in the other model (e.g., trustworthiness and dominance).

RC task (for more details see Brinkman et al., 2017).

Chapter III. Relationships between Dimensions

In this chapter, we highlight the distinctions between the person perception and the social face perception dimensions. In doing so, we focused on the warmth-by-competence model of person perception for two reasons. First, because this model and the trustworthiness-by-dominance share the same functional interpretation of the dimensions: Warmth and trustworthiness are both interpreted as reflecting an initial assessment of someone's intentions, whereas competence and dominance are interpreted as reflecting a subsequent assessment of someone's ability to implement those intentions (Fiske et al., 2007; Oosterhof & Todorov, 2008). And second, because both the warmth-by-competence and the trustworthiness-by-dominance models focus on specific trait concepts as the ones that best represent the dimensions, in contrast with the scientifically devised constructs of communion and agency which do not spontaneously emerge in perceivers' descriptions of their impressions about others. Nevertheless, we will discuss throughout this first section of the chapter how the distinction between these dimensions across models are captured in the recent facet model of the Big Two.

In the second section of this chapter, we shift our focus to the relationship between the two dimensions *within* each model, while paying special attention to the role played by valence in shaping these relationships.

Dominance versus Competence

Both dominance and competence have been considered dimensions related with an ability to implement intentions (Fiske et al., 2007; Oosterhof & Todorov, 2008). Despite of having this feature in common, these traits exhibit differences at the conceptual and evaluative levels that are worth considering before assuming their interchangeability in the social perception domain. A first distinction that can be made is at the conceptual level. Dominance is a relational trait in the sense that it reflects a perception of how one is positioned in a social hierarchy, and also reflects the degree to which one is motivated to advance the self over others in a given context (e.g., Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013; Maner & Case, 2016). By contrast, competence is related with the possession of skills and talents that may be used to achieve certain tasks—often involving cognitive ability (e.g., traits relevant for intellectual activities; Rosenberg et al., 1968)—that do not

necessarily entail an underlying motivation to pursue social rank, nor require a relational context. For instance, the execution of a task does not necessarily require a relationship with others (e.g., Carrier et al., 2014; Maner & Case, 2016). A second distinction pertains to the valence of behaviors associated with these traits. Displays of dominance are typically associated with negative behaviors such as intimidation, coercion, or aggression. On the contrary, competence is associated with positively perceived behaviors such as providing solutions to problems, or sharing knowledge with others (Cheng et al., 2013; Henrich & Gil-White, 2001; Maner & Case, 2016). In evolutionary grounded theories of status attainment, dominance and prestige (i.e., competence-related) are contrasted as two routes to status attainment. Whereas status is attained in a forceful manner by dominant individuals, it is freely conferred to prestigious individuals by others who acknowledge or benefit from their talents (Cheng et al., 2013; Henrich & Gil-White, 2001). And a third distinction pertains to the valence associated with the facial content that conveys these two traits: Whereas dominant-looking faces structurally resemble negative emotional expressions (i.e., anger), the structural properties of competent-looking faces resemble positive emotional expressions (i.e., happiness) (Said, Sebe, & Todorov, 2009). This is also reflected in the small negative correlations found between dominance and valence/trustworthiness (Chen, Jing, & Lee, 2014; Oosterhof & Todorov, 2008) Lastly, in light of the recent facet model of the Big Two, one may notice that dominance and competence are also traits captured by the different facets of agency. Whereas dominance is an assertiveness-related trait, competence itself represents the the other agency-related facet (Abele et al., 2016).

These conceptual and evaluative[1] differences challenge the assumption that dominance and competence would overlap in either the person perception or social face perception models. Not only do these traits differ in their behavioral and physical expressions, as they also vary in the degree of potential threat that they convey to a perceiver's goals. Therefore, one should expect perceivers to differentiate between the two, despite their common association with ability.

[1] Here we refer to the "evaluative" meaning of a trait, in the sense of "valence associated with a trait" (e.g., Osgood et al., 1957; Rosenberg & Olshan, 1970).

Trustworthiness versus Warmth

Despite being both understood as intentionality-related dimensions, trustworthiness and warmth also exhibit some distinctions that captured the interest of social perception researchers. One distinction that can be made between these traits pertains to the extent to which each is informative about someone's moral character versus someone's sociability. Brambilla and colleagues (Brambilla et al., 2011 Study 1 pretest) found that whereas trustworthiness was perceived as a morality-related trait, warmth was perceived as a sociability-related trait. And in a clarification of why this distinction is important, Goodwin and colleagues (Goodwin et al., 2014) further demonstrated that whereas traits such as trustworthy, honest, or loyal were perceived as relevant for morality only, other traits such as warm, sociable, or agreeable were perceived as relevant for social warmth only, while there were also traits seen as relevant for both morality and social warmth (e.g., humble, kind, cooperative). Drawing on these findings, Goodwin and colleagues (Goodwin, 2015; Goodwin et al., 2014) proposed a branching of the warmth dimension (referring to Fiske and colleagues' model; Fiske et al., 2002, 2007) into a sub-dimension of morality and a sub-dimension of warmth/sociability, arguing that whereas morality was related with someone's intentions, warmth reflected instead an ability to attract supporters for one's own intentions. This branching would be consistent with previous findings showing that perceivers differentiate between morality and sociability information, in such a way that they attribute higher relevance to morality-related information, in both person (Brambilla et al., 2011) and group perception (Leach et al., 2007) contexts. Moreover, these findings are supportive of the branching of the communion dimension into its morality and warmth facets, in the recently proposed facet model of the Big Two (Abele et al. 2016).

But although these findings support a conceptual differentiation of trustworthiness and warmth, a distinction at the evaluative level (i.e., regarding valence) seems to be less likely, as they are both prosocial traits. The evaluative overlap between these traits was especially apparent in the domain of social face perception. In their seminal work, Oosterhof and Todorov (2008; see Table S3 in this article's Supporting Information) found that both trustworthiness (morality-related) and sociability (warmth-related) ratings of faces were highly correlated between themselves and with general valence. This suggests that trustworthy-looking faces are highly similar to sociable-looking faces (at least in their stimulus set), thus making it difficult for perceivers to differentiate between the two traits.

Thus far, our focus has been on the comparison of dimensions between models of different domains. The literature reviewed so far provides the basis for a better understanding of our empirical work in Chapters VI and VII. Next, we will turn our focus to how the two dimensions within each of these models relate with one another. While doing so, we will pay special attention to the literature suggesting that valence plays an important role in shaping the relationship between dimensions (e.g., Abele, Uchronski, Suitner, & Wojciszke, 2008; Suitner & Maass, 2008) as already became apparent in our review of the evaluative distinctions between dominance and competence—or lack thereof between trustworthiness and warmth.

Relationship of Dimensions within Models of Social Perception

Although two-dimensional models imply an independence between their constituent dimensions, this independence is seldom observed in the content of personality impressions or stereotypes. Research focusing on the relationship dynamics between the two dimensions (i.e., social vs. intellectual, warmth vs. competence, or communion vs. agency), in both the contexts of person and group perception, has been reporting inconsistent results throughout the past decades.

Sometimes, the relationship between the dimensions is positive, as found by Rosenberg and colleagues (1968) in their seminal work. A positive correlation between the social and intellectual dimensions showed a tendency of perceivers to pursue evaluative consistency in their personality impressions of other individuals, a well-documented phenomenon known as the *halo effect* (Kelley, 1950; Nisbett & Wilson, 1977; Thorndike, 1920). But other times, a negative relationship between the dimensions emerges. For instance, Judd and colleagues (Judd, James-Hawkins, Yzerbyt, & Kashima, 2005) found that when participants judged a target as positive in one dimension, they tended to compensate it by judging the same target more negatively on the other dimension. They called this compensatory tendency the *compensation effect*, while noting that it is especially triggered by situations where a contrast between the two dimensions is made salient in judgments about different targets (i.e., individuals or groups) (Kervyn, Yzerbyt, & Judd, 2010, 2011; Kervyn, Yzerbyt, Judd, & Nunes, 2009; Yzerbyt, Provost, & Corneille, 2005; Yzerbyt, Kervyn, & Judd, 2008). This compensatory tendency also provides a potential explanation for the predominance of stereotypes whose content is mixed in valence (e.g., elderly as warm but

incompetent; rich people as competent but cold; Cuddy et al., 2009; Fiske et al., 2002), which in turn were already suggestive of a negative correlation between the dimensions.

Regarding the independence of the two dimensions, empirical support can be found in the emergence of a two-dimensional solution itself when dimensionality reduction analyses (e.g., MDS, PCA, or Factor Analysis) are applied to trait ratings' data; in zero correlations between self-ratings of agency and communion (Abele, 2003); or in studies showing that when we perceive a target to vary along one dimension, without varying on the other, our impression of the target shifts along the varying dimension while remaining stable in regard to the other (Cislak & Wojciszke, 2008; Zanna & Hamilton, 1972).

And lastly, in the domain of social face perception, the independence between the trustworthiness/valence and dominance dimensions suggested by the PCA solution, seems to obscure the small negative correlation of dominance with trustworthiness/valence (Chen et al., 2014; Oosterhof & Todorov, 2008). In fact, this negative correlation between the dimensions of facial impressions is reflected in a compensation effect found in the judgments of single target faces (see discussion in Dotsch & Todorov, 2012), which interestingly, occurred in the absence of an explicit comparative context. Although more studies specifically designed to address compensatory patterns in facial impressions are desirable to reliably conclude about the stability of compensatory patterns in facial judgments, the evidence available so far seems to suggest that the negative relationship between the dimensions of facial impressions may be more stable than the relationship between the dimensions of person perception, which in turn exhibit either a halo or a compensatory pattern.

Making Sense of the Inconsistent Relationship between Dimensions

The inconsistent findings regarding the relationship dynamics of the two fundamental dimensions of social judgment triggered subsequent investigations on the potential causes of such inconsistency. Calling attention to the fact that both of the two dimensions exhibited an evaluative component, Suitner and Maass (2008) argued that the implied common relationship established between valence and each of the dimensions could be playing a determinant role in the correlation found between them. Using a statistical metaphor, Suitner and Maass (2008) argued that the interpretation of a positive relationship between the two dimensions could be understood as a 'Type I error' as it may merely be indicating common variance with valence. At the same time, a 'Type II error' would reflect an inability to detect

the ‘true’ association between the two dimensions due to the confounding effect of their common variance with valence. To demonstrate this, they asked participants to rate 130 trait-adjectives on how each was perceived to be related with communion, agency, and valence. And as they suspected, their results indicated that the positive correlation between communion and agency became negative once they controlled for their common association with valence. The positive relationship is related with the perception of both communion- and agency-related traits as desirable qualities *per se*, which is reflected in the correlation of each dimension with valence (although communion is more strongly correlated with valence than agency; see next section). On the other hand, the finding of a negative relationship complements the previous finding of a compensation effect in intergroup and interpersonal judgments (e.g., Judd et al., 2005). However, we must pay attention to the fact that Suitner and Maass (2008, Study 1) worked with evaluations of trait-adjectives, rather than evaluations of individuals or groups, which precludes the attribution of their results to halo and compensation effects. Moreover, it remains a question whether these findings can be extended to the domain of social face perception. In this domain, valence is highly and positively correlated with the trustworthiness dimension, and both trustworthiness and valence are slightly negatively correlated with the dominance dimension (Oosterhof & Todorov, 2008). Thus, the negative relationship between trustworthiness and dominance is more clearly apparent without controlling for valence, comparatively to the findings for communion and agency. In addition, as we outlined in Chapter II (see Facial Cues of Trustworthiness and Dominance and their Meaning section), communicating valence is one of the main functions of the face, and thus, controlling for valence in judgments of faces may result in relationships between dimensions that do not reflect naturalistic impressions of faces. Nevertheless, the work by Suitner and Maass (2008) highlights the importance of taking the relationship with valence into account when examining the complex relationship between the fundamental dimensions of social perception.

An additional account for the inconsistent relationship between the dimensions relates to the nature of the relationship these dimensions establish with valence, and its implications for the measurement and analysis of trait dimension relationships. The possibility that dimensions may not be linearly related to valence (or between themselves) seems to have been overlooked in most of the social perception literature. This does not mean, however, that all researchers failed to take into consideration such an important detail. More than half a century ago, Lemann and Solomon (1952) took notice of this possibility and proposed a

distinction between traits in regard to their relationship with valence. They called *alpha-traits* to traits establishing a linear relationship with valence (e.g., trait scale ranging from positive to negative, or from negative to positive); and *beta-traits* to traits establishing a curvilinear relationship (U-shaped or inverted U-shaped) with valence (e.g., trait scale ranging from positive to negative to positive, or from negative to positive to negative). Despite of the crucial implications that this distinction has on the measurement of trait relationships, the work of Lemann and Solomon (1952) remained a surprisingly obscure reference in the literature. To the best of our knowledge, the work in this thesis is the first to revisit and take into consideration the alpha-beta trait distinction in the pursuit of our research aim of assessing relationships between dimensions (see Chapter VII).

More recently, though, some researchers took notice of the importance of questioning the assumption of linearity in studies aiming to measure the relationship between dimensions of social perception. While seemingly unaware of the work of Lemann and Solomon (1952), Imhoff and Koch (2017) questioned the linear nature of the relationship between the Big Two. Imhoff and Koch (2017) built upon the Aristotelian idea that virtuous behavior lies somewhere in-between deficiency and excess (i.e., the 'golden mean'; see Grant & Schwartz, 2011), and findings showing that many relationships between variables in the world follow curvilinear trends (e.g., Carter et al., 2014; Rhodes & Wood, 1992), to test the hypothesis that impressions of communion peaked at average levels of agency, or that high and low agency were associated with low communion. And indeed, they found evidence for an inverted U-shaped relationship in the contexts of interpersonal, intergroup, and interspecies perception. They further suggested that the inconsistent relationship between these dimensions across the literature could be resulting from the use of stimulus sets (e.g., social groups) whose content is capturing different ranges of the inverted U-shaped curve that defines the relationship between communion and agency. As a result, a positive (negative) correlation would be found for stimulus sets whose features ranged from low agency (mid-agency) to mid-agency (low agency).

An intriguing question that arises from the alpha-beta trait distinction proposed by Lemann and Solomon (1952), and the findings obtained by Suitner and Maass (2008) and Imhoff and Koch (2017), is whether both the relationship with valence and nature of the relationship between dimensions may be integrated in the assessment of the perceived valence of the dimensions in order to provide a more complete understanding of the complex

dynamics between the Big Two. This question is addressed in the empirical work presented in Chapter VII.

Interim Summary

The literature reviewed throughout this chapter provides the basis for a better understanding of the relationships between core dimensions of social judgments, both *between* the domains of person perception and social face perception and *within* each domain in particular. This review is helpful for a more complete understanding of the questions addressed in the empirical section of the thesis (Chapters VI, VII, and VIII).

However, regarding the relationship of dimensions within a model, two other questions have been capturing the attention of researchers. One of the questions is (a) whether one of the dimensions is perceived as more relevant than the other in impression formation. The other pertains to (b) how these dimensions are integrated into a unitary impression. One of the goals of the present work was to address these question within the domain of social face perception (in Chapter VIII). For this reason, in the next chapter, we turn our attention to the literature providing insight into these questions.

Chapter IV. Integrating Dimensions

Our review on the relationship between dimensions so far suggests that when we form an impression about someone's personality, the impression may be positive (or negative) on both warmth and competence—reflecting a halo effect—, or positive on one dimension but negative on the other—reflecting a compensation effect. But there are two other important aspects influencing the features of the final impression we form about someone: priority of processing attributed to a dimension, and integration of dimensions into unitary impressions. As we will see in this chapter, these aspects are important to consider if we intend to gain further insight into the processes influencing the relationship between dimensions within a two-dimensional model, and how this relationship is ultimately reflected in the content of an impression of personality. For clarity, we will first focus on how these aspects have been investigated in the domain of person perception, and subsequently review how they have been approached in the domain of social face perception. This literature review complements the one supporting our empirical work in Chapter VIII, and intends to emphasize the theoretical relevance of an extension of the current knowledge about trait integration in person perception to the domain of facial impressions.

The Primacy of Morality in Person Perception

The first aspect pertains to the question of whether perceivers prioritize one dimension over the other while forming an impression. Put differently, and drawing on the functional interpretation of the two dimensions adopted by Fiske and colleagues (Fiske et al. , 2007) and Oosterhof and Todorov (2008): Do perceivers prioritize gathering information about someone's intentions (i.e., warmth, trustworthiness)? Or do they prioritize information about someone's capabilities (i.e., competence, dominance) instead?

A number of studies found that perceivers exhibit preferential processing for morality-related information (Abele & Bruckmüller, 2011; Wojciszke et al., 1998; Ybarra et al., 2001). Morality-related information is processed faster than competence-related information (Abele & Bruckmüller, 2011; Ybarra et al., 2001); given priority over competence-related information when participants are tasked with gathering information about individuals to form a global impression about them (Wojciszke et al., 1998), or after being informed they would later interact with them (De Bruin & Van Lange, 2000); and is more predictive of global

impressions (Wojciszke et al., 1998). This body of findings is thus consistent in suggesting that perceivers are primarily concerned with gathering information about others' intentions.

This "primacy of morality" effect may be a result of a difference between the dimensions regarding their strength of association with valence. Several studies have shown that morality-related (or communion-related) information is more strongly related with valence than competence-related (or agency-related) information (Abele, Uchronski, et al., 2008; Suitner & Maass, 2008; Wojciszke et al., 1998), and people seem to exhibit greater emotional reactivity to others' communion-related traits comparatively to agency-related information (Wojciszke, 2005a; Wojciszke & Szymkow, 2003). When we take into account the primary role played by valence in the perception of our environment (Lewin, 1935; Osgood et al., 1957; Zajonc, 1980) and in guiding our approach-avoidance behavior (Chen & Bargh, 1999; Duckworth, Bargh, Garcia, & Chaiken, 2002), it should become less surprising that any social dimension strongly associated with valence would carry equivalent relevance.

We have outlined above that communion-related content is given preferential processing and is weighted more heavily than competence in global impressions of others. This suggests that when we receive information about a target on both dimensions, more often than not, it is the communion-related information and its valence that will serve as the context within which we interpret the subsequent agency-related information. This leads us to the second aspect we intended to address in this last section, which pertains to the question of how the information gathered on the two dimensions is subsequently integrated into a unitary global impression of the target.

Trait Integration in Person Perception

Two influential approaches tackled the question of how people integrate trait information into unitary global impressions about others. One was the *gestaltic* approach to impression formation proposed by Asch (1946), according to which the meaning of a trait changes in function of the context of trait interrelationships within which it is embedded. According to this perspective, for instance, the trait intelligent assumes different meanings depending on whether it is paired with friendly or unfriendly in the description of a target person. Thus, although intelligent carries a meaning in isolation, this meaning is modulated by the trait context when intelligent is incorporated into a unitary impression along with other traits. As a result, an impression that is other than the mere sum of its component traits is said to emerge.

The other approach was the linear integration of information proposed by Anderson (Anderson, 1962, 1968a), according to which the evaluative content of a final global impression corresponds to the sum (or the average) of the evaluative values associated with all of its component traits. To illustrate, our global impression about someone exhibiting four positive traits and one negative trait would be expected to be overall more positive, than if the person exhibited three positive traits and two negative traits. Thus, unlike the gestaltic approach, the linear approach places the focus on the evaluative component of traits—considered to be stable and context-independent—, while disregarding the denotative component of traits.

Of these two approaches, Wojciszke and colleagues (Wojciszke et al., 1998) hypothesized that the linear additive approach would fit better with their theoretical predictions. Arguing that the main function of judging someone's morality is to locate the person in an approach-avoidance dimension, these authors expected morality-related information to predict the valence of the global impression, and competence-related information to secondarily influence the extremity of the impression along the valence dimension. Specifically, they expected that the impression of a moral and competent (immoral and incompetent) person would be more positive (negative) comparatively to that of a moral and incompetent (immoral and incompetent) person.

Their overall results (Wojciszke et al., 1998; Study 4) confirmed these predictions, albeit inconsistently. When behavioral information was presented in a format that integrated morality- and competence-related information into the same behavior, the valence of the final impressions confirmed the expected pattern—i.e., moral and competent as more positive than moral and incompetent; and immoral and competent as more negative than immoral and incompetent. But when the same morality- and competence-related information was separately presented as two behavioral acts (of the same target person), their results were in every way identical to those in the "same act" condition, with the exception that immoral targets were equally disliked regardless of their level of competence. Put differently, morality-related information operated as the context within which competence-related information was subsequently evaluated, but the same was not true of competence-related information. And interestingly, this pattern of findings did not fit the linear additive model, as the evaluative value of the traits violated the assumption of their stability and context-independency. Instead, the degree of valence of a trait shifted as a function of the trait context (i.e., isolated from other traits in a separate behavior vs. paired with another trait in the same behavior) (Wojciszke et al., 1998; Study 4). Although the authors made no mention to Asch's gestaltic

approach, this pattern in their findings should at least invite its consideration in subsequent research. And as we will see, both the gestaltic and the linear integration approaches were taken into consideration in our work presented in Chapter VIII.

More recently, similar research further showed that when both warmth- and competence-related information describing a target person are mixed in valence (i.e., warmth-but-incompetent, cold-but-competent), the resulting impressions exhibit an amplification effect such that describing a person as cold (incompetent) increases perceived competence (warmth) to a higher extent than describing her as warm (competent) (Kervyn, Bergsieker, Grignard, & Yzerbyt, 2016). The same set of studies showed that this amplification was more evident in ambivalent (i.e., mixed valence), rather than univalent descriptions of warmth and competence. These findings provide further insight into the underpinnings of the compensation effect by demonstrating that it does not necessarily require an explicit comparative context, as previously thought. In addition, they nicely complement the findings of Wojciszke and colleagues (Wojciszke et al., 1998) by directly examining more diverse conditions where the information on one dimension interacts with information on the other.

Primacy of Morality and Trait Integration in Social Face Perception

Thus far, our review was entirely focused on the domain of person perception. Turning our focus to the domain of social perception, we find several studies suggesting that a similar primacy of morality effect emerges when faces alone are judged in trait dimensions. In several studies, Todorov and colleagues found that trustworthiness/valence judgments of faces occurred as fast as between 33 to 100 ms after exposure to a face; were the most stable across time since their inception; and preceded competence-related inferences (Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006). Although these studies were not designed to directly test whether trustworthiness or dominance received preferential processing, they fit well with the primacy of morality found in the person perception domain, especially if we also take into account the fact that trustworthiness is a morality-related trait, and that trustworthiness/valence judgments account for the largest portion of the variability in facial impressions (Oosterhof & Todorov, 2008).

Regarding the question of how trustworthiness and dominance are integrated in social face perception, we found no studies directly addressing how the trait integration processes examined in the context of person perception map onto the domain of facial impressions. To the best of our knowledge, only the work by Todorov and colleagues (Oosterhof & Todorov, 2008; Todorov et al., 2008) tangentially approached this question. Specifically, as we

previously described in Chapter II (see The Trustworthiness-By-Dominance Face Space section) Oosterhof and Todorov (2008) artificially combined face models of dominance and trustworthiness into face models representing the quadrants of the trustworthiness-by-dominance face space, and found that the resulting face models were highly correlated with perceptions of threat (high dominance and low trustworthiness quadrant to low dominance and high trustworthiness quadrant), and perceptions of attractiveness, competence and likeability (for the remaining two quadrants). However, these models might not be reflecting how perceivers spontaneously integrate trustworthiness and dominance information in their minds, as they do not take into account the priority of valence information (e.g., Todorov et al., 2009; Willis & Todorov, 2006) and potential asymmetries in the weighting of dimensions (e.g., Abele & Bruckmüller, 2011; Wojciszke et al., 1998).

In sum, most of the research concerned with investigating the asymmetries between dimensions and how they are integrated into unitary impressions in the minds of perceivers has been conducted in the domain of person perception, with targets consisting mostly of verbal person descriptors (e.g., behavioral information, trait-based descriptions). In the domain of social face perception, however, these two topics have not received as much attention, and as a result it remains a question whether information about intentions and ability is similarly processed when the target of social judgment is facial appearance alone (instead of behaviors or trait categories). Nevertheless, the research on these two domains converges in one aspect: Morality-related information and valence carry more weight and are given priority over other types of information in impression formation (when no specific context is provided), although this phenomenon is more clearly supported in research involving verbal person descriptors than in research involving evaluations of facial appearance alone. If this is so, then valence information may serve as the context within which all subsequent judgments are interpreted throughout the impression formation process, as supported by the findings of Wojciszke and colleagues (Wojciszke et al., 1998). This possibility highlights the importance of understanding the nature (i.e., linear or curvilinear) and direction of the relationship that valence establishes with other dimensions. The different relationships that competence and dominance establish with valence, also hint at potential differences between the person perception and the social face perception models regarding how these dimensions are integrated. However, these are questions that remain unanswered in the literature.

This chapter concludes our review of the literature that we consider relevant to understand the work reported in this thesis. Next, we highlight some of the questions that arise

from our review of the literature, and which we empirically addressed in the three papers that constitute the core of this thesis.

Chapter V. The Present Thesis

One of the main goals of the present thesis is to clarify and describe the nature of the relationships between the core dimensions of social perception found in the contexts of facial impressions and person perception. Within this broader goal, we identified three questions in our review of the literature that we sought to empirically address in the three papers that constitute the empirical section of this thesis (Chapters VI to VIII). Here, we will present each one of these questions separately before discussing how they relate to each other.

Relating Social Face Perception with Person Perception

The first question pertains to the extent to which the trustworthiness-by-dominance model of social face perception overlaps with the warmth-by-competence model of person perception. As outlined in Chapters I and III, these models are remarkably similar in their structure and functional meaning. Both warmth and trustworthiness are thought to reflect information about intentionality, whereas both dominance and competence are thought to reflect information about the ability to implement intentions (Fiske et al., 2007; Oosterhof & Todorov, 2008). Despite of, or perhaps due to this resemblance, little attention has been devoted to a systematic investigation of how the basic dimensions of social face perception are related with the basic dimensions of person perception—with a notable exception of recent work by Sutherland and colleagues (Sutherland et al., 2016) on the relationship between the dimensions of facial impressions and dimensions of inter-group perception (i.e., warmth and competence). But why should the relationship between these models be systematically investigated? Why not assume their redundancy based on their structural and functional resemblance?

First, it is important to note that person perception models were mainly derived from methodologies where verbal person descriptors (e.g., trait-adjective lists, behavioral descriptions, group membership) tapped into participants' conceptual knowledge about trait or social categories (e.g., lay theories of personality; stereotypical knowledge). In this context, the target features on which participants relied to make their judgments were less restrained (e.g., behavior or appearance, or both of these could be underlying a judgment; see Skowronski, McCarthy, & Wells, 2013) comparatively to what happens within the context of social face perception, where they are asked to specifically focus on appearance features of

the target. It is conceivable that an evaluation of personality based on physical features alone may color the interpretation of ability- and intention-related information. For instance, ability may be interpreted as physical ability (vs. cognitive ability), and the perception of intentions may be restrained by the emotional signal that is possible to be conveyed by facial appearance alone. Put simply, whereas perceived ability may take the form of dominance when we judge someone based on their appearance, this perception turns into competence when we judge individuals in their multiple features. And when we judge someone's intentions based on facial appearance we may not be able to differentiate between trustworthiness and warmth, as these dimensions may be more easily differentiated in terms of behavioral information than in terms of facial features.

Second, as our review of the literature in Chapter III suggests, while dominance and competence exhibit noteworthy distinctions at both the conceptual and evaluative (i.e., valence) levels, the trustworthiness and warmth dimensions appear to diverge more clearly at the conceptual than at the evaluative level. Altogether, the pieces of evidence found in our literature review appear to suggest that the warmth-by-competence model of person perception and the trustworthiness-by-dominance model of social face perception may not be entirely redundant, as their common functional interpretation and structural resemblance may initially suggest. The distinctions between the dimensions of different domains call for a systematic investigation of the extent to which trustworthiness and warmth, and dominance and competence, are similarly represented in the minds of perceivers in the context of social face perception.

This question was empirically addressed in our first paper (Chapter VI), entitled "Dominance and competence face to face: Dissociations obtained with a reverse correlation approach". In this paper, we tested the hypothesis that the person perception and social face perception models are not entirely redundant due to evaluative and conceptual divergences between their analogous dimensions. To examine this question, we employed reverse correlation methods to assess the degree of similarity between how perceivers (approximately) mentally represent dominance and competence, and of trustworthiness and warmth, in the domain of social face perception.

The Nature of the Relationship with Valence

The second question pertains to the nature of the relationship that the dimensions of person perception and social face perception establish with valence. Throughout our review of the literature, we found that the dominance and competence dimensions tend to exhibit opposite relationships with valence. Dominance is (slightly) negatively related with valence in facial impressions (Oosterhof & Todorov, 2008), whereas competence tends to be positively correlated with valence in judgments of individuals (e.g., Anderson, 1968b; Judd et al., 2005; Rosenberg et al., 1968; Rosenberg & Sedlak, 1972). The studies that led to these findings have, however, largely overlooked the distinct relationship established between these two traits and valence, often conflating them in composite measurements of agency and competence. Given the role that the association with valence plays on how the fundamental dimensions of social perception are found to be related (e.g., communion and agency; Suitner & Maass, 2008), such distinctions at the evaluative level become important to consider in endeavors to further understand the complex dynamics between fundamental dimensions.

Another limitation of most studies measuring perceptions on the fundamental dimensions, has been their tendency to implicitly assume that the relationship between dimensions of social judgment is linear in nature, despite of previous work calling attention to the importance of inspecting the linearity of the relationship that traits establish with important dimensions such as valence (viz. linear alpha vs. curvilinear beta traits; Lemann & Solomon, 1952). This trend was recently challenged by the work of Imhoff and Koch (2017) who demonstrated that the relationship between the Big Two (i.e., communion and agency dimensions) follows an inverted U-shaped trend, with communion peaking at average levels agency. Such findings raised awareness about the need to consider the nature of the relationship between dimensions of social judgment. However, the studies by Imhoff and Koch (2017) did not include competence judgments in their agency measures, which in turn sampled only assertiveness-related traits of the agency dimension (see Abele et al., 2016) such as dominance. In doing so, these studies overlooked the impact that the inclusion of competence in agency measures could have on the relationship found between the Big Two. Imhoff and Koch's (2017) finding suggests that any agency-related trait will exhibit a curvilinear relationship with communion (and likely with valence too, given the higher correlation between communion and valence). Although this curvilinear trend may be the reason why small correlations were found between dominance and valence (e.g., Oosterhof & Todorov, 2008), it is incompatible with the commonly found positive correlation between

competence and valence (e.g., Judd et al., 2005; Kervyn, Fiske, & Yzerbyt, 2013; Rosenberg et al., 1968). Moreover, the recent branching of the agency dimension into an assertiveness facet (e.g., dominance, confidence) and a competence facet (e.g., competence, intelligence) (Abele et al., 2016), further underline the importance of considering both facets in composite measures of agency. Comparatively, unlike the agency-related traits, the traits representing the two facets of the communion dimension—warmth and trustworthiness—both exhibit a strong positive correlation with valence (e.g., Kervyn et al., 2013; Oosterhof & Todorov, 2008; Rosenberg et al., 1968), which suggests that their relationship is more clearly linear in nature. Together with the evaluative and conceptual distinctions identified in our comparison of the dimensions of person perception with the dimensions of social face perception (see Chapter III), these considerations call for a closer examination of the nature of the relationship that valence establishes with the trait content of the Big Two, in both the contexts of person perception and social face perception.

In our second paper (Chapter VII), entitled "Good to bad or bad to bad? What is the relationship between valence and the trait content of the Big Two?", we aimed to resolve some of the issues identified in previous research by clarifying the nature of the relationship that valence establishes with relevant trait dimensions in both the contexts of person perception and social face perception. In doing so, we hypothesized that the direction and nature of the relationship established with valence would be more heterogeneous for the trait content of the agency than for the trait content of communion. To examine this question, we developed a new paradigm that takes into account both the role of valence and the nature of a relationship between constructs, in the assessment of the perceived valence of trait dimensions.

Integrating Dimensions in Social Face Perception

The third and final question pertains to how the trustworthiness and dominance dimensions are naturally integrated in the context of social face perception. Since Asch's (1946) seminal work on impression formation, most of the research concerned with the question of how trait information is integrated into unitary impressions of personality worked mainly with verbal person descriptors (e.g., trait-based or behavior-based descriptions). The output of this research informed us that morality-related information was given preferential processing over ability-related information in impression formation, and carried more weight in the final impressions we form about someone (Abele & Bruckmüller, 2011; Wojciszke et

al., 1998; Ybarra et al., 2001). However, the topic of trait integration has been largely overlooked in the domain of social face perception.

Although some studies in social face perception tangentially approach this topic, they never explicitly focused on it. For instance, as we outlined in Chapter IV, the work by Oosterhof and Todorov (2008) linearly combined facial models of dominance with facial models of trustworthiness to produce synthetic faces representing the quadrants of the trustworthiness-by-dominance face space, and subsequently explore how these were perceived on several candidate dimensions such as threat. However, their studies were not designed to directly approach the question of trait integration in facial impressions. The balanced weighing of dimensions in Oosterhof and Todorov's (2008) approach may not represent how perceivers spontaneously integrate the two dimensions in their facial impressions, as suggested by the findings showing that morality-related information (trustworthiness/valence) carries a higher weight in impressions of personality. To the best of our knowledge, no studies so far have directly addressed how perceivers integrate trait information in the domain of social perception. Thus, it remains a question whether trustworthiness or dominance are equally weighted in facial impressions, or if morality-related information act as a context within which ability-related information is interpreted, similarly to what was found in person perception research (Wojciszke et al., 1998).

In our third paper, entitled "Combining traits into a face: A reverse correlation approach" we aimed to directly address the question of how perceivers integrate trustworthiness and dominance simultaneously in terms of facial appearance, by examining how this integration fits with the linear and gestaltic approaches to impression formation. Taking into account the literature suggesting the primacy of morality in impression formation, we hypothesized that trustworthiness information would predominate over dominance information in facial impressions. To examine this question, we employed reverse correlation methods to assess the visual facial content of impressions resulting from four different combinations of trustworthiness and dominance information, each representing a quadrant of the trustworthiness-by-dominance face space.

Overview of the Empirical Section

Altogether, the three papers presented in the empirical section of this thesis, provide answers that contribute to the clarification of the relationship between fundamental dimensions of personality perception, not only across the domains of social face perception and person perception, but also within the two-dimensional models found in each of these domains. The first paper (Chapter VI) aims to clarify whether the dimensions of social face perception and the dimensions of person perception are redundant. The second paper (Chapter VII) further clarifies the nature of the relationship that these dimensions establish with valence, while calling attention to the importance of the role played by valence in shaping the relationships between trait dimensions. Finally, the third paper (Chapter VIII), clarifies how the two dimensions of social face perception are naturally integrated into unitary impressions of facial appearance, and whether this integration is similar to that observed in the domain of person perception. The answers to the questions examined in these papers have the ability to cross-pollinate each other in order to provide a deeper understanding of the complex dynamics of the relationships established between the dimensions, between and within two parallel domains of social perception.

In the next three chapters we present the three published papers that constitute the empirical section of this thesis.

Section II

Empirical Section

Chapter VI

Dominance and competence face to face: Dissociations obtained with a reverse correlation approach *

Manuel Oliveira[1†], Teresa Garcia-Marques[1], Ron Dotsch[2], Leonel Garcia-Marques[3]

Abstract

The article explores whether the traits representing the dimensions underlying the structure of facial and non-facial impressions are similarly mapped in the face space. Two studies examine whether the trustworthiness-by-dominance and the warmth-by-competence two-dimensional models overlap in face perception. In Study 1 (N = 200), we used a reverse-correlation task to obtain classification images (CIs) reflecting how each dimension is mapped onto a face. Results show that the similarity between CIs was higher between warmth and trustworthiness than between competence and dominance. In Study 2 (N = 31) the evaluations of each CI on each social dimension show a higher dissociation between dominance and competence than between trustworthiness and warmth. These results, obtained at both perceptual and judgment levels, suggest that there is only a partial correspondence between the two models that seems to be driven by the relationship that the competence and dominance dimensions establish with valence.

Key-words: competence; dominance; face perception; person perception; reverse correlation

---

[1] William James Center for Research, ISPA – Instituto Universitário, Lisbon, Portugal.

[2] Utrecht University, Utrecht, The Netherlands.

[3] Faculdade de Psicologia, Universidade de Lisboa, Lisbon, Portugal.

* Article published in the *European Journal of Social Psychology*, *49*(5), 888-902, doi: 10.1002/ejsp.2569

[†] Correspondence to: Manuel Oliveira, William James Center for Research, ISPA—Instituto Universitário, R. Jardim do Tabaco 34, 1149-041 Lisbon, Portugal.; E-mail: manueljbo@gmail.com

Dominance and competence face to face: Dissociations obtained with a reverse correlation approach

Despite the abundance of traits that we are able to infer from a person's face, two trait dimensions were found to be central in face-driven impressions of personality: trustworthiness and dominance (Oosterhof & Todorov, 2008; Sutherland et al., 2013). Trustworthiness judgments capture facial features associated with the perceived intentions of a person; whereas dominance judgments capture facial features associated the target's ability to implement benevolent or malevolent intentions towards the perceiver (e.g., cues to physical strength; Oosterhof & Todorov, 2008; Toscano, Schubert, & Sell, 2014). When combined, these dimensions inform approach-avoidance decisions regarding social interactions (Krumhuber et al., 2007; Todorov, Baron, & Oosterhof, 2008).

Interestingly, the *trustworthiness-by-dominance* model strongly resembles other two-dimensional models of social perception such as the classical *social-by-intellectual* model of person perception (Rosenberg, Nelson, & Vivekananthan, 1968), or the more recent *warmth-by-competence* model of stereotype content (akin to Rosenberg's person perception model; see Cuddy, Fiske, & Glick, 2008; Fiske, Cuddy, & Glick, 2007; Fiske, Cuddy, Glick, & Xu, 2002). Warmth (cf. Rosenberg et al.'s social dimension) is a dimension that encapsulates traits relevant for social functioning (Fiske et al., 2007). In turn, the competence dimension (cf. Rosenberg et al.'s intellectual dimension) encompasses ability-related traits. From a functional perspective, these two dimensions are thought to reflect a primary concern of the social perceiver to gather information about someone's intentions (warmth) and their ability to enact them (competence). Identically to the trustworthiness and dominance dimensions, warmth and competence are thought to inform approach-avoidance behavioral decisions (Fiske et al., 2007). This view suggests an overlap in the functional meaning attributed to personality inferences derived from facial and non-facial social information (i.e., conceptual knowledge about traits). But are the dimensions of both models represented equally in terms of physical facial appearance in the minds of perceivers?

The current article aims to go beyond the apparent functional overlap of the two impression formation models and address how these models overlap in terms of facial content. Specifically, our aim was to assess the facial content expected by perceivers to be associated with warmth and competence and then compare it with the expected facial content of the trustworthiness and dominance dimensions. We have done this by testing how the facial content of these four dimensions objectively (physical similarity between face images) and

subjectively (trait judgments of faces) overlaps. While doing so, we further explored how valence may be playing an important role in shaping the relationship between the trustworthiness-by-dominance and the warmth-by-competence models, based on previous literature suggesting evaluative distinctions between dominance and competence.

Warmth and Trustworthiness versus Competence and Dominance

Although the warmth dimension was initially viewed as unidimensional (e.g., Fiske et al., 2007; Rosenberg et al., 1968), there is growing evidence in person perception suggesting that sociability- and morality-related traits are conceptually distinguished and differently weighted by perceivers (Brambilla & Leach, 2014; Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Goodwin, Piazza, & Rozin, 2014; Leach, Ellemers, & Barreto, 2007). As a result, some authors recently proposed a branching of the warmth dimension into two facets: morality and warmth/sociability (e.g., Abele et al., 2016; Fiske, 2018). This becomes relevant to our research aims when coupled with previous studies showing that whereas trustworthiness is perceived as morality-related trait, warmth is perceived as a sociability-related trait (Brambilla et al., 2011, Study 1). A more nuanced distinction between warmth and trustworthiness may, however, not easily emerge in face-driven impressions. The shorter time window associated with spontaneous inferences from faces (e.g., Willis & Todorov, 2006), or the information that physical features of faces are able to convey, may not allow for more fine-grained impressions. Previous studies have shown that both face judgments of trustworthiness (morality-related) and of sociability (warmth-related) share valence as their latent construct (see Table S3 in Oosterhof & Todorov, 2008). In addition, Walker and Vetter (2016) found a positive correlation between trustworthiness and communion (a dimension considered to be akin to Fiske and colleagues' warmth dimension; Abele & Wojciszke, 2014) in face judgments. Thus, with regard to face-driven impressions, we did not expect warmth to be easily discriminated from trustworthiness.

A distinction between dominance and competence may, however, be more likely. There is a body of research suggesting that dominance and competence are distinguishable not only at a conceptual level, but also at an evaluative level. At the conceptual level, it was recently proposed that competence, along with assertiveness, can be regarded as two distinct facets of a superordinate dimension called agency (i.e., a dimension encapsulating traits related with goal-achievement and task-functioning, akin to Fiske and colleagues' competence dimension; Abele et al., 2016; Abele & Wojciszke, 2014; Carrier, Louvet, Chauvin, & Rohmer, 2014). This branching is consistent with research showing that dominance and

competence are associated with distinct perceptions of status (Carrier et al., 2014; Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013; Henrich & Gil-White, 2001). Traits implying an underlying motivation for self-promotion (e.g., assertiveness, dominance) are better predictors of perceived status than traits purely related with task-functioning ability (e.g., intelligence, competence) (Carrier et al., 2014). The distinction is also apparent in work showing that socioeconomic success is related to dominance, but not competence (Koch, Imhoff, Dotsch, Unkelbach, & Alves, 2016). In personality impressions, Rosenberg and Sedlak (1972a, p. 264) have shown that these dimensions are not entirely redundant in the underlying structure of unrestricted person descriptions. At the evaluative level, attaining status via dominance implies more negative behaviors (e.g., coercion, threat) than those associated with the competence route (i.e., prestige) (Henrich & Gil-White, 2001). Moreover, bodily expressions of pride perceived as dominant were found to be less likeable compared to bodily expressions perceived as competent (Lange & Crusius, 2015). In face evaluation, facial dominance has been found to resemble more negative emotional expressions than facial competence (Said, Sebe, & Todorov, 2009). This is in line with evidence showing that facial dominance is perceived as threatening (Chen, Jing, & Lee, 2014; Said et al., 2009; Todorov, Said, Engell, & Oosterhof, 2008). Preliminary evidence in the work of Oosterhof and Todorov (2008; see supplementary Table S3) also shows that face judgments of intelligence load higher in the valence/trustworthiness component than in the dominance/power component. Moreover, Todorov, Said, et al. (2008) reported that faces representing one diagonal dimension of the trustworthiness-by-dominance face space tracked judgments of likeability, extraversion, and competence, thus suggesting that facial competence and valence-related dimensions such as likeability are positively related. Altogether this body of research suggest that (1) dominance and competence should be easily distinguished in terms of facial content and (2) should lead to judgments of opposite valence (i.e., competent faces judged as more positive in general than dominant faces).

Overall, these arguments lead us to hypothesize a higher overlap between the diagnostic facial content of warmth and trustworthiness, and a lower overlap between the diagnostic facial content of competence and dominance. Preliminary information about these hypotheses has been provided by previous studies (Sutherland, Oldmeadow, & Young, 2016; Todorov, Said, et al., 2008). In Sutherland et al. (2016) participants were asked to rate a set of 1000 highly varied face photographs on warmth, competence, trustworthiness and dominance. These authors found that although there was a positive correlation between warmth and trustworthiness ratings and between dominance and competence ratings, the correlation was

significantly lower between dominance and competence. The current article aimed to conceptually replicate and extend Sutherland and colleagues' (2016) work, by assessing how participants map traits onto a face based on their prior expectations about how conceptual trait information is associated with facial content. To do this we used a reverse correlation (RC) approach.

Reverse Correlation Methodology

RC methods have proven their usefulness in social perception research as a data-driven tool to probe perceiver's a priori expectations (or internal representations) about their social world (for a review see Brinkman, Todorov, & Dotsch, 2017; Jack & Schyns, 2017). These methods allow the identification of face configurations that are diagnostic of specific social judgments. In essence, the methodology requires the perceiver to report the extent to which the features of a target stimulus (e.g., face) match with the content of her internal representation of the target concept (e.g., competence). During this process, the internal representation activated by the target concept serves as the reference to which the incoming stimuli are compared with. As a result, this method and its output are highly sensitive to the content of the representation. In a typical RC task, participants select from a pair of face images (that convey a great number of random variations of a base face stimulus) the one that elicits a particular trait. This allows researchers to obtain a classification image (CI): an image that isolates the relevant stimulus features that predict the target trait (e.g., face image whose features predict a 'dominant' judgment).

One advantage of RC over traditional methods of identifying diagnostic facial features of traits based on ratings of photographs is the possibility to directly perform analyses with the output pixel data of CIs. By comparing the resulting CIs researchers can obtain an objective measure of the degree of overlap between the visual content of trait representations, and analyze inter-CI agreement based on pixel information (e.g., Dotsch & Todorov, 2012). Additionally, subsequent trait ratings of the CIs inform about the trait signal contained in the face, and how much independent judges agree on the face features that elicit a given trait dimension. Previous studies already tapped into the visual facial content of both the trustworthiness and dominance dimensions (Dotsch & Todorov, 2012), and the warmth and competence dimensions (Imhoff, Woelki, Hanke, & Dotsch, 2013) showing that the CIs successfully elicited their intended traits. However, these studies' CIs used different base faces, thus preventing their comparison. In the present article we generated CIs for all four dimensions using the same base face, to be able to investigate their overlap.

The Present Research

Our aim was to reveal whether perceivers associate similar facial content with trait dimensions sharing the same functional interpretation (i.e., trustworthiness and warmth, and dominance and competence), considering that similar facial patterns are known to elicit similar judgments (Secord & Bevan, 1956). In terms of facial content, we expected to find a higher overlap between the trustworthiness and warmth dimensions, and a lower overlap between the dominance and competence dimensions, due not only to conceptual differences, but also to the different relationships established between these dimensions and valence.

In our approach, we selected the warmth and competence model on the basis of (1) the similarity of its labels with those of the seminal social and intellectual model of person perception (Fiske et al., 2007), (2) previous findings suggesting that friendly (warmth-related) and intelligent (competence-related) are highly frequent traits in spontaneous person descriptions (Rosenberg & Sedlak, 1972b), and (3) the functional interpretation shared with the trustworthiness-by-dominance model. In Study 1 we used a RC paradigm to generate CIs representing our four target dimensions and addressed their objective similarity (via pixel correlations) to examine the degree of overlap in facial content between their corresponding trait poles (e.g., dominant vs. competent). In Study 2, we asked independent raters to judge all the CIs in traits previously identified as relevant in the person perception and face evaluation domains to validate the CIs, analyze the structure underlying these judgments, and examine their relationship with valence.

Data Availability and Supporting Information

All the data, scripts of the analyses, and supporting information are publicly available online in the Open Science Framework and can be accessed via the URL: https://osf.io/hr5pd/

(Supporting Information can be found in Appendix A of this thesis.)

Study 1

Participants and Design

For the sake of clarity, we divided this study in two moments, each represented by a sample. This division results from a decision to collect additional data for the competence dimension. This was done with the purpose of clarifying whether the trait-adjective words we first used to represent competence (see Target traits section for more details) were adequately capturing the intended construct (see Results section).

Sample 1a was composed by 160 university students and employees (135 females; M_{age}= 21.7 years, SD_{age}= 5.90), who participated either in exchange for course credits or to earn a chance to win a lottery, in a between-participants design defined by 4 (Trait dimension: trustworthiness vs. warmth vs. dominance vs. competence) × 2 (Dimension pole: high vs. low).

Sample 1b was composed by 40 university students (39 female; M_{age}= 20.1 years, SD_{age}= 5.08) who were allocated to either the *Competent* or *Incompetent* trait condition.

Following the recommendations of previous RC studies (e.g., Dotsch & Todorov, 2012), twenty participants participated in each condition, and each participant provided 300 decisions (i.e., trials). Although this is not recommended to analyze CIs at an individual level (see Brinkman et al., 2018), our focus was on the so-called group-level CIs—henceforth simply referred to as CIs (for CIs analyses at the individual level, please see sample size requirements presented in Brinkman et al., 2018). We followed a common approach in RC studies and analyzed data at the group-level (e.g., Dotsch & Todorov, 2012; Imhoff, Dotsch, Bianchi, Banse, & Wigboldus, 2011). Thus, we ignored dispersion at the level of individual image creators, and assumed that the averaged data retain the "true" signal (i.e., facial features that are highly predictive of judgment). Consequently, our interpretations are made at the group level, and are derived only from the visual content of trait representations. This visual content is common across perceivers and evades the idiosyncratic content of a single perceiver's representation. However, as we discuss in the Results section, we rely on the error

component of the averaged data as a measurement of participants' agreement regarding the content of each CI.

Face Stimuli

The face stimuli used as targets on the main task were generated with the R package rcicr version 0.3.0 (Dotsch, 2015). Each stimulus consisted of a base-face image on which random visual noise was superimposed. The noise generation procedure was identical to the one reported in Dotsch and Todorov (2012). Further details can be found in the online supporting information. The base image was the grayscale average male face (Karolinska Face Database; Lundqvist, Flykt, & Ohman, 1998), resized to 256 × 256 pixels (Dotsch & Todorov, 2012). By repeatedly applying these visual noise patches to the same base-face image (e.g., each iteration yields a face image with superimposed noise), we essentially created many different versions of the same face, as a result of the distortions introduced by the noise in random locations of the face stimulus. The randomness of the task stimuli circumvents *a priori* assumptions made about the facial content associated with any of the target trait dimensions (Gosselin & Schyns, 2003; Jack & Schyns, 2017).

Target Traits

Data from the first sample (1a) were collected using traits selected from Oosterhof and Todorov's (2008) dimensions labels and Rosenberg and Sedlak (1972b) study showing that the trait-adjectives friendly and intelligent occur frequently in spontaneous person descriptions and represent the warmth and competence dimensions (also known as social and intellectual dimensions; Rosenberg et al., 1968). Our second sample (1b) yielded CIs associated with the competence labels themselves (i.e., competent and incompetent), instead of the (competence-related) intelligence traits.

For each trait dimension we used two trait-adjectives representing the dimension's poles (e.g., friendly and unfriendly). To represent the high-pole of each trait dimension we used: *Dominant*, *Trustworthy*, *Friendly*, *Intelligent,* in sample 1a, and *Competent* in sample 1b. For the low-pole we used: *Submissive*, *Untrustworthy*, *Unfriendly*, *Unintelligent,* in sample 1a, and *Incompetent* in sample 1b.

The trait-adjectives Competent and Incompetent, were used to generate additional CIs representing the competence dimension. Although Rosenberg and Sedlak's (1972b) work suggests that intelligent may be a trait-adjective that is spontaneously used by perceivers in their evaluations of competence—and thus, that it should adequately capture the construct—we decided to further clarify the extent to which the constructs of intelligence and competence overlap. Specifically, we were concerned about whether competence would be broader in meaning than intelligence[1].

Procedure

Participants were invited to participate in a study about personality impressions based on faces. All participants were randomly distributed across the trait conditions. The experiment ran on desktop computers stationed inside one-person laboratory cubicles. Instructions provided on the computer screen emphasized that they should not spend too much time on a trial and that they should follow their 'gut feeling' whenever they felt they were taking too long. The 2IFC image classification task consisted of 300 trials and included a forced pause of 2 minutes after the 150th trial. Trials were randomly presented without replacement, and the side of the screen on which the image with the superimposed inverted noise (i.e., negative image) appeared was counterbalanced within-participants (i.e., left image with inverted noise for half of the trials; noise generation details can be found in the online Supporting Information). In each trial, a pair of faces was presented side-by-side in the center of the screen with a question asking the participant to select the face that, in their opinion, best elicited the target trait. To give their response they pressed "E" on the keyboard to select the left image, and "I" for the right image. Before each trial a blank screen with a centered fixation cross was exhibited for 1000 ms. After completing the task, participants were thanked and debriefed.

[1] A similar question could be asked regarding the friendliness trait-adjectives representing warmth. The concept of friendliness may also be more specific than warmth. Nevertheless, according to the perspective that warmth encapsulates two facets (see Introduction), friendliness is a sociability-related trait, whereas trustworthiness is a morality-related trait. As a result, using a more specific trait to represent warmth (i.e., friendliness), seems to be acting against our hypothesis that warmth and trustworthiness are harder to disentangle in face evaluation. Thus, we consider the comparison between the two traits as a more conservative test. A choice to generate additional CIs for the trait-adjectives Warm and Cold would be likely to promote a finding that fits with our hypothesis, if indeed it is a broader concept than friendliness. On the other hand, by choosing to generate additional CIs for a broader concept such as competence, we are also being conservative, in the sense that we may be promoting a lower differentiation between competence and dominance (e.g., a dominant-looking bouncer can be judged as competent), which acts against our hypothesis of low overlap. Under these circumstances, any findings supporting our hypotheses may be considered as more compelling.

Results and Discussion

Classification Image Generation

A CI for each dimension was assembled by first averaging all the noise patterns selected by all the participants in a trait condition (i.e., average of the noise patches of the images selected by participants in the task), and subsequently superimposing the averaged noise on the original base-face image. Although a CI can be computed for each participant using the participant's 300 responses (i.e., individual CI), our focus was on the group-level CIs. Each group-level CI (see Figure 1) reflects the average of all individual CIs' noise patterns, keeping only the most commonly expected face features associated with a target dimension (i.e., inter-individual differences are averaged away).

Upon a brief visual inspection of Figure 1, the trustworthiness- and friendliness-CIs appear to share more resemblances overall than both competence-related CIs share with the dominance-CIs. Moreover, the competence-related CIs appear to be highly similar to the friendliness- and trustworthiness-CIs. A more objective analysis of the similarities between the CIs is reported ahead.

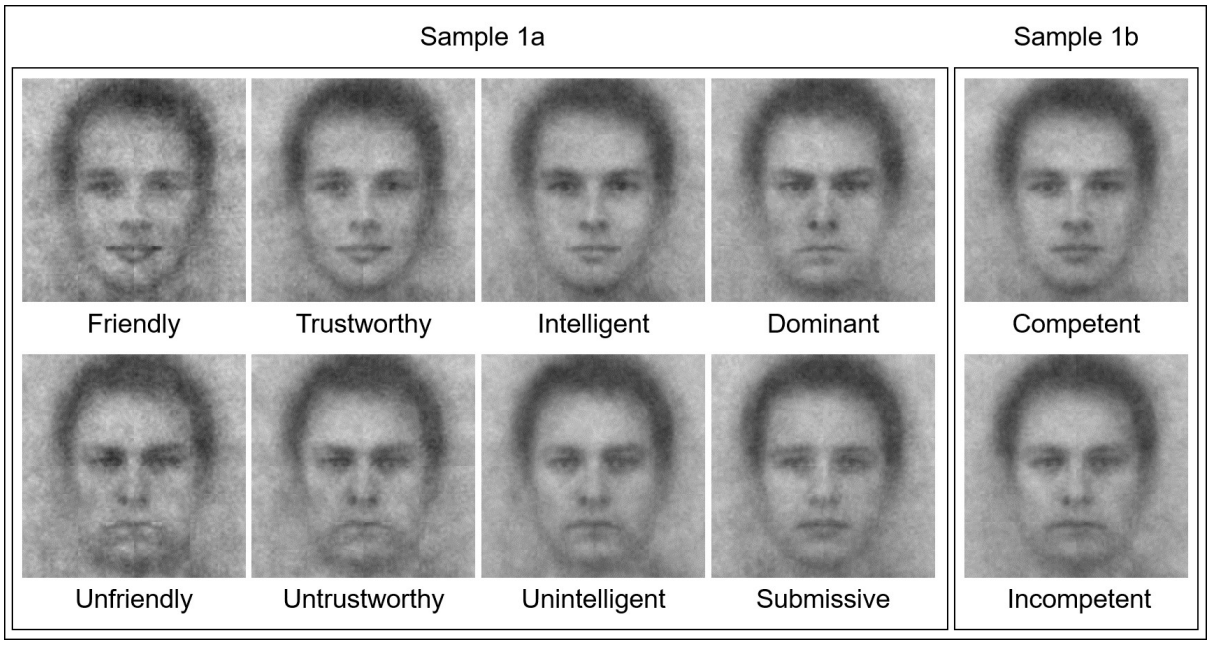

Figure 1. Group-level CIs by trait condition, for samples 1a and 1b. Each CI was computed using the responses of all participants of a trait condition.

Inter-rater Agreement

The extent to which participants showed agreement in their task responses in a given CI condition, and the extent to which the pixel data between the individual-CIs' results were similar across participants of the same condition were assessed with intra-class correlation coefficients (ICCs). Table 1 lists the ICCs for every group-level CI. The pattern of results was practically identical for the two types of data (i.e., response-based vs. pixel-based). Overall, ICCs indicate that there was higher inter-rater agreement for the trustworthiness- and warmth-CIs, and lower agreement for the dominance- and intelligence-CIs, especially in their low poles (submissive and unintelligent). Noticeably, the ICCs exhibit low values overall, possibly as the result of a guessing response strategy in trials where both target images were equivalent in signal strength (e.g., both high or low in signal). The data also suggest that it may have been harder for the participants to map dominance and competence-related traits onto a face space (i.e., a space defined by physical face features) than for the warmth and trustworthiness traits, which could have decreased agreement by promoting random responses. Another possibility is that participants indeed agree less on the facial content expected to represent these dimensions, especially at the low poles of competence and dominance. Low ICCs may lead to noisier group-level CIs, in that they capture less common features across participants' individual CIs. This pattern is consistent with findings by Hehman, Sutherland, Flake, and

Slepian (2017) showing that competence-related judgments depend more on the perceiver's characteristics than on the target's facial features.

Table 1

Participants' Agreement Measures for 2IFC Task Judgments, and Agreement between Classification Images' Pixel Data, in Each CI Trait Condition Quantified as Intra-class Correlation.

	CI	2IFC response data		CI pixel data	
		ICC (2, *k*)[a]	95% C.I.	ICC (2, *k*)[a]	95% C.I.
Sample 1a	Trustworthy	.60***	[.53, .66]	.69***	[.69, .69]
	Untrustworthy	.56***	[.48, .63]	.61***	[.61, .62]
	Dominant	.44***	[.35, .53]	.51***	[.50, .51]
	Submissive	.35***	[.24, .45]	.29***	[.29, .30]
	Friendly	.79***	[.76, .82]	.81***	[.81, .81]
	Unfriendly	.71***	[.67, .76]	.74***	[.74, .75]
	Intelligent	.41***	[.31, .51]	.54***	[.54, .55]
	Unintelligent	.21**	[.08, .34]	.24***	[.23, .25]
Sample 1b	Competent	.31***	[.19, .42]	.30***	[.30, .31]
	Incompetent	.22**	[.08, .34]	.19***	[.18, .20]

Note: CI = classification image; C.I. = confidence interval; [a] *k* represents the number of raters and *k* = 20 for each group-level CI, and applies to both the 2IFC response and pixel data sets. Degrees of freedom of the *F*-statistic were (299, 59501) for 2IFC response data, and (65535, 1245165) for pixel data. ** $p < .01$; *** $p < .001$

Objective Similarities

Similarity between CIs was assessed using the correlation between their pixel luminance values. All CIs were masked beforehand with an oval shape that preserved only the pixels of the face, hair and ear regions. Positive correlations correspond to higher physical similarities between the images and negative correlations indicate that the images are physically opposite (i.e., the darker a pixel is in one image, the lighter it is on the other image). Close to null correlations indicate that the images share little to no similarities.

The correlations of interest are between: trustworthy and friendly, untrustworthy and unfriendly, dominant and intelligent, dominant and competent, submissive and unintelligent,

and submissive and incompetent. The correlation values for these comparisons correspond to the bold values in Table 2, where we report the Pearson pixel-wise correlations between all possible pairs of CIs (for a color version of the correlation matrix see Figure S1 in online supporting information). Results suggest a stronger overlap between the friendliness and the trustworthiness CIs, than between the intelligence and the dominance and competence CIs (close to zero correlations).

The almost null correlation found between the submissive, unintelligent and incompetent CIs, indicates that they have little to no similarities. These smaller correlations may have been promoted by their lower inter-rater agreement and higher noise. A closer inspection of the matrix of all CI inter-correlations shows, however, that CIs with low ICCs still exhibit correlations between .30 and .40 with the CIs representing their opposite pole, which suggests that they contain signal of their intended trait (see Figure S1 in online supporting information). This was subsequently confirmed by trait judgments of these CIs (see Figure 3).

To assess how the facial content associated with the dominance and competence-related dimensions is related with the facial content associated with the warmth and trustworthiness dimensions, we computed a dimensional-CI for each of these five dimensions by subtracting the (masked) low-pole CI from the (masked) high-pole CI of the same dimension (e.g., subtracting the pixel matrix of the submissive-CI from the pixel matrix of the dominant-CI). Thus, the pixel information of each dimensional-CI reflects the difference between the facial content its two CI poles, which can be understood as the information that is common across the dimension poles. Next, we ran separate multiple regression analyses where we entered either the dominance and competence or the dominance and intelligence dimensional-CIs as predictors of either the warmth or the trustworthiness dimensional-CI. Regarding the models with the warmth-CI as the outcome, the model estimates were negative for the dominance-CI and positive for competence-CI ($b_{\text{dominance-CI}} = -0.75$; $b_{\text{competence-CI}} = 1.09$, $R^2 = .46$, all ps $< .001$). Replacing the competence-CI with the Intelligence-CI in the model yielded practically the same results ($b_{\text{dominance-CI}} = -0.80$; $b_{\text{intelligence-CI}} = 0.79$; $R^2 = .41$, all ps $< .001$). Similar relationships were found for models with the trustworthiness-CI as an outcome ($b_{\text{dominance-CI}} = -0.50$; $b_{\text{competence-CI}} = 0.92$, $R^2 = .44$, all ps $< .001$; and $b_{\text{dominance-CI}} = -0.54$; $b_{\text{intelligence-CI}} = 0.71$; $R^2 = .41$, all ps $< .001$). The pixel-wise correlations between all dimension CIs can be found online in the supporting information (Figure S2). In other words, these data suggest that the physical dissimilarity between the dimensional CIs of competence and dominance may be related with the opposite relationship that these dimensions establish with

valence (i.e., dimension representing how positively a target is evaluated), while taking into account that trustworthiness highly overlaps with valence (see Oosterhof & Todorov, 2008).

In line with our hypotheses, these results suggest that dominance and competence-related traits were expected (by the participants) to be associated with very distinct facial content; whereas the facial content expected to be associated with trustworthiness and warmth-related traits (i.e., friendliness) is highly similar. Moreover, the lower overlap in the former comparison may stem from the different relationship that the concepts of competence and dominance establish with valence.

Table 2

Pearson Correlations between Group-Level CIs' Pixel Luminance Values (Lower Matrix Triangle) and respective 95% Confidence Intervals (Upper Matrix Triangle).

	Dominant	Untrustw.	Unfriendly	Incompetent	Unintellig.	Submissive	Intelligent	Competent	Trustworthy	Friendly
Dominant	-	[.52, .54]	[.60, .62]	[.03, .05]	[.20, .22]	[-.41, -.40]	**[.02, .04]**	**[-.23, -.21]**	[-.38, -.36]	[-.50, -.49]
Untrustw.	.53 ***	-	**[.73, .74]**	[.30, .32]	[.47, .49]	[-.29, -.27]	[-.30, -.28]	[-.51, -.49]	[-.57, -.56]	[-.68, -.67]
Unfriendly	.61 ***	**.74** ***	**-**	[.26, .27]	[.42, .43]	[-.24, -.22]	[-.17, -.15]	[-.43, -.41]	[-.59, -.58]	[-.77, -.77]
Incompetent	.04 ***	.31 ***	.27 ***	-	[.34, .36]	**[.01, .03]**	[-.43, -.41]	[-.22, -.20]	[-.34, -.32]	[-.35, -.33]
Unintellig.	.21 ***	.48 ***	.42 ***	.35 ***	-	**[-.07, -.05]**	[-.39, -.37]	[-.35, -.33]	[-.47, -.45]	[-.49, -.47]
Submissive	-.41 ***	-.28 ***	-.23 ***	**.02** ***	**-.06** ***	-	[-.15, -.13]	[-.08, -.07]	[-.02, .00]	[.08, .10]
Intelligent	**.03** ***	-.29 ***	-.16 ***	-.42 ***	-.38 ***	-.14 ***	-	[.32, .34]	[.30, .32]	[.36, .38]
Competent	**-.22** ***	-.50 ***	-.42 ***	-.21 ***	-.34 ***	-.07 ***	.33 ***	-	[.39, 41]	[.48, .49]
Trustworthy	-.37 ***	-.56 ***	-.59 ***	-.33 ***	-.46 ***	-.01 *	.31 ***	.40 ***	-	**[.68, .69]**
Friendly	-.50 ***	-.68 ***	-.77 ***	-.34 ***	-.48 ***	.09 ***	.37 ***	.48 ***	**.69** ***	-

Note: Values in bold refer to the comparisons of interest between CIs of correspondent poles of the person perception and social face perception dimensions. Each correlation can be interpreted as an objective measure of similarity between the CIs. * $p < .05$; *** $p < .001$

Study 2

The conclusions derived from Study 1's results hinge on the assumption that the facial content expressed by the CIs indeed predicts the trait judgments they are intended to represent. In this study we investigated how the CIs previously obtained in Study 1 were subjectively evaluated on several relevant trait dimensions. This allowed us to: (1) ascertain whether the CIs contained diagnostic information of the trait used to generate them (e.g., obtaining ratings of trustworthiness for the trustworthy-CI); (2) to replicate Study 1's findings at an evaluative, subjective level; (3) to clarify how the CIs were perceived on valence and additional traits representing dimensions (e.g., warmth) and dimension facets (e.g., morality; sociability) previously identified as relevant in person perception research; and (4) to examine the relationships established between these dimensions in social face perception.

Participants and Design

Thirty-one individuals (11 female; M_{age}= 32.7 years, SD_{age}= 8.92) with English as their first language were recruited online via Prolific Academic to judge a set of 10 CIs on a total of 9 dimensions with two poles each (18 trait judgments).

Face Stimuli

The target faces were all the CIs obtained in Study 1 (see Figure 1).

Trait Judgments

The selected traits included the high and low poles of each trait-dimension and included: (a) traits used to generate the CIs (*Friendly*, *Unfriendly*, *Intelligent*, *Unintelligent*, *Dominant*, *Submissive*, *Trustworthy*, *Untrustworthy*, *Competent*, and *Incompetent*); (b) warmth traits (*Warm*, *Cold*) to ascertain whether the friendliness CIs were perceived as intended in warmth; (c) traits selected from the social perception literature to represent the two facets of warmth (*Sociable*, *Unsociable*, *Honest*, and *Dishonest*); and (d) traits representing a general valence measure (*Likeable*, *Unlikeable*). The first two sets of traits (i.e., a and b) were chosen to validate the CIs. The trait sets listed in (c) and (d) were chosen to explore the relationship of the four target dimensions (i.e., trustworthiness, warmth, dominance, and competence) with the previously identified facets of warmth (Abele et al., 2016; Brambilla et al., 2011; Goodwin et al., 2014) and with valence.

Procedure

The task was conducted online using Qualtrics software. After informed consent the participants were instructed to rate the CIs on multiple personality traits, using a seven-point scale ranging from 1 (*Not at all*) to 7 (*Very* [trait]). In each trial a face was positioned at the center of the screen with a question above it ("How [trait] does this person look to you?"), and the rating scale below. The faces were blocked by trait judgment (i.e., all faces were judged on the same trait before being judged on the next trait). Both the order of faces and judgment blocks was randomized. Upon task completion, participants were thanked, debriefed, and received compensation for their participation.

Results and Discussion

Inter-rater Agreement

The inter-rater agreement for each trait subjective judgment across all 10 CIs was computed using ICCs. All ICCs were significant ($p < .001$) and ranged from .93 to .99 (see Table S1 in online supporting information). Albeit indicating high inter-rater agreement of these subjective ratings, the lowest values corresponded to the competence-related dimensions.

CI Ratings

To test whether each CI was judged as intended in its correspondent trait (e.g., dominant-CI judged as dominant), and to examine the extent to which CI pairs from two corresponding poles across the person and social face perception models had similar trait ratings, we ran two separate analysis. In the first analysis—CI validation—we compared any two CIs representing two opposing poles of one trait dimension (e.g., friendly-CI vs. unfriendly-CI). To do this, we computed the standardized difference (Cohen's *d*) between the judgments obtained for each pair of these CIs (e.g., judgment of competent for friendly-CI and unfriendly-CI). Each Cohen's *d* represents how differently the two CIs were judged in one and the same trait (e.g., how competent each one is perceived to be). The second analysis—CI comparisons—was identical to the first except that it compared between CIs corresponding to traits of the different models (i.e., CI of a person perception dimension vs. CI of a social face perception dimension). These results are summarized in the heatmap depicted in Figure 2 (effect size values can be inspected in the interactive heatmap available in our online data repository). The CI comparisons were the same as those previously done in Study 1(see comparisons associated with bold values in Table 2).

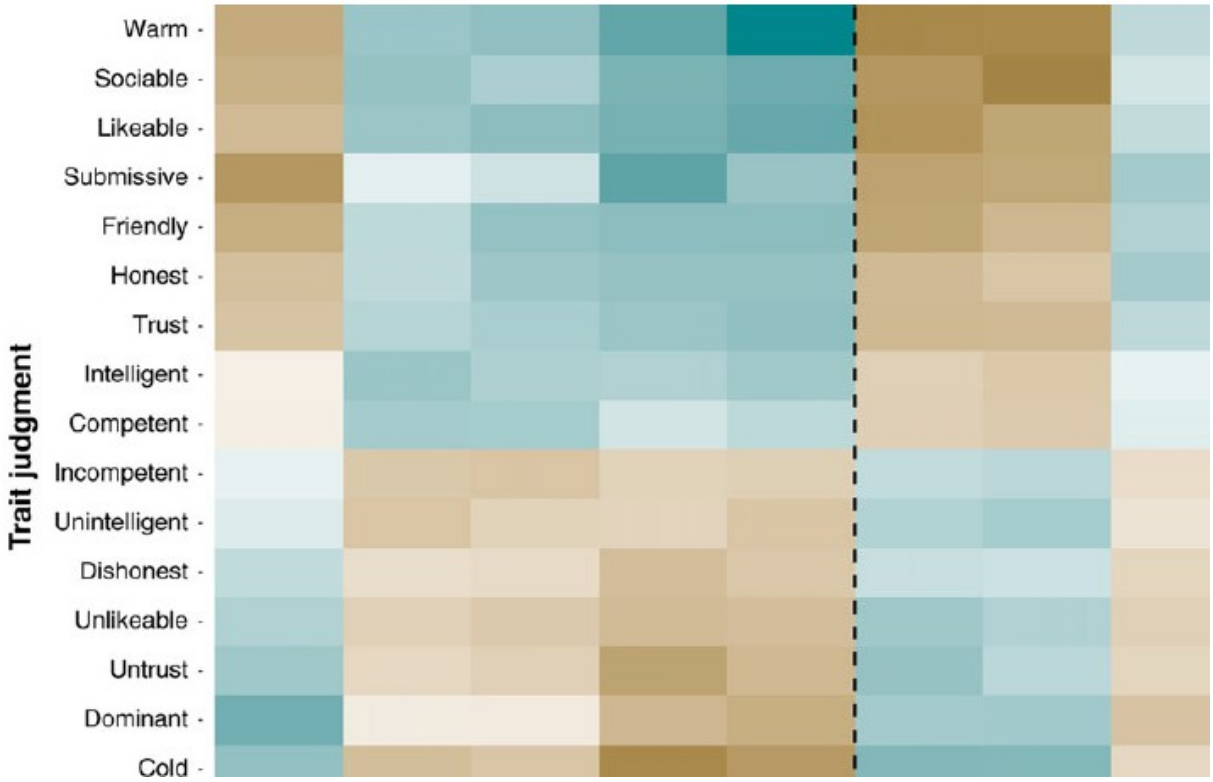

Figure 2. Each effect size (Cohen's *d*) corresponds to a difference between the mean trait judgment ratings of two CIs under comparison (*x*-axis). The higher the effect size, the higher the difference. Columns on the left side of the vertical dashed line show the comparisons between two CIs of opposite poles of the same dimension. Columns on the right side of the dashed line represent comparisons between the correspondent pole CIs of different models (i.e., trustworthiness-by-dominance model vs. warmth-by-competence model). Positive values (green cells) indicate that the first CI (i.e., CI 1) was rated higher than the second CI (i.e., CI 2) in a given trait. Negative values (brown cells) indicate that CI 2 was rated higher in a trait than CI 1. The fainter the colors (*d*'s closer to zero), the less are the differences in how the two CIs were perceived in each trait. The color version of this figure can be found in the article online, or be reproduced using the R script included in our online data repository.

CI validation. The overall pattern for comparisons of CIs of opposing poles of the same dimension (Figure 2) indicates that, in general, they were all judged higher in the trait that originated them compared to its counterpart (e.g., friendly-CI judged as higher in friendliness than unfriendly CI; unfriendly-CI judged higher in unfriendliness than the

friendly-CI). An important feature of these data is that, whereas the high pole CIs of competence, intelligence, and friendliness (i.e., warmth) were all evaluated as positive in valence (i.e., perceived as likeable), the high pole CI of dominance was evaluated as negative. This helps to clarify the negative correlation found between the competent- and dominant-CIs in Study 1, by suggesting that the dominance and competence dimensions exhibit an opposite relationship with valence. In other words, it is good to look competent, but bad to look dominant.

Regarding the comparisons between CIs of opposing poles of the same dimension, the results indicate that each CI is judged as intended in the trait that was used to generate it. For instance, the friendly-CI is judged as more friendly, and the unfriendly-CI is judged as more unfriendly. Nevertheless, each pair of CIs also tends to exhibit strong differences in other trait dimensions. A notable exception to this trend is the case of dominance-CIs, with the dominant-CI and the submissive-CI maximally differing in their respective traits. This suggests that the facial content of the dominance CIs predicted more successfully their intended trait judgment. Moreover, it also suggests that the facial content of the competence, warmth, and trustworthiness CIs may share more associations with the same trait judgments. This is in line with Study 1's findings showing that there is a high overlap between the facial content of warmth and trustworthiness CIs, and further suggests that these similarities are extended to the competence-CIs.

CI comparisons. Data from Figure 2 suggest that the warmth and the trustworthiness CIs were judged more similarly across all trait dimensions (smaller effect sizes) than were the dominance and competence-related CIs. Such a pattern adds to our claim that, unlike trustworthiness and warmth, competence and dominance are not redundant in social face perception, and replicates Study 1's objective findings at the subjective level.

Comparing Studies 1 and 2: Objective and Subjective Similarities between CIs

We submitted the objective and subjective data from both studies to a Multidimensional Scaling (MDS) and property fitting analyses (ProFit; Chang & Carroll, 1969; Rosenberg et al., 1968) to visualize how trait judgments were predicted by the pattern of physical similarities between CIs.

Multidimensional Scaling

MDS is a dimensionality reduction technique that allows us to represent the degree of similarity between a set of objects (e.g., CIs) in terms of spatial distances in an *n*-dimensional

Euclidean space. Because the MDS requires dissimilarities (between objects) as input data, we converted the CI correlation matrix into a dissimilarity matrix. These dissimilarities were then submitted to an MDS (using the R package smacof version 1.8; De Leeuw & Mair, 2009). Next we selected the optimal number of dimensions characterizing the differences between the CI pixel data, based on the *Stress-I* goodness-of-fit index and multiple R^2 computed for several dimensionalities (Kruskal & Wish, 1978). A two-dimensional solution proved to be the most parsimonious solution with a good fit to the data (*Stress-I* = .082; R^2 = .952; see Table S2 in the online supporting information for higher dimensionality results). A 2D solution also has the advantage of yielding a more intelligible 2D plot, depicted in Figure 3.

Figure 3 represents a two-dimensional Euclidean space across which the dots representing the CIs (CI objects) are scattered according to their (dis)similarity along two dimensions. That is, similarly to the pixel correlations, the distances between CI objects in the 2D space represent their degree of similarity (correlation). The closer the CI objects are in space, the more physically similar they are.

The axes representing the two dimensions were not labeled, as this was not the goal of our analysis. Nevertheless, the differentiation between CIs along Dimension 1 suggests that this dimension can be interpreted as valence. Dimension 2 contrasts CIs of low ability with high ability (i.e., submissive vs. dominant or intelligent). In this regard, this 2D solution is in agreement with the structure of both the trustworthiness-by-dominance and warmth-by-competence models.

Figure 3 shows that CIs of opposing poles of the same dimension are located farther apart from each other, with trustworthiness and warmth CIs differentiated along the first dimension and competence-related and dominance CIs differentiated along both the first and second dimensions. Moreover, trustworthiness and warmth CIs practically overlap in physical similarity, unlike competence-related and dominance CIs.

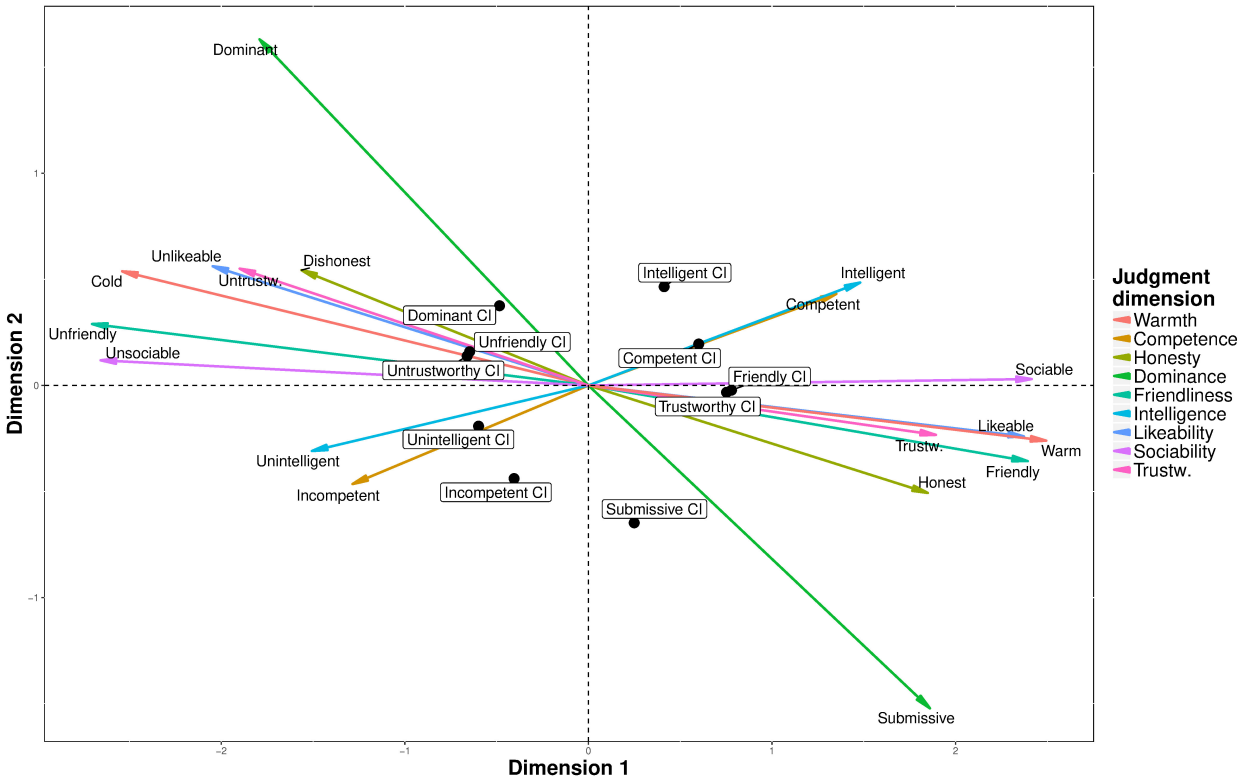

Figure 3. Two-dimensional Euclidian space yielded by the MDS. Black dots correspond to CIs. The higher the distance between CIs, the more physically dissimilar they are. Vectors correspond to trait judgments. Vector arrows indicate the direction towards which CIs were judged higher in a trait, and vector length indicates the magnitude of the judgment. The color version of this figure can be found in the article online, or be reproduced using the R script included in our online data repository.

Property Fitting

A ProFit analysis using the trait judgments allowed us to examine how the coordinates of CIs in the 2D space predicted specific trait judgments, thus informing about the relationships between those trait dimensions in social face perception. In separate multiple regressions, the raw ratings of each trait judgment (e.g., dominant; submissive) across all CIs were regressed on the CI coordinates on the two dimensions (yielded by the MDS). The output of each regression was subsequently plotted as a vector in the 2D space (see Figure 3).

The goodness-of-fit values obtained from fitting the 18 trait judgments to the 2D configuration are available in the online supporting information (Table S3), and provide a measure of how the spatial coordinates of the CIs in the 2D space account for each judgment

(n = 310 observations per judgment). The models were highly significant across all the selected trait dimensions suggesting that they were all relevant to interpret the pattern of physical similarity between CIs. Judgments of sociability, warmth, and likeability yielded the highest values (R^2's between .45 and .67), suggesting that these traits explain most of the discriminability between CIs. Dominance and morality-related judgments (i.e., trustworthiness and honesty) showed slightly lower values (R^2's between .50 and .33), but still account for considerable variability. Finally, the competence-related judgments exhibit the lowest values (R^2's between .35 and .28), suggesting that (1) the variability in these judgments may be overlapping with other trait judgments (e.g., valence-related), (2) that additional dimensions would be needed to better accommodate competence judgments, or (3) simply be reflecting that there is low agreement between participants regarding face judgments of competence.

The angle between any two vectors is a geometric indicator of their correlation (see Rodgers & Nicewander, 1988). Therefore, the angles between the vectors plotted in Figure 3 inform about the relationship between trait judgments. Acute-angles indicate positive correlations, 90º angles indicate orthogonality, and obtuse-angles indicate negative correlations. The perceptual map in Figure 3 clarifies that the vectors of all trait pairs chosen to represent opposite poles of a dimension are related by approximately 180º (mean angle = 175º, SD = 3.02º), corroborating that they indeed define one dimension.

Relevant for our goals, we found a positive correlation between warmth and trustworthiness judgments (1º at high pole, 4 º at low pole) suggesting that these dimensions tend to be conflated in face impressions. Moreover, the low mean angle between warmth, trustworthiness, friendliness, sociability, and honesty, and likeability (mean angle = 6º, SD = 4.68º, for the high-poles and, mean angle = 8º, SD = 5.16º for the low-poles) suggests that a strong common factor (likely valence) underlies all these judgments. Consistent with an evaluative dissociation between dominance and competence, dominance and competence-related judgments were negatively correlated (120º for Dominant-Competent; 120º for Dominant-Intelligent; 143º for Submissive-Incompetent, 121º for Submissive-Unintelligent).

The relationship established between likeability and dominance judgments was opposite to the relationship between likeability and competence-related judgments. The angles between likeability and competence indicate a high positive correlation (24º at high pole, 35 º at low pole). However, the correlation was highly negative between likeability and dominance (143º at high pole, 156 º at low pole). We further examined how these differences in the relationship with valence played a role in the negative correlation found between the

Dominant and Competent judgments. We partialled out the relationship with likeability from the equation, and found that the relationship between the dominance and competence judgments becomes orthogonal (partial $r(307) = .045$, *ns*, for high poles; partial $r(307) = .062$, *ns*, for low poles). This suggests that the dissociation between these dimensions is driven by their relationship with valence.

Figure 3 also informs that all judgments are more differentiated at their negative (i.e., as they are perceived in valence) than at their positive poles. This pattern is consistent with the density hypothesis according to which negative information is more differentiated than positive information in memory (Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008). This density effect has been previously generalized across hundreds of positive and negative words, images and daily events (Koch, Alves, Krüger, & Unkelbach, 2016), and agrees with findings by Bruckmüller and Abele (2013) showing that the positive poles of the Big Two (i.e., communion and agency dimensions) are perceived as more similar than their negative poles. This higher discrimination underlying negative traits should also be taken into account to understand the lower inter-rater agreement values observed in both studies (see Tables 1 and Table S1 in supporting information).

In sum, the overall results clearly suggest that competence and dominance did not overlap neither in terms of expected facial content, nor in their relationship with valence. On the other hand, warmth and trustworthiness seem to overlap in both facial content and in their relationship with valence. Moreover, competence and intelligence overlapped to a greater extent with warmth and trustworthiness, in both their expected facial content and relationship with valence. In sum, these inter-relationships appear to stem mainly from the relationship that each of these dimensions establishes with valence.

General Discussion

Our research offered data suggesting that the basic dimensions of person perception and of social face perception should not be expected to be redundant in the context of face-driven impressions. As hypothesized, we observed an overlap between warmth and trustworthiness and a dissociation between the dominance and competence dimensions. Moreover, our results suggest that the dominance and competence dissociation is driven by their relationship with valence. In this way, our results add to other findings supporting the hypothesis that there is a lower overlap between dominance and competence than between warmth and trustworthiness in face impressions. First, they add to indirect evidence offered by Koch et al.'s (2016) work showing that socioeconomic success overlaps with agency

(dominance) but not competence, and by Lange and Crusius's (2015) work showing that bodily displays of dominance versus competence (hubristic vs. authentic pride) lead to distinct perceptions (malign vs. benign envy) which are dissociated in valence. Second, they add to more direct evidence offered by Oosterhof and Todorov's (2008) work, by specifically focusing on the relationship between dimensions. Finally, our findings are consistent with Sutherland et al.'s (2016). Note, however, that our approach inverts the logic used by Sutherland et al. (2016) to test the hypothesis. Sutherland et al. (2016) asked their participants to attribute trait ratings to faces and subsequently examined how these ratings were correlated. In our research, we asked participants to select the physical features that elicited a given trait and subsequently examined the physical and perceived overlap of these features across trait dimensions. In this way, the present work derives the dissociation between dominance and competence, not only from how traits are communicated by a face, but also from how facial features are communicated by traits, via a methodology that is highly sensitive to the content of internal representations of trait concepts: the RC methodology. This methodology allowed us to explore physical similarities between the facial content selected by participants to represent relevant traits (Study 1). The subsequent judgments of the CIs by independent judges (Study 2) showed that: (a) the CIs obtained in Study 1 were perceived as intended in the traits from which they were derived (i.e., contained trait signal); (b) competence-related and dominance judgments were especially differentiated in their relationship with valence; and lastly, (c) negative trait judgments were more clearly differentiated than positive trait judgments. These data offer compelling evidence that the person perception and face evaluation 2D models, although related, are not redundant in face impressions.

Although other studies offer data similar to ours, there is at least one study (Walker & Vetter, 2016) that found an overlap between facial dominance and a dimension that has been argued to be akin to the competence dimension: perceived agency (see Cuddy et al., 2008). This evidence raises the question of what may be the relevance of the models selected to examine the relationship between dimensions. Both Sutherland et al.'s (2016) studies and ours anchored in ratings of warmth and competence based on the model's labels, whereas Walker and Vetter's (2016) data were based on the communion and agency (Big Two) model (Abele & Wojciszke, 2007; Bakan, 1966) and were obtained, not via direct trait ratings, but via a questionnaire (viz. GEPAQ; Runge, Frey, Gollwitzer, Helmreich, & Spence, 1981). This questionnaire measures perceptions of masculinity-femininity, and its adoption is assumed to be based on the overlap found between masculinity and femininity and the agency and communion constructs, respectively (Abele & Wojciszke, 2007). But more importantly, the

questionnaire assesses agency/masculinity with items that directly measure perceived dominance, competitiveness, and aggressiveness, but never competence. As a result, Walker and Vetter's (2016) studies could only find an overlap between the perceived agency/masculinity and the perceived dominance of faces. In this regard, they are not informative about the status of an overlap between dominance and competence, because they never assess competence.

Other relevant information is obtained from our data. First, it shows a halo pattern between face judgments of trustworthiness and warmth that seems to be extended to competence, but not to dominance judgments. Whereas a competent face is judged as positive (e.g., likeable), a dominant face is judged as negative (e.g., unlikeable). This shows that competence and dominance face judgments will diverge given their opposite relationship with valence. This valence-driven dissociation also seems to extend beyond the present context of face impressions. Previous studies on group perception showed that competence and potency (i.e., power-related dimension of the semantic differential model; Osgood, Suci, & Tannenbaum, 1957) exhibit opposite correlations with warmth and valence (Kervyn et al., 2013). Second, although inter-rater agreement differs across dimensions, it is especially low for the competence-related traits. This may be informative about the perceivers' difficulty in mapping their representation of competence onto a face space, and/or about the variability in perceivers' expectations about the diagnostic facial features of competence (e.g., see Hehman et al., 2017). Although our data suggest that dominant and competent faces do not overlap at the group-level, their relationship may be more malleable due to low inter-rater agreement, especially at the low poles. For instance, an overlap between the CIs may emerge at the individual level in cases where the perceiver construes dominance as overlapping with competence. Third, our data also contributes with information about how trait-space elements (i.e., trait representations) take shape in the face-space (i.e., diagnostic face features), which is highly relevant for current theoretical approaches (Over & Cook, 2018; Stolier, Hehman, & Freeman, 2018; Tamir & Thornton, 2018) aiming to integrate face perception and personality inferences. Specifically, our results clarify that valence plays an important structuring role in shaping the relationship between the trustworthiness-by-dominance and the warmth-by-competence models.

Our studies are, however, not free of limitations, one being the unbalanced gender materials and participants' distribution. It may be argued that the fact that we conducted our studies only with male base faces and a predominantly female sample, does not offer the ideal setting to test our hypothesis. However, this is not incompatible with our aims. From a

Popperian point of view, corroborations are logically always less informative than refutations (Popper, 1985). And although we cannot conclude that our results will necessarily generalize to the ratings of male participants and female targets, we can claim that competence and dominance do not always map very close onto each other. And that is indeed informative. The possibility that a dissociation between dominance and competence occurs in at least one of the cases suggests the equivalence of the models derived from traits and faces is false. But our results do not exclude the hypothesis that gender plays an important role in the possibility of great overlap between conceptual and face models in the case of female respondents. This is a possibility since picturing a dominant man may—for a female respondent—evoke very different associations than a dominant woman. Fortunately, Sutherland et al.'s (2016) work already clarifies this possibility. These authors balanced participant and face target gender and found that the dissociation between dominance and competence occur in both cases; just that it is lower for male targets comparatively to female targets. Nevertheless, in fields of research where this moderation by target gender proves to be relevant, it may also be relevant to address it with data-driven methods that place the least amount of restraints on the target stimuli and are sensitive to top-down biases (e.g., Dotsch, Wigboldus, & van Knippenberg, 2011), such as RC methods.

A second criticism we may make to our studies is that in order to assess internal face representations of personality dimensions, we need to label them. Because our hypotheses anchor in the warmth-by-competence and social-by-intellectual models (Fiske et al., 2007; Rosenberg et al., 1968) we labeled the two dimensions as warm (friendliness) and competence (intelligence). There was no guarantee that the naïve use of these labels would capture such dimensions. In addition, with a distinct theoretical background we could have used other labels to assess the two dimensions. Unlike the warmth and competence trait-based labels, the communion and agency constructs are assumed to capture a broader range of trait content (Abele & Wojciszke, 2007) and can be branched into two facets each: Communion into warmth/sociability and morality, and agency into assertiveness and competence (Abele et al., 2016). If future studies replicate our data using a facets approach, we may discover that face impressions are better captured by the assertiveness facet of agency, and by a general valence dimension that aggregates the competence, warmth/sociability, and morality facets, taking into account the results obtained by us, Sutherland et al. (2016) and Walker and Vetter (2016).

A third criticism pertains to the absence of prior power calculations. We must take two aspects into consideration to discuss it. The first is related with a limitation of the field itself regarding power implications for the reliability of the CIs. Although attempts have been made

to overcome this limitation (see Brinkman et al., 2018) the current use of noise-based reverse correlation is still "hampered by the lack of methodological work addressing validity, reliability, and guidelines for best practice" (Brinkman et al., 2017, p.352). Until an optimal power analytical approach is established, the most viable option is to adopt the task parameters used in previous RC studies, as we have done here. A suggestion advanced by Brinkman et al., (2018) after our data collection, was that researchers should include at least 500 trials in a RC task to obtain reliable individual CIs (measured with a new index, infoVal, which assesses the amount of signal in a CI relative to CI data derived from random responses). But only future approaches would show how this is relevant to studies like ours that rely on CIs generated from 6000 trials[2] provided by 20 different individuals. The second point is that our statistical inference was not based on participant as the unit of analysis. Participants can be understood as the judges whose judgments allowed us to estimate the pixel luminance values used in our analysis. Thus, the units of analysis in Study 1 were pixel luminance values (n = 38958 per masked CI) and in Study 2 we assessed the properties of the CIs with multiple regressions and property fitting analyses based on 310 ratings (per trait judgment).

Finally, future research should account for how our impressions of personality are *gestaltic* despite being drawn from different stimulus modalities and qualitatively different information.

In sum, the current studies present compelling evidence supporting a divergence between dominance and competence in social face perception, which appears to be driven by the relationship that each dimension establishes with valence. Furthermore, it demonstrates the usefulness of RC methods in assessing and comparing approximations of socially meaningful visual information associated with fundamental social dimensions.

Acknowledgments

This research had the support of Fundação para a Ciência e Tecnologia (FCT) through the doctoral grant PD/BD/113471/2015 awarded to Manuel Oliveira, and the strategic project UID/PSI/04810/2013 granted to the William James Center for Research. We thank Elizabeth Collins and Rui Costa-Lopes for their feedback during various stages of this project and Anneloes Kip and Márcia Oliveira for their comments on earlier drafts of this paper.

[2] Please notice that a single trial in a 2IFC RC task is less informative (i.e., contains less signal) when compared to, for instance, a single item in a typical survey. This should be taken into consideration when discussing the amount of trials used in RC studies.

References

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*. doi: 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*, 751–763. doi: 10.1037/0022-3514.93.5.751

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. *Advances in Experimental Social Psychology*, *50*, 195–255. doi: 10.1016/B978-0-12-800284-1.00004-7

Bakan, D. (1966). *The duality of human existence*. Reading, PA: Addison-Wesley.

Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition*, *32*, 397-408. doi: 10.1521/soco.2014.32.4.397

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, *41*, 135-143. doi: 10.1002/ejsp.744

Brinkman, L., Goffin, S., van de Schoot, R., van Haren, N., Aarts, H., & Dotsch, R. (2018). Quantifying the informational value of classification images. Retrieved from https://psyarxiv.com/e5h8y/

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*, 333–361. doi: 10.1080/10463283.2017.1381469

Bruckmüller, S., & Abele, A. E. (2013). The density of the Big Two: How are agency and communion structurally represented? *Social Psychology*, *44*, 63–74. doi: 10.1027/1864-9335/a000145

Carrier, A., Louvet, E., Chauvin, B., & Rohmer, O. (2014). The primacy of agency over competence in status perception. *Social Psychology*, *45*, 347–356. doi: 10.1027/1864-9335/a000176

Chang, J. J., & Carroll, J. D. (1969). *How to use PROFIT, a computer program for property fitting by optimizing nonlinear or linear correlation*. Murray Hill, NJ: Bell Telephone Laboratories.

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. doi: 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*, 103–125. doi: 10.1037/a0030398

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. *Advances in Experimental Social Psychology*, *40*, 61–149. doi: 10.1016/S0065-2601(07)00002-0

De Leeuw, J., & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, *31*, 1–30. doi: 10.18637/jss.v031.i03

Dotsch, R. (2015). rcicr: Reverse correlation image classification toolbox (R package version 0.3.0) [Computer program package]. Retrieved from https://cran.r-project.org/web/packages/rcicr/index.html

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, *3*, 562–571. doi: 10.1177/1948550611430272

Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of Personality and Social Psychology*, *100*, 999–1014. doi: 10.1037/a0023026

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, *27*, 67–73. doi: 10.1177/0963721417738825

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*, 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*, 878–902. doi: 10.1037//0022-3514.82.6.878

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*, 148–168. doi: 10.1037/a0034726

Gosselin, F., & Schyns, P. G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychological Science*, *14*, 505-509. doi: 10.1111/1467-9280.03452

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. doi: 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*, 165–196. doi: 10.1016/S1090-5138(00)00071-4

Imhoff, R., Dotsch, R., Bianchi, M., Banse, R., & Wigboldus, D. H. J. (2011). Facing Europe: Visualizing spontaneous in-group projection. *Psychological Science*, *22*. doi: 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*. doi: 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. doi: 10.1146/annurev-psych-010416-044242

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*, 673–681. doi: 10.1002/ejsp.1978

Koch, A., Alves, H., Krüger, T., & Unkelbach, C. (2016). A general valence asymmetry in similarity: Good is more alike than bad. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*, 1171–1192. doi: 10.1037/xlm0000243

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*, 675–709. doi: 10.1037/pspa0000046

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, *7*, 730–735. doi: 10.1037/1528-3542.7.4.730

Kruskal, J., & Wish, M. (1978). *Multidimensional Scaling*. Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781412985130

Lange, J., & Crusius, J. (2015). The tango of two deadly sins: The social-functional relation of envy and pride. *Journal of Personality and Social Psychology*, *109*, 453–472. doi: 10.1037/pspi0000026

Leach, C., 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*, 234–249. doi: 10.1037/0022-3514.93.2.234

Lundqvist, D., Flykt, A., & Ohman, A. (1998). The Karolinska directed emotional faces (KDEF) [Face database]. Stockholm, Sweden: Karolinska Institutet. Retrieved from http://kdef.se/

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*, 11087–11092. doi: 10.1073/pnas.0805664105

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Oxford, England: University of Illinois Press.

Over, H., & Cook, R. (2018). Where do spontaneous first impressions of faces come from? *Cognition*, *170*, 190–200. doi: 10.1016/j.cognition.2017.10.002

Popper, K. (1985). *Realism and the aim of science: From the postscript to the logic of scientific discovery*. London: Routledge. doi: 10.4324/9780203713969

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *American Statistician*, *42*, 59–66. doi: 10.1080/00031305.1988.10475524

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*, 283–294. doi: 10.1037/h0026086

Rosenberg, S., & Sedlak, A. (1972a). Structural representations of implicit personality theory. *Advances in Experimental Social Psychology*, *6*, 235–297. doi: 10.1016/S0065-2601(08)60029-5

Rosenberg, S., & Sedlak, A. (1972b). Structural representations of perceived personality trait relationships. In A. K. Romney, R. N. Shepard, & S. B. Nerlove (Eds.), *Multidimensional scaling* (pp. 133–162). New York, NY,: Seminar Press.

Runge, T. E., Frey, D., Gollwitzer, P. M., Helmreich, R. L., & Spence, J. T. (1981). Masculine (instrumental) and feminine (expressive) traits: A comparison between students in the United States and West Germany. *Journal of Cross-Cultural Psychology*, *12*, 142–162. doi: 10.1177/0022022181122002

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, *9*, 260–264. doi: 10.1037/a0014681

Secord, P. F., & Bevan, W. (1956). Personalities in faces III: A cross-cultural comparison of impressions of physiognomy and personality in faces. *Journal of Social Psychology*, *43*, 283–288. doi: 10.1080/00224545.1956.9919224

Stolier, R. M., Hehman, E., & Freeman, J. B. (2018). A dynamic structure of social trait space. *Trends in Cognitive Sciences*, *22*, 197–200. doi: 10.1016/j.tics.2017.12.003

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*, 105–118. doi: 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. doi: 10.1016/j.cognition.2016.09.006

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in Cognitive Sciences*, *22*, 201–212. doi: 10.1016/j.tics.2017.12.005

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social Cognitive and Affective Neuroscience*, *3*, 119–27. doi: 10.1093/scan/nsn009

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*, 455–60. doi: 10.1016/j.tics.2008.10.001

Toscano, H., Schubert, T. W., & Sell, A. N. (2014). Judgments of dominance from the face track physical strength. *Evolutionary Psychology*, *12*, 1–18. doi: 10.1177/147470491401200101

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology*, *95*, 36–49. doi: 10.1037/0022-3514.95.1.36

Walker, M., & Vetter, T. (2016). Changing the personality of a face: Perceived Big Two and Big Five personality factors modeled in real photographs. J*ournal of Personality and Social Psychology*, *110*, 609–624. doi: 10.1037/pspp0000064

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

Chapter VII

Good to bad or bad to bad? What is the relationship between valence and the trait content of the Big Two?*

Manuel Oliveira[1†], Teresa Garcia-Marques[1], Leonel Garcia-Marques[2], Ron Dotsch[3],

Abstract

In this article we directly assessed the relationship between valence and relevant traits of the Big Two dimensions (i.e., communion and agency). Drawing on previous research, we expected that the relationship with valence would be less monotonous and more variable in direction across agency-related traits, compared to communion-related traits. In three repeated measures studies we assessed the perceived valence of each trait dimension on a continuum of seven points. Studies 1 and 2 defined each continuum verbally. In Study 3 each continuum was defined by facial features. Results across these studies show that valence is linearly and more consistently related with communion-related traits than with agency-related traits. Within agency, however, competence established a positive linear relationship with valence, whereas dominance showed a target-sensitive relationship with valence: quadratic in evaluation of trait concepts, and negative and linear in face evaluation. We discuss the implications of these data for Big Two-related research.

Key-words: Big Two, communion and agency, competence and dominance, face perception, person perception, valence

---

[1] William James Center for Research, ISPA – Instituto Universitário, Lisbon, Portugal.

[2] Faculdade de Psicologia, Universidade de Lisboa, Lisbon, Portugal.

[3] Utrecht University, Utrecht, The Netherlands.

* Article published in the *European Journal of Social Psychology*, (Advance online publication). doi: 10.1002/ejsp.2618

[†] Correspondence to: Manuel Oliveira, William James Center for Research, ISPA—Instituto Universitário, R. Jardim do Tabaco 34, 1149-041 Lisbon, Portugal.; E-mail: manueljbo@gmail.com

Good to bad or bad to bad? What is the relationship between valence and the trait content of the Big Two?

Traits such as competence, warmth, honesty, dominance, and trustworthiness play a highly relevant role in how we perceive each other (for reviews see Abele & Wojciszke, 2014; Cuddy, Fiske, & Glick, 2008). This was shown to be the case for inferences about people's behavior (Hastie & Kumar, 1979; Wojciszke, 1994, 2005), trait concepts (Rosenberg, Nelson, & Vivekananthan, 1968; Rosenberg & Sedlak, 1972), groups (Fiske, Cuddy, & Glick, 2007), or people's facial appearance (Oosterhof & Todorov, 2008). These traits were identified as highly representative of the content of two primary and relatively independent dimensions consistently found to underlie social judgments: *communion* and *agency*, or the *Big Two* of social perception (Abele, Cuddy, Judd, & Yzerbyt, 2008; Abele & Wojciszke, 2007). Because all of these traits are inherently imbued with an evaluative meaning (i.e., positive or negative connotation; Kim & Rosenberg, 1980; Peabody, 1970), they are tightly intertwined with interpersonal attitudes (Wojciszke, Abele, & Baryla, 2009). As a result, these trait inferences influence how much we like a person (Anderson, 1968; Wortman & Wood, 2011) or how desirable her personality is (e.g., Hampson, Goldberg, & John, 1987; Rosenberg et al., 1968; Wortman & Wood, 2011). But how traits inform our interpersonal attitudes depends on the nature of their relationship with valence (i.e., positive or negative evaluation attached to an entity; Fiske & Taylor, 2017).

In this article, we review current knowledge on this relationship and offer empirical data able to directly inform about the nature of this relationship within the domains of person perception (Studies 1 and 2) and social face perception (Study 3).

Traits and their Relationship with Valence

The Big Two dimensions underlying social perception have been repeatedly found across time, cultures and different lines of research (for reviews see Abele & Wojciszke, 2014; Cuddy et al., 2008). From a functional perspective, communion encapsulates traits related with the appraisal of intentions and social connection (e.g., trustworthiness, honesty, warmth, sociability). Whereas agency captures traits related with perceived ability and motivation for goal achievement (e.g., competence, dominance, confidence). Given the pervasiveness of valence in personality impressions, it should not be surprising that valence is correlated with both of the Big Two (e.g., Abele & Wojciszke, 2007; Kim & Rosenberg, 1980; Suitner & Maass, 2008). However, previous research has consistently shown that

communion overlaps with valence to a greater extent than agency does. The high positive correlations of valence with communion (Kervyn, Fiske, & Yzerbyt, 2013; Rosenberg et al., 1968; Suitner & Maass, 2008) strongly suggest that communion-related traits and valence express the same underlying evaluative dimension (i.e., social evaluation in terms of perceived positivity). This would suggest that the task of evaluating someone's trustworthiness or warmth is practically indistinguishable from expressing how positive or negative is our global evaluation of the person. This is further substantiated by the fact that valence and trustworthiness are interchangeably used as interpretations of the same primary dimension of face impressions (Oosterhof & Todorov, 2008). Overall, these findings support the prediction that, although agency traits are not independent of valence, valence is strongly and positively related with communion traits.

Furthermore, research suggests that the Big Two are further branched into two facets each (Abele et al., 2016; Brambilla & Leach, 2014; Carrier, Louvet, Chauvin, & Rohmer, 2014). Communion encapsulates a warmth (e.g., warmth, sociability, friendliness) and a morality facet (e.g., trustworthiness, honesty, benevolence). In turn, agency encapsulates a competence (e.g., competence, intelligence) and an assertiveness facet (e.g., dominance, confidence). Two dimensions analogous to the Big Two have also emerged in social face perception. But here, they seem to be less multifaceted, and best represented by trustworthiness (communion morality) and dominance (agency assertiveness) (Oosterhof & Todorov, 2008). This branching suggests that the Big Two may be concealing a more complex relationship with valence, given that their facets vary in their relationship with valence. For instance, previous research has shown that the valence of personality impressions is more strongly determined by morality rather than warmth (e.g., Brambilla & Leach, 2014; Goodwin, 2015): Whereas morality-related traits inform about whether someone's intentions are good or bad, warmth-related traits inform about someone's proficiency in recruiting social support for their intentions (Landy, Piazza, & Goodwin, 2016). Moreover, competence and dominance diverge in their relationship with valence, despite of their common association with agency. Competence-related traits (e.g., competence, intelligence) are perceived as highly positive and likeable, whereas dominance is perceived as slightly negative (close to neutral) and highly unlikeable (Abele, Uchronski, Suitner, & Wojciszke, 2008; Anderson, 1968). Similar findings emerged in social face perception showing that whereas competence is positively correlated with valence, dominance is (slightly) negatively correlated with trustworthiness and valence (Chen, Jing, & Lee, 2014; Oliveira, Garcia-Marques, Dotsch, & Garcia-Marques, 2019; Oosterhof & Todorov, 2008). This opposite relationship with valence

is in agreement with what evolutionary theories of status attainment (Chapais, 2015; Henrich & Gil-White, 2001) would predict: Dominant individuals act in ways that inflict costs on others to benefit themselves, whereas competent individuals act in ways that are beneficial to them through helping others.

A clarification of the relationship between traits and valence is highly relevant for a more complete understanding of the person perception space. Any assumption regarding the independence between the Big Two (e.g., Cislak & Wojciszke, 2008) is challenged at the evaluative level, given the common variance that the Big Two share with valence (Suitner & Maass, 2008). Similarly to communion-related traits, agency-related traits are polarized in valence (e.g., *intelligent* is more likeable than *unintelligent*; Anderson, 1968). However, unlike communion-related traits, the literature suggests that agency-related traits exhibit inconsistent relationships with valence. And within agency, two traits stand out as potential promoters of such inconsistency: competence and dominance.

Although the research reviewed so far documents the basic dimensions underlying personality impressions, it does not address the nature of the relationship that these dimensions establish with valence.

Nature of the Relationship between Valence and the Dimensions of Personality Impressions

Although a few studies have addressed the nature of the relationship between the Big Two (e.g., Imhoff & Koch, 2017), none, to the best of our knowledge, have explicitly focused on the nature of the relationship between the Big Two and valence. It is relevant to understand if the absent or weak linear relationships with valence found in previous research are indirectly expressing the presence of curvilinearity in the data. Lemann and Solomon (1952) provided early evidence about how we perceive the relationship between valence and traits. Their work acknowledged that the nature of that relationship may be either linear or quadratic. To take this into account in the assessment of trait perceptions they proposed two types of scales: *alpha-trait* and *beta-trait* scales. Alpha-trait scales are used when the trait is assumed to exhibit a linear relationship with valence, such that the increased (or decreased) presence of the trait in a target reflects an increase (or decrease) in perceived positivity (i.e., valence). In turn, beta-trait scales are used when the trait is assumed to establish a curvilinear relationship with valence. These are the traits exhibiting an inverted U-shaped relationship with valence. Specifically, the perceived positivity of a beta-trait increases from one extreme to the mid-

point of its dimension continuum, where it reaches a positivity peak, and then starts decreasing towards the other extreme.

The strong positive relationship that communion-related traits establish with valence suggests they fit well with the definition of alpha-traits. However, the nature of agency-related traits such as competence or dominance is less clear. The divergence between competence and dominance regarding their relationship with valence may be signaling that these two diverge in the type of trait (alpha or beta) they best fit with. Several reasons lead us to expect that dominance has a beta nature (i.e., curvilinear relationship). First, the linear relationship between dominance and valence, although negative, tends to be close to neutral (e.g., Oosterhof & Todorov, 2008). Second, the two extremes of dominance seem to be much more similar in valence than a linear relationship would lead us to expect. Submissiveness may be undesirable for being associated with vulnerability (e.g., invites exploitation; Richards, Rollerson, & Phillips, 1991), whereas high dominance conveys a threatening image (e.g., Chen et al., 2014). The same argument may apply to other agency-related traits such as assertiveness or confidence. Too little assertiveness results in passiveness, while too much of it may project aggressiveness (Ames, Lee, & Wazlawek, 2017). And low-to-moderate confidence (cf. self-enhancement) is more socially attractive than overconfidence (Dufner et al., 2013). These considerations agree with the Aristotelian idea that virtue results from a balance between excess and deficiency (e.g., Grant & Schwartz, 2011; Imhoff & Koch, 2017).

Previous work by Imhoff and Koch (2017) offered empirical support for the idea that agency-related traits are more desirable in moderate amounts. At least if we take into account that likeability, warmth and trustworthiness (communion-related traits) can serve as proxies of valence given their strong positive correlation (e.g., Abele, Uchronski, et al., 2008; Anderson, 1968). In their work, Imhoff and Koch (2017) found that social targets are perceived as more likeable and warm at average levels of status, power, and dominance; and as less likeable or warm at extreme levels of agency. That is, they found an inverted U-shaped relationship between communion and agency.

However, the generalizability of that finding across the trait content of agency is put into question by the fact that those studies did not include competence traits. As noted earlier, competence dissociates from dominance with regard to their relationship with valence. Competent individuals are perceived as substantially more positive than incompetent ones (Rosenberg et al., 1968). Therefore, we would expect, not an inverted U-shaped, but a linear relationship between competence and valence. Competence may be beneficial not only to the trait holder—who gains prestige and admiration by others—but also to others who benefit

from collaborations with competent individuals (Henrich & Gil-White, 2001). It may be argued, nevertheless, that competence is a beta-trait. This is suggested by historical records revealing that intellectual giftedness used to be perceived as leading to morally deviant behavior (Hegarty, 2011), or by the existence of the "nerd" stereotype, which blends task-oriented competence with social inability.

Present Research

Our goal is to clarify Imhoff and Koch's (2017) data regarding the alpha and beta nature of communion- and agency-related traits, by directly relating them with perceived valence (instead of relating the Big Two with each other). We also aim to extend those data by clarifying whether competence and dominance are both beta-traits due to their common association with agency, or if competence is instead an alpha-trait, as suggested by previous research.

The three studies presented here focused on how the perceived variability in the expression of a trait along its continuum is related with perceived valence. Our approach relies on a new experimental task. Methods traditionally used in person perception research are purely correlational and do not assess valence independently of the traits themselves. The detection of the relationship requires a method that ensures the capture of the variability of the traits themselves, by directly assessing the perceived valence of the different levels spanning their respective continuum. If this variability is not captured, the result may truncate the "true" relationship or disguise a quadratic relationship as a weak linear one (see Imhoff & Koch, 2017, p. 124). Taking this into account, we developed a new paradigm that allowed us to directly assess the relationship. We used the degree of expression of a trait—semantically defined in Studies 1 and 2, and defined by faces in Study 3—along its dimension continuum points as our independent variable, and valence as the dependent variable.

Our target sample included representative traits of the Big Two and their respective facets (i.e., agency assertiveness, agency competence, communion morality and communion warmth; Abele et al., 2016). Taking into account the high overlap between communion and valence, we expected (i) to offer new data supporting a consistent linear relationship with valence across communion-related traits; and (ii) to conceptually replicate Imhoff and Koch's (2017) work by finding an inverted U-shaped relationship between agency-related traits and valence (instead of communion). However, unlike Imhoff and Koch's (2017) findings would seem to suggest, we do not expect the curvilinearity to be consistent across different agency-

related traits. Specifically, we expect competence-related and assertiveness-related traits to exhibit distinct relationships with valence.

Although it is yet an empirical question whether the same relationships with valence are found when traits are inferred from faces, here we expected the same pattern to emerge regardless of whether the target stimuli were trait words or face stimuli. Of the two independent dimensions hypothesized to underlie social face perception—trustworthiness and dominance (Oosterhof & Todorov, 2008)—only trustworthiness highly overlaps with valence. But also in this domain, competence and dominance were found to dissociate, although by establishing opposite linear relationships with valence (Oliveira et al., 2019). Here, we expected to clarify all these relationships by integrating them in the same study using a new experimental paradigm that directly assesses the relationship of these traits with valence. In Study 3, we directly explore this relationship, expecting a linear relationship for competence and a curvilinear relationship for dominance.

Data Availability and Supporting Information

All the data, materials, and supporting information associated with this work are publicly available online at: https://osf.io/gc4d8/

(Supporting Information can be found in Appendix B of this thesis.)

Study 1

Participants and Design

Forty native English speakers (95% male, 5% female; M_{age} = 33 years, SD_{age} = 10.29) were recruited online via Prolific Academic and participated in the study in exchange for payment (£2.50). Forty-two participants had been initially recruited, but two participants were excluded before the analyses for showing signs of rushing through the task (i.e., invariant responses across blocks). The study was defined by a 12 (Trait dimension) x 7 (Target persons representing trait continuum levels) within-participants design with a valence-score, based on three different ratings (desirability for self, likeability, and valence), as the dependent measure.

Power considerations. Without a basis for effect size estimation, sample size was calculated for a within repeated measures GLM, to estimate a linear and a quadratic contrast of medium size for the 7-point continuum (f = 0.25; G*Power; Faul, Erdfelder, Lang, & Buchner, 2007), for the total of traits analyzed (12 measurements), with an error probability of

α = .05 and 99% power. The calculation suggested $N = 28$ as the optimal sample size, which was increased on the basis of available resources.

Trait Selection and Trait Continuum Design

Our trait sample included trait scales previously used to calculate aggregated scores of communion and agency in Imhoff and Koch (2017), or found to best represent the dimensions and sub-dimensions/facets of person and group perception (see Abele et al., 2016; Brambilla et al., 2011; Fiske, Cuddy, Glick, & Xu, 2002; Rosenberg et al., 1968). Following these criteria, the selected communion-related traits were: warm-cold (*warmth),* sociable-unsociable *(sociability)*, trustworthy-untrustworthy (*trustworthiness)*, honest-dishonest (*honesty)*, sincere-insincere (*sincerity)*, and benevolent-malevolent (*benevolence)*. The selected agency-related traits were: dominant-submissive (*dominance)*, confident-unconfident (*confidence)*, competent-incompetent (*competence)*, intelligent-unintelligent (*intelligence)*, powerful-unpowerful (perceived *power)*, and high status-low status (perceived *status).* Status and power may not be traits *per se*, but they are thought to overlap with competence and dominance (Fiske et al., 2007; Oosterhof & Todorov, 2008) and were previously used in agency-scores (e.g., Imhoff & Koch, 2017).

We created hypothetical target persons to represent the different levels of a trait continuum (see Figure 1). Each trait continuum was composed of seven points (i.e., levels), corresponding to seven target persons. Each point of the continuum corresponded to an explicit quantification of a trait by means of an adverb (e.g., Somewhat; Extremely; see Cliff, 1959) or a verbal quantification (e.g., Much more [trait] than average). For the mid-point we used the label “About average” (in the target dimension). The continuum was bipolar. The points below the mid-point used the low-pole trait of its dimension (e.g., submissive), and the points above the mid-point used the high-pole trait (e.g., dominant).

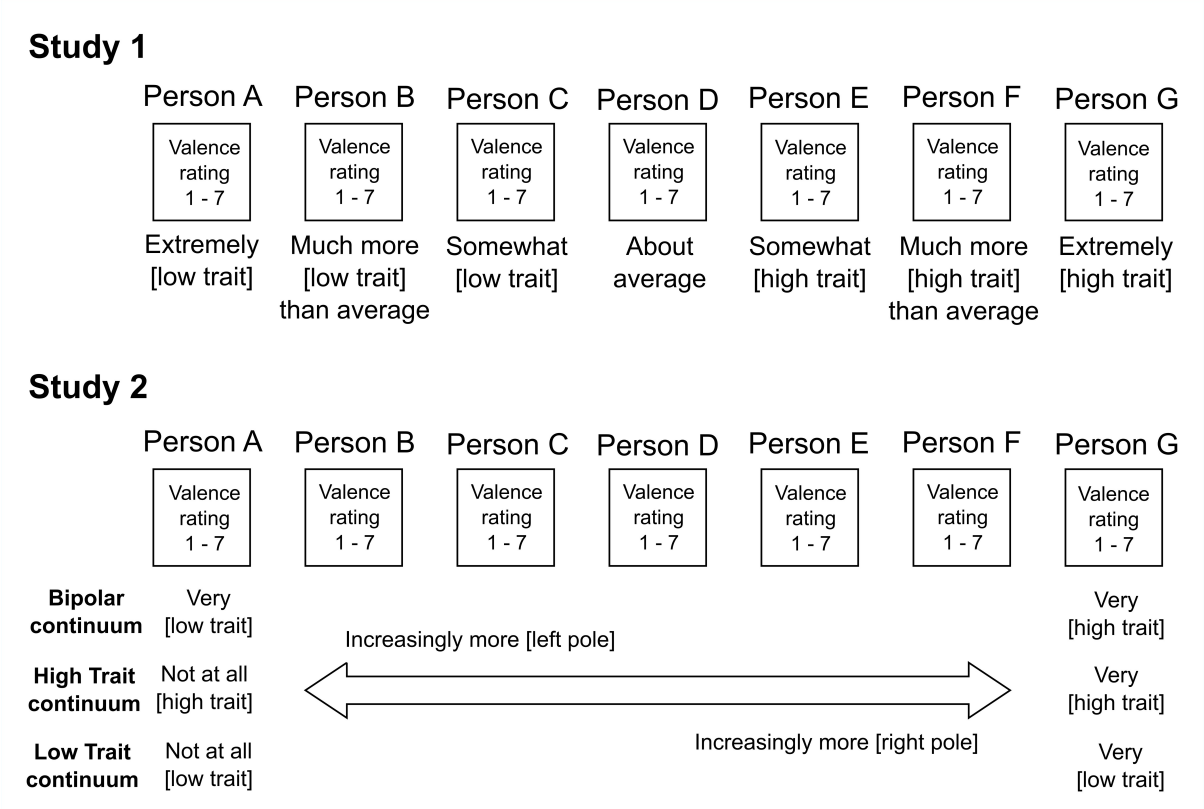

Figure 1. Structure of the trait continua used in Studies 1 and 2. Each continuum point was operationalized as a hypothetical target person to be judged in a valence measure (each target was rated on desirability for self, likeability, and valence). In Study 1, all continuum points were labeled. In Study 2, only the continuum endpoints were labeled, and task instructions emphasized that the degree of expression of the trait increased (in Bipolar and High Trait continuum types), or decreased (in Low Trait continuum type), at each step from the left endpoint to the right endpoint of the continuum.

Dependent Measures

Valence is our main dependent measure. In seeking validity and reliability for our measure, we attended to the fact that in person perception research valence is often measured as perceived likeability (e.g., Anderson, 1968; Wortman & Wood, 2011), goodness/badness or desirability of a trait under particular circumstances (Rosenberg et al., 1968), or positivity associated with the target (i.e., pure valence; Abele, Uchronski, et al., 2008). Assuming that a more general evaluative dimension (i.e., valence) underlies these constructs, each of these three dimensions was assessed on a 7-point rating scale. The valence and likeability scales ranged from 1 (Very bad/unlikeable) to 4 (Neutral or Neither good/likeable nor bad/unlikeable) to 7 (Very good/likeable). The "desirability for self" scale ranged from 1

(Very undesirable in myself / I wouldn't want to be this person) to 4 (Neither desirable nor undesirable / Indifferent) to 7 (Very desirable in myself / I would definitely want to be this person). Each dependent measure had its own specific instruction. All task instructions are available in our online repository. Because the target trait continuum points were operationalized as "Person [A to G]," we adjusted the instruction of the "desirability for the self" block so that the participants had to indicate how much they would desire to be that "particular hypothetical person."

Procedure

The task was programmed using Qualtrics software. Participants were invited to participate in an online study aimed at "understanding how people evaluate several personality characteristics of other people." The task was composed of three blocks, each for a dependent measure (i.e., valence, likeability, desirability for self), presented in randomized order. Trait continua were randomly presented one-by-one within each block. In each trial, participants were shown verbal descriptions (quantifications of the target trait) for each of the seven different persons varying in the degree to which they expressed a given trait. Each target person corresponded to one of the seven continuum points (see Figure 1). To ensure adequate comprehension of the task, a practice trial was first presented using the trait "extraversion" as an example (data not analyzed). In this example, participants were asked to indicate how much they would like: an "Extremely Introverted" person "A"; a "Much Less Introverted Than Average" person "B"; a "Somewhat Introverted" person "C"; a person "D" who is "About Average" in extraversion; a "Somewhat Extraverted" person "E"; a "Much More Extraverted Than Average" person "F"; and finally, an "Extremely Extraverted" Person "G". In the main task, trials were identical to the practice trials, except for the target trait.

To prevent participants from directly mapping, in a linear fashion, the seven points of the rating scale onto the seven continuum points, we emphasized that they should focus their evaluation on what they thought about each target in isolation. At the end of the task, we asked participants whether they were aware that they "could use the same number for two or more people differing in the amount of the same trait?" Together with an inspection of each participant's data, this check served as a criterion to exclude participants who failed to understand the instructions and simply mapped one dimension onto the other. After finishing the task, participants were thanked, debriefed, and compensated.

Results and Discussion

Valence-score

We submitted all three ratings to a principal component analysis (PCA) to ascertain that a single component optimally accounted for their variance. The data points corresponded to the raw response values of each rating for each of the points within a trait continuum. An oblimin rotation was applied to allow for non-zero correlations between the components, and a parallel analysis (Horn, 1965) revealed that one component was sufficient to account for 84% of the variability in the data. As expected, all three ratings loaded highly on that component (desirability for self, likeability, and valence yielded loadings of .91, .92, and .93, respectively), which we interpreted as "general valence". On the basis of these results, we computed a valence-score by averaging the responses of the three measures for every continuum point of every trait dimension.

Inter-rater Agreement

Inter-rater agreement for the valence-score was assessed with two indexes: intra-class correlation coefficients (*ICC*s; see Shrout & Fleiss, 1979) and average inter-rater correlations (*AIC*s; see Brand & Bradley, 2012). Because ICCs are inflated by sample size, we additionally computed the AICs to complement the ICCs and obtain more nuanced and conservative results. Results are listed in Table 1. Both indexes indicate that the lowest agreement occurred for agency assertiveness-related traits. This low agreement is also apparent in the wider dispersion of the valence peak distributions obtained for these traits (see Table 1 and Figure 2).

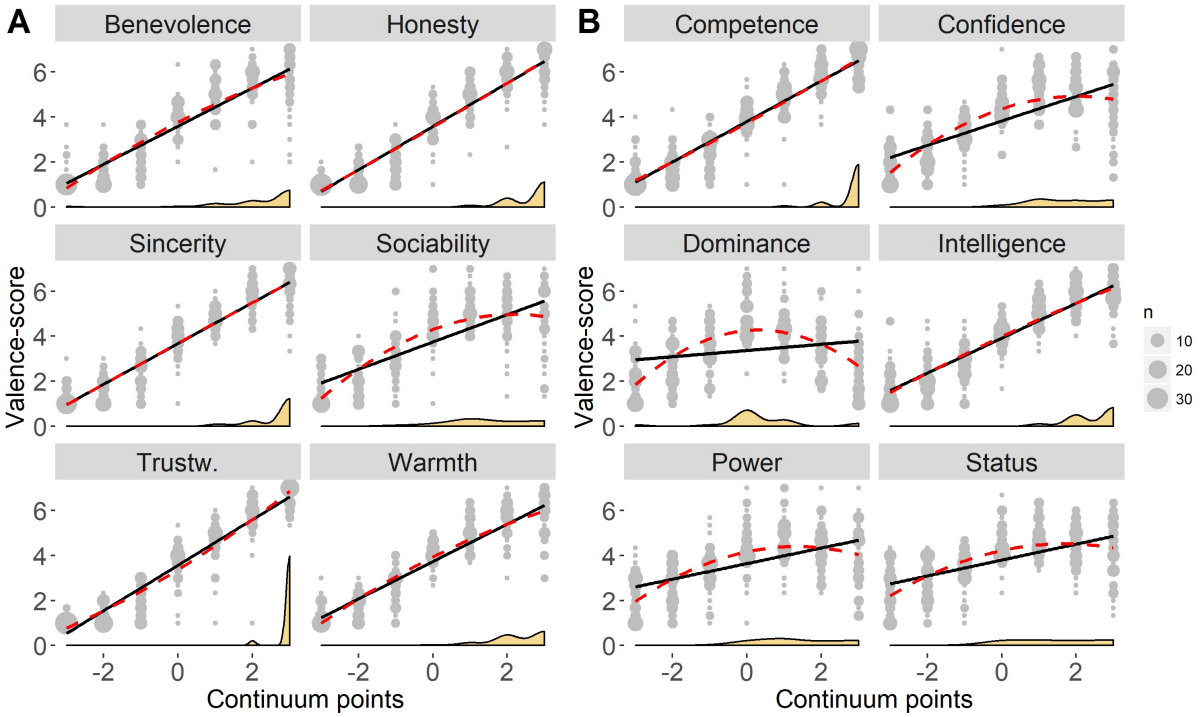

Figure 2. Linear and quadratic regression lines fitted to the valence judgments of (panel A) Communion-related trait dimensions and (panel B) Agency-related trait dimensions, in Study 1. Black lines represent the linear fit. Red dashed lines represent the quadratic fit. The gray dots represent the number of observed ratings (n) and their density along the valence-score scale, for every continuum point (ranging from -3 to +3) of a trait dimension. The density distributions of valence-score peaks across the seven continuum points of a trait dimension are shown at the bottom of each plot (in yellow).

Linear and Quadratic Fits

To examine whether a linear or a quadratic trend better predicted the valence-score on a given trait dimension, we used a linear mixed-effects models approach (LMM; Pinheiro & Bates, 2000). These analyses were conducted in R (version 3.3.2) using the *lme4* and *lmerTest* packages (Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2017). We ran separate LMMs for each trait with valence-score as the outcome variable. To correct for multiple testing we applied a false discovery rate (FDR) correction to all estimates' *p*-values (Benjamini & Hochberg, 1995). In all models, we entered the continuum points as fixed-effect predictors in a quadratic polynomial form, where the first term corresponded to the linear predictor (i.e., *continuum points*) and the squared term (*squared continuum points*) corresponded to the quadratic predictor (curvilinear trend predictor). The seven levels of the

continuum points' factor were set to range from -3 to +3 (and squared values of this range for the quadratic predictor) in the analysis. Additionally, we entered participants as a random-intercept effect, to obtain estimates of the variability of the mean valence-score across participants. Results are listed in Table 1, and plotted in Figure 2. Additionally, we examined where the valence judgments peaked for each trait. For each trait, we performed a local polynomial regression fit on the valence-scores (by participant) and used the fitted model to estimate the location of the valence peaks in the continuum. The density distributions of the estimated valence peaks per trait are plotted in Figure 2. Means of valence peak locations are listed in Table 1. This analysis complements the LMM results by providing a more nuanced description of the relationships established with valence.

Results showed that all the relationships between traits and valence were significantly predicted by the linear component. However, they differed in both the strength of the linear component and whether the relationship exhibited a (significant) quadratic component. Only two communion traits—honesty and sincerity—were not significantly predicted by the quadratic component. This suggests that these traits fit better with the definition of alpha-traits than other traits such as warmth, benevolence, trustworthiness, and sociability. Our review of the literature did not lead us to expect that warmth-related traits such as sociability would exhibit a strong quadratic component. Yet, our results suggest that moderate-to-high sociability is generally preferred to very high sociability. Nevertheless, these data seem to support a distinction between warmth- and morality-related traits previously endorsed in the literature (e.g., Brambilla et al., 2011).

In agreement with previous research, we found that two agency traits—competence and intelligence—are better defined as alpha-traits, as suggested by their non-significant quadratic components. Dominance, in turn, exhibited the expected inverted-U shaped relationship with valence. Importantly, dominance was the only trait for which the quadratic coefficient was higher than its linear counterpart, and the one exhibiting the lowest linear and highest quadratic coefficients of all traits. This strongly suggests that dominance is the only agency-related beta-trait of our sample. Despite of exhibiting higher quadratic coefficients than most communion-related traits, power, status, and confidence showed clear linear relationships with valence, accompanied by weaker curvilinear trends.

Our data suggest that, as expected, not all agency-related traits share the same relationship with valence. The stark contrast between the dominance and the competence-related plots suggests that, despite being related with the same agency dimension, dominance

is a beta-trait, whereas competence is an alpha-trait. Finally, against our prediction, not all communion-related traits exhibited a pure alpha nature.

Table 1
Study 1 Results Including Linear and Quadratic Unstandardized Regression Coefficients, Mean Peak Locations, and Inter-Rater Agreement for the Valence-Score by Trait Dimension.

Big Two	Trait	Mean Peak	Intercept	Linear *b*	Quadratic *b*	ICC(2, *k*)	*AIC*
Communion	Benevolence	2.30 (1.18)	3.77 (0.37)	**0.85** ***	**-0.04** **	.994***	*0.95*
	Honesty	2.65 (0.58)	3.55 (0.31)	**0.96** ***	0.01	.997***	*0.96*
	Sincerity	2.71 (0.59)	3.69 (0.29)	**0.91** ***	0.00	.997***	*0.96*
	Trustw.	2.95 (0.22)	3.37 (0.38)	**1.01** ***	**0.05** ***	.997***	*0.97*
	Sociability	1.52 (1.16)	4.30 (0.40)	**0.61** ***	**-0.14** ***	.986***	*0.83*
	Warmth	2.35 (0.77)	3.93 (0.24)	**0.83** ***	**-0.05** ***	.996***	*0.94*
Agency	Confidence	1.80 (0.97)	4.36 (0.24)	**0.54** ***	**-0.13** ***	.986***	*0.81*
	Dominance	0.32 (1.25)	4.25 (0.46)	**0.14** ***	**-0.22** ***	.966***	*0.63*
	Power †	1.40 (1.17)	4.16 (0.33)	**0.35** ***	**-0.13** ***	.967***	*0.65*
	Status †	1.45 (1.22)	4.22 (0.32)	**0.35** ***	**-0.11** ***	.971***	*0.66*
	Competence	2.85 (0.43)	3.74 (0.35)	**0.90** ***	0.02	.996***	*0.96*
	Intelligence	2.50 (0.64)	3.99 (0.21)	**0.78** ***	-0.02	.995***	*0.93*

Note: Valence-score values range from 1 to 7. Mean peak locations for the valence-score range from -3 to 3 (i.e., continuum point values). Significant unstandardized regression coefficients are in bold, and correspond to the fixed effect predictors (Linear *b* = continuum points; Quadratic *b* = squared continuum points). Intercept and its standard deviation refer to the random-intercept by participant effect and represent the between-participant variability of the mean valence-score per trait. *ICC*s indicate inter-rater agreement for the target judgments (*k* = 40, i.e. number of raters). *AIC* = Average inter-rater correlation (i.e., zero-order correlation of all possible raters within trait). † Power and Status are not traits per se, but have been used to measure agency in previous research (see Study 1 Method section). A Benjamini-Hochberg (*FDR*) correction was applied to all *p*-values of linear and quadratic estimates. $\alpha = .05$; $^{**}p < .01$; $^{***}p < .001$

Correlation between Traits' Valence

To better understand how the relationship between traits and valence may be interfering with the relationship between the fundamental dimensions, we computed the correlations between all traits' raw valence ratings (including all three measures). As expected, alpha-traits exhibited stronger correlations between them than beta-traits (see Table 2). Overall, the correlational pattern between communion-related traits supports that their common variance should be expected if they share an underlying dimension (except for sociability). The same occurred for all agency-related traits, with the exception of competence and intelligence. Competence and intelligence exhibited a strong positive correlation between themselves and, remarkably, with other communion-related alpha-traits. This suggests that valence promotes the association between competence and communion. This interferes with the assumption of their independence (as claimed in the literature), which may only emerge when this relationship with valence is partialled out (but see Suitner & Maass, 2008, who found a negative relationship between communion and agency after partialling out valence).

Table 2

Pearson Correlations between All Trait Dimensions' Valence Ratings.

	1	2	3	4	5	6	7	8	9	10	11
1. Benevolence											
2. Honesty	.83										
3. Sincerity	.82	.89									
4. Trustw.	.82	.91	.89								
5. Sociability	.65	.67	.67	.64							
6. Warmth	.81	.85	.85	.85	.71						
7. Competence	.79	.85	.88	.88	.66	.83					
8. Intelligence	.75	.78	.82	.79	.63	.79	.86				
9. Confidence	.64	.64	.65	.59	.68	.64	.67	.72			
10. Dominance	.29	.24	.25	.20	.43	.31	.27	.35	.53		
11. Power	.51	.46	.48	.42	.61	.52	.50	.56	.70	.61	
12. Status	.51	.47	.50	.43	.61	.52	.51	.59	.70	.57	.75

Note: Communion-related traits numbered from 1 to 6. Agency-related traits numbered from 7 to 12. All correlations' *p*s < .001, α = .05.

Study 2

Study 2 addresses three limitations of Study 1. One caveat of Study 1 relates to the presentation style and specific properties of the continuum. An over-specification of continuum point labels, the bipolar nature of the continuum, or the linguistic properties of traits selected to represent opposite poles of a continuum could have induced the observed evaluations. In Study 2, we addressed this possibility by manipulating the presentation style of continua. A second limitation of Study 1 is the unbalanced number of traits across Big Two facets. In this study, we counterbalanced the number of traits per facet. This allowed us to conduct additional exploratory analyses at the facet level. Finally, we now counterbalanced participant gender to overcome a possible gender bias in Study 1.

Participants and Design

Sixty native English speakers (50% female, 50% male, M_{Age} = 34.02 years, SD_{Age} = 11.25) were recruited via Prolific Academic to participate in the study in exchange for payment (£4.16). All 60 participants were included in the analyses. The study was defined by a 3 (Continuum type) × 8 (Trait) × 7 (Continuum points) within-participants design with a valence-score, identical to Study 1's, as the dependent measure.

Power considerations. A use of G*Power similar to the procedure described in Study 1, now with 8 traits and with the addition of the within-participants manipulation of the three continuum types (24 measurements), suggests a sample size of 21 participants. We additionally conducted a statistical power simulation using the R package *simr* (see Brysbaert & Stevens, 2018; Green & Macleod, 2016). This analysis, informed by the 95% confidence intervals obtained for agency assertiveness-related traits in Study 1, suggests a sample size of N = 60 for a well-powered study. This sample size allowed us to detect whether unstandardized regression coefficients as small as ± 0.10 were significantly different from zero (α = .05), with a statistical power of 99%, for either the linear or quadratic predictors.

Continuum Manipulation

To assess the influence of continuum presentation style on the valence ratings, we created three different types of trait continua: a bipolar continuum (BC), a high-pole trait continuum (HTC), and low-pole trait continuum (LTC). The structure of these three types of continuum is illustrated in Figure 1. Unlike in Study 1, all three continuum types only exhibited labels for their two extreme endpoints. Like the bipolar continuum in Study 1, the BC uses two trait words (e.g. submissive for low pole, dominant for high pole). The other two

unipolar types used only one target trait each. In the HTC, the trait represented the high pole of its dimension (e.g., dominant). In the LTC, the trait represented the low pole of its dimension (e.g., submissive).

Dependent Measures

Valence measures were identical to Study 1's.

Procedure

The procedure was in every way identical to Study 1's, with some exceptions. This time there were three blocks defined by trait continuum type (BC, HTC, and LTC). Block order was counterbalanced between-participants. In-between blocks, participants were instructed to pay attention to the upcoming changes regarding the target trait and valence rating. Because there were no labels for intermediate continuum points, the instructions additionally clarified that the degree of expression of a trait increased (BC and HTC), or decreased (LTC), step-by-step from the left endpoint to the right endpoint of the continuum. All task instructions are available in our online repository. After finishing the task, participants were thanked, debriefed and compensated.

Results and Discussion

The continuum points of the LTC were reverse-scored to match the direction rationale of the other two continuum types (e.g., *Very Submissive* as -3 and *Not at all Submissive* as +3). We closely followed the analytical procedure in Study 1, but adapted it to this study's goals. Again, a PCA revealed one component accounting for 84% of the variability in ratings (desirability for self, likeability, and valence loadings were .91, .91, and .93, respectively). Thus, we computed a valence-score exactly as in Study 1.

Continuum Type and Participant Gender Analyses

To examine the effect of continuum type and participant gender on the perceived valence of traits we conducted a 3 (Continuum Type) × 8 (Trait) × 2 (Participant Gender) mixed ANOVA with the last factor between-participants and valence-score as the dependent variable. All effects are reported with Greenhouse-Geisser corrections. The three-way interaction between all factors was non-significant, $F(8.8, 511.3) = 1.35$, $p = .21$. No effects involving Participant Gender were significant, suggesting it had no influence on the results. The significant interaction between Continuum Type and Trait, $F(8.8, 511.3) = 2.29$, $p = .017$,

η_G^2 = .007, indicates that the perceived valence of traits differed across continuum types. Bonferroni post-hoc comparisons clarified that only ratings of dominance differed across continuum types (p < .001), and specifically between the BC and LTC. To understand the impact of the Continuum Type × Trait interaction on the linear and quadratic components, we conducted LMM analyses by trait as in Study 1, separately for each continuum type. An inspection of the LMM results clarified that the linear component of dominance was stronger than its quadratic counterpart but only for the LTC (b_{linear} = 0.30, p_{FDR} < .001; $b_{quadratic}$ = -0.14, p_{FDR} < .001). Nevertheless, even in this continuum the quadratic (linear) component of dominance remained the highest (lowest) of all traits. Overall, and independently of continuum type, the results replicated those obtained in Study 1 and clarify that they cannot be entirely explained by continuum presentation style. LMM results and plots for the BC type are shown in Table 3, and Figures 3 and 4 (but see next section). Additional results for all continuum types, and plots by participant gender, can be found in our online repository (in Supporting Information).

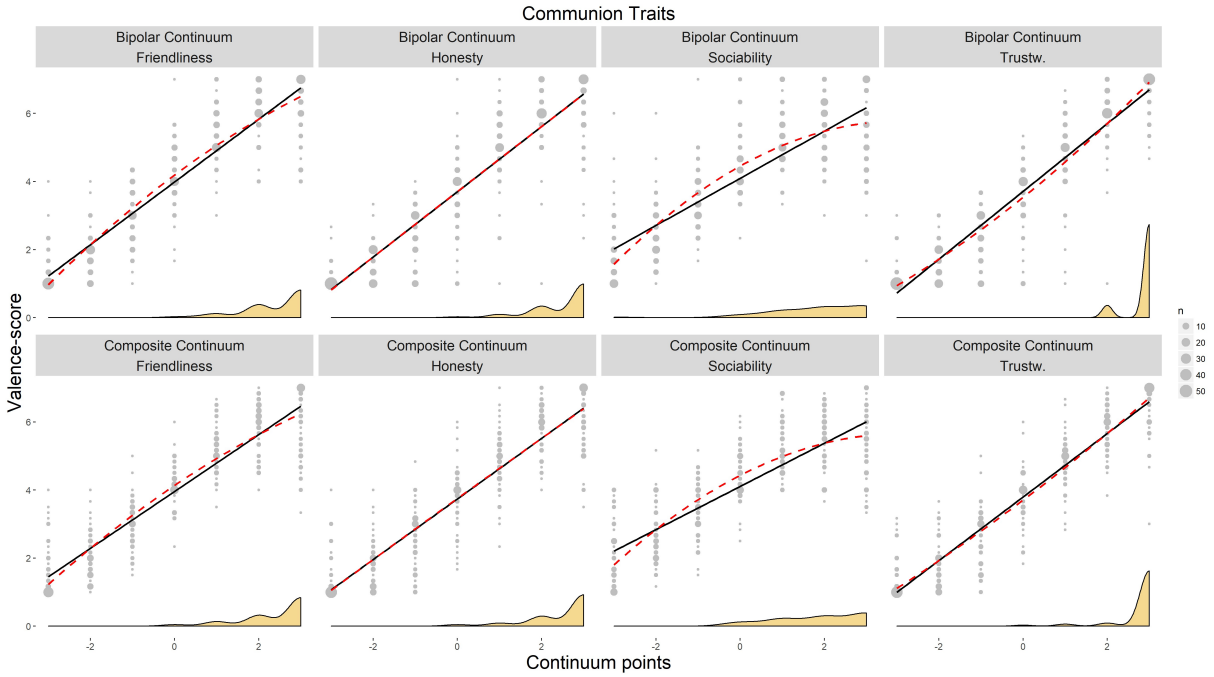

Figure 3. Linear and quadratic regression lines for each communion-related trait dimension for the Bipolar and Composite Bipolar Continuum types. Black lines represent the linear fit. Red dashed lines represent the quadratic fit. The gray dots represent the number of observed ratings (n) and their density along the valence-score scale, for every continuum point (ranging from -3 to +3) of a trait dimension. The density distributions of valence-score peaks across the seven continuum points of a trait dimension are shown at the bottom of each plot (in yellow).

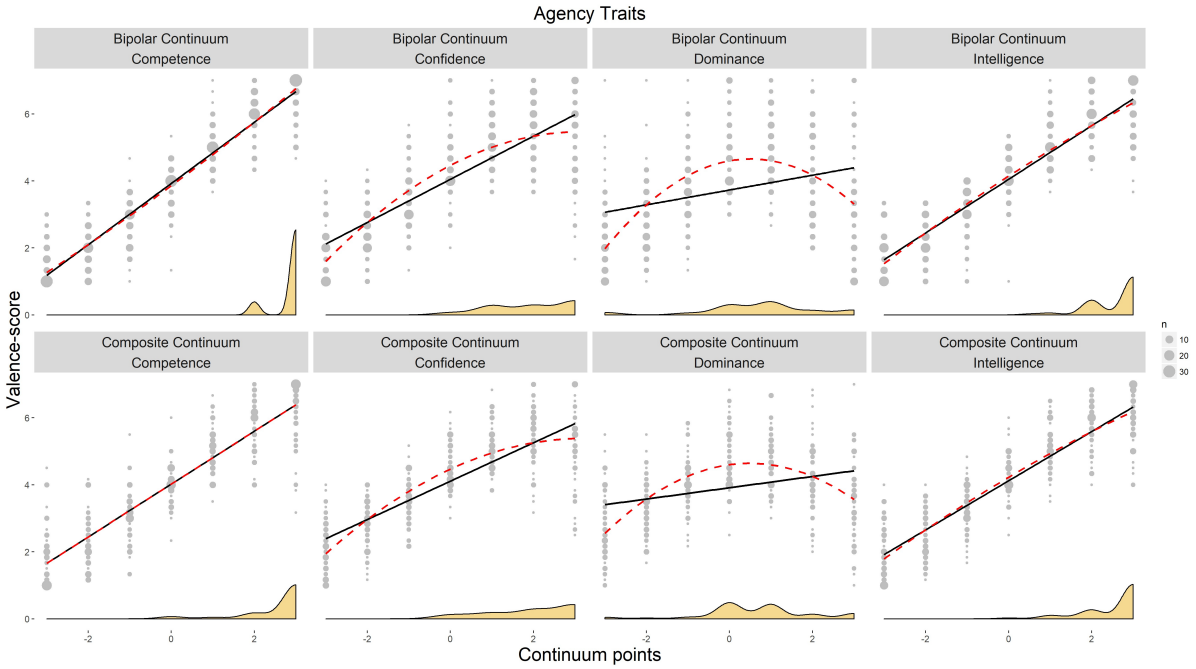

Figure 4. Linear and quadratic regression lines for each agency-related trait dimension for the Bipolar and Composite Bipolar Continuum types. Black lines represent the linear fit. Red dashed lines represent the quadratic fit. The gray dots represent the number of observed ratings (n) and their density along the valence-score scale, for every continuum point (ranging from -3 to +3) of a trait dimension. The density distributions of valence-score peaks across the seven continuum points of a trait dimension are shown at the bottom of each plot (in yellow).

Bipolar Continua Analyses

To better understand how people mapped their evaluations onto a continuum, and facilitate the comparison of results across continuum type and across studies, we converted the HTC and LTC into a Composite Bipolar Continuum (CBC), by averaging the valence-scores of each continuum point for each trait across these two continuum types (note that we found no significant differences between these types). The CBC can be understood as a synthetic bipolar continuum that circumvents some of the methodological issues discussed earlier. For instance, there is no contrast between traits representing opposite poles, nor between their linguistic features, underlying the CBC data. Any impact of continuum design on the ratings should thus be observable by comparison with the BC.

We obtained linear and quadratic estimates per trait for the CBC using separate LMMs exactly as we did earlier for the BC. Additionally, we conducted four additional exploratory LMMs at the level of Big Two facet for the BC and CBC types. LMM results for the BC and

CBC are listed in Table 3. Plots of the results for both bipolar continuum types are shown in Figures 3 and 4.

Table 3

Study 2 Linear and Quadratic Unstandardized Regression Coefficients for the Valence-Score by Type of Bipolar Continuum, Big Two Facet and Trait Dimension.

		Bipolar Continuum			Composite Bipolar Continuum		
Facet	Trait	Intercept (*SD*)	Linear *b*	Quadratic *b*	Intercept (*SD*)	Linear *b*	Quadratic *b*
CM	Honesty	3.71 (0.32)	**0.96** ***	0.00	3.75 (0.28)	**0.89** ***	0.00
	Trustw.	3.53 (0.42)	**1.00** ***	**0.04** ***	3.70 (0.30)	**0.93** ***	**0.02** *
CW	Friendliness	4.19 (0.34)	**0.92** ***	**-0.05** ***	4.14 (0.15)	**0.84** ***	**-0.04** ***
	Sociability	4.45 (0.39)	**0.69** ***	**-0.09** ***	4.43 (0.13)	**0.63** ***	**-0.08** ***
AA	Confidence	4.47 (0.32)	**0.64** ***	**-0.10** ***	4.47 (0.19)	**0.57** ***	**-0.09** ***
	Dominance	4.60 (0.45)	**0.22** ***	**-0.22** ***	4.60 (0.26)	**0.17** ***	**-0.17** ***
AC	Competence	3.86 (0.11)	**0.91** ***	0.02	4.02 (0.15)	**0.79** ***	0.00
	Intelligence	4.14 (0.23)	**0.80** ***	**-0.02** *	4.22 (0.29)	**0.73** ***	**-0.03** *
CM	-	3.62 (0.40)	**0.98** ***	**0.02** **	3.73 (0.35)	**0.91** ***	0.01
CW	-	4.32 (0.37)	**0.81** ***	**-0.07** ***	4.28 (0.21)	**0.73** ***	**-0.06** ***
AA	-	4.53 (0.38)	**0.43** ***	**-0.16** ***	4.53 (0.24)	**0.37** ***	**-0.13** ***
AC	-	4.00 (0.20)	**0.86** ***	0.00	4.12 (0.27)	**0.76** ***	-0.01

Note: CM = Communion Morality; CW = Communion Warmth; AA = Agency Assertiveness; AC = Agency Competence. Valence-score values range from 1 to 7. Significant unstandardized regression coefficients are in bold, and correspond to the fixed effect predictors (Linear *b* = continuum points; Quadratic *b* = squared continuum points). Intercept and its standard deviation refer to the random-intercept by participant effect and represent the between-participant variability of the mean valence-score per trait. A Benjamini-Hochberg (*FDR*) correction was applied to all *p*-values of linear and quadratic estimates. $\alpha = .05$; *$p < .05$, **$p < .01$; ***$p < .001$

Table 3 shows that the overall pattern of results was identical across bipolar continuum types. These results replicated the pattern observed in Study 1, with the exception that, this time, the linear component for dominance was stronger, and practically identical[1] to the quadratic component. Nevertheless, the quadratic (linear) component of dominance remained

[1] A reviewer raised a concern about the effect of including desirability ratings in our valence-score, given Study 3's results with faces. That does not seem to be the case. When desirability ratings were analyzed separately from a valence-score aggregating likeability and valence (as in Study 3), results led to the same conclusions. Nevertheless, our procedure may have inflated the linear component of dominance (and other agency-related traits). When desirability was dropped, its linear component became lower than its quadratic component, while remaining the lowest across all traits. This analysis is available in our online repository.

the highest (lowest) across all traits. In contrast, competence and honesty were the only traits for which only the linear component was significant.

Facet Analyses

The analysis at the facet level further showed a divergence between the two facets of agency. Although the linear components were stronger across all facets, the quadratic component of assertiveness remained the highest of its class, and competence was the only facet with a purely linear component (additional Big Two facet plots available in online repository). To test if the relationship valence-assertiveness facet is significantly less linear and more inverted U-shaped than the relationship valence-competence facet, we ran additional exploratory LMMs including two interaction terms, each specifying an interaction between facet (e.g., assertiveness and competence) and either a linear or an inverted-U trend predictor (i.e., -3, -2, -1, 0, -1, -2, -3), with intercepts and slopes by participant as random effects. For simplicity, LMMs were computed separately for each pair of agency- and communion-related facets, and each bipolar continuum type. Significant interactions in the following analyses indicate a significant difference between the (linear or inverted-U) slopes of any two facets. We expected the difference between linear slopes to be positive (and negative for inverted-U slopes) for competence compared to assertiveness (reference facet level). And indeed, for both bipolar continuum types, we found the expected interactions involving the linear predictor (BC: b = 0.42; CBC: b = 0.39; both $ps_{FDR} < .001$), and the inverted-U predictor (BC: b = -0.53; CBC: b = -0.40; both $ps_{FDR} < .001$). The same pattern of results was found for the slope difference between the warmth (reference facet level) and morality facets, for both the linear (BC: b = 0.17; CBC: b = 0.18; both $ps_{FDR} < .001$) and inverted-U predictors (BC: b = -0.31; CBC: b = -0.24; both $ps_{FDR} < .001$). These results provide stronger support for our hypothesis that the nature of the relationship with valence differs across facets within agency, and additionally suggest a similar, though unexpected, difference within communion.

Inter-rater Agreement and Valence Peaks

Inter-rater agreement and valence peaks were computed for both bipolar continuum types (see Table 4). The overall pattern of results at the trait level replicated the one observed in Study 1. At the facet level, inter-rater agreement was lower for the assertiveness and warmth facets. These were also the facets for which valence peaked closer to the continuum's mid-point. In contrast, the competence and morality facets both exhibited the highest inter-rater agreement and valence peaks at their high poles. Overall, this supports that extreme expressions of assertiveness- or warmth-related traits are less positively evaluated, whereas the more competence or morality one expresses the better, at least when no specific context is provided.

Table 4

Study 2 Mean Peak Locations and Inter-Rater Agreement Results for the Valence-Score by Type of Bipolar Continuum, Big Two Facet and Trait Dimension.

		Bipolar Continuum			Composite Bipolar Continuum		
Facet	Trait	Mean Peak (*SD*)	ICC (2, *k*)	AIC	Mean Peak (*SD*)	ICC (2, *k*)	AIC
CM	Honesty	2.58 (0.70)	.997***	.978	2.55 (0.77)	.997***	.963
	Trustw.	2.88 (0.32)	.998***	.994*	2.83 (0.56)	.998***	.983*
CW	Friendliness	2.46 (0.77)	.998***	.979	2.46 (0.81)	.997***	.968
	Sociability	1.90 (1.16)	.994***	.925	1.93 (1.04)	.994***	.912
AA	Confidence	2.00 (0.97)	.993***	.863	2.00 (1.05)	.993***	.884
	Dominance	0.74 (1.47)	.972***	.593	0.70 (1.31)	.973***	.581
AC	Competence	2.87 (0.34)	.998***	.984*	2.65 (0.78)	.997***	.969
	Intelligence	2.62 (0.61)	.997***	.961	2.61 (0.70)	.996***	.964
CM	-	2.65 (0.66)	.997***	.933**	2.73 (0.60)	.997***	.950**
CW	-	2.30 (0.82)	.996***	.888**	2.37 (0.76)	.996***	.903**
AA	-	1.29 (1.13)	.986***	.666*	1.41 (1.03)	.987***	.699*
AC	-	2.73 (0.52)	.998***	.936**	2.75 (0.60)	.996***	.933**

Note: CM = Communion Morality; CW = Communion Warmth; AA = Agency Assertiveness; AC = Agency Competence. Mean peak locations for the valence-score range from -3 to 3 (i.e., continuum point values). *ICC*s indicate inter-rater agreement for the target judgments ($k = 60$, i.e. number of raters). *AIC* = Average inter-rater correlation (i.e., zero-order correlation of all possible raters within trait/facet). A Benjamini-Hochberg (*FDR*) correction was applied to all *p*-values of linear and quadratic estimates. $\alpha = .05$; *$p < .05$, **$p < .01$; ***$p < .001$

Altogether, these data replicate Study 1's while circumventing some of its methodological limitations.

Study 3

In Study 3, we adapted the paradigm developed in Studies 1-2 to social face perception. This study is, thus, a conceptual replication of Studies 1-2 using face stimuli in place of verbal stimuli: The independent variable was defined by a continuum of faces known to vary in the target trait, rather than by verbal descriptions of the trait continuum. By exposing participants to sets of seven faces representing a continuum of a trait dimension, we expected the evaluation of these faces to represent the evaluation of the trait continuum itself. To minimize the influence of gender stereotypes, we balanced participant and target gender.

Participants

Forty native English speakers [2] (50% female, 50% male, M_{age} = 30.50 years, SD_{age} = 7.02) were recruited via Prolific Academic to participate in the study in exchange for payment (£1.70). All 40 participants were included in the analyses.

Power considerations. The same power considerations discussed in Study 1 apply to this study. For consistency, we pre-specified a participant sample size identical to Study 1's.

Face Continua

In this study we focused on a shorter range of trait dimensions due to difficulties[3] in obtaining stimuli that would entirely correspond to the traits used in Studies 1-2. Thus, we focused on the basic dimensions of social face perception (i.e., trustworthiness and dominance), and an equal number of Big Two-related dimensions for which we found previously validated face continua (competence and warmth). We generated a set of seven face images per continuum, where faces varied along a given trait dimension. Each set was bounded by two faces whose features conveyed the low and high poles of the dimension. Materials were created from two types of face stimuli sets: a widely known face photograph database (Karolinska face database; Lundqvist, Flykt, & Ohman, 1998), and continua of synthetic "FaceGen" faces previously generated and validated by Todorov, Dotsch, Porter, Oosterhof, and Falvello (2013). The inclusion of FaceGen faces served the purpose of validating our custom photograph-based continua, which were expected to convey the same traits. Using both materials allowed us to assess influential trait labels used in Big Two

[2] In total we recruited 49 participants, but nine were excluded from the analyses as a result of using mobile devices with small screens incapable of displaying an entire face continuum, which we considered a crucial requirement in our study. These participants were subsequently replaced to achieve the intended sample size.

research such as competence (Fiske et al., 2007); as we were only able to manipulate intelligence using photographs.

Face photograph-based continua. These continua were generated using stimuli from the Karolinska database and its correspondent ratings collected by Oosterhof and Todorov (2008). These ratings included traits such as dominance, trustworthiness, intelligence (competence-related), and caring (warmth-related). For each trait, we generated a male and female version of the continuum. First, we used PsychoMorph (Version 5; Tiddeman, Burt, & Perrett, 2001) to generate two average-faces representing the high and low poles of a continuum. Each average-face was derived from the 10 face photographs with the highest (or lowest) trait ratings. Next, we used Webmorph (Version 0.0.0.9001; DeBruine, 2017) to generate continua with seven face images each, to match the seven-point continua used in Studies 1-2. Along each continuum, the features of the low-pole average-face (e.g., submissive) gradually shifted toward the features of the high-pole average-face (e.g., dominant) at each step (for details see Sutherland, Rhodes, & Young, 2017). The resulting face continua are shown in Figure 5. These materials are available in our online repository.

FaceGen continua. Todorov et al. (2013) generated, validated, and made available several sets of face continua composed of FaceGen stimuli that partially corresponded to our target dimensions. From these sets, we selected the dominance, trustworthiness, competence, and likeability face continua. The likeability continuum was the only one available to serve as a proxy for the warmth/communion dimension, as likeability and communion/warmth are highly positively correlated (e.g., Oliveira et al., 2019; Wojciszke et al., 2009). However, please notice that this continuum matches with our likeability measure (aggregated in the valence-score), and constitutes yet another example of how valence and communion-related traits are highly conflated. To control for gender, we slightly modified the face continua by dropping the most distant faces from the continuum's mid-point (e.g., +3 SD and -3 SD faces, or others closer to the continuum's mid-point as deemed necessary), at which point the faces started to clearly convey a gender transformation. These modifications were only necessary for the dominance and trustworthiness continua. Using Webmorph (DeBruine, 2017), we generated a replacement for any image dropped from the original continuum. Specifically, we generated a 3-face continuum using the two faces that bounded the removed image, and subsequently extracted the mid-point face to use as the replacement image. These continua are shown in Figure 5.

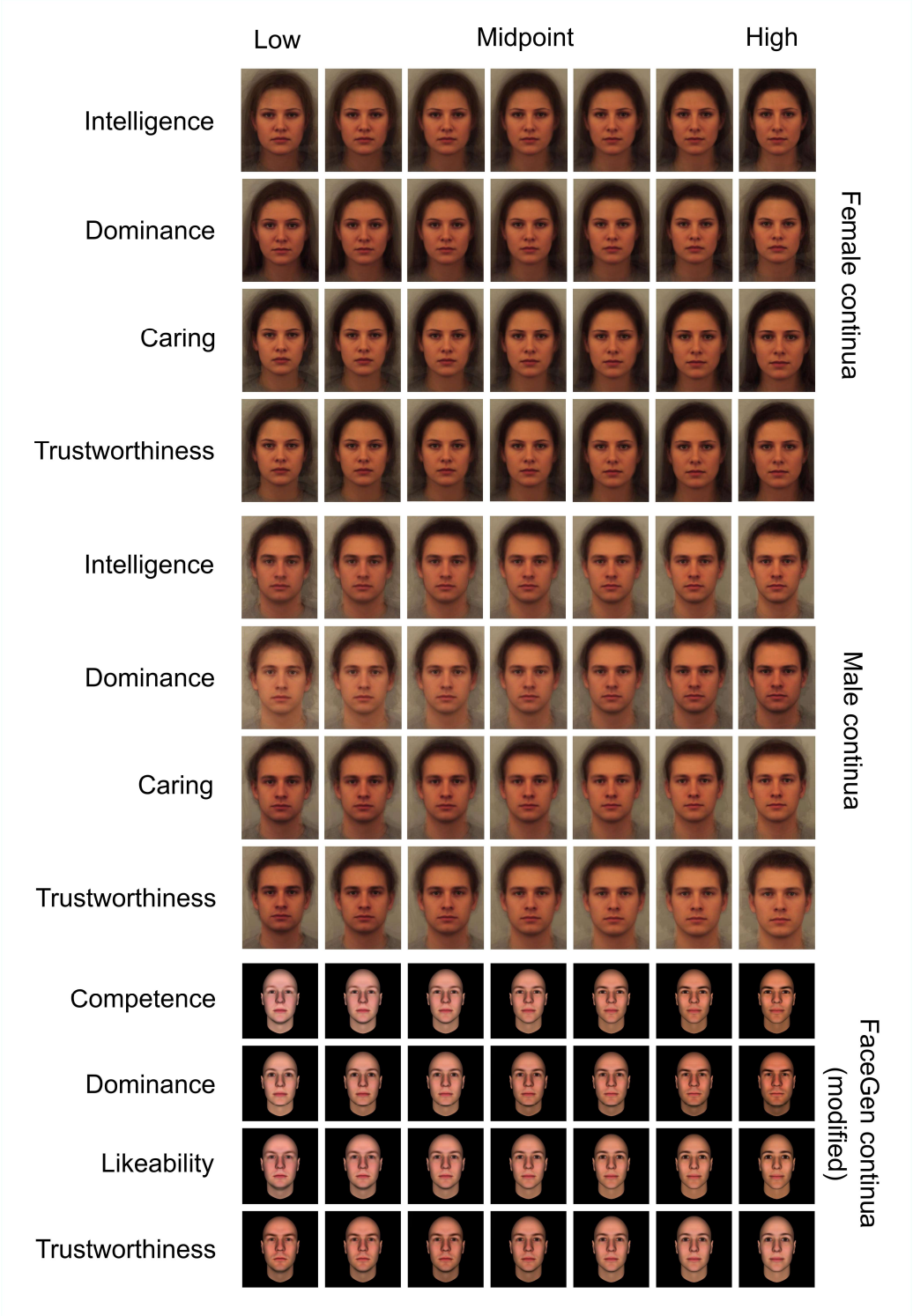

Figure 5. Face continua used in Study 3. The bottom set of FaceGen faces correspond to a slightly modified version of the original face continua generated by Todorov et al. (2013). Specifically, only the trustworthiness and dominance FaceGen continua were modified. All stimuli were made available in our online repository.

Dependent Measures

As in Studies 1-2 we used valence, likeability, and desirability ratings to measure perceived valence. To forestall the possible interaction between participant and target gender, the desirability (for self) ratings' block only included face stimuli that matched the participant's gender. Consequently, only half of the targets (male or female faces) were rated on desirability, which resulted in half the observations for this measure compared with any of the others. Therefore, we had two separate measures: valence-score (aggregating likeability and valence) and desirability for self.

Procedure

The task was programmed using Qualtrics software. Participants were invited to participate in a study about "how people perceive and evaluate faces." Participant gender was filtered via the Prolific Academic website to randomly assign the participants to their appropriate condition. All participants rated the perceived likeability and valence of all the faces of the continua presented in their assigned condition. Only continua matching the participant's gender were rated on desirability for self (excluding the masculine-looking FaceGen continua). Blocks of trials with different targets (detailed in Figure 5) were defined by the target rating, target gender, and stimulus' type (photograph-based vs. FaceGen). Trial order was randomized within each block. The task structure was such that the participants always started by evaluating the female and male face target blocks before the FaceGen targets' blocks. The order of target gender blocks was counterbalanced, but varied with participants' gender. Male (female) participants started with either: two blocks of female (male) faces, one for likeability and another for valence ratings; or, three blocks of male (female) faces, each for one of the three different ratings. In each trial, the whole face continuum was displayed in the center of the screen along with instructions tailored to the specific rating of the block (all instructions available in our online repository). In the valence rating blocks, participants were asked, "How Good or Bad in general is the impression you get from each face?" In the likeability rating blocks, participants were asked, "How Likeable does each face seem to you?" In the "desirability for self" blocks, participants were asked to "imagine they were going to be a character in a Virtual Reality setting" and indicate "how likely you would be to choose each face to represent you in the virtual world, in order to create the best impression." Responses were given on a 7-point rating scale, ranging between 1 (Very Bad/Very Unlikeable/Would never choose) to 7 (Very Good/Very Likeable/Would definitely choose; for valence, likeability and desirability, respectively). Each of the seven

faces in each continuum was associated with a response box where participants entered their response using numerical keys. After completing the task, participants were thanked, debriefed, and compensated.

Results and Discussion

Valence Measures

Because the desirability ratings were only assessed in half of the data points comparatively to the other ratings, we analyzed them separately. Note that PCA requires an equal amount of observations per measure. Therefore, we submitted only the likeability and valence ratings to a PCA using the same criteria as in Studies 1-2. The PCA yielded one component, interpreted as general valence that accounted for 77% of the variance in the ratings. Loadings for likeability and valence were both .88. We then calculated a valence-score by aggregating the ratings of likeability and valence for each point of each face continuum.

Inter-rater Agreement

We calculated ICCs and AICs for the valence-score (see Table 5) and desirability (see Table 6) ratings, using the values of each point of each face continuum. Similarly to Studies 1-2, high agreement was not observed for all traits. The overall pattern suggests that participants agreed less on the perceived valence of agency-related male face continua. In contrast, agency-related female continua exhibited high inter-rater agreement for both the valence-score and desirability ratings. Thus, unlike in the Studies 1-2, the low agreement was now also observed for competence, and exclusively for male targets.

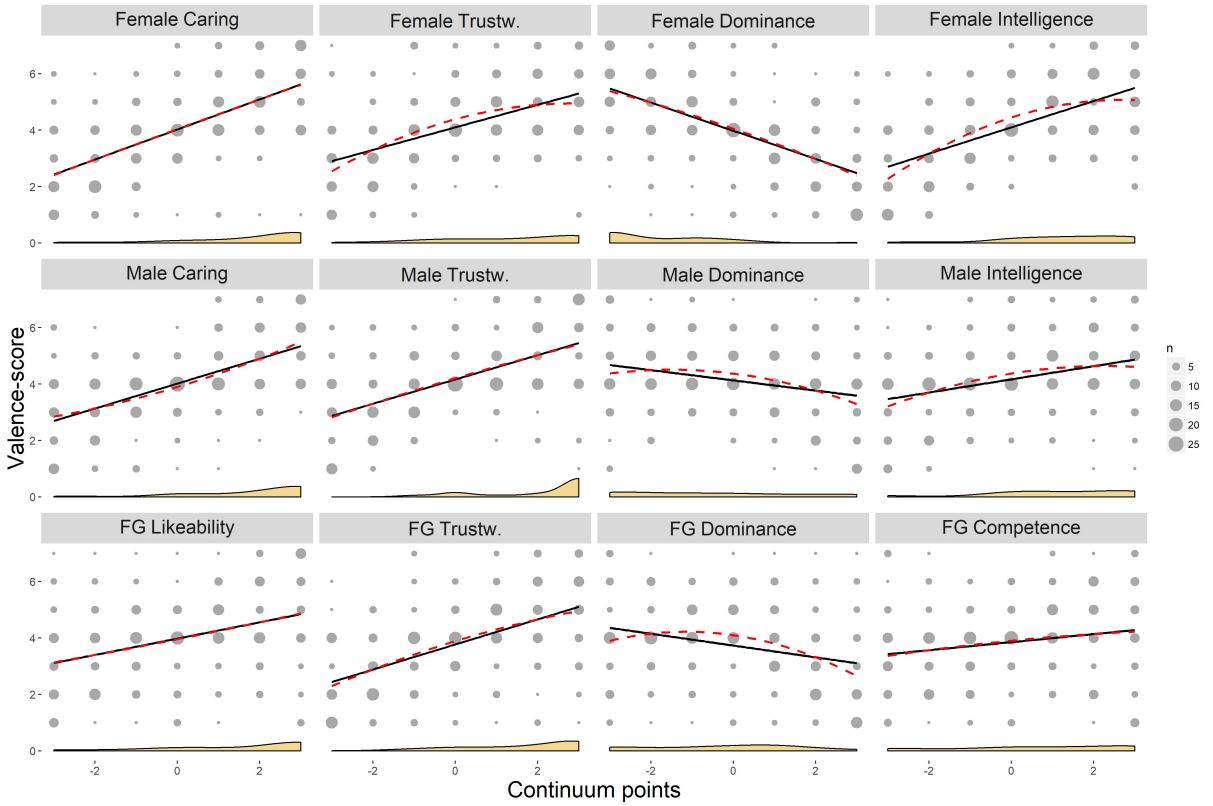

Figure 6. Linear and quadratic regression lines fitted to the valence-scores of each face continua (by target gender and trait dimension) for the valence-score, in Study 3. Black lines represent the linear fit. Red dashed lines represent the quadratic fit. The gray dots represent the number of observed ratings (n) and their density along the valence-score scale, for every continuum point (ranging from -3 to +3) of a trait dimension. The density distributions of valence-score peaks across the seven continuum points of a trait dimension are shown at the bottom of each plot (in yellow). FG = FaceGen face continua.

Linear and Quadratic Fits

Valence-score. Again, we ran separate LMMs (same fixed and random effects as in Studies 1-2) by face continuum with valence-score as the outcome variable. All estimates' *p*-values were FDR-corrected. Face continua were defined by stimulus type (photograph-based vs. FaceGen), target gender (photograph-based continua only) and trait dimension (see Figure 5). Results are listed in Table 5, and data are plotted in Figure 6. As in Studies 1-2, the valence-scores of all traits were significantly predicted by the linear component. Regarding communion-related traits, as expected, we found stronger linear components for all the trustworthiness and warmth-related continua, regardless of stimulus type.

The results obtained in Studies 1-2 for agency-related traits were, however, not entirely replicated. Instead of a stronger quadratic component for dominance, we found that

the linear component was the strongest predictor for all agency-related continua, especially for female continua. The relationship of valence with female facial dominance was more clearly linear, and stronger, than the observed for male facial dominance. Nevertheless, only the dominance-related continua established a negative relationship with valence, and especially the female one. As expected, the FaceGen competence continuum showed a purely linear relationship with valence. However, the same did not occur for the photograph-based intelligence continua, which exhibited a small but significant quadratic component. It remains unclear, however, whether this resulted from higher noise in our custom continua, or from actual differences between facial features across competence and intelligence.

Table 5

Study 3 Results Including Linear and Quadratic Unstandardized Regression Coefficients, Mean Peak Locations, and Inter-Rater Agreement for the Valence-Score by Face Continuum.

Stimuli	Face Continuum	Mean Peak (*SD*)	Intercept (*SD*)	Linear *b*	Quadratic *b*	ICC (2, *k*)	*AIC*
Photograph-based	Female Caring	2.31 (1.02)	4.19 (0.62)	**0.52** ***	-0.03	.977 ***	0.82
	Male Caring	2.42 (1.17)	3.94 (0.56)	**0.46** ***	0.02	.972 ***	0.75
	Female Trustw.	1.91 (1.59)	4.27 (0.50)	**0.40** ***	**-0.04** *	.961 ***	0.66
	Male Trustw.	2.15 (1.42)	4.18 (0.52)	**0.41** ***	**0.00** *	.962 ***	0.68
	Female Dominance	-1.68 (1.61)	4.22 (0.65)	**-0.46** ***	-0.03	.966 ***	0.7
	Male Dominance	-0.26 (2.13)	4.29 (0.44)	**-0.16** **	**-0.06** **	.811 ***	0.16
	Female Intelligence	1.90 (1.19)	4.34 (0.64)	**0.47** ***	**-0.07** ***	.972 ***	0.76
	Male Intelligence	1.10 (1.63)	4.37 (0.52)	**0.21** **	**-0.05** **	.892 ***	0.4
FaceGen	Likeability	1.90 (1.65)	3.91 (0.80)	**0.33** ***	-0.01	.925 ***	0.53
	Trustw.	1.85 (1.67)	3.72 (0.94)	**0.34** ***	-0.01	.916 ***	0.59
	Competence	1.10 (1.93)	3.94 (0.83)	**0.17** ***	-0.03	.740 ***	0.25
	Dominance	-0.47 (2.11)	4.00 (0.85)	**-0.17** ***	**-0.07** ***	.809 ***	0.27

Note: Valence-score values range from 1 to 7. Mean peak locations for the valence-score range from -3 to 3 (i.e., continuum point values). Significant unstandardized regression coefficients are in bold, and correspond to the fixed effect predictors (Linear *b* = continuum points; Quadratic *b* = squared continuum points). Intercept and its standard deviation refer to the random-intercept by participant effect and represent the between-participant variability of the

Desirability for self. LMMs were run separately by participant gender with desirability ratings as the outcome variable. Results are listed in Table 6, and plotted in Figure 7. Regardless of participant gender, communion-related continua showed a stronger linear relationship with desirability. Results for the agency-related continua were less consistent across traits and suggest sensitivity to participant gender. Only the results for male faces replicated Studies 1-2's findings: a stronger quadratic component for dominance, and a stronger linear component for intelligence. For female faces, the linear components of intelligence and dominance were both stronger than their quadratic counterparts and exhibited a clear opposite relationship with valence. These results must, however, be read with caution given the lower sample size. Nevertheless, they may be informative to future research focusing on actor-observer differences in face perception.

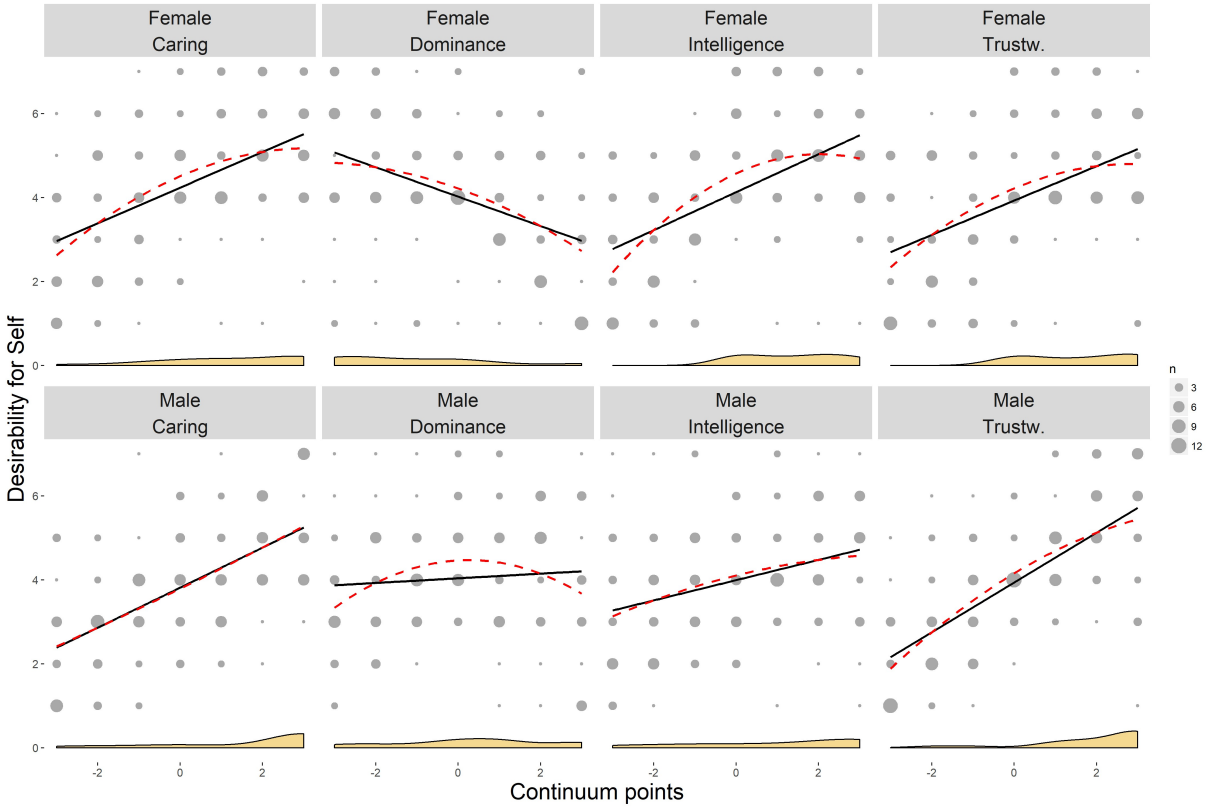

Figure 7. Linear and quadratic regression lines for each trait dimension by participant gender for the 'desirability for the self' ratings (ranging from low (1) to high desirability (7)), in Study 3. Black lines represent the linear fit. Red dashed lines represent the quadratic fit. The gray dots represent the number of observed ratings (n) and their density along the valence-score scale, for every continuum point (ranging from -3 to +3) of a trait dimension. The density distributions of valence-score peaks across the seven continuum points of a trait dimension are shown at the bottom of each plot (in yellow).

Table 6

Study 3 Results Including Linear and Quadratic Unstandardized Regression Coefficients, Mean Peak Locations, and Inter-Rater Agreement for the Desirability For Self Ratings by Face Continuum and Participant Gender.

Face Continuum	Participant Gender	Mean Peak (*SD*)	Intercept (*SD*)	Linear *b*	Quad. *b*	*ICC*(2,*k*)	*AIC*
Caring	Female	1.35 (1.76)	4.51 (0.56)	**0.43** ***	-0.07	.881 ***	.421
	Male	1.82 (1.94)	3.79 (0.61)	**0.48** ***	0.01	.923 ***	.670
Trustw.	Female	1.60 (1.24)	4.22 (0.91)	**0.41** ***	**-0.07** *	.875 ***	.604
	Male	2.02 (1.44)	4.15 (0.36)	**0.59** ***	-0.05	.953 ***	.651
Dominance	Female	-1.20 (1.91)	4.22 (0.71)	**-0.35*****	-0.05	.788 ***	.212
	Male	0.37 (1.91)	4.46 (0.53)	0.06	**-0.11** **	.491	.129
Intelligence	Female	1.42 (1.16)	4.57 (0.63)	**0.45** ***	**-0.11** **	.924 ***	.534
	Male	1.12 (2.18)	4.10 (0.51)	**0.24** ***	-0.03	.652 **	.201

Note: Desirability ratings range from 1 to 7. Mean peak locations for the desirability ratings range from -3 to 3 (i.e., continuum point values). Significant unstandardized regression coefficients are in bold, and correspond to the fixed effect predictors (Linear *b* = continuum points; Quadratic *b* = squared continuum points). Intercept and its standard deviation refer to the random-intercept by participant effect and represent the between-participant variability of the mean valence-score per trait. *ICC*s indicate inter-rater agreement in the target trait judgments (k = 20, i.e. number of raters). *AIC* = Average inter-rater correlation (i.e., zero-order correlation of all possible raters within face continuum). A Benjamini-Hochberg (*FDR*) correction was applied to all *p*-values of linear and quadratic estimates. $\alpha = .05$; *$p < .05$; **$p < .01$; ***$p < .00$

General Discussion

In three studies we assessed the relationship between valence and traits that have been identified as central in person perception and social face perception. Studies 1 and 2 focused on the conceptual knowledge about the traits themselves. Study 3 focused on how these traits were (likely to be) inferred from faces. Results from Studies 1-2 show that the relationship established with valence is less stable in nature across agency-related traits comparatively to communion-related traits, a phenomenon that was particularly noticeable when comparing dominance and competence. Study 3 assessed the same relationships using face stimuli instead of verbal trait labels and found that the curvilinearity between dominance and valence observed in Studies 1-2 shifted to a negative linear relationship, whereas all the other relationships replicated the pattern of results obtained with trait words. Moreover, the

desirability ratings in Study 3 suggest that the nature of the relationship between dominance and valence is sensitive to perceivers' gender. Our results are thus informative regarding how core traits of the Big Two dimensions relate with valence and additionally raise questions that are relevant to the field. We summarize them below.

A piece of information directly offered by these results is that one chief difference between the communion and agency dimensions is the extent to which their traits are homogeneous in the relationship established with valence: Whereas communion-related traits (except for sociability) overlapped extensively with valence, agency-related traits exhibited a more heterogeneous and context-dependent relationship with valence (see also Bruckmüller & Abele, 2013). All agency-related traits, except for dominance, consistently exhibited an alpha nature (i.e., linearity with valence). Dominance was the only trait more clearly exhibiting a beta-trait nature (i.e., curvilinearity with valence), although only in the evaluation of verbally described targets. Moreover, all agency-related traits, except for competence, exhibited a quadratic component. The curvilinear trend (unexpectedly) found for the communion-related trait of sociability suggests that high sociability can be negatively evaluated (for a similar finding see Landy et al., 2016). Future research should seek to understand why sociability shares this feature with dominance, as both show strong inverted-U relationships with valence. For instance, future studies could investigate whether the curvilinearity observed for these two traits is related with social perspective. The impact of social perspective (having a trait myself vs. interacting with someone who expresses that trait) on trait evaluations may be more crucial for agency-related traits, and more relevant for sociability than for other communion-related traits. Although our studies were not optimally designed to examine social perspective, we must note that our PCA results speak against that hypothesis given the high correlation found between desirability for self and likeability (of others).

Our data also clarify and add to Imhoff and Koch's (2017) data regarding the relationship between the Big Two. With a new paradigm where traits themselves are defined as independent variables, we show that not all agency-related traits exhibit a curvilinear relationship with valence. For instance, competence-related traits showed a purely linear relationship with valence. Our results additionally clarify that although other agency-related traits (and power and status dimensions) exhibited significant quadratic relationships with valence, these were a weaker feature of that relationship (except for dominance). It is conceivable that any differences between Imhoff and Koch's (2017) results and ours may have derived from their definition of agency in terms of assertiveness- but not competence-

related traits (e.g., Koch, Imhoff, Dotsch, Unkelbach, & Alves, 2016), coupled with examining social entities other than traits and faces *per se*.

The divergence between traits regarding their relationship with valence may also constitute a reason to branch the Big Two into facets. In light of the results obtained in Study 2, one may notice that the facets within each Big Two dimension can be characterized as diverging in the nature of the relationship that their traits establish with valence. Whereas competence-related traits tend to more strongly exhibit an alpha nature, assertiveness-related traits tend to more strongly exhibit a beta nature. Moreover, the beta nature trend observed for sociability (warmth-related) is also supportive of the branching of communion into the morality and warmth facets (see Brambilla & Leach, 2014; Goodwin, 2015). This, however, remains a hypothesis to be rigorously tested in the future, since the present studies focused primarily on relationships at the trait level, and our analyses at the facet level were exploratory in nature.

It is important to stress that our results do not allow us to state that one specific trait is immutably alpha or beta. Our results only suggest that, by default, some traits are more likely to show the quadratic component relatively to others. Thus, although we consistently found a positive linear relationship between competence and valence, it is conceivable that a curvilinear or even negative relationship may emerge under different circumstances. For instance, in a competitive environment, highly intelligent individuals may be perceived as "bad" (e.g., Carrier, Dompnier, & Yzerbyt, 2019). Further studies may clarify how changes in the meaning of a trait modulate its perceived valence across different contexts. This hypothesis of flexibility in valence perception emerges as a possibility in light of our results showing that, although dominance was perceived as a beta-trait, it also exhibited low inter-rater agreement. This suggests that some perceivers may perceive dominance more as an alpha trait than others—an individual differences hypothesis worthy to be pursued in the future (see also Stolier, Hehman, Keller, Walker, & Freeman, 2018, for a similar idea regarding trait interrelationships).

We find our face evaluation data from Study 3 to be highly relevant. First, because they clarify that the conceptual variability perceived in a trait dimension does not necessarily map onto the variability perceived in a face. And second, because they suggest that these perceptions are modulated by participant gender and type of judgment. Evaluations of face continua did not exhibit the beta patterns previously observed in Studies 1-2. All relationships between valence and face-trait dimensions tended to be linear. And the curvilinearity between dominance and valence was only found under circumstances where male perceivers evaluated

the desirability of male faces. Moreover, the nature of that relationship was more inconsistent across agency- than across communion-related faces. Why is this so, is yet an empirical question, but once more, this suggests that the valence of a trait is flexible and context-dependent. For instance, competence- and dominant-looking faces match positive and negative emotional expressions, respectively (Said, Sebe, & Todorov, 2009). This may constitute one reason why they exhibited an opposite linear relationship with valence.

Our results also suggest that participant gender interfered with the perceived valence of targets. Females preferred more submissive-looking faces comparatively to males, who instead preferred a low-to-moderate dominance appearance. This may have occurred because counter-stereotypical facial features (e.g., masculine-looking female faces) lead to amplified negative evaluations (Oh, Dotsch, Porter, & Todorov, 2019; Sutherland, Young, Mootz, & Oldmeadow, 2015). These results concerning participant gender raise the possibility that self-perception may have modulated the perceived valence of traits—another interesting hypothesis to be addressed in future studies.

Another source of variability in our data may have been the standard used to support each trait rating. One possibility was that participants were making self-reference ratings (O'Mahony, 1984). This may have been more likely in Studies 1-2 than in Study 3, where targets were faces of "others". The actor-observer differences predicted by the Dual Perspective model proposed by Abele and Wojciszke (2014), according to which agency-related traits are more desirable for oneself (actor perspective), whereas communion-related traits are more desirable in others (observer perspective), could be underlying the divergent results found with faces (Study 3). This hypothesis should also be further explored.

Considerations Regarding the Relationship between Dimensions

Two traits that are clearly and linearly related should establish similar linear relationships with a third dimension (e.g., valence). Our results clearly indicate that not all of the traits encapsulated by the same Big Two dimension establish the same relationship with valence, as one would expect. Our findings would suggest that results regarding the relationship between agency and communion may critically depend upon the trait(s) selected in a study to represent the agency dimension. This is consistent with the stimulus-sampling explanation proposed by Imhoff and Koch (2017, p.124) to account for the inconsistent relationship between the fundamental dimensions reported in the extant literature (Judd, James-Hawkins, Yzerbyt, & Kashima, 2005; Kervyn, Yzerbyt, & Judd, 2010; Rosenberg et al., 1968). Although Imhoff and Koch (2017) provide an example for societal groups, the

same can be applied to traits. If alpha-traits (e.g., competence-related) are oversampled in a composite-score of agency, the resulting relationship with communion should be more linear than it would be if the trait set suffered from an oversampling of beta-traits (e.g., assertiveness-related) (but see also Judd, Garcia-Marques, & Yzerbyt, 2019). Regarding face evaluations, the distinct relationship that dominance and trustworthiness established with valence in Study 1 (Table 2) may be contributing to the orthogonality found between these two dimensions (Oosterhof & Todorov, 2008). If the independence between dominance and trustworthiness is an artifact resulting from a non-linear relationship, results will be contingent on how well a face set represents all the levels of dominance. Targets biased toward the lower (higher) levels of the dimension could promote a positive (negative) correlation between the dimensions. This could also explain why we did not detect a curvilinear, but instead a negative relationship between dominance and valence in Study 3.

Limitations

Although the paradigm used in the present work adds to the correlational data in the literature, its methodology is not free of limitations and confounds. Our results could have been affected by the psychological tendency to dislike extremes of either type (Grant & Schwartz, 2011), which suggests that a non-linear relationship would be found by default in all relationships where extremity is represented. Even if this explains the results we obtained for dominance, it would not explain why extreme competence did not suffer from the same tendency. In addition, our procedure may lack the natural variability of a trait, truncating it at some level. This could lead us to believe that some traits are alpha traits when in fact they are beta. This problem is likely to be greater in Study 3 where no clear curvilinear relationships were found. Perhaps more extreme submissive faces, along with more differentiated intermediate stimuli, would increase sensitivity to any existent curvilinearity. Still, when we attend to Studies 1-2, we may conclude that when trait variability was described by the same labels, curvilinearity emerged more clearly for dominance than for competence. While we believe that our paradigm offers more compelling data than pure correlation measures for inferring the relationship between traits and valence, we also think that more data should be collected using different materials and labels than those used in our studies.

Other limitations of our approach include the possibility that our participants' features are moderating the effects, and the fact that our sample is far from being representative in terms of all relevant variables that can affect trait perception on the perceiver side such as, for

instance, political ideology (Olivola, Sussman, Tsetsos, Kang, & Todorov, 2012) and self-perception (Srivastava, Guglielmo, & Beer, 2010).

Conclusion

The present research examined the relationship between valence and core traits involved in personality impressions based on conceptual knowledge and facial appearance. We found that: (i) the majority of core agency- and communion-related traits exhibited a linear relationship with valence; and, (ii) dominance was the only trait establishing a clear quadratic relationship with valence, although this pattern was more evident in assessments of conceptual knowledge than in face evaluation. Our findings add to the current literature by clarifying the relationships between valence and the trait content of the Big Two using two different modalities of social stimuli (i.e., verbal and visual). Importantly, the present findings caution against assuming only linearity or curvilinearity in studies concerned with the relationship between the Big Two.

Acknowledgments

This research had the support of Fundação para a Ciência e a Tecnologia (FCT) through the doctoral grant PD/BD/113471/2015 awarded to Manuel Oliveira, and the strategic project UID/PSI/04810/2013 granted to the William James Center for Research.

References

Abele, A. E., Cuddy, A. J. C., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, *38*, 1063–1065. doi: 10.1002/ejsp.574

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: 10.3389/fpsyg.2016.01810

Abele, A. E., Uchronski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *European Journal of Social Psychology*, *38*, 1202–1217. doi: 10.1002/ejsp.575

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*, 751–763. doi:

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 J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 50, pp. 195–255). Academic Press. doi: 10.1016/B978-0-12-800284-1.00004-7

Ames, D., Lee, A., & Wazlawek, A. (2017). Interpersonal assertiveness: Inside the balancing act. *Social and Personality Psychology Compass*, *11*, 1–16. doi: 10.1111/spc3.12317

Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, *9*, 272–279. doi: 10.1037/h0025907

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*, 289–300. doi: 10.2307/2346101

Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition*, *32*, 397–408. doi: 10.1521/soco.2014.32.4.397

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, *41*, 135–143. doi: 10.1002/ejsp.744

Brand, A., & Bradley, M. T. (2012). More voodoo correlations: When average-based measures inflate correlations. *Journal of General Psychology*, *139*, 260–272. doi: 10.1080/00221309.2012.703711

Bruckmüller, S., & Abele, A. E. (2013). The density of the big two: How are agency and communion structurally represented? *Social Psychology*, *44*, 63–74. doi: 10.1027/1864-9335/a000145

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*, 1–20. doi: 10.5334/joc.10

Carrier, A., Dompnier, B., & Yzerbyt, V. (2019). Of nice and mean: The personal relevance of others' competence drives perceptions of warmth. *Personality and Social Psychology Bulletin*. Advance online publication. doi: 10.1177/0146167219835213

Carrier, A., Louvet, E., Chauvin, B., & Rohmer, O. (2014). The primacy of agency over competence in status perception. *Social Psychology*, *45*, 347–356. doi: 10.1027/1864-9335/a000176

Chapais, B. (2015). Competence and the evolutionary origins of status and power in humans. *Human Nature*, *26*, 161–183. doi: 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. doi: 10.1016/j.jesp.2013.10.008

Cislak, A., & Wojciszke, B. (2008). Agency and communion are inferred from actions serving interests of self or others. *European Journal of Social Psychology*, *38*, 1103–1110. doi: 10.1002/ejsp.554

Cliff, N. (1959). Adverbs as multipliers. *Psychological Review*, *66*, 27–44. doi: 10.1037/h0045660

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. *Advances in Experimental Social Psychology*, *40*, 61–149. doi: 10.1016/S0065-2601(07)00002-0

DeBruine, L. (2017). Webmorph (Version v0.0.0.9001) [Computer software]. Zenodo. Retrieved from https://doi.org/10.5281/zenodo.1073696

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*, 621–633. doi: 10.1002/per.1934

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. doi: 10.3758/BF03193146

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*, 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*, 878–902. doi: 10.1037//0022-3514.82.6.878

Fiske, S., & Taylor, S. (2017). *Social cognition: From brains to culture* (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Goodwin, G. P. (2015). Moral character inperson perception. *Current Directions in Psychological Science*, *24*, 38–44. doi: 10.1177/0963721414550709

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*, 61–76. doi: 10.1177/1745691610393523

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*, 493–498. doi: 10.1111/2041-210X.12504

Hampson, S. E., Goldberg, L. R., & John, O. P. (1987). Category-breadth and social-desirability values for 573 personality terms. *European Journal of Personality*, *1*, 241–258. doi: 10.1002/per.2410010405

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology*, *37*, 25–38. doi: 10.1037/0022-3514.37.1.25

Hegarty, P. (2011). Sexuality, normality and intelligence.What is queer theory up against? *Psychology and Sexuality*, *2*, 45–57. doi: 10.1080/19419899.2011.536314

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*, 165–196. doi: 10.1016/S1090-5138(00)00071-4

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185. doi: 10.1007/BF02289447

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*, 122–137. doi: 10.1177/1745691616657334

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. doi: 10.1016/j.jesp.2019.01.008

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*, 899–913. doi: 10.1037/0022-3514.89.6.899

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*, 673–681. doi: 10.1002/ejsp.1978

Kervyn, N., Yzerbyt, V., & 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*, 155–187. doi: 10.1080/13546805.2010.517997

Kim, M. P., & Rosenberg, S. (1980). Comparison of two structural models of implicit personality theory. *Journal of Personality and Social Psychology*, *38*, 375–389. doi: 10.1037/0022-3514.38.3.375

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*, 675–709. doi: 10.1037/pspa0000046

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. doi: 10.18637/jss.v082.i13

Landy, J. F., Piazza, J., & Goodwin, G. P. (2016). When it's bad to be friendly and smart: The desirability of sociability and competence depends on morality. *Personality and Social Psychology Bulletin*, *42*, 1272–1290. doi: 10.1177/0146167216655984

Lemann, T. B., & Solomon, R. L. (1952). Group characteristics as revealed in sociometric patterns and personality ratings. *Sociometry*, *15*, 7–90. doi: 10.2307/2785447

Lundqvist, D., Flykt, A., & Ohman, A. (1998). The Karolinska directed emotional faces (KDEF) [Face database]. Stockholm, Sweden: Karolinska Institutet. Retrieved from http://kdef.se/

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2019). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General*. Advance online publication. doi: 10.1037/xge0000638

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*, 888-902. doi: 10.1002/ejsp.2569

Olivola, C. Y., Sussman, A. B., Tsetsos, K., Kang, O. E., & Todorov, A. (2012). Republicans prefer republican-looking leaders: Political facial stereotypes predict candidate electoral success among right-leaning voters. *Social Psychological and Personality Science*, *3*, 605–613. doi: 10.1177/1948550611432770

O'Mahony, J. F. (1984). Knowing others through the self—Influence of self-perception on perception of others: A review. *Current Psychological Research & Reviews*, *3*, 48–62. doi: 10.1007/BF02686558

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings*

of the National Academy of Sciences, *105*, 11087–11092. doi: 10.1073/pnas.0805664105

Peabody, D. (1970). Evaluative and descriptive aspects in personality perception: A reappraisal. *Journal of Personality & Social Psychology*, *16*, 639–646. doi: 10.1037/h0030259

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer.

Richards, L., Rollerson, B., & Phillips, J. (1991). Perceptions of submissiveness: Implications for victimization. *The Journal of Psychology*, *125*, 407–411. doi: 10.1080/00223980.1991.10543302

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*, 283–294. doi: 10.1037/h0026086

Rosenberg, S., & Sedlak, A. (1972). Structural representations of implicit personality theory. *Advances in Experimental Social Psychology*, *6*, 235–297. doi: 10.1016/S0065-2601(08)60029-5

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, *9*, 260–264. doi: 10.1037/a0014681

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428. doi: 10.1037/0033-2909.86.2.420

Srivastava, S., Guglielmo, S., & Beer, J. S. (2010). Perceiving others' personalities: Examining the dimensionality, assumed similarity to the self, and stability of perceiver effects. *Journal of Personality and Social Psychology*, *98*, 520–534. doi: 10.1037/a0017057

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*, 9210-9215. doi: 10.1073/pnas.1807222115

Suitner, C., & Maass, A. (2008). The role of valence in the perception of agency and communion. *European Journal of Social Psychology*, *38*, 1073–1082. doi: 10.1002/ejsp.525

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*, 105–118. doi: 10.1016/j.cognition.2012.12.001

Sutherland, C. A. M., Rhodes, G., & Young, A. W. (2017). Facial image manipulation. *Social*

Psychological and Personality Science, *8*, 538–551. doi: 10.1177/1948550617697176

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*, 186–208. doi: 10.1111/bjop.12085

Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications*, *21*, 42–50. doi: 10.1109/38.946630

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion*, *13*, 724–738. doi: 10.1037/a0032335

Walker, M., Schönborn, S., Greifeneder, R., & Vetter, T. (2018). The Basel Face Database: A validated set of photographs reflecting systematic differences in Big Two and Big Five personality dimensions. *PLoS ONE*, *13*, 1-20. doi: 10.1371/journal.pone.0193190

Wojciszke, B. (1994). Multiple meanings of behavior: Construing actions in terms of competence or morality. *Journal of Personality and Social Psychology*, *67*, 222–232. doi: 10.1037/0022-3514.67.2.222

Wojciszke, B. (2005). Morality and competence in person- and self-perception. *European Review of Social Psychology*, *16*, 155–188. doi: 10.1080/10463280500229619

Wojciszke, B., Abele, A. E., & Baryla, W. (2009). Two dimensions of interpersonal attitudes: Liking depends on communion, respect depends on agency. *European Journal of Social Psychology*, *39*, 973–990. doi: 10.1002/ejsp.595

Wortman, J., & Wood, D. (2011). The personality traits of liked people. *Journal of Research in Personality*, *45*, 519–528. doi: 10.1016/j.jrp.2011.06.006

Chapter VIII

Combining traits into a face: A reverse correlation approach*

Manuel Oliveira[1†], Teresa Garcia-Marques[1], Ron Dotsch[2],

Abstract

The integration of multiple traits into a unitary impression has been extensively investigated in impression formation research. However, because the focus has typically been on the verbal output of the formed impressions, little is known about how impressions resulting from different trait combinations impact perceivers' expectations about facial content. Here, we offer initial evidence about how trait integration occurs in social face perception. In two studies we used a reverse correlation paradigm to obtain face images reflecting participants' expectations about facial content for different trait combinations of dominance and trustworthiness. Analyses of the physical and perceived content of these images suggest that: (a) trustworthiness information outweighs dominance information in expectations about facial content; (b) the face content derived from any trait combination contains information that goes beyond the content associated with each separate trait. These findings extend the research on trait integration to social face perception.

Key-words: Face Perception; Reverse Correlation; Trait Integration

---

[1] William James Center for Research, ISPA – Instituto Universitário, Lisbon, Portugal.

[2] Utrecht University, Utrecht, The Netherlands.

* Article published in *Social Cognition*, *37*(5), 516-545. doi: 10.1521/soco.2019.37.5.516

[†] Correspondence to: Manuel Oliveira, William James Center for Research, ISPA—Instituto Universitário, R. Jardim do Tabaco 34, 1149-041 Lisbon, Portugal.; E-mail: manueljbo@gmail.com

Combining traits into a face: A reverse correlation approach

Forming an impression of someone's personality involves the acquisition and integration of multiple information sources. Most of this information corresponds to traits inferred from the person's behavior and appearance (for reviews see Gilbert, 1998; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015; Uleman & Kressel, 2013).

After Asch's (1946) seminal work, researchers have strived to understand the cognitive operations underpinning the integration of traits into a unitary evaluative impression (e.g., Anderson, 1965; Asch & Zukier, 1984; Bruner, Shapiro, & Tagiuri, 1958; Hampson, 1990; Himmelfarb, 1972; Rosenberg, Nelson, & Vivekananthan, 1968; Rosenberg & Jones, 1972). Two theoretical perspectives emerged from those studies: *gestaltic*, which posits that an impression emerges from the perceived interrelations between traits (Asch, 1946; Asch & Zukier, 1984; see also Hamilton & Sherman, 1996); and *linear integration*, according to which evaluative impressions are better predicted by additive or averaging models of the traits' valence and weight (e.g., Anderson, 1965). Whatever the integration process and whatever the number of behaviors or traits we have access to while forming an impression, the final outcome will be a unitary impression. Importantly, moreover, the content of this impression will be characterized by how warm and how competent a person seems (Abele & Wojciszke, 2014; Cuddy, Fiske, & Glick, 2008; Rosenberg et al., 1968). Specifically, according to the warmth-by-competence model (for a review see Cuddy et al., 2008), the content of an impression formed about a person/group can be parsimoniously described by its positioning in a two-dimensional space formed by the two perpendicular axes of warmth and competence. In essence, research suggests that perceivers integrate multiple traits into an impression that is characterized by the combination of two "more fundamental" traits.

Previous research has mainly investigated the process of trait integration in terms of abstract mental representations of the target persons (for a review see Hamilton & Sherman, 1996). Here, we investigate the same process within the social face perception domain. Specifically, we examine, and offer initial evidence about, how the integration of different traits (conceptual knowledge) influences expectations about the facial content (physical appearance) of a target person.

Facial appearance is highly relevant to our impressions (Zebrowitz, 2006). We infer traits from faces as quickly as 34 to 100 ms (Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006). Furthermore, not only do we read traits "from faces," we also read them "into faces" (Hassin & Trope, 2000), meaning that we are able to visualize someone's face based on

a trait description (e.g., visualizing a book character). Despite the many personality traits we are able to infer from the face alone, the resulting impression is ultimately mapped onto a two-dimensional space defined by trustworthiness and dominance (Oosterhof & Todorov, 2008). That is, the impression is characterized by how trustworthy and how dominant a person appears to be.

The trustworthiness-by-dominance model of social face perception was identified by Oosterhof and Todorov (2008) via a data-driven approach (from face ratings of a target). This allowed them to (a) identify trustworthiness and dominance as the trait dimensions that accounted for most of the variance in personality inferences from faces, and to (b) capture the variance in facial structure leading to the inference of each one of those traits (for convergent results see Kleisner, Chvátalová, & Flegr, 2014; Sutherland et al., 2013; Walker, Jiang, Vetter, & Sczesny, 2011; Walker & Vetter, 2009). With this information, Oosterhof and Todorov (2008) generated novel synthetic faces whose physical features resulted from an averaging (i.e., linear combination) between the features associated with inferences of dominance and the features associated with inferences of trustworthiness. Because the trustworthiness-by-dominance 2D-space encompasses four quadrants, these faces conveyed four different combinations of these dimensions (i.e., dominant with trustworthy or untrustworthy; submissive with trustworthy or untrustworthy). The specific content of these faces was subsequently found to be associated with traits such as threat, attractiveness, competence, and likeability (Todorov, Said, Engell, & Oosterhof, 2008; see their Figure 2). Hence, Oosterhof and Todorov‘s (2008) approach followed an inverse pathway to that initially used to identify the two-dimensional space. Instead of deriving the facial content reflecting the space’s diagonals from judgments that explicitly combine the two dimensions (e.g., how “dominant and trustworthy” is the person?), the approach relies on synthetic faces artificially derived from the trustworthiness-by-dominance face space. That is, to create faces that integrated the two traits, Oosterhof and Todorov (2008) linearly combined (averaged) the physical features previously identified to be associated with each separate dimension (i.e., trustworthiness and dominance), and mapped them into a face that was subsequently judged on several other traits.

We wondered whether people indeed integrate these two traits in such a way. In the present work, we ask if the linear perspective assumed by Oosterhof and Todorov (2008) reflects how people themselves make such an integration.

Integration of trustworthiness and dominance traits

The process of integrating traits into a face can be thought of as following an equal or unequal weighting linear model (e.g., Anderson, 1968) or to be (trait-) context dependent (e.g., Asch, 1946). In the general field of impression formation, we find evidence suggesting that trait integration is not predicted by a linear model with equal weights for the traits (e.g., weight average). The main argument has been that traits are unequally weighted. Social perceivers are generally more sensitive to traits related with a morality dimension, such as honesty and trustworthiness, than to traits related with an ability dimension, such as competence and dominance (Abele & Bruckmüller, 2011; Brambilla & Leach, 2014; Goodwin, Piazza, & Rozin, 2014; Wojciszke, Bazinska, & Jaworski, 1998). This is supported by evidence suggesting that perceivers tend to prioritize the detection of morality-related traits over ability-related information (e.g., Abele & Bruckmüller, 2011; Ybarra, Chan, & Park, 2001), and to attribute higher informational weight to morality-related traits (Wojciszke et al., 1998). Altogether, these studies suggest when perceivers integrate traits into an impression and map this impression onto a face space, they will attribute a higher weight to trustworthiness than to dominance information. Thus, if the same priority of morality information occurs in social face perception, and traits are being combined in a linear fashion, then we should not expect them to be equally weighted.

The higher weight of morality information found in the general impression formation domain has also echoed in the social face perception literature. However, comparatively less evidence and focus can be found in the latter domain regarding the differential weighting of dimensions. For instance, Oosterhof and Todorov (2008)'s data show that the component of the trustworthiness (also interpreted as valence) dimension was the one that accounted for most of the variability in face impressions, compared to the component of the dominance (or power) dimension. In addition, studies showing that valence and trustworthiness judgments of faces occur as fast as between 33 and 167 ms after stimulus presentation (Todorov et al., 2009), may also be suggesting that those dimensions have priority over others. Nevertheless, these studies did not offer data that allow for a direct comparison between the weighting of trustworthiness and dominance in face impressions.

Altogether, the reviewed evidence suggests that if a linear combination reflects how trustworthiness and dominance are simultaneously mapped onto a face, the first will outweigh the second. However, although some of the reviewed evidence shows that a linear integration assumption has predictive power, there is no study addressing the validity of the linear

assumption in explaining trait integration. This assumption would be challenged if the traits' weights change with the context.

Here, we aim to test whether a personality trait is similarly weighted when the other trait (i.e., of the other dimension) changes. If that is not the case (i.e., traits are differently weighted in different contexts), the integration is not reflecting a simple linear compound. This hypothesis is supported by Asch's (1946) work showing that the trait-context modulates the meaning/weight attributed to a specific trait.

Reverse Correlating Trait Integration with Faces

To answer this question we needed a method capable of capturing the facial content associated with an impression previously formed by the perceiver. This possibility was offered by the so-called psychophysical reverse-correlation (RC) method, also known as the classification image technique (for a review see Brinkman, Todorov, & Dotsch, 2017; Jack & Schyns, 2017). Psychophysical RC allowed us to visualize how social dimensions such as trustworthiness or dominance are mapped into a face by the perceiver (Dotsch & Todorov, 2012). At the same time, it makes as few assumptions as possible about which features of the stimulus are associated with the target judgment. Hence, this method taps into perceivers' naïve theories about how a trait is mapped onto a face by capturing the strategies they used in a task in which they were instructed to detect a trait (signal) in noisy face stimuli. These strategies are assumed to be correlated with the visual content of a perceiver's mental representation of the target construct (see Brinkman et al., 2017).

Previous studies using RC methods have only inquired about how individuals separately represent each trait (e.g., tasks with a single target trait; see Dotsch & Todorov, 2012; Imhoff, Woelki, Hanke, & Dotsch, 2013). Here, we used this method to assess how people visually integrated the two traits into a face. Specifically, our aim was to assess how people mentally represent a face reflecting a combination of two different trait poles of the basic dimensions of social face perception (i.e., trustworthiness and dominance). By relying on this methodology, we aimed to obtain new empirical data documenting how more specific and complex personality judgments are integrated into a face.

The Current Research

We approached the question of how trait combinations are mentally integrated into a face by using a psychophysical RC method. Visual proxies of mental representations are obtained in the form of classification images (CIs). To obtain these CIs, we created a large set

of variations of the same face image by a repeated process of applying visual noise to that image. We asked participants to perform a two-image forced-choice task (2IFC) (e.g., Dotsch & Todorov, 2012; Dotsch, Wigboldus, & van Knippenberg, 2011), in which they were instructed to select from a pair of face images with superimposed visual noise the one that most resembled a person described by a specific trait combination. These data were then used to estimate a CI, by averaging across all the visual noise patches selected by the participant—indicating signal detection—and re-superimposing them on the base face image. In other words, the CI is a face image that represents how participants mapped the trait combination onto a face.

Our first approach (Study 1) aimed to examine the extent to which the visual content of each trait contributes to a final impression, and whether the degree of this contribution can be argued to be context dependent. Our main analysis addressed this question with an objective similarity approach (i.e., pixel correlations between CIs). Specifically, we examined the relationship between CIs generated with two-trait combinations (i.e., henceforth referred to as two-trait CIs or 2TCIs) and CIs generated using only one of the traits (henceforth referred to as one-trait CIs or 1TCIs, obtained in a previous study by Oliveira, Garcia-Marques, Dotsch, and Garcia-Marques (2019). By analyzing how much variance in a 2TCI can be explained by the two 1TCIs reflecting its component traits, we assessed the fit of a linear trait integration assumption. All CIs used in these analyses were subsequently validated with subjective judgments of dominance and trustworthiness performed by an independent sample of judges.

Study 2 replicated and consolidated the previous results obtained in Study 1 by (a) using more stable estimates for each CI (relying on bigger samples), (b) counter-balancing word sequence in instructions, and (c) adding two tasks in which the default 2IFC task stimuli were replaced by actual CIs. These two additional tasks allowed us to examine the degree of overlap between the signal contained in 2TCIs and 1TCIs.

Power Considerations of Statistical Analyses

The participant sample size of each study is relevant to inform about the stability of CI estimation. Although it has an implication on statistical power, it is not the participant sample size that is directly determining the power associated with the majority of the statistical analyses in this article. The units of analysis in these analyses are the pixel luminance values of each CI. Thus, it is the pixel sample size that is of relevance (65536 per unmasked CI, or 38958 pixels per masked CI). Because of these large sample sizes, effects may more easily

achieve significance at the nominal type I error rate (.05) regardless of their meaningfulness (e.g., Johnson, 1999). To help us evaluate the meaningfulness of our results, our main analysis was complemented by simulations aimed at identifying the range of values that we could expect to obtain by chance alone (see *R*-squared simulations section) (see also Murayama, Pekrun, & Fiedler, 2014).

The analyses of the CI ratings in Study 1 involved an independent sample of participants (i.e., not involved in CI generation), whose size was shown *a priori* to be sufficient to detect a medium effect size with 80% power and an error probability of $\alpha = .05$ (G*Power v.3.1.9.2; Faul, Erdfelder, Lang, & Buchner, 2007). Nevertheless, in these analyses we additionally report the power achieved in the detection of the effects ($1-\beta$).

Data Availability

The data required to generate the CIs and results reported in all studies, including additional supporting information, were made publicly available online at: https://osf.io/4zncm/

(Supporting information can be found in Appendix C of this thesis.)

Study 1

In this study we assessed composites of face images that result from perceivers' expectations about the facial appearance of someone possessing either a specific combination of traits of the poles of the dominance and trustworthiness dimensions (2TCIs), or just one of those traits separately (1TCI). To address whether or not the 2TCIs result from a linear combination of the single-trait information that perceivers map onto a 1TC1, we analyzed how the pixel information in each 2TCI was predicted by the pixel information of its correspondent 1TCIs using a multiple linear regression procedure. This analysis allowed us to examine whether the two predictive dimensions (i.e., of the 1TCIs) were symmetrically (as expected by a linear integration perspective) or asymmetrically (as expected by a gestaltic perspective) weighted in different trait combination contexts (i.e., different 2TCIs). We further approached these different hypotheses by generating CIs using the residual variance of our regression analyses. These residual CIs inform about the extent to which a 2TCI contains facial information that goes beyond that predicted by each of its correspondent 1TCIs, thus offering the opportunity to examine whether any gestaltic integration occurred. Subsequently, with the aim of validating the obtained CIs, we assessed how 2TCIs and 1TCIs were rated on dominance and

trustworthiness by independent judges. Furthermore, these ratings allowed us to test for differences in signal strength between the two traits of a 2TCI, in a bottom up fashion. This additionally informs about how the signal strength of any of these traits might have been modulated by the trait combination context in the 2TCI.

Participants and Design

CI generation sample. Eighty individuals, including undergraduate students and university staff (65 female, M_{Age} = 23.71 years, SD_{Age} = 9.38), participated in the main task of this study either in exchange for course credits, or to take part in a lottery to win gift vouchers. Participants were randomly assigned to one of four conditions, each of which corresponded to a combination of two traits: one trait represented one of the poles of dominance (i.e., dominant or submissive) and another represented one of the poles of trustworthiness (i.e., trustworthy or untrustworthy). Previous studies using RC tasks to estimate CIs of social traits (e.g., Dotsch & Todorov, 2012) established a sample size of 20 per CI as providing stable estimates (see also Power considerations of statistical analyses).

CI trait ratings sample. An independent sample of 29 participants (20 female, M_{Age} = 24.79 years, SD_{Age} = 5.21; undergraduate students in laboratory setting) was additionally recruited for a subsequent CI trait ratings task (see CI ratings section).

Materials and Procedure

Face stimuli. Stimuli were generated with the R package rcicr 0.3.0 (Dotsch, 2015). Each stimulus consisted of a base face image with a superimposed layer of visual noise. The base face image corresponded to the gray-scale average male face from the Karolinska Face Database (Lundqvist & Litton, 1998), resized to 256 × 256 pixels. The noise patterns superimposed on the base image consisted of truncated sinusoid patches in six orientations (0º, 30º, 60º, 90º, 120º, and 150º), five spatial frequency scales (2, 4, 8, 16, and 32 cycles per image), and two phases (0, π/2) with random amplitudes (for more details see Dotsch & Todorov, 2012). Three hundred pairs of stimuli were generated for a 2IFC (i.e., two-image forced-choice) task. Each stimulus pair was designated to be presented in one trial of the 2IFC task. Therefore, the task consisted of 300 trials. For each pair of images, the noise pattern of one image was the negative (i.e., opposite values of pixel luminance) of the other, thus maximizing the differences between the images and circumventing the possibility of highly similar image pairs inflating guessing responses throughout the task.

Two-trait combinations. The traits selected for the task included the poles of the dominance and the trustworthiness dimensions. These poles were combined in four different pairings in such a way as to represent the four quadrants of the two-dimensional trustworthiness-by-dominance space. Therefore, the resulting trait combinations were: *Dominant & Trustworthy* (DomTrust), *Dominant & Untrustworthy* (DomUntrust), *Submissive & Trustworthy* (SubTrust), and *Submissive & Untrustworthy* (SubUntrust).

Reverse correlation task. Participants were recruited to take part in a face evaluation study. The experiment was conducted in a lab setting, where participants were randomly assigned to individual cubicles equipped with desktop computers. They were told that they would see pairs of face images exhibiting a type of noise similar to that of "old television screens" and that one of the aims of the study was to understand how image quality affected face evaluation. This last detail was included to serve as a cover story explaining why the target faces were noisy. The 2IFC task had a total of 300 trials, with a forced break of one minute after the 150th trial. In each trial, a pair of face images was presented side-by-side (i.e., location of image with negative noise was counterbalanced) on the center of the screen and the participants' task was to select the face from the pair that, in their opinion, most resembled the face of a person defined by a specific combination of two traits (e.g., dominant and trustworthy). After completing the task, participants were thanked and debriefed.

Classification image generation. We computed a grand-mean two-trait classification image (2TCI) for each of the four conditions. To compute each 2TCI, we first averaged all the noise patterns selected by all the participants in a condition, and subsequently superimposed the averaged noise pattern on the base image. The resulting 2TCIs are depicted in Figure 1.

CIs generated using only one trait (i.e., 1TCIs), were included in our studies to serve as a reference for the visual inspection of the 2TCIs, and to be entered as predictor variables in subsequent analyses (see Results and Discussion). All the 1TCIs shown in Figure 1 were generated in a previous study by Oliveira et al. (2019) who used an identical CI generation procedure, the same dimensions for the CIs (256 × 256 pixels; N = 65536 pixels per CI) , and 20 participants per CI condition sampled from the same population.

CI ratings. The four resulting 2TCIs and the 1TCIs (the latter obtained from Oliveira et al., 2019) were submitted to dominance and trustworthiness judgments by an additional independent sample of judges (sample details in Participants and design section). Participants judged the 2TCIs and the 1TCIs on a scale ranging from 1 (not at all [trustworthy; dominant]) to 4 (not very [trustworthy; dominant] nor not at all [trustworthy; dominant]), to 7 (very [trustworthy, dominant]). Each CI was rated on all target traits before the next CI. The order

of CIs and the two target trait judgments within the CI blocks were randomized. Additional trait ratings (e.g., threat, competence, and others related with perceptions of morality and sociability) were collected for these CIs for exploratory purposes but are not reported in here for the sake of simplicity. However, because some of these additional ratings (i.e., threat and competence) may be of interest to researchers interested in comparisons with related studies (i.e., Oosterhof & Todorov, 2008; Todorov et al., 2008) we made them available as supplementary data in our online data repository.

Results

In our analyses the data corresponded to the classification images (i.e., 2TCIs and 1TCIs) generated at the group-level in each trait combination condition. We assessed how these CIs were related using their pixel luminance values as the units of analysis. First, we tested whether traits were linearly (symmetrically) integrated by addressing how each 2TCI was related to their correspondent 1TCIs using multiple regression analyses. Subsequently, we assessed the meaningfulness of these models' results by conducting simulations to identify the range of values that we could expect to obtain by chance given these data (see *R*-squared simulations section). These simulations informed us about the validity of the conclusions suggested by the regression models' results.

A second follow-up analysis addressed the impact of our methodological decision to include all pixels of the CIs (vs. pixels of the face region only) on the explained variance of our regression models. Specifically, we repeated our analysis using only the pixels of the face region (i.e., masked CIs). To further test whether a simple linear trait combination is able to entirely account for a 2TCI, we explored the residual images of our CI regressions. Residual images displaying any face information, especially if it resembles the 2TCI itself, would suggest that perceivers integrate traits in such a way that they give rise to an impression that goes beyond the mere sum of its parts (e.g., Asch, 1946; Asch & Zukier, 1984).

Finally, to validate the subjective content of each CI, we analyzed ratings of the CIs on the two traits used to generate them, performed by an independent sample of judges. Moreover, this analysis allowed us to test the results of the integration process, as the 2TCIs' trait also inform about the extent to which one trait outweighed the other in different trait combination contexts.

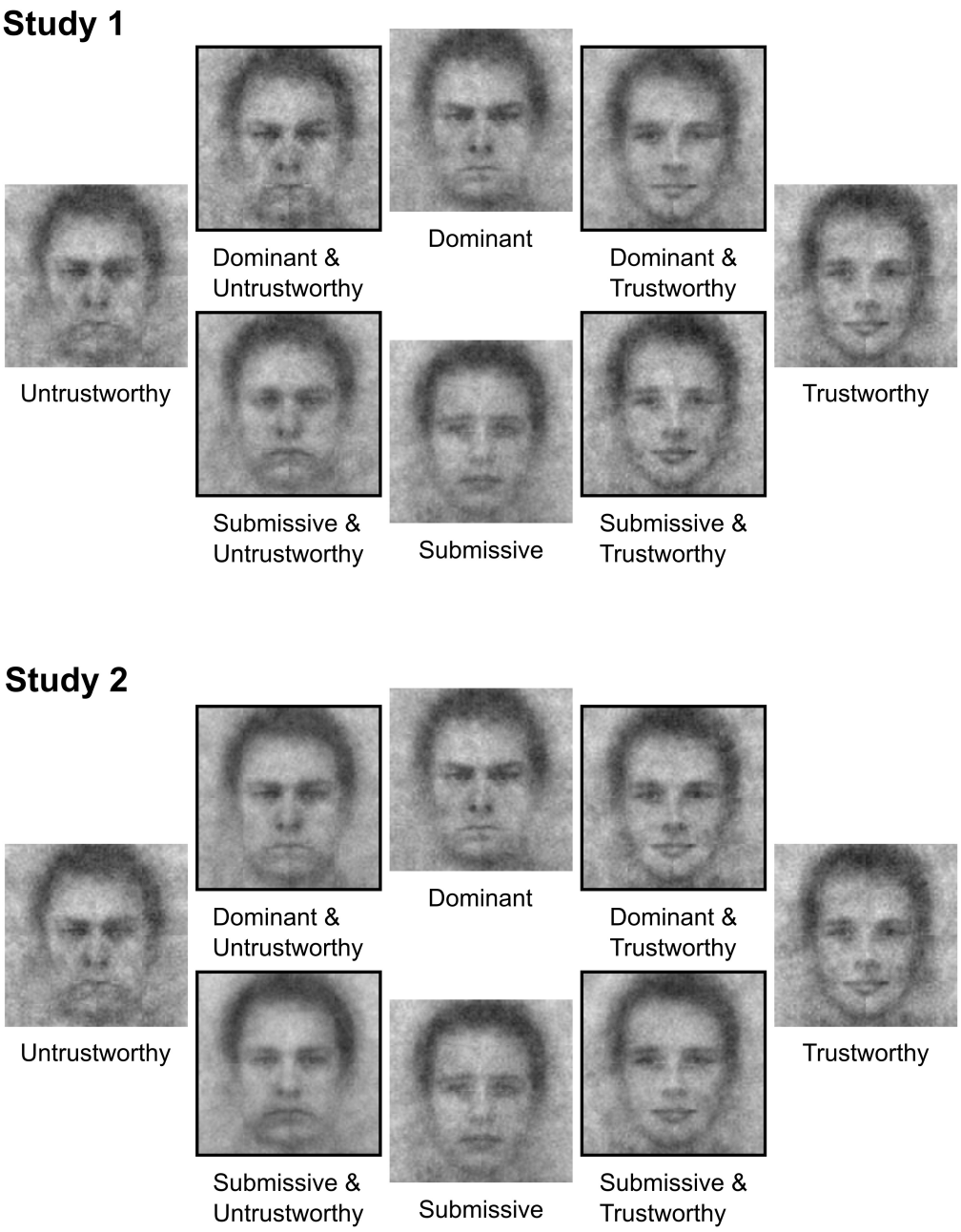

Figure 1. 2TCIs (framed pictures) obtained in Studies 1 and 2, for each of the four two-trait combinations of the poles of the dominance and trustworthiness dimensions, and 1TCIs (unframed pictures) extracted from Oliveira et al. (2019).

Classification Images

A visual inspection of the upper panel of Figure 1 (Study 1) suggests that there is not a full overlap between any of the 2TCIs and the 1TCIs. This hints that the participants used the two traits in their judgments during the task. A greater similarity between 1TCIs and 2TCIs generated with positive trustworthiness suggests that this trait (i.e., trustworthy) weighted more in the 2TCIs compared to other traits. However, 2TCIs that included negative trustworthiness resulted in a higher differentiation of dominance-related features between the 2TCIs.

Although CIs are by themselves informative because of their visual nature, our interpretation of the CIs is limited by subjectivity. To circumvent this limitation, we submitted the CIs to a more objective analysis of similarity, described below.

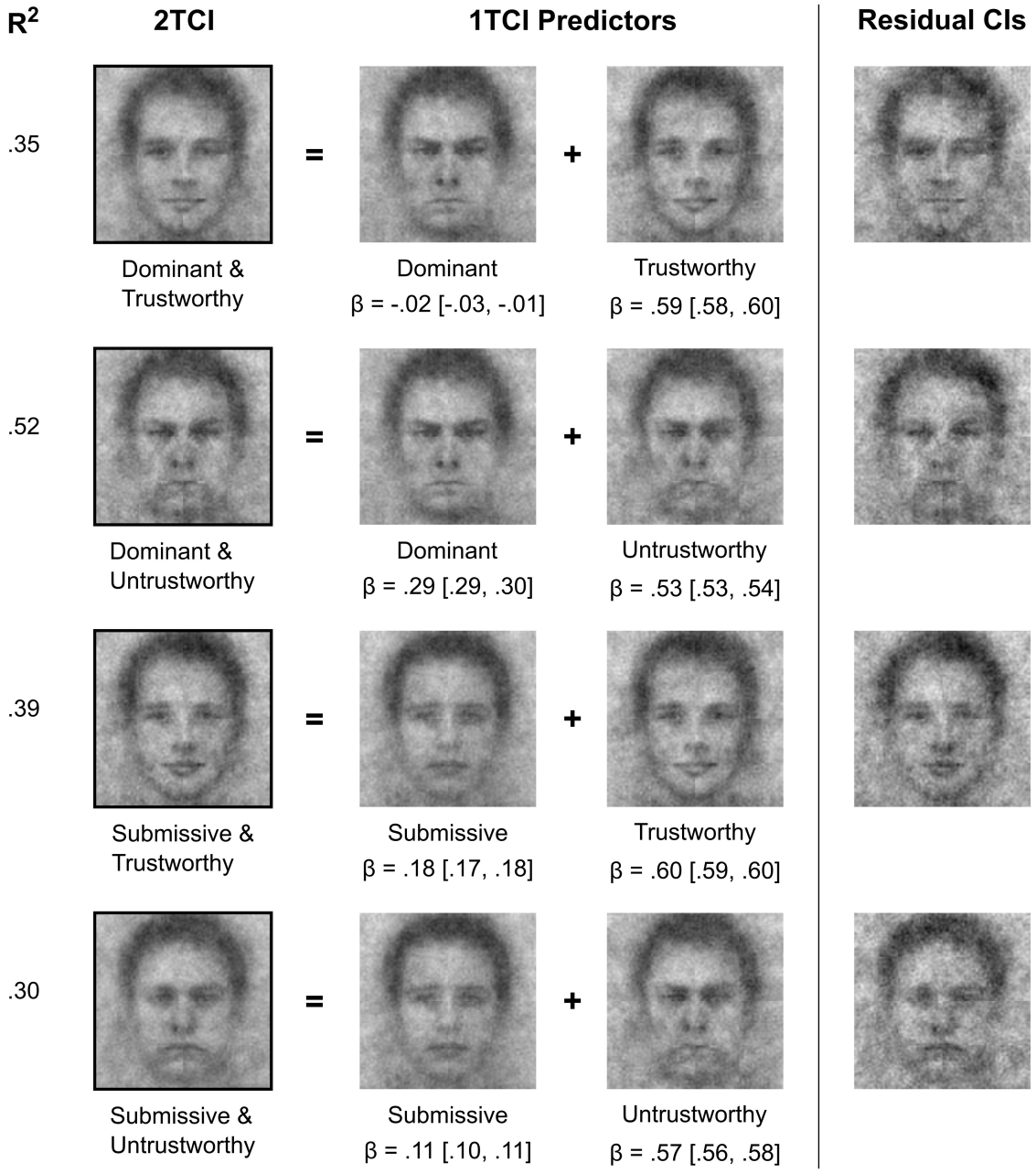

Figure 2. 2TCIs (framed pictures) obtained in Study 1 for each of the four two-trait combinations of the poles of the dominance and trustworthiness dimensions, and 1TCIs (unframed pictures) from Oliveira et al. (2019). All results presented here are for the unmasked CIs. Each row illustrates a regression model with a 2TCI as the outcome and the trait combinations of correspondent 1TCIs. Residual CIs represent the variance in the pixel data left to be explained by each model. Reported values correspond to standardized beta coefficients [95% confidence intervals]. All coefficients were significantly different from zero, all *ps* < .001.

Main Analysis: Objective Similarity between 2TCIs and 1TCIs

An objective measure of similarity between two CIs is obtained by computing how the pixel luminance values of the two images co-vary. The stronger a positive (negative) correlation is between two CIs, the more similar (dissimilar) they are, whereas close to null correlations indicate that the CIs share little to no similarities. Taking this into account, we relied on regression analyses to test our trait integration hypotheses, as they are more informative than a correlation matrix. Regression analyses allowed us to predict a 2TCI from its two correspondent 1TCIs, and to examine how much variance each predictor CI shares with (i.e., is objectively similar to) an outcome 2TCI.

In four separate regression analyses we analyzed how the pixel values of each 2TCI (N = 65536 pixels per CI) would be predicted by the pixel values of 1TCIs (N = 65536 pixels per CI) generated for each of the traits that made part of the 2TCI trait combination (e.g., Dominant 1TCI and Trustworthy 1TCI entered as predictors of the Dominant & Trustworthy 2TCI). For that purpose, we used 1TCIs previously generated for the poles of the dominance and trustworthiness dimensions. For this analysis, no masks (i.e., isolation of specific regions of the image) were applied to the CIs, and thus, all the information contained in the CIs was included in the analyses. Results of the multiple regression models are shown in Figure 2. All the F-ratios of the models depicted in Figure 2 were significant (all ps < .001) and ranged from $F(2, 65533) = 14370$ (Submissive & Untrustworthy) to $F(2, 65533) = 35950$ (Dominant & Untrustworthy).

Importantly for our aims, the contribution of each specific trait for the 2TCI varied as a function of the trait combination context. The results consistently showed that the trustworthiness-related 1TCIs primarily predicted the 2TCI across all four trait combinations. However, the weight of a 1TCI trait in the 2TCI also changes depending on the combination context. Whereas practically only trustworthiness is present in the DomTrust 2TCI, the contribution of trustworthiness and dominance becomes conjoined in other 2TCIs.

The amount of explained variance in each model also informs us about the extent to which each 2TCI shares visual information with both of its 1TCI predictors. So, the higher the R^2, the more the visual information of a 2TCI can be explained by its two 1TCI predictors. A theoretical assumption regarding the linear integration of the two traits would lead us to expect similar R^2 values across the four regression models. Our data suggest, instead, that the contribution of the 1TCIs to the 2TCI was disproportionate and varied depending on the trait context (see Figure 2 and Table 1). As shown in Table 1 (see Pure 2TCI model), the

confidence intervals for the R^2 values do not overlap between 2TCI conditions, thus suggesting that they are not similar.

An alternative methodological approach to assess this explained variance is to create artificial 2TCIs for each of the four trait combinations, by averaging the correspondent pairs of 1TCIs, and assess their squared correlation with the "pure" 2TCIs. Although this approach ignores differences in variability associated with each 1TCI overall, the R^2 values were lower than those associated with the multiple regression analyses. As Table 1 shows, this analysis leads to a similar conclusion: A general weak overlap between a pure and its correspondent artificial 2TCI, which varies depending on the trait context. Thus, each 2TCI is unlikely to result from a simple linear combination of the 1TCIs.

Our next analysis aimed to assess the validity of the conclusions suggested by these data.

Table 1

Variance of Pure 2TCI Accounted For by Artificial 2TCI with Pure 2TCI Model Values as Reference for Unmasked and Masked 2TCIs, in Studies 1 and 2.

Study	CI masking	Trait combination	R^2 [95% C.I.]	
			Pure - Artificial 2TCIs	Pure 2TCI model
1	Unmasked	Dom. & Trust.	.15 [.15, .16]	.35 [.35, .36]
		Dom. & Untrust.	.51 [.51, .52]	.52 [.52, .53]
		Sub. & Trust.	.35 [.35, .36]	.39 [.38, .40]
		Sub. & Untrust.	.22 [.22, .23]	.30 [.30, .31]
	Masked	Dom. & Trust.	.20 [.19, .21]	.43 [.42, .43]
		Dom. & Untrust.	.53 [.52, .54]	.55 [.54, .56]
		Sub. & Trust.	.41 [.41, .42]	.45 [.45, .46]
		Sub. & Untrust.	.27 [.26, .28]	.34 [.33, .35]
2	Unmasked	Dom. & Trust.	.21 [.20, .22]	.43 [.42, .43]
		Dom. & Untrust.	.47 [.47, .48]	.49 [.48, .50]
		Sub. & Trust.	.40 [.39, .40]	.41 [.41, .42]
		Sub. & Untrust.	.18 [.18, .19]	.18 [.18, .19]
	Masked	Dom. & Trust.	.24 [.23, .25]	.45 [.44, .46]
		Dom. & Untrust.	.49 [.48, .50]	.51 [.51, .52]
		Sub. & Trust.	.41 [.40, .41]	.42 [.41, .42]
		Sub. & Untrust.	.18 [.18, .19]	.18 [.18, .19]

Note: Artificial 2TCIs result from the artificial averaging of two 1TCIs for a given trait combination (e.g., Artificial DT 2TCI is the average of the Dominant and Trustworthy 1TCIs). The Pure-Artificial 2TCI R^2 values represent the variance in a pure 2TCI accounted for by the artificial 2TCI derived from the same trait combination. The column Pure 2TCI model R^2 provides the values obtained in the regression models where pure 2TCIs were predicted by the 1TCIs, as a reference to compare with the Pure-Artificial 2TCI R^2 values.

R-squared simulations. To ascertain whether the observed R^2 values are meaningful or could have been obtained by chance alone, we ran a simulation for each 2TCI model. In each simulation, we substituted the two 1TCI predictors by two randomly generated CIs. Unlike the 1TCIs, the randomly generated CIs did not, in principle, contain any specific trait signal, since they were composed of random visual noise. Thus, in these simulation models, each 2TCI was essentially being predicted by random noise. As a result, we expected the explained variance of these models to drop to close to null values. For each simulation, we generated two different stimuli sets, using a different randomization seed for each set (specifically, 1 and 2). Each set consisted of 300 pairs of stimuli, which is the same amount of pairs used in the study's main task (i.e., 2IFC task), to simulate the 2IFC task's 300 trials. Next, we generated 300 random responses (i.e., coded as -1 or 1, corresponding to the

selection of one or the other image of a pair) per participant for a total of 20 participants per set, as this was the sample size of the real data. This resulted in 20 individual CIs per set. Finally, for each 2TCI, we ran a total of 20 regression models. In these models one of the 2TCIs was fixed as the outcome variable, and the predictors were two of the randomly generated individual CIs (substituting the 1TCIs). At each iteration of the simulation, each of the individual CI predictors was sampled from a different set than that of the other predictor.

The R^2 ranges obtained for each predicted 2TCI simulation are described in Table 2. The resulting ranges represent the interval of R^2 values that could be expected to be obtained by CIs without trait signal (i.e., random noise). Since none of the R^2 values of the four 2TCI regression models falls within the simulations' R^2 ranges, they are unlikely to be due to chance, and thus, we can be more confident that the trait signal contained in the 1TCIs accounts for the observed variance in the outcome 2TCI.

Table 2

R^2 Values Observed in Study 1 Versus R^2 Values Expected to be Obtained by Chance based on Simulated Data, for each Predicted 2TCI.

		Simulation R^2				
				Quantile		
Predicted 2TCI	Observed R^2	*M*	*SD*	50%	95%	100%
Dom. & Trustw.	0.35	0.011	0.010	0.008	0.030	0.033
Dom. & Untrustw.	0.52	0.013	0.013	0.010	0.041	0.050
Sub. & Trustw.	0.39	0.015	0.013	0.015	0.036	0.043
Sub. & Untrustw.	0.30	0.014	0.014	0.011	0.042	0.054

Masked versus unmasked CIs. Our main analysis was performed using all the pixel information in a CI. It could be argued that our main analysis' results were driven by the inclusion of pixel information outside the face region of a CI, which in turn could be inflating explained variance. Alternatively, it could also be the case that these outer pixel regions were themselves a source of variability. To examine this possibility we repeated our analysis by now restricting it to the most meaningful information in the CIs (i.e., face region). Both the outcome 2TCI and the 1TCI predictors were masked. The applied mask was oval-shaped and

kept only the area containing the full face (including hair) in the analyzed image (mask and scripts of the analyses are provided in our online data repository). The results of the regression models using masked CIs, including the R^2 of the models using unmasked CIs as a reference, are shown in Table 3. Results indicate that the explained variance was generally higher in models using masked CIs, thereby ruling out the possibility that the inclusion of outer pixel regions was inflating explained variance and showing that the use of unmasked CIs in fact decreased it. Finally, the observed differences in explained variance varied in extent across 2TCIs. 2TCIs including high trustworthiness exhibited a larger gap between the explained variance of masked and unmasked models, compared to 2TCIs including untrustworthiness. This, however, may just be reflecting that high trustworthiness 1TCIs explain more variance in the outcome 2TCI compared to its untrustworthiness counterparts (see Figures 2 and 4).

Table 3

R^2 Values and Standardized Coefficients (With 95 % Confidence Intervals) for Regression Models of Masked Classification Images, and Reference R^2 Values of Models of Unmasked Classification Images, for Studies 1 and 2.

		Masked 1TCI Predictors		Masked CIs model		
Study	Masked 2TCI	β Dominance [95% C.I.]	β Trustw. [95% C.I.]	F (2, 38955)	R^2	Unmasked model R^2
1	DT	0.00 [-0.01, 0.00]	0.65 *** [0.64, 0.66]	14410 ***	.43	.35
	DU	0.25 *** [0.24, 0.26]	0.58 *** [0.57, 0.58]	23690 ***	.55	.52
	ST	0.20 *** [0.19, 0.20]	0.65 *** [0.64, 0.65]	16140 ***	.45	.39
	SU	0.15 *** [0.15, 0.16]	0.61 *** [0.60, 0.62]	10100 ***	.34	.30
2	DT	0.04 *** [0.03, 0.05]	0.68 *** [0.68, 0.69]	15820 ***	.45	.43
	DU	0.21 *** [0.21, 0.22]	0.58 *** [0.57, 0.59]	20530 ***	.51	.49
	ST	0.27 *** [0.26, 0.28]	0.59 *** [0.58, 0.60]	13920 ***	.42	.41
	SU	0.28 *** [0.27, 0.29]	0.41 *** [0.41, 0.42]	4417 ***	.18	.18

Note: DT = Dominant & Trustworthy; DU = Dominant & Untrustworthy; ST = Submissive & Trustworthy; SU = Submissive & Untrustworthy. $\alpha = 0.05$; *** $p < .001$

Residual CIs: Visualizing Beyond the Two Traits

The variance in each model that is not explained by the two 1TCI predictors can be interpreted not only as error but also, and more interestingly, as additional information mapped in the 2TCI that goes beyond a linear combination of the two separate traits.

To visualize the unexplained variance for each model, we computed additional CIs from the residuals. These residual CIs are shown in Figure 2; they appear to be very similar to their correspondent 2TCIs. Such similarity suggests that there is a large portion of additional trait information in the 2TCI beyond that which is accounted for by the 1TCIs and error. Otherwise, if the 1TCIs accounted for all of the 2TCIs' variance, the resulting residual CI would be an image mostly filled with pixels that are uniformly one value with no variation (e.g., all pixels would exhibit the same color, and thus, only the underlying base image would remain in the image; for a demonstration see the R scripts regarding the artificial 2TCIs analyses in our online data repository).

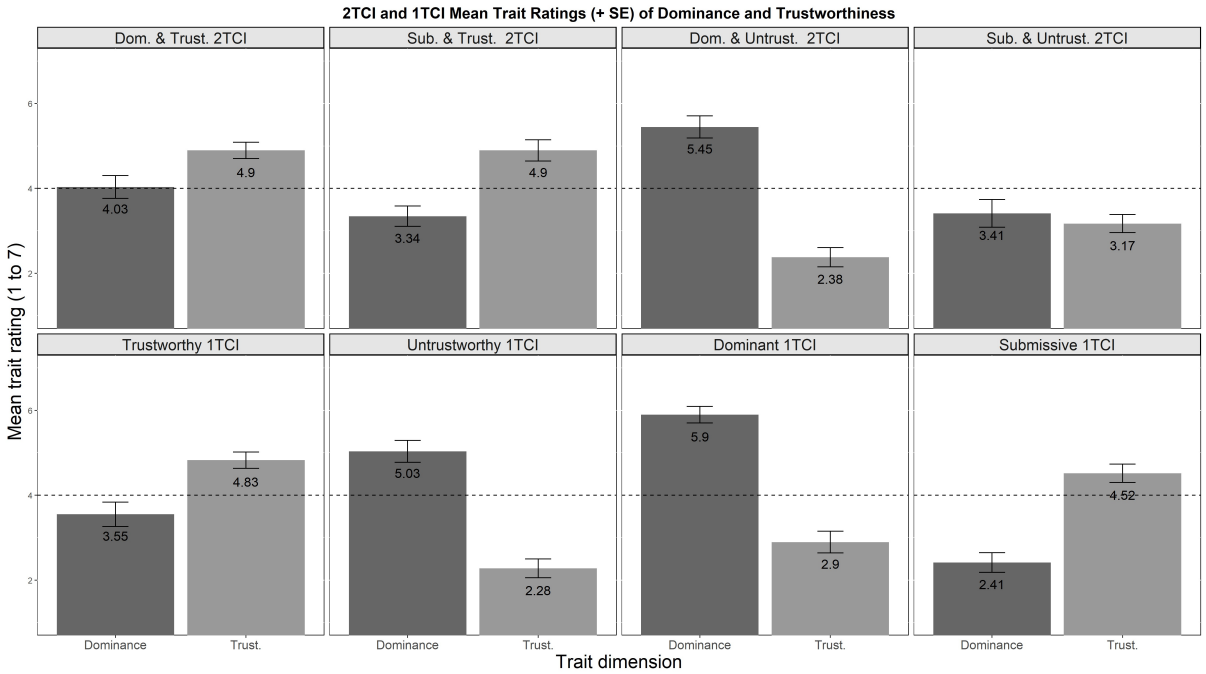

Figure 3. Mean trait ratings of dominance and trustworthiness obtained for the 2TCIs generated in Study 1, and for the 1TCIs obtained from Oliveira et al. (2019). The horizontal line intercepting the y-axis at value 4 represents the midpoint of the rating scale which was labeled in the task as "not very [trait] nor not at all [trait]". Error bars denote within-subjects standard errors.

CI Trait Ratings

Figure 3 presents the results for the CI ratings task, informing about whether the CIs were perceived as intended in the trait(s) used to generate them. For the sake of simplicity, we submitted the 1TCI ratings and the 2TCI ratings to separate within-participants ANOVAs, and applied a Benjamini-Hochberg correction to the omnibus tests' p-values to control for the false discovery rate (FDR, Benjamini & Hochberg, 1995). Thus, we additionally report FDR corrected *p*-values for every omnibus test. Effect sizes were calculated following the recommendations described in Lakens (2013) for within-participants designs. Probability of Type I error (α) was set to .05 for all analyses.

1TCI ratings. The overall pattern of the 1TCI ratings presented on Figure 3 shows that each 1TCI was rated higher in its intended trait (e.g., trustworthy 1TCI rated as high on trustworthiness). We tested the significance of these differences, by submitting the 1TCI

ratings to a 2 (Dimension: dominance vs. trustworthiness) × 2 (Pole: high vs. low) within-subjects ANOVA for each separate trait rating (i.e., dominance or trustworthiness).

Trustworthiness ratings yielded the expected Dimension × Pole interaction, $F(1, 28) = 99.45$, $p < .001$ ($p_{FDR} = .002$), $\eta_G^2 = .44$, $1\text{-}\beta = 1.00$, indicating that the difference in perceived trustworthiness occurred as intended between the two trustworthiness-related 1TCIs, $t(28) = 9.30$, $p_{tukey} < .001$, $M_{diff} = 2.55$, M_{diff} 95% CI [1.98; 3.11], Cohen's $d_z = 1.73$ (see Figure 3). However, suggesting that the two dimensions are not perceived as independent, the two dominance-related 1TCIs also differed in perceived trustworthiness, $t(28) = 6.02$, $p_{tukey} < .001$, $M_{diff} = 1.62$, M_{diff} 95% CI [1.06; 2.17], Cohen's $d_z = 1.12$, with the dominant 1TCI perceived as less trustworthy than the submissive 1TCI (cf. Oosterhof & Todorov, 2008).

For the dominance ratings, the ANOVA yielded a Dimension × Pole interaction, $F(1, 28) = 82.40$, $p < .001$ ($p_{FDR} = .002$), $\eta_G^2 = .48$, $1\text{-}\beta = 1.00$, indicating that the dominant 1TCI was rated as more dominant than the submissive 1TCI as intended, $t(28) = 12.29$, $p_{tukey} < .001$, $M_{diff} = 3.48$, M_{diff} 95% CI [2.90; 4.06], Cohen's $d_z = 2.28$. Again suggesting that the two dimensions are not perceived as independent, the perceived dominance of the trustworthiness-related 1TCIs also differed, $t(28) = 3.61$, $p_{tukey} < .003$, $M_{diff} = 1.48$, M_{diff} 95% CI [0.64; 2.32], Cohen's $d_z = 0.67$, with the trustworthy 1TCI being rated as less dominant than the untrustworthy 1TCI (see Figure 3). The negative relationship between perceived dominance and trustworthiness was further corroborated by a Pearson correlation coefficient computed across all 1TCIs ($r = -.36$, $p < .001$).

Overall, these results indicate that not only each 1TCI contained their intended trait signal but also that this signal modulates the perception of dominance.

2TCI ratings. To examine how each 2TCI was evaluated on each of the two traits used to generate it, we submitted their dominance and trustworthiness ratings to a 2 (Trustworthiness: 2TCI with trustworthy vs. 2TCI with untrustworthy) × 2 (Dominance: 2TCI with dominant vs. 2TCI with submissive) within-subjects ANOVA for each separate trait rating. Although we expected to corroborate that participants perceived the 2TCIs as intended on the two traits (e.g., DomUntrust 2TCI rated high on dominance and low on trustworthiness), we also assumed that the integration of the two traits during the 2TCI generation task would modulate how they are perceived in a face (e.g., any differential weighting of traits underlying the 2TCI would be reflected on the ratings).

Results for the trustworthiness ratings yielded the expected main effect for Trustworthiness, $F(1, 28) = 84.28$, $p < .001$ ($p_{FDR} = .002$), $\eta_G^2 = .45$, $1\text{-}\beta = 1.00$, indicating

that 2TCIs generated with trustworthy were rated as more trustworthy than 2TCIs generated with untrustworthy (see Figure 3). However, the integration of the traits promoted a Trustworthiness × Dominance interaction, $F(1, 28) = 6.65$, $p = .015$ ($p_{FDR} = .015$), $\eta_G^2 = .03$, 1-β =.70, indicating that the difference in trustworthiness ratings between the DomTrust and DomUntrust 2TCIs was greater than between the SubTrust and SubUntrust 2TCIs. Post-hoc comparisons further clarified that the DomTrust and SubTrust 2TCIs did not differ in perceived trustworthiness (equal means), whereas the DomUntrust 2TCI was perceived as less trustworthy than the SubUntrust 2TCI, $t(285) = 2.85$, $p_{tukey} = .005$, $M_{diff} = 0.79$, M_{diff} 95% CI [0.22; 1.36], Cohen's $d_z = 0.53$. This suggests that 2TCIs generated with untrustworthy were more differentiated than 2TCIs generated with trustworthy. This could not have resulted from a simple addition or subtraction of the information about the other trait. The specific trait combination creates a context that changes their perception.

For the dominance ratings, besides the expected main effect for Dominance, $F(1, 28) = 30.78$, $p < .001$ ($p_{FDR} = .002$), $\eta_G^2 = .18$, 1-β = 1.00, the ANOVA also yielded a Dominance × Trustworthiness interaction, $F(1, 28) = 9.75$, $p = .004$ ($p_{FDR} = .005$), $\eta_G^2 = .05$, 1-β = .85. This interaction indicates that the difference in perceived dominance between the DomUntrust and SubUntrust 2TCIs, $t(28) = 4.88$, $p_{tukey} < .001$, $M_{diff} = 2.03$, M_{diff} 95% CI [1.18; 2.88], Cohen's $d_z = 0.91$, was greater than the difference between the DomTrust and SubTrust 2TCIs, $t(28) = 3.46$, $p_{tukey} = .131$, $M_{diff} = 0.69$, M_{diff} 95% CI [0.28; 1.10], Cohen's $d_z = 0.64$, suggesting an inflation of dominance evaluations when dominance traits are combined with untrustworthiness.

This analysis demonstrates that the trait combination context modulated the strength of the expression of a trait in a 2TCI. Importantly, this pattern of results indicates that the trait combination context distorted the negative correlation between dimensions observed in the 1TCIs. This distortion was more apparent in the DomTrust and SubUntrust 2TCIs whose trait combinations do not reflect a negative relationship between dimensions, like the DomUntrust and SubTrust 2TCIs do. Although more clearly in the DomTrust 2TCI than in the SubUntrust 2TCI, it was trustworthiness, rather than dominance information, that prevailed in the CI. This seems to suggest that being informed that someone is untrustworthy promotes visualizing their face as dominant, whereas being informed that someone is dominant does not promote visualizing them as untrustworthy. However, this interpretation should be taken with caution as this study was not designed to directly test that possibility.

Nevertheless, although our pattern of results cannot be entirely explained by a simple negative correlation between the two trait dimensions, a negative correlation between

dimensions still emerges when it is computed across all 2TCIs (r = -.30, p = .001), as this aggregation counterbalances any distortions induced by the trait context.

Study 2

Participants and Design

One hundred sixty undergraduate students from ISPA – Instituto Universitário (148 female, M_{Age} = 20.29 years, SD_{Age} = 4.67) participated in the study in exchange for course credits. In total, there were eight between-participants conditions defined by the 2IFC task instruction. As in Study 1, each condition (n = 40) corresponded to a unique combination of two trait poles: one from dominance and another from trustworthiness. The position of a trait in the task instruction was counterbalanced between-participants. This resulted in two instruction conditions for every specific two-trait combination (e.g., "trustworthy and dominant" and "dominant and trustworthy"). Still, it must be noted that we expected both traits of the combination to be activated simultaneously in the mind of a perceiver during a task trial, as opposed to being sequentially processed across time.

Extended Reverse Correlation Task

The 2IFC task stimuli and procedure was identical to that in Study 1 until the 300th trial. After the 300th trial, there were five additional trials in which we presented strategic pairings of the previously obtained CIs (i.e., 1TCIs and 2TCIs) instead of the usual pair of images generated to serve as task stimuli. The specific pairings are listed in Table 4. The five additional CI pairs were presented in random order. The side of the screen on which each CI from a pair was shown was counterbalanced between-participants. Importantly, for these additional trials, the task instructions did not change; they were exactly the same instructions that were given for the previous 300 trials of the task. That is, the participant's task was to select from a pair of face images the one that most resembled a person who possessed both of the two target traits. There were two different (hidden) tasks in these additional trials: one that involved a decision that could be considered "correct," and another in which any decision taken was "partially incorrect." Henceforth, we refer to them as Task I and Task II, respectively.

Task I. This task included trials in which each group of participants continued to perform judgments for one of the four combinations of traits (see Table 4). In each of these

four trials, the 2TCI—previously generated in Study 1—that matched the current combination of traits was paired with each of the four 1TCIs. Therefore, each group of participants in a 2TCI condition performed four trials. For example, for the group of participants who were instructed to detect the "dominant and trustworthy" combination, the 2TCI was always the DomTrust 2TCI and the 1TCI paired with it changed across these four trials. Importantly, we considered the selection of the 2TCI as the correct response in all of these trials, since the 2TCI was expected to contain signal of the target trait combination of the condition.

Task II. This task corresponded to only one trial (see Table 4). In this trial, each group of participants in a 2TCI condition was presented with a pair of 1TCIs. Each 1TCI corresponded to one of the traits in the target combination. For example, for the group of participants who were instructed to detect the "dominant and trustworthy" combination, this trial would exhibit the dominant 1TCI and the trustworthy 1TCI side by side. Since each 1TCI was expected to contain the signal of only one of the two traits in the task instruction, the decision to select either 1TCI was always (partially) incorrect. This task was devised to simulate a scenario in which both target stimuli exhibited partial signal, and to examine whether the participant exhibited some bias toward one of the traits.

After completing all of the tasks, participants were thanked and debriefed.

Table 4

CI Stimuli Pairings for the Five Last Trials of the 2IFC Task, by 2TCI Condition

Task	2TCI Condition			
	Dom. & Trust.	Dom. & Untrust.	Sub. & Trust.	Sub. & Untrust.
I	DT vs. D	DU vs. D	ST vs. D	SU vs. D
	DT vs. T	DU vs. T	ST vs. T	SU vs. T
	DT vs. S	DU vs. S	ST vs. S	SU vs. S
	DT vs. U	DU vs. U	ST vs. U	SU vs. U
II	D vs. T	D vs. U	S vs. T	S vs. U

Note: D = Dominant 1TCI; S = Submissive 1TCI; T = Trustworthy 1TCI; U = Untrustworthy 1TCI; DT = Dominant & Trustworthy 2TCI; DU = Dominant & Untrustworthy 2TCI; ST = Submissive & Trustworthy 2TCI; SU = Submissive & Untrustworthy 2TCI; DT, DU, ST, and SU correspond to the 2TCIs obtained in Study 1.

Results

Our analytical approach was similar to Study 1's. We expected to find that (a) the 1TCIs are meaningful predictors of the 2TCIs, (b) that trustworthiness plays a stronger predictive role than dominance, and (c) that these linear models would differ across trait combination conditions. Thus, our main analysis consisted in analyzing the objective similarity between each 2TCI and both of its correspondent 1TCIs. Again, we conducted this analysis using either unmasked or masked CIs to examine the impact of CI masking procedure on our results; and computed the residual CIs to visualize any information unaccounted for by the 1TCIs in each 2TCI regression model.

Next, we examined the strategies used by participants to integrate the traits during CI estimation in the 2IFC task. First, we examined participants' performance in detecting the signal of the two-trait combination during Task I with a binomial logistic regression analysis. Finally, we addressed the bias towards any of the two dimensions that could be occurring during the trait integration process, with a Log linear analysis of the response data obtained in Task II.

Classification Images

The 2TCIs depicted on the bottom panel of Figure 1 (Study 2) were generated using the same procedure described in Study 1, and excluded the five additional trials of Tasks I and II (intended for separate analyses). The overall pattern of similarities (based on our subjective assessment) among these CIs appears to remain, and thus, we maintain the same view reflected in our previous comments (see Study 1 Results).

Main analysis: Objective Similarity Between 2TCIs and 1TCIs

As in Study 1, we conducted a regression analysis for each 2TCI to examine how much variance is explained by each of the two correspondent 1TCIs (see Figure 4). In these analyses, the data points corresponded to CI pixel values (N = 65536 pixels per CI). The results indicated, again, a higher contribution of the trustworthiness 1TCIs to all the final 2TCIs. All the F-ratios of these models were significant (all ps < .001) and ranged from $F(2, 65533) = 7256$ (Submissive & Untrustworthy) to $F(2, 65533) = 31490$ (Dominant & Untrustworthy).

Again, we found that the explained variance varied across the four 2TCI linear models. Overall, the pattern of explained variance was similar to the one found in Study 1 (see Figure 4 and Table 1). Once more, we analyzed how the 2TCIs obtained in this study (i.e., pure

2TCIs) were related to their correspondent artificial 2TCIs (i.e., average of two 1TCIs) (see Table 1). Overall, this analysis yielded similar results to Study 1's, and suggests that the assumption of a context-independent linear integration of individual traits is unlikely to have occurred.

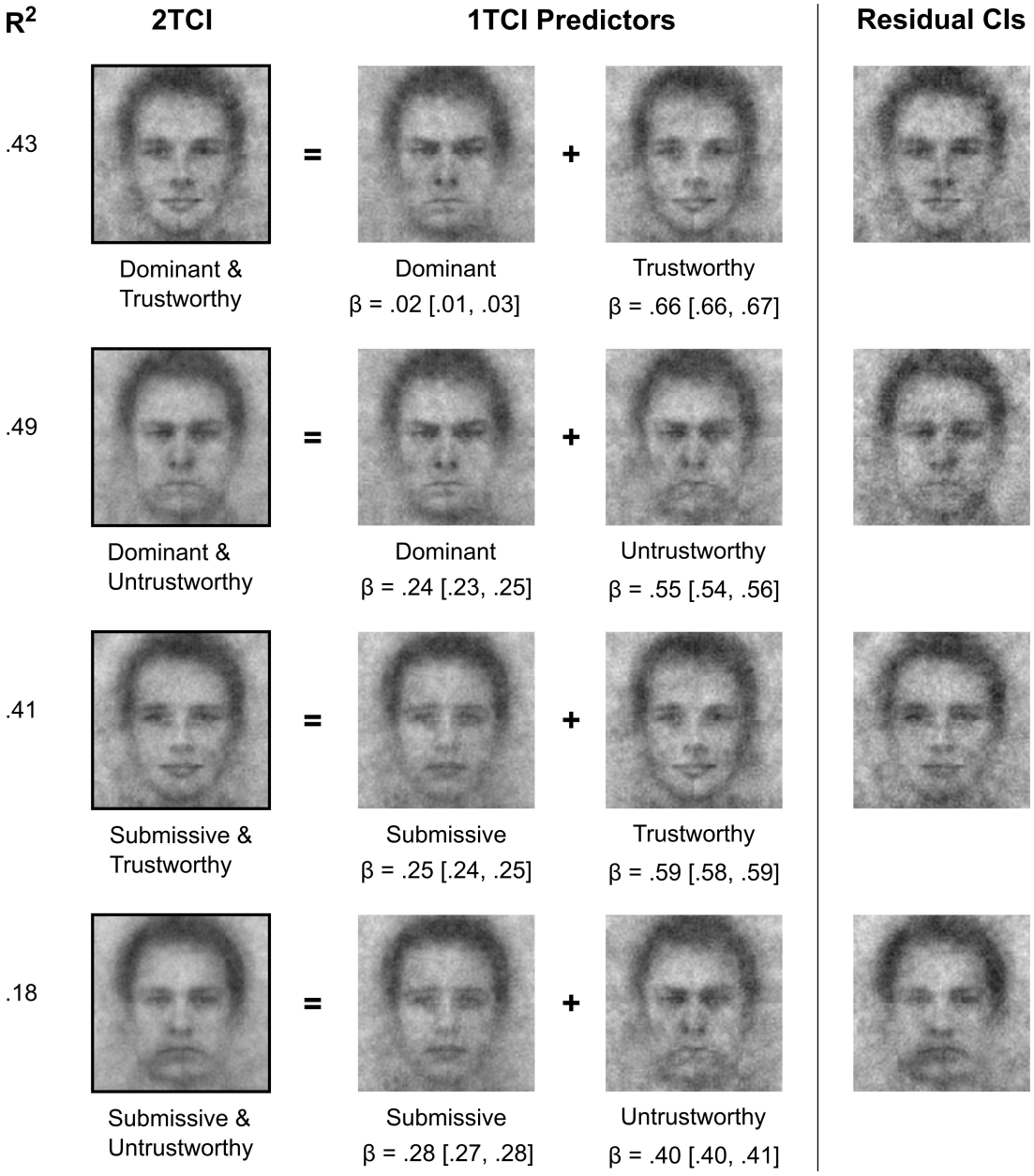

Figure 4. 2TCIs (framed pictures) obtained in Study 2 for each of the four two-trait combinations of the poles of the dominance and trustworthiness dimensions, collapsed across order of traits in the task instruction, and 1TCIs (from Oliveira et al., 2019). All results presented here are for the unmasked CIs. Each row illustrates a regression model with a 2TCI as the outcome and the trait combinations of correspondent 1TCIs. Residual CIs represent the variance in the pixel data left to be explained by each model. Reported values correspond to standardized beta coefficients [95% confidence intervals]. All coefficients were significantly different from zero, all *ps* < .001.

Masked versus unmasked CIs. As in Study 1, we also ran the regression models using masked CIs (see Study 1 Results' section for masking details). Results are listed in Table 3. The overall pattern of results is consistent with those found in Study 1, with some exceptions. Applying a mask to the SubTrust and SubUntrust 2TCIs add a lower impact on the explained variance of these models, especially for the SubUntrust 2TCI. This suggests that the explained variance of these 2TCIs was less affected by the inclusion of pixels outside the face region, and thus, that the pixel information contributing to the explained variance in the unmasked CI models was located within the boundaries of the applied mask. Nevertheless, the explained variance of the masked versions of these 2TCIs was not lower than that of their unmasked counterparts, thus demonstrating that the use of unmasked CIs was not inflating our models' explained variance.

Residual CIs

Again, we generated CIs using the residuals of each model (see Figure 4). As in Study 1, these CIs highly resemble their correspondent 2TCIs. This suggests that additional information emerged from the integration of the two traits, as these CIs cannot be accounted for by any of the 1TCIs in the model. We further discuss these results in the General Discussion.

Task I: Discriminability between 2TCIs and 1TCIs in 2IFC Task

In this analysis we examined participants' performance and response strategies during the 2IFC task. This informs us about whether participants were attending to both traits simultaneously or were instead strategically changing from one trait to the other during the task. If participants were attending to the presence of both traits in a target face, we would expect to observe a higher probability of correctly selecting the 2TCI regardless of the 1TCI with which it was paired with. However, if one trait in the target combination (e.g., Dominant & Trustworthy) was being more heavily weighted than the other, then we would expect the 1TCI containing that signal (e.g., Trustworthy 1TCI) to overlap with the signal of the 2TCI (which reflects that asymmetrical weighting). The resulting competition between the signals of the 2TCI and the 1TCI should lead to an equal probability of selecting any of these CIs. This would additionally suggest that participants could rely only on that (more heavily weighted) trait to make their decisions in the 2IFC task.

To examine this, we conducted a binomial logistic[1] regression with Selected CI Type (2TCI vs. 1TCI) as the outcome variable, and 2TCI Condition and Reference 1TCI (i.e., 1TCI paired with 2TCI) as predictors. The model including both the main effects and interaction was significant, χ^2 (15) = 106.6, p < .001, *AIC* = 656, $R^2_{\text{McFadden's}}$ = .15 (vs. main effects only model, *AIC* = 713, $R^2_{\text{McFadden's}}$ = .04). In our results, the odds ratios (*OR*s) represent the ratio between the probability of identifying the 2TCI (hit) and the probability of identifying a 1TCI as representing the two traits (false alarm). An *OR* of 1 indicates equal probability of hit and false alarm (i.e., chance level). *OR*s higher (lower) than 1 indicate that the probability of hits (false alarms) are superior to those of false alarms (hits).

A significant 2TCI Condition × Reference 1TCI interaction, χ^2 (9) = 75.5, p < .001, indicated that the differences in the probability of selecting a 2TCI (i.e., hit) across 2TCI-1TCI pairings, also differed across 2TCI condition. We report FDR corrected p-values for the multiple comparisons. A plot of these results can be found in Figure S4 of the Supporting Information file in our online data repository. The results show a higher overlap in signal between the Trustworthy 1TCI and both 2TCIs whose trait combinations included the trait "trustworthy". Specifically, the probability of a hit was the same (*p* = .50) for the DomTrust 2TCI and the SubTrust 2TCI when they were paired with the Trustworthy 1TCI (*OR* = 1.00, p_{FDR} = 1.00, 95% CI: 0.53, 1.85; in both 2TCI conditions). Moreover, comparatively to the Trustworthy 1TCI, the probability of a hit was always higher when the 2TCI was paired with the Dominant, Submissive, and Untrustworthy 1TCIs in both the 2TCIs conditions (*OR*s were, respectively, 9.00, 12.33, and 19.00, all ps_{FDR} < .001, in the DomTrust condition; and 4.71, 4.71, and 9.00, all ps_{FDR} < .001, in the SubTrust condition).

A signal overlap was also apparent between the Dominant 1TCI and 2TCIs generated with "untrustworthy" (see Figure S4 in online Supporting Information). Specifically, the probabilities of a DomUntrust 2TCI hit (*p* = .57) and a SubUntrust 2TCI hit (*p* = .50) when these were paired with the Dominant 1TCI (DomUntrust condition: *OR* = 1.35, p_{FDR} = .489, 95% CI: 0.72, 2.53; SubTrust condition: *OR* = 1.00, p_{FDR} = 1.00, 95% CI: 0.54, 1.85), did not differ across the 2TCI conditions (*OR* = 0.74, p_{FDR} = .656, 95% CI: 0.31, 1.78). In the

[1] This analysis is akin to a signal detection analysis where a *d*-prime score for two-alternative forced-choice tasks can be obtained (see Macmillan & Creelman, 2005). We relied on the binomial logistic regression as not only it provided the same information as a signal detection analysis, but also facilitated a more clear and concise reporting of inferential statistics for the relevant comparisons in the context of our discussion. Nevertheless, we additionally provide the results of a *d*-prime analysis of our data in our online data repository (see Supporting Information). The analysis of the *d*-prime scores is consistent with the binomial logistic regression results as it shows an overlap (lower discrimination) in signal between the Trustworthy 1TCI and both 2TCIs whose trait combinations included the trait "trustworthy", as well as an overlap between Dominant 1TCI and the 2TCIs generated with "untrustworthy."

DomUntrust 2TCI condition, comparatively to the Dominance 1TCI, 2TCI hits were more likely for any other reference 1TCIs (*OR*s were, respectively, 9.12, p_{FDR} =.002, and 5.17, p_{FDR} =.008, for the Submissive and Trustworthy 1TCIs), except for the Untrustworthy 1TCI, for which the probability of a hit did not significantly differ from Dominant 1TCI's (*OR* = 1.72, p_{FDR} =.382, 95% CI: 0.68, 4.34). In the SubUntrust 2TCI condition, 2TCI hits were only more likely when the 2TCI was paired with the Trustworthy 1TCI (*OR* = 5.67, p_{FDR} =.002) comparatively to the Dominance 1TCI. However, the probability of 2TCI hits in the pairing with Dominant 1TCI did not significantly differ from that in the pairings with the Submissive and Untrustworthy 1TCIs (respectively, *OR* = 2.08, p_{FDR} =.194, 95% CI: 0.84, 5.14, and *OR* = 0.82, p_{FDR} =.794, 95% CI: 0.34, 1.97).

From those results we can infer that the generally high probabilities for 2TCI hits (*OR*s > 1) across all 2TCI conditions (see Figure S4 in online Supporting Information), in cases where the signal did not overlap with the 1TCI, resulted either from participants focusing only on one trait, or in both, to make their decisions.

Overall, these results support the claim that 2TCIs are reflecting a cognitive integration of both traits by the participants. However, they also suggest that individuals may rely on only one trait to make their decisions in the 2IFC task, which would contribute to the higher weight of that trait in the resulting 2TCI. Despite of the methodological relevance of this evidence, these results are also suggesting that the weight that participants attribute to traits is dependent on the trait combination context. When the context is one of high trustworthiness, dominance loses relevance. On the other hand, high dominance only seems to be meaningful when coupled with untrustworthiness.

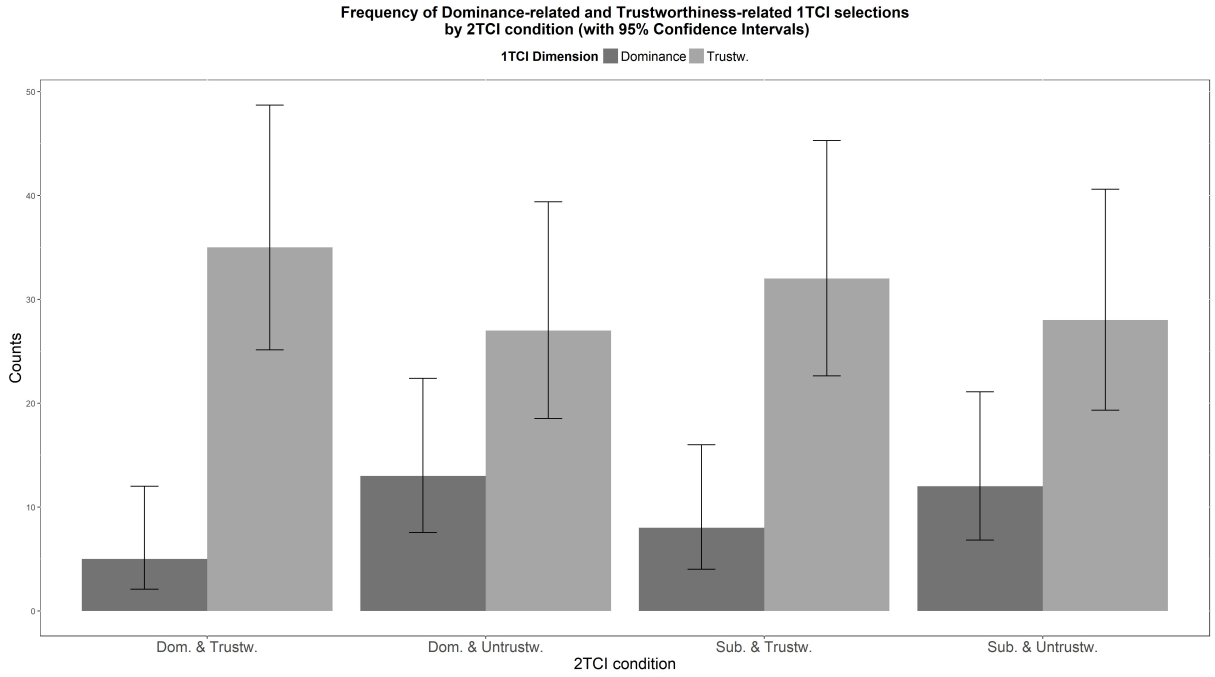

Figure 5. Plot exhibits absolute frequencies (counts) of dominance-related and trustworthiness-related 1TCI selections in the trial of the 2IFC extended reverse correlation task in which only 1TCIs were presented (task II), by 2TCI condition. Error bars denote 95% confidence intervals.

Task II: Response Tendency Analysis

In this analysis we assessed the participants' tendency to rely on one dimension versus the other, in a trial where each of the two target faces corresponded to only one of the traits in the instruction (e.g., Trustworthy 1TCI vs. Dominant 1TCI, in Dominant & Trustworthy condition). Performance on this task should inform about whether there is a general bias to rely on one trait dimension over the other during the 2IFC task. For instance, in the Dominant & Trustworthy 2TCI condition participants may be inclined to choose a trustworthy-looking face rather than a dominant-looking face. The absence of a particular inclination towards any of the two traits would promote a random selection in this task.

To examine response tendency, we submitted the frequencies of dominance-related and trustworthiness-related 1TCI selections in each 2TCI condition to a loglinear regression analysis. Specifically we entered 2TCI Condition and 1TCI Dimension as variables in the regression. Results are shown in Figure 5. There was no difference between the saturated model including the interaction (χ^2 (7) = 52.3, $p < .001$, AIC = 53.2) and the main effects only model (χ^2 (4) = 46.4, $p < .001$, R^2_{McF} = .89, AIC = 53.1) which indicates a non-significant

interaction between 2TCI condition and the dimension of the selected 1TCI (χ^2 (3) = 5.93, p = .115). We kept the saturated model with the interaction to account for all the variability in the data. The main effect of 2TCI condition was non-significant (χ^2 (3) = 4.59, p = .204) indicating that the difference between frequencies of dominance- and trustworthiness-related 1TCIs did not differ across 2TCI condition. Only the main effect for 1TCI Dimension was significant (χ^2 (1) = 46.4, p < .001), indicating that, across all 2TCI conditions, the rate of selection of a trustworthiness-related 1TCI was 3.21 times the rate of selection of a dominance-related 1TCI (b = 1.17, p < .001, rate ratio = 3.21, 95% CI: 2.23, 4.62). In other words, participants showed a tendency to rely on trustworthiness-related facial content in Task II's trial, regardless of the 2TCI instruction. This is consistent with previous results suggesting that perceivers ascribe more weight to the trustworthiness dimension (see introductory section). These results help to clarify that a bias to perceive trustworthiness in a face may have contributed to the predominance of trustworthiness information in the 2TCIs.

General Discussion

The present studies offer initial evidence about how impressions derived from two trait combinations are mapped into a face, thereby extending Asch's (1946) seminal work on trait integration to the domain of social face perception. The first relevant aspect of our data is that it shows that this integration occurs. It suggests that participants were matching an impression derived from two different trait dimensions to particular face content. Moreover, it suggests that the content of the resulting faces—or 2TCIs—went beyond a mere linear compound of the features associated with each separate dimension. Although subsequent ratings of these 2TCIs by an independent sample of participants showed that the face content of the 2TCIs conveyed information on the two trait dimensions, participants were not equally weighting them and even changed the weight of a trait depending on its trait context. For instance, our results suggest that the weight of dominance changes depending on the perceived trustworthiness of a person.

Our data are consistent with previous findings regarding how trait information is integrated in impressions formation, that show an unequal weighting of traits (e.g., Anderson, 1968), and different trait centrality and context effects (e.g., Asch, 1946). To the best of our knowledge, these are the first studies that extend those findings to the domain of social face perception.

The results from both studies are also consistent with data previously reported in the impression formation literature supporting that morality-related information such as trustworthiness weights more in our impressions (e.g., Abele & Bruckmüller, 2011; Goodwin et al., 2014; Wojciszke et al., 1998). They suggest that when perceivers receive information about the trustworthiness and dominance of a target person, trustworthiness information will weigh more than dominance information on their expectations about the target's facial appearance. Our studies only suggest that this is what occurs by default, when no specific context is provided. Future studies could further investigate if a specific context where dominance is a relevant dimension would promote a higher contribution of this dimension to the final visual face impression.

Our results additionally document that when we evaluate the two dimensions simultaneously in a face, not only do we ascribe a higher weight to trustworthiness, as we also contrast the evaluation of both dimensions, thereby promoting a compensation effect (see Judd, Garcia-Marques, & Yzerbyt, 2019). Although this type of compensation strategy is identified, by itself it does not seem to account for the context effects observed in our data.

The results from Tasks I and II in Study 2 also show that the different weights of trustworthiness found in trait integration (see CI regression models) are detected in the integration process itself as differences in detection of the presence of the trustworthiness trait itself, and as a bias that favors the detection of trustworthiness relatively to dominance. These data are in line with evidence suggesting that valence and trustworthiness information are automatically assessed by perceivers in social face perception (Todorov et al., 2009), and that they may have priority over other dimensions.

Importantly for our goals, although the findings of our conceptually-driven perception approach are consistent with those obtained using synthetic faces, they extend their interpretation. As stressed above, in agreement with Oosterhof and Todorov's (2008) results, ours show that trustworthiness outweighs dominance in face impressions of personality. However, what our approach adds to that knowledge is that the specific weight of a trait dimension in a face impression (which integrates several traits) is not predicted by a linear model relying in a fixed set of weights attributed to each individual trait. Our data show instead that those weights are trait context-dependent. The regression analyses using 1TCIs as predictors of the 2TCIs clarify that the trait, dominant, weights more in a face representation when paired with untrustworthy than with trustworthy. This is further corroborated by the following factors : (a) 1TCIs did not account for all of the variance in the 2TCIs, thus allowing the residual CIs to preserve face information that bore a strong resemblance to the

2TCI; (b) evaluating a face on a trait was dependent on other available trait information associated with the same face; (c) the results of the signal detection analysis in Study 2, which suggest that dominance is more heavily mapped in a face when coupled with untrustworthiness, than when it is coupled with trustworthiness. Hence, our data fits well with Asch's (1946) claims that "characteristics forming the basis of an impression do not contribute each a fixed, independent meaning, but that their content is itself partly a function of the environment of the other characteristics, of their mutual relations" (p. 268).

Besides offering new evidence supporting a gestaltic integration of trait information in face perception, we believe that this paper is methodologically innovative: First, it demonstrates the feasibility of using RC methods to visualize face content associated with more complex personality judgments; and second, it describes how the output of these methods can be analyzed (i.e., CI regression analyses) to quantify the impact of conceptual trait knowledge on representations of face content (i.e., how trait-space information shapes face-space information; see Over & Cook, 2018).

This methodological approach opens a new perspective to the study of the process of how information is integrated in our minds, which can go beyond the focus of the discussion in this paper. Take for instance the residual CIs that resulted from the CI regressions (see Figures 2 and 4). These residual CIs exhibit emotional expressions that were not accounted for by the trustworthiness- and dominance-related features of the 1TCIs. It is an empirical question what these emotional expressions are reflecting. It could be specific deviations from the predominant trustworthiness-related features in the 2TCIs, or alternatively, an additional emotional state expected by participants as a result of the trait integration. Future studies could further investigate these possibilities by, for instance, examining how similar 2TCIs are predicted by 1TCIs generated for different emotions and obtaining trait and emotion ratings for all the CIs, including the residual CIs.

To our knowledge, our studies represent the first attempt to investigate how perceivers associate physical face content with impressions derived from different combinations of trait information. Our results are consistent with Asch's configural (1946) perspective, which argues that the content of an impression is determined by the specific interrelations between the traits from which it originated. In this way, our approach is theoretically relevant by bringing the discussion promoted by Asch's (1946) work into the social face perception domain.

Acknowledgments

This research had the support of Fundação para a Ciência e a Tecnologia (FCT) through the doctoral grant PD/BD/113471/2015 awarded to Manuel Oliveira, and the strategic project UID/PSI/04810/2013 granted to the William James Center for Research. Special thanks to Pedro Figueira and Alexandre Fernandes for helping with data collection.

References

Abele, A. E., & Bruckmüller, S. (2011). The bigger one of the "Big Two"? Preferential processing of communal information. *Journal of Experimental Social Psychology*, *47*, 935–948. doi: 10.1016/j.jesp.2011.03.028

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 50, pp. 195–255). Academic Press. doi: 10.1016/B978-0-12-800284-1.00004-7

Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology*, *70*, 394–400. doi: 10.1037/h0022280

Anderson, N. H. (1968). Application of a linear-serial model to a personality-impression task using serial presentation. *Journal of Personality and Social Psychology*, *10*, 354–362. doi: 10.1037/h0026816

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*, 258–290. doi: 10.1037/h0055756

Asch, S. E., & Zukier, H. (1984). Thinking about persons. *Journal of Personality and Social Psychology*, *46*, 1230–1240. doi: 10.1037//0022-3514.46.6.1230

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*, 289–300. doi: 10.2307/2346101

Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition*, *32*, 397–408. doi: 10.1521/soco.2014.32.4.397

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*, 333–361. doi: 10.1080/10463283.2017.1381469

Bruner, J. S., Shapiro, D., & Tagiuri, R. (1958). The meaning of traits in isolation and in

combination. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 277–288). Stanford: Stanford University Press.

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. *Advances in Experimental Social Psychology*, *40*, 61–149. doi: 10.1016/S0065-2601(07)00002-0

Dotsch, R. (2015). rcicr: Reverse correlation image classification toolbox (R package version 0.3.0) [Computer program package]. Retrieved from https://cran.r-project.org/web/packages/rcicr/index.html

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, *3*, 562–571. doi: 10.1177/1948550611430272

Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of Personality and Social Psychology*, *100*, 999–1014. doi: 10.1037/a0023026

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. doi: 10.3758/BF03193146

Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 89–150). New York, NY,: McGraw-Hill.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*, 148–168. doi: 10.1037/a0034726

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, *103*, 336–355. doi: 10.1037/0033-295X.103.2.336

Hampson, S. E. (1990). Reconciling inconsistent information: Impressions of personality from combinations of traits. *European Journal of Personality*, *4*, 157–172. doi: 10.1002/per.2410040208

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*, 837–852. doi: 10.1037/0022-3514.78.5.837

Himmelfarb, S. (1972). Integration and attribution theories in personality impression formation. *Journal of Personality and Social Psychology*, *23*, 309–313. doi: 10.1037/h0033126

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. doi: 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. doi: 10.1146/annurev-psych-010416-044242

Johnson, D. H. (1999). The insignificance of statistical significance testing. *The Journal of Wildlife Management*, *63*, 763–772. doi: 10.2307/3802789

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. doi: 10.1016/j.jesp.2019.01.008

Kleisner, K., Chvátalová, V., & Flegr, J. (2014). Perceived intelligence is associated with measured intelligence in men but not women. *PLoS ONE*, *9*, 1-7. doi: 10.1371/journal.pone.0081237

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 1-12. doi: 10.3389/fpsyg.2013.00863

Lundqvist, D., & Litton, J. E. (1998). *The Averaged Karolinska Directed Emotional Faces: AKDEF* [Face database]. Stockholm, Sweden: Karolinska Institutet. Retrieved from http://kdef.se/

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

Murayama, K., Pekrun, R., & Fiedler, K. (2014). Research practices that can prevent an inflation of false-positive rates. *Personality and Social Psychology Review*, *18*, 107–118. doi: 10.1177/1088868313496330

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*, 888–902. doi: 10.1002/ejsp.2569

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*, 11087–11092. doi: 10.1073/pnas.0805664105

Over, H., & Cook, R. (2018). Where do spontaneous first impressions of faces come from? *Cognition*, *170*, 190–200. doi: 10.1016/j.cognition.2017.10.002

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*, 283–294. doi: 10.1037/h0026086

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*, 372–386. doi: 10.1037/h0032891

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*, 105–118. doi: 10.1016/j.cognition.2012.12.001

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*, 519–545. doi: 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*, 813–833. doi: 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*, 455–460. doi: 10.1016/j.tics.2008.10.001

Uleman, J., & Kressel, L. M. (2013). A brief history of theory and research on impression formation. In D. Carlston (Ed.), *The Oxford Handbook of Social Cognition* (pp. 53–73). New York: Oxford University Press. doi: 10.1093/oxfordhb/9780199730018.013.0004

Walker, M., Jiang, F., Vetter, T., & Sczesny, S. (2011). Universals and cultural differences in forming personality trait judgments from faces. *Social Psychological and Personality Science*, *2*, 609-617. doi: 10.1177/1948550611402519

Walker, M., & Vetter, T. (2009). Portraits made to measure : Manipulating social judgments about individuals with a statistical face model. *Journal of Vision*, *9*, 1–13. doi: 10.1167/9.11.12

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

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*, 1251–1263. doi: 10.1177/01461672982412001

Ybarra, O., Chan, E., & Park, D. (2001). Young and old adults' concerns about morality. *Motivation and Emotion*, *25*, 85–100. doi: 10.1023/A:1010633908298

Zebrowitz, L. A. (2006). Finally, faces find favor. *Social Cognition*, *24*, 657–701. doi: 10.1521/soco.2006.24.5.657

Section III

General Discussion

The main goal of the present thesis was to clarify and gain further insight into the relationships established between personality dimensions previously identified as being central in the domains of person perception and social face perception. Within this broader topic, we empirically addressed three questions. The first question pertained to the degree of overlap between the dimensions of social face perception (i.e., trustworthiness and dominance) and the dimensions of person perception (i.e., warmth and competence) (Chapter VI). The second questioned pertained to the nature of the relationship that valence establishes with core dimensions of social perception in the domains of person perception and social face perception (Chapter VII). And lastly, the third question pertained to the integration of the dimensions of social face perception into unitary impressions in terms of visual facial content (Chapter VIII). We start with a brief summary of our findings regarding each one of these questions. Subsequently, we discuss how these findings relate with one another and how they add to the literature on the fundamental dimensions of social perception. Finally, we conclude with a discussion of additional topics and limitations associated with our studies, and provide suggestions for future directions building upon the current work.

Summary of Findings

In our first paper (Chapter VI), we investigated the extent to which the dimensions of person perception—warmth and competence—overlapped with the dimensions of social face perception—trustworthiness and dominance—in the context of facial impressions. In our review of the literature, we found pieces of evidence suggesting that whereas dominance and competence exhibit differences at both the conceptual and evaluative levels, trustworthiness and warmth are more clearly distinguishable at the conceptual level than at the evaluative level. In turn, these differences and similarities between dimensions should lead to differences and similarities in how perceivers mentally represent them in terms of facial appearance. Using a reverse correlation paradigm, we tested the hypothesis that the dominance and competence dimensions would overlap to a lesser extent than the trustworthiness and warmth dimensions in terms of visual facial content. Our findings supported this hypothesis, by demonstrating that perceivers expected the facial content of trustworthiness and warmth to be less distinguishable than the facial content of dominance and competence. Subsequently, we demonstrated that the perceived valence (measured as likeability) of the facial content associated with each of our target four dimensions was likely driving the degree of overlap between the dimensions. Whereas perceptions of facial dominance and competence diverged

in their correlation with valence (negative for dominance, positive for competence), perceptions of trustworthiness and warmth converged in their positive correlation with valence. Overall, these findings suggest that (a) the models of person perception and social face perception only partially overlap, and (b) that their divergence seems to be driven by the relationship that each of the dimensions establishes with valence.

In our second paper (Chapter VII), we aimed to examine and clarify the nature of the relationship that valence establishes with a number of dimensions identified as central in the domain of person perception and social face perception. In doing so, we expected to advance our knowledge on the potential reasons (e.g., relationship with valence) underlying the divergence between the person perception and the social face perception dimensions. In our review of the literature we found that composite measures of communion and agency dimensions not only were inconsistent across the literature in regard to the traits selected to represent them, but also tended to overlook how different agency-related traits (e.g., dominance and competence) established opposite correlations with valence. Moreover, previous research had shown the relevance of taking into account both the relationship with valence (Suitner & Maass, 2008), as well as the nature of a relationship between traits (Imhoff & Koch, 2017; Lemann & Solomon, 1952), in investigations of the relationship between dimensions of social perception. However, despite of the relevance of these two aspects, we found no studies focused on directly assessing both the perceived valence and the nature that it establishes with the trait content associated with the Big Two (i.e., umbrella term referring to communion and agency, assuming these to be akin to the warmth and competence dimensions; Abele & Wojciszke, 2007). We filled this this gap in the literature, by developing a new paradigm that allowed us to assess the perceived valence of different degrees of expression of a trait along its dimensions continuum. We applied this paradigm in both the contexts of person perception—using only verbal person descriptors—and social face perception—using only faces whose features varied along a given trait dimension. Drawing on the literature, as well as on the findings of our first paper (Chapter VI), we expected to find that the relationship established with valence would be more stable in nature for communion-related traits (e.g., trustworthiness and warmth), and more differentiated for agency-related traits (e.g., dominance and competence).

Overall, our findings supported our expectations. Agency-related traits exhibited clearer differences in the relationship that they establish with valence, across the person perception and social face perception contexts. However, we also found differences between

these contexts. Regarding agency-related content, we found that competence established a strong positive and linear relationship with valence in both the person perception and social face perception contexts. Dominance, on the other hand, exhibited a clear inverted U-shaped relationship with valence in the context of person perception, but a negative linear relationship in the context of social face perception. Thus, although the linearity between competence and valence seems to be independent of the modality of the target stimulus (conceptual knowledge vs. facial appearance), the curvilinear relationship between dominance and valence emerged more clearly when perceivers relied on their conceptual knowledge of traits than when they relied on facial appearance information. Nevertheless, these findings replicate the divergent relationship that dominance and competence establish with valence found in our first paper, thus offering support to our conclusion that the relationship of dimensions with valence is driving the divergence between the person perception and social face perception two-dimensional models of personality impressions. In addition, the findings in this paper also clarify that morality- and competence-related traits are the ones who establish a stronger linear and positive relationship with valence, whereas assertiveness-related traits such as dominance and—unexpectedly—warmth-related traits such as sociability exhibit stronger curvilinear components in their relationship with valence.

Altogether, the findings in our second paper add to the literature by clarifying the nature of the relationship that the most relevant trait content of the communion (i.e., warmth and morality) and agency (i.e., competence and assertiveness) dimensions establishes with valence. Moreover, these findings appear to suggest that the relationship established between valence and dominance is more strongly linear in the context of facial impressions, thus raising the question of whether valence information weights more in judgments of facial appearance, especially when we take into account that valence and trustworthiness are practically indistinguishable dimensions in social face perception (e.g., Oosterhof & Todorov, 2008).

In our third paper (Chapter VIII), we turned our focus exclusively to the domain of social face perception, to investigate how trustworthiness and dominance were integrated into unitary impressions of personality in terms of facial appearance. Although the topics of the relationship between the two dimensions (i.e., warmth and competence, communion and agency) and of how they are integrated into unitary impressions received a great deal of attention in the person perception literature, the same cannot be said of the social face perception literature. Thus far, the literature on social face perception only tangentially

addressed these questions, with findings suggesting a priority of valence and trustworthiness processing in facial impressions (Todorov et al., 2009; Willis & Todorov, 2006), or the facial models that artificially combine the two dimensions in a linear fashion to represent the quadrants of the trustworthiness-by-dominance face space (Oosterhof & Todorov, 2008; Todorov et al., 2008). In this paper, we aimed to fill the gap in the literature by examining how the two dimensions were spontaneously integrated by perceivers in the visual face space. To do so, we employed reverse correlation methods to assess the facial content expected by perceivers to reflect impressions emerging from combinations of trustworthiness and dominance information.

Our results demonstrate that (a) trustworthiness (or valence) information overweighs dominance information in facial impressions; (b) a simple linear model is not able to entirely account for the facial output of impressions integrating the two dimensions, with a noticeable amount of visual residual variance—the residual CIs—suggesting that more facial information other than trustworthiness and dominance emerges from their integration (in agreement with the gestaltic perspective on trait integration; Asch, 1946); and that (c) trustworthiness information appears to modulate the expression of dominance information in a face when perceivers integrate these dimensions into unitary impressions. Altogether, these findings provide initial insights into how old questions extensively investigated within the context of person perception may be extended to the domain of social face perception, and have the potential to trigger a new line of research within the social face perception literature. But more importantly, they add to our understanding of the role played by valence (or trustworthiness) in facial impressions, by suggesting that impressions implying a negative correlation between trustworthiness and dominance (i.e., untrustworthy and dominant, or trustworthy and submissive) are associated with the attribution of a higher weight to dominance information.

The Structuring Role of Valence

Altogether, the findings reported throughout this thesis converge on a common theme: the central role played by valence in structuring the relationships between core dimensions of social perception. The nature and direction of the relationship that valence establishes with each of any two trait dimensions have a great impact on how these two trait dimensions are found to be related, and on how they are integrated into unitary impressions of personality. The findings across all of our studies clearly support that a wide range of core dimensions of

social judgment are saturated with valence. This further supports the predominant role of valence in social judgments, as once argued by Osgood and colleagues (Osgood, 1964; Osgood et al., 1957) in light of their *Evaluation-Potency-Activity* semantic differential model. Building upon prior research (e.g., Abele & Bruckmüller, 2011; Klapper, Dotsch, van Rooij, & Wigboldus, 2016; Todorov et al., 2009; Willis & Todorov, 2006; Wojciszke et al., 1998), our findings also attest to the primacy of valence in both the person perception and social face perception domains, by showing that trait dimensions strongly and linearly related with valence (such as communion-related traits; see Chapter VII) weight more in visual facial impressions (see Chapter VIII) comparatively to trait dimensions exhibiting weaker linear (or more curvilinear) relationships with valence. Although previous work in person perception advocated for a disentanglement between the evaluative (i.e., valence) and descriptive components of trait judgments (Peabody, 1967, 1984), that separation proved hard to accomplish (e.g., Rosenberg & Olshan, 1970), with the growing body of findings supporting that both evaluative and descriptive are aspects of comparable importance in trait judgments (Peabody, 1970; Rosenberg & Olshan, 1970). Thus, although person perception models such as the social-by-intellectual (Rosenberg et al., 1968) and warmth-by-competence (Fiske et al., 2002) models argue against unidimensional valence-based judgments by distinguishing between two dimensions of descriptive meaning (i.e., traits relevant for social vs. intellectual activities; or intentionality vs. ability), the fact is that these dimensions remain both saturated with evaluative information that distinguishes a negative pole from a positive pole (see also Suitner & Maass, 2008, for a similar argument regarding the communion-by-agency model). In fact, this confound between evaluative and descriptive aspects of traits is even explicit in the labels "social good-bad" and "intellectual good-bad" proposed for the dimensions by Rosenberg and colleagues (Rosenberg et al., 1968) themselves. However, despite of the well-recognized ubiquity of valence in social perception (e.g., Osgood, 1964; Osgood et al., 1957; Rosenberg & Olshan, 1970; Suitner & Maass, 2008), the literature preceding the present work offered no clear empirical evidence informing about how valence related with all the core dimensions underlying person perception and social face perception, or how the features of the relationship (i.e., nature and direction) of those dimensions with valence is shaping their association within two-dimensional models (e.g., communion-by-agency, warmth-by-competence, trustworthiness-by-dominance) and between two-dimensional models obtained in different domains of research (e.g., person perception vs. social face perception). The three papers comprising this thesis help to fill this gap in the literature and additionally emphasize the importance of taking into consideration the role of the relationship of dimensions with

valence in social perception-related research, especially, but not exclusively, when the focus of research involves two different classes of social stimuli (i.e., verbal person descriptors and/or faces).

Our findings offer contributions to the literature at two different *foci* of discussion. First, they contribute to a clarification of the relationship "between" models of personality impressions associated with different social stimuli. And second, they contribute to a deeper understanding of how the two dimensions "within" each two-dimensional model relate to each other. For clarity, we will branch the discussion on the basis of these two different foci. We start with a discussion of how valence structures the relationship between dimensions of different two-dimensional models (e.g., relationship between the competence dimension of person perception and its correspondent dominance dimension of social face perception). Subsequently, we move the focus to the structuring role of valence within each model (e.g., relationship between trustworthiness and dominance, both dimensions of social face perception). Throughout the discussion of these two themes we also elaborate on some of the questions raised by our findings and related sub-topics.

The Role of Valence in the Overlap "between" Face and Person Perception Models

Our findings in Chapter VI suggest that the relationship that trait dimensions establish with valence is a crucial aspect to consider when examining the degree of overlap between the dimensions of person perception and the dimensions of social face perception. Although both dominance and competence establish a relationship with valence, the direction of the relationship is opposite, thus leading to a more clear differentiation between these dimensions in terms of facial signals. This pattern of findings demonstrates that although dominance and competence (and trustworthiness and warmth) can be differentiated at a conceptual level (either theoretically or by perceivers' beliefs), the degree to which they overlap in terms of visual facial content is more clearly determined by associations at the evaluative level (i.e., associations with valence). In this way, trait-valence associations can be said to play a central role in structuring the relationship between the dimensions of social face perception and the dimensions of person perception.

While the evaluative divergence between dominance and competence found in Chapter VI is sufficient to demonstrate that the person perception and social face perception models are not redundant, it remained unclear whether the same relationships of dimensions

with valence could be extended to the domain of person perception. By examining the relationships of valence with core dimensions in each domain separately in Chapter VII, we were able to detect any differences in the nature and direction of these relationships across domains. One of the main distinctions we found between domains was that the nature of the relationship of dominance with valence was more strongly curvilinear in the person perception domain (Studies 1 and 2, Chapter VII) than in the domain social face perception, where it became more linear (or at least less curvilinear) and sensitive to target and perceiver gender (Study 3, Chapter VII). On a broader level, this pattern of results suggests a higher impact of social categorization and self-construal in the evaluation of faces compared to the evaluation of (more abstract) trait concepts. At least at the target level, this should not be surprising, since faces are naturally imbued with information regarding social category membership (i.e., sex, race, age). By contrast, the targets we used to assess relationships with valence in the context of person perception (Studies 1 and 2, Chapter VII) did not specify any association of the targets with social categories (i.e., targets were only defined as "persons"). What is more surprising however, is how the difference in the relationship with valence across domains is more apparent for dominance, an agency assertiveness-related trait. The results of our studies suggest that dominance exhibits the highest inconsistency in the nature of its association with valence across stimulus modality (i.e., clearly curvilinear in conceptual trait space vs. stronger linearity in face space), target gender (more strongly linear for female targets in face space), and perceiver gender (curvilinear in conceptual trait space regardless of gender vs. more strongly linear for female than male perceivers in face space). By comparison, the relationship of competence with valence appears to be more stable across domains and perceiver/target characteristics. These findings (in Chapter VII) further clarify that the positive relationship between competence and valence found in the context of social face perception in Chapter VI is not limited to inferences based on perceptual features of faces or their resemblance to specific emotional signal. Nevertheless, competence showed small signs of being more sensitive to target and perceiver gender in the domain of facial impressions (i.e., small but noticeable curvilinear trend for female intelligence) compared to all communion-related dimensions (e.g., trustworthiness, warmth). Unlike agency-related traits (i.e., competence and dominance), communion-related traits (i.e., trustworthiness, warmth) exhibited a highly stable relationship with valence across domains and target/perceiver characteristics. Valence, thus, seems to promote the convergence of person perception and social face perception models in terms of communion-related content, and to drive the divergence between models in terms of agency-related content.

Together, the findings in Chapter VI and VII expand our knowledge on the differences between models of person perception and social face perception, and on how the relationship between these models is structured by the associations of dimensions with valence. More specifically, our present findings

- provide important clarifications to the social perception literature in general, by allowing to predict how different fundamental dimensions of personality perception will be found to be related based on their association with valence;
- further emphasize the need to empirically address their degree of redundancy, especially when we take into account that opposite relationships of dimensions with valence (e.g., competence vs. dominance) predict different behavioral outcomes in social settings (e.g., approach or avoidance); and
- clarify that the opposite relationship that competence and dominance establish with valence occurs not only in the social face perception domain but also in the person perception domain (see also Rosenberg & Sedlak, 1972 who present some initial evidence of this divergence in person perception).

And lastly, they raise a question regarding the factors affecting the consistency in the nature of trait-valence relationships across our studies reflecting the contexts of person perception (Studies 1-2, Chapter VII) and social face perception (Study 3, Chapter VII).

Inconsistency in the Nature of the Relationship with Valence across Domains

Why did a curvilinear relationship between dominance and valence emerge in the context of person perception, but not in the context of facial impressions?

With the finding of an inverted U-shaped relationship between dominance and valence in the context of person perception, the possibility arises that a curvilinear relationship is underlying the orthogonality between the dominance and trustworthiness (or valence) dimensions of facial impressions suggested by the trustworthiness-by-dominance model (Oosterhof & Todorov, 2008). But our results in Chapter VII suggest, instead, that the dimensions of facial impressions are linearly and negatively related between themselves, considering their opposite linear relationship with valence. This negative relationship is not entirely surprising given its consistent emergence in both the ratings of classification images of our studies in Chapters VI and VIII, and prior research in social face perception (e.g., Chen et al., 2014; Dotsch & Todorov, 2012; Oh et al., 2019; Oosterhof & Todorov, 2008).

Moreover, it draws attention to the small, but nevertheless, negative correlation between dominance and trustworthiness found by Oosterhof and Todorov (2008) themselves in the data supporting the trustworthiness-by-dominance model. However, this is not to say that a curvilinear relationship is unlikely. A small negative correlation between dominance and valence does not preclude the existence of a "true" inverted U-shaped relationship, as it might also be reflecting that the peak of valence in such a curvilinear trend is leaning towards the "submissive" pole of the dimension continuum. What our results in Chapter VII (Studies 1-3) are able to inform is that, at least under similar circumstances of measurement (i.e., dimension continuum with 7 levels, identical valence measures), a curvilinear relationship between dominance and valence was more easily detected when dimensions were presented as verbal trait continua than when they were conveyed by face stimuli alone. This leads us to wonder why any curvilinear relationship was harder to find in the context of face evaluation.

One possibility suggested by our results (Studies 1-3, Chapter VII) relates to the permeability of the judgments of each stimulus modality (i.e., verbal trait-based descriptions vs. faces) to the influence of stereotypical information. As we have seen, the pattern of results from our study using faces as targets (Study 3, Chapter VII) indicate that the relationships of dimensions with valence were sensitive to perceiver and target gender, social perspective (i.e., judging oneself vs. judging others) and self-presentational goals (i.e., judging how the target face would favor one's own public image). And importantly, our results suggest that the detection of a curvilinear relationship between dominance and valence in the context of facial evaluations is either facilitated, or hindered, by particular circumstances defined by the combination of the three aforementioned factors. When female perceivers judged the desirability of female faces for themselves, a strong negative linear relationship between dominance and valence emerged. In contrast, when both perceivers and targets were male, a stronger curvilinear relationship emerged. As discussed elsewhere (see General Discussion in Chapter VII), these results are likely being driven by gender stereotypes (Oh et al., 2019; Sutherland, Young, Mootz, & Oldmeadow, 2015), and by the natural confounds between multiple social categories and emotion which are inextricably linked with the human face (e.g., Bijlstra, Holland, Dotsch, Hugenberg, & Wigboldus, 2014; Bijlstra, Holland, & Wigboldus, 2010; Brooks, Stolier, & Freeman, 2018; Fischer, Rodriguez Mosquera, Van Vianen, & Manstead, 2004; Plant, Hyde, Keltner, & Devine, 2000). In sum, it would be reasonable to expect that the high saliency and natural co-occurrence of cues to social categories and emotion in a face, is contributing to a higher permeability of face judgments to

the influence of stereotypical information activated by those cues. In turn, this stereotypical information is likely interacting with perceivers' self-construal and self-presentational goals, to ultimately shape the dimension-valence relationships we observed.

In our studies reflecting the context of person perception (Studies 1 and 2, Chapter VII), however, the judgments may have been less permeable to the influence of stereotypical information, since no information was provided regarding the social category membership of the verbally described targets. With our current data, however, it is impossible to tell whether our participants incorporated any information about social categories in their judgments. The task in our verbal trait continua studies imposed no restraints on the person knowledge relied upon by participants to make their judgments. Thus, we cannot rule out the possibility that a participant inferred different social categories for different levels (i.e., target persons) of the same trait continuum. For instance, by relying on the stereotype that men are more agentic than women (e.g., Abele, 2003; Abele & Wojciszke, 2007), one could be inferring that the target persons on the lower range of dominance are likely to be women, while the target persons on the higher range of dominance are likely to be men. A future conceptual replication of our Study 2 in Chapter VII, could clarify whether our pattern of results is sensitive to target gender by including a description of the target persons' gender in the target descriptions, and creating a male and female version of the trait continua similarly to our manipulation in the study using face continua. This additional manipulation would increase the comparability of findings across the person perception and social face perception domains, and further clarify whether dominance establishes a more curvilinear trend with valence, regardless of target gender.

Altogether, the considerations discussed so far suggest that one potential and more broad explanation for the inconsistency in the nature of trait-valence relationships across domains involves an asymmetry in the salience of social category information between the two modalities of target stimulus (i.e., faces vs. trait-based person descriptions), and the role that a higher saliency of this information plays in the early stages of impression formation (e.g., Brewer, 1988; Fiske & Neuberg, 1990; Kunda & Thagard, 1996). Both the *serial* (Brewer, 1988; Fiske & Neuberg, 1990) and *connectionist* (Kunda & Thagard, 1996) models of impression formation, predict that the stereotypical information activated at the initial stage of impression formation will override more individuating trait-based information, and thus dominate the resulting impression, especially when the trait-based information is not salient (see Kunda & Thagard, 1996)—as was the case in our face continua study. In light of these

models, the naturally occurring higher saliency of social category information in faces (vs. saliency in our verbal trait-based person descriptors), should promote an initial impression that is predominantly stereotype-driven. This aligns well with the results of our face continua task showing a higher sensitivity to gender differences. By contrast, the results regarding the verbal trait-based continua, may have been less driven by social category-related information and might be reflecting evaluations of the trait concepts themselves. In other words, it could be that the trait-valence relationships found in the context of our face continua study were colored by stereotypical information to a greater extent than the relationships found in the context of verbal trait continua. At this point we should reiterate that the asymmetric salience of social category information between face-based and verbal trait-based information is promoted by the absence of social category information in the person descriptions of our verbal trait continua stimuli. If our verbal person descriptors included explicit social category information (e.g., male/female with a given amount of dominance), then we could expect the same influence of stereotypical information in the results for verbal continua. That is, we do not intend to claim that the influence of stereotypes in impressions is exclusive to any of the stimulus modalities.

The possibility that the trait-valence relationships found in the verbal trait continua studies were less colored by stereotype information than the relationships found in the face continua study remains, however, to be tested. Future research could clarify whether the asymmetric salience of stereotypical information in the target stimulus was indeed driving the inconsistency of dominance-valence relationships across domains, by manipulating the focus on trait-level versus social category-level in valence judgments of trait and face continua (e.g., by revealing the trait dimension along which faces vary in a continuum, and as previously suggested, by adding gender labels in verbal trait continua). In addition, it would also be interesting to examine how facial impressions are updated by trait-implying information (e.g., trait-based or behavioral descriptions), and vice-versa, by for instance, manipulating the order of presentation of stimulus modality. While doing so, it could also be interesting to examine whether the valence of traits exhibiting curvilinear relationships with valence (e.g. dominance) is more susceptible to change in the presence of the new information (facial or verbal, depending on the order condition) comparatively to traits establishing strong linear relationships with valence (e.g., trustworthiness).

An alternative explanation for the inconsistency in the nature of trait-valence relationships across domains lies in a potential methodological limitation related with the

range of the face continua used in Study 3 of Chapter VII. It is conceivable that the shift of the dominance-valence relationship from curvilinear (in the context of person perception) to linear and negative (in the context of face evaluation), was promoted by an inability of our dominance face continua to match the dimensional range of the verbal dominance continuum. If our dominance face continua were truncated at the low pole of dominance (i.e., most extreme submissive-looking face), we would not be able to detect any "real" curvilinear relationship in our data. For this reason, we advise caution in the interpretation of our finding. Nevertheless, the negative relationship between dominance and valence does appear to be corroborated by the results obtained in Chapters VI and VIII (i.e., trait ratings of classification images), as well as by findings in the literature (e.g., Dotsch & Todorov, 2012; Oh et al., 2019; Oosterhof & Todorov, 2008), thus raising our confidence in our result. Still, the shift in nature of the dominance-valence relationship across domains deserves a further examination. We highly recommend a replication of our Study 3 in Chapter VII, using more finely tuned face materials. One potential way to enhance the face continua would be to employ RC methods—ideally face space-based as they produce more realistic faces—to generate faces perceived as highly extreme, and very balanced on dominance (or any other trait dimension of interest). This would allow for an increased control over the range of a face continua and increase the sensitivity of the face materials to any "real" curvilinear trend.

Lastly, another account of our findings relates to the stability of the data. While we replicated the finding of a curvilinear relationship between dominance and valence in the context of person perception (Studies 1 and 2, Chapter VII), we did not conduct a second study to confirm the stability of the relationships found between dominance and valence found in the context of facial impressions (Study 3, Chapter VII). In this regard, we should take into account that our pattern of findings is corroborated by more recent work conducted by Oh and colleagues (Oh et al., 2019), who also found a stronger linear negative correlation between dominance and trustworthiness for impressions of female faces (vs. lower, close to null, negative correlation found for male faces). However, because these authors did not manipulate social perspective (i.e., judging oneself vs. judging others), their work is not informative about the stability of our findings regarding the desirability for self ratings. With this in mind, a future replication of our face continua study should additionally place a higher focus (i.e., larger sample, more theory-driven hypotheses) on the desirability for self ratings to clarify whether self-referential judgments of facial appearance may indeed be promoting a curvilinearity between dominance and valence.

The Role of Valence in the Relationship and Integration of the Two Dimensions "within" Face and Person Perception Models

In this section, we discuss the role that valence plays in the relationship and integration of the two dimensions of each separate model/domain. In regard to person perception, our results suggest that the warmth and competence dimensions are positively related due to a common positive relationship with valence (although warmth-related traits such as sociability exhibited a somewhat strong curvilinearity, see Chapter VII). This positive relationship between warmth and competence is in agreement with the majority of previous findings in person perception research (e.g., Rosenberg et al., 1968; Rosenberg & Sedlak, 1972; Suitner & Maass, 2008; but see discussion below). In regard to social face perception, our results indicate that the trustworthiness/valence and dominance dimensions consistently exhibit a negative relationship, in agreement with previous findings in social face perception (e.g., Chen et al., 2014; Dotsch & Todorov, 2012; Oosterhof & Todorov, 2008).

This difference between person perception and social face perception regarding the relationship that dimensions establish with one another, has implications on (a) how the stability of the relationship of dimensions within a domain (e.g., warmth and competence) impacts the relationship of dimensions between models of different domains (e.g., competence vs. dominance), and on (b) how these dimensions are integrated within each domain (e.g., trustworthiness with dominance). Below, we discuss these implications and questions associated with them.

Stability of the Relationship between Dimensions for each Domain

In this section we discuss the stability of the relationship between the two dimensions of person perception (i.e., warmth and competence, or communion and agency), and between the two dimensions of social face perception (i.e., trustworthiness and dominance). Before we move on, we must clarify that whenever we refer to *stability*, we are referring to the stability of the relationship across different studies at the *between-individuals* level of analysis, since there is evidence in the literature showing that while the content of social knowledge structures (e.g., stereotype content; interrelationships between traits) may exhibit stability at the between-individuals level (i.e., aggregate or group-level), it can at the same time show instability at the within-individuals level (i.e., test-retest reliability) (see Garcia-Marques, Santos, & MacKie, 2006; Garcia-Marques, Santos, Mackie, Hagá, & Palma, 2017).

In person perception, the instability of the relationship between dimensions—typically defined as warmth and competence, or communion and agency—across the literature is a well-known conundrum among contemporary social psychologists (Imhoff & Koch, 2017; Judd, Garcia-Marques, & Yzerbyt, 2019). Although a halo pattern is typically found between dimensions (e.g., Rosenberg et al., 1968; Rosenberg & Sedlak, 1972; Suitner & Maass, 2008), sometimes a compensation effect emerges (e.g., Judd et al., 2005; Kervyn et al., 2010; Yzerbyt et al., 2005; Yzerbyt et al., 2008).

In social face perception, however, a similar instability has not been reported so far. The negative relationship between trustworthiness and dominance in facial impressions has been showing stability across social face perception studies (e.g., Chen et al., 2014; Dotsch & Todorov, 2012; Oh et al., 2019; Oliveira et al., 2019; Oosterhof & Todorov, 2008), at least at the between-individuals level of analysis since, to the best of our knowledge, no studies in this domain have, so far, examined the within-individual stability of this relationship (see Garcia-Marques et al., 2017).

One clear distinction between these two literatures is that whereas social face perception research is guided by a single model (trustworthiness-by-dominance), person perception research is guided by multiple models with distinct theoretical backgrounds (e.g., warmth-by-competence, communion-by-agency, social-by-intellectual, just to name the most influential). One of the reasons for the inconsistent relationship between dimensions in person perception research may be related with differences between the trait sets used by different authors in regard to how the content of these sets is related, on average, with valence. The selection of different trait sets to represent the target fundamental dimensions is usually guided by the author's preferred model and theoretical framework. In this regard, it becomes important to strive for consensus and search for an integration of person perception models to better understand the real stability of the relationship between the two dimensions of person perception (but see Abele & Wojciszke, 2007).

Nevertheless, and importantly, our findings (in Chapter VII) suggest the relevance of distinguishing traits on the basis of their relationship with valence. Future studies investigating the relationship of the two dimensions, should not only counterbalance the number of traits representing each facet of a Big Two dimension, but also take into account their relationship with valence, in order to prevent oversampling issues. This recommendation is especially relevant for agency-related measurements, given that the competence and assertiveness facets of the agency dimension strongly diverge in the nature of the relationship

they establish with valence (see Chapter VII). Regarding the communion dimension, although both the morality and warmth facets of the communion dimension exhibited strong positive linear relationships with valence, we should take into account that the communion warmth facet (and especially the trait sociability) showed a somewhat strong curvilinear trend in its relationship with valence (see Chapter VII). We believe these recommendations will be helpful in the long run to better understand *when* and *why* the literature is finding an inconsistent relationship between dimensions within person perception models.

The Integration of the Two Dimensions into One Impression

Although the topic of the integration of the two dimensions of person perception had been extensively investigated prior to the present work (e.g., Abele & Bruckmüller, 2011; Wojciszke & Abele, 2008; Wojciszke et al., 1998; Ybarra et al., 2001; Zanna & Hamilton, 1972), little or no information was available on how the dimensions of facial impressions were integrated. The present work aimed to fill this gap in the literature (see Chapter VIII).

Our results in Chapter VIII indicate that global impressions of faces had a higher weight of trustworthiness information—which we here consider to represent valence on the basis of their strong positive linear relationship (see Chapter VII, and Oosterhof & Todorov, 2008)—than dominance-related information. In addition, they indicate that trustworthiness information modulates the expression of dominance information in expected facial content. Coupled with previous findings supporting that trustworthiness judgments are spontaneous (Todorov et al., 2009) and precede other judgments (e.g., competence; see Willis & Todorov, 2006), our findings may be suggesting that trustworthiness information serves as the context within which dominance information is interpreted. For instance, whenever positive information on trustworthiness (trustworthy) was paired with dominance information (dominant or submissive), the resulting face was perceived as more positive (higher on trustworthiness). But when negative information on trustworthiness (untrustworthy) was paired with dominance information (dominant or submissive), the resulting face was perceived as more negative. Moreover, when untrustworthiness was paired with high dominance it increased the dominance signal conveyed by the output face.

One potential explanation for these findings could be that perceivers only pay attention to dominance-related information when they perceive the target to be untrustworthy. An untrustworthy person signals negative intentionality, which in turn may trigger a concern in

the perceiver to evaluate the degree of potential threat that person may pose (e.g., Fiske et al., 2007; Oosterhof & Todorov, 2008; Sutherland et al., 2013). By contrast, this concern may be reduced, or even absent, when the target person conveys positive intentionality (i.e., looks trustworthy). This, however, remains a hypothesis to be tested. Future studies examining such a hypothesis could, for instance, assess the extent to which (the valence of) trustworthiness information mediates the mobilization of attentional resources to facial cues exclusively diagnostic of dominance information, possibly using an eye-tracking methodology.

A New Light on the Interrelationships between Dimensions of Person and Social Face Perception

In this section we compile all the knowledge obtained throughout the three papers (Chapters VI to VIII) that constitute the empirical core of this thesis. Our aim is to provide a clearer picture on how the relationship between trait dimensions within person perception and within social face perception may be shaped by the association they establish with valence. At the same time, we hope to shed light on the main differences across the two domains regarding the relational structure of dimensions within each of them.

One useful framework to organize this knowledge is the Big Two facets model (Abele et al., 2016). This is the only model of social perception we are aware of that is able to accommodate all the different trait dimensions identified in both person perception and social face perception research. For instance, the agency assertiveness and communion morality facets of the Big Two encapsulate, respectively, the dominance and trustworthiness dimensions of facial impressions. At the same time, the warmth and competence dimensions proposed by Fiske and colleagues (Cuddy et al., 2008; Fiske et al., 2007) are encapsulated, respectively by the communion warmth/sociability and agency competence facets; while the agency assertiveness facet and both the communion morality and warmth facets accommodate the agency-by-communion model (Abele & Wojciszke, 2007, 2014) according to how its dimensions have been conceptualized. Within a facets model mindset, our findings strongly suggest that the dimensions of facial impressions may be understood as less multifaceted due to a lower differentiation between trustworthiness and facets sharing the same relationship with valence (see Figure 3). On the other hand, a large body of person perception studies (e.g., Abele & Wojciszke, 2007; Brambilla et al., 2011; Carrier et al., 2014; Goodwin et al., 2014) supports the conclusion that all four facets are relevant to a more comprehensive modelling of person perception (see also Fiske, 2018, who implicitly acknowledges these facets).

Bearing in mind that the overlap or divergence between two trait dimensions is driven by a common or a distinct relationship with valence, respectively, our findings allow us to hypothesize how trait dimensions will be related according to the nature and direction of their relationship with valence. Below in Figure 1, we present a diagram showing how our findings predict the relationships between all pairs of facets associated with the Big Two, within the context of person perception. Figure 2 shows an identical diagram for the domain of social face perception. And finally, Figure 3 shows the predicted associations between dimensions across the domains of person perception and social face perception, on the basis of their common relationship with valence (based on the combined findings of Chapter VI and VII).

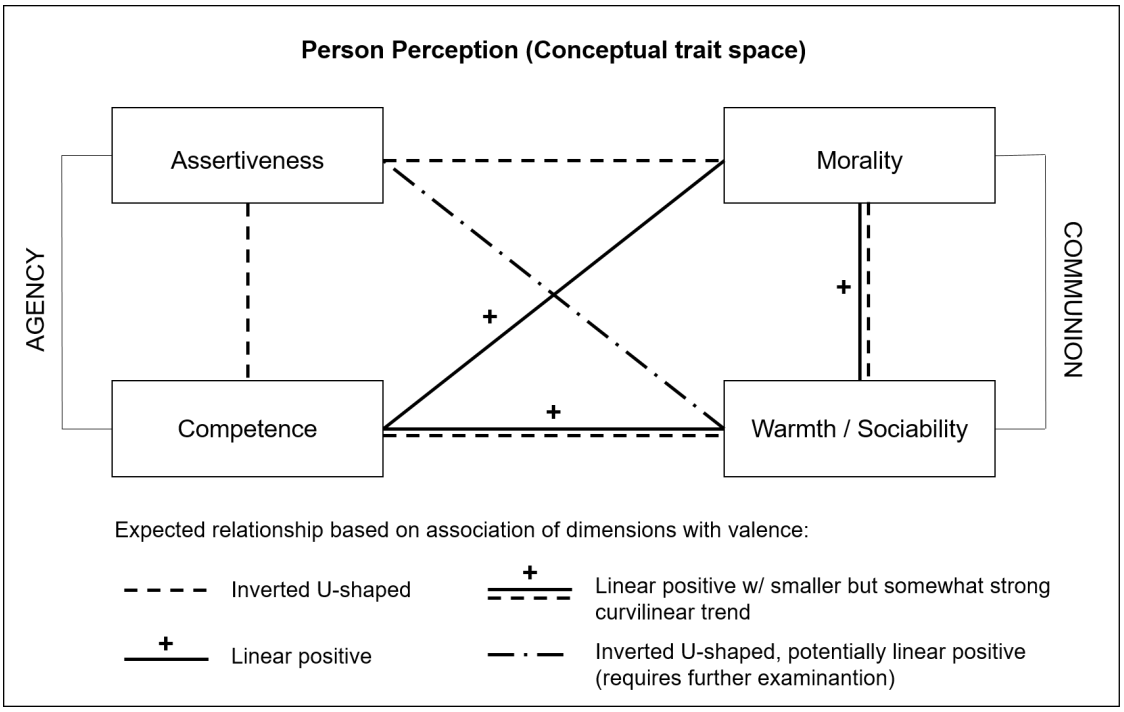

Figure 1. Hypothetical relationships between Big Two facets within the context of person perception, predicted on the basis of the relationship that each of these trait dimensions establish with valence, according to the findings reported in Chapter VII.

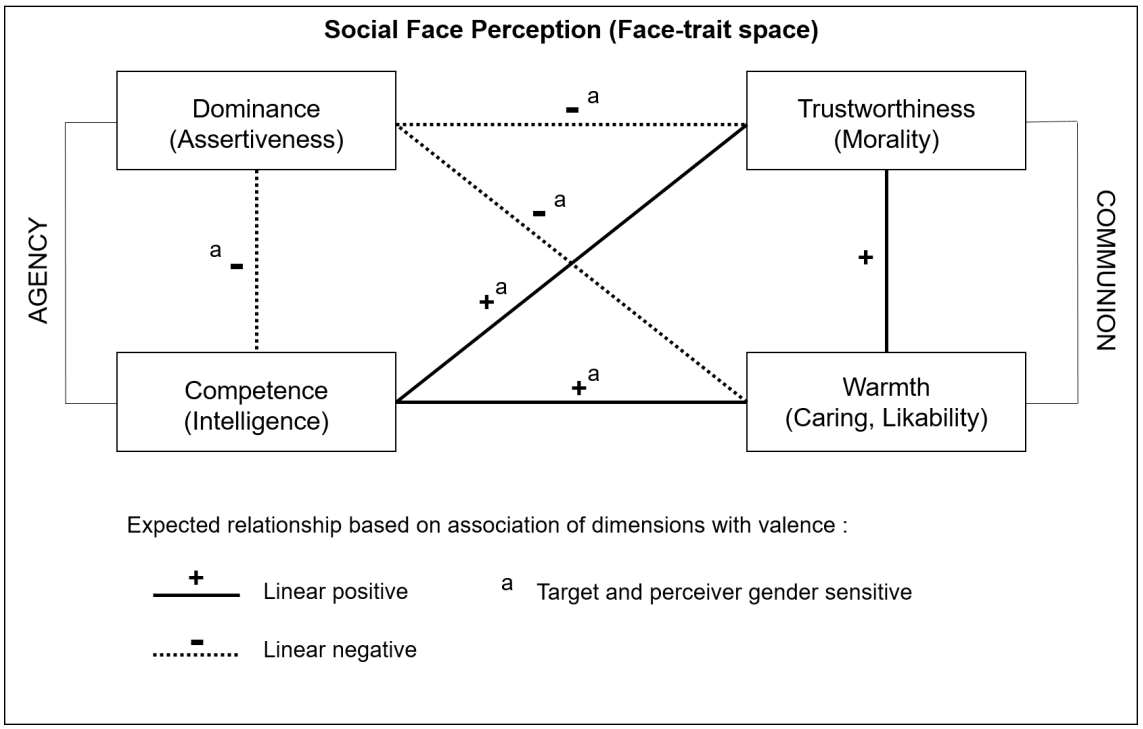

Figure 2. Hypothetical relationships between trait dimensions within the context of social face perception, predicted on the basis of the relationship that each of these trait dimensions establish with valence, according to the findings reported throughout Chapters VI to VIII. For simplicity, no curvilinear-related information was incorporated, since curvilinear trends were all substantially smaller than linear trends in facial impressions, with the exception of desirability for self in dominance evaluations by males (see Chapter VII for more details).

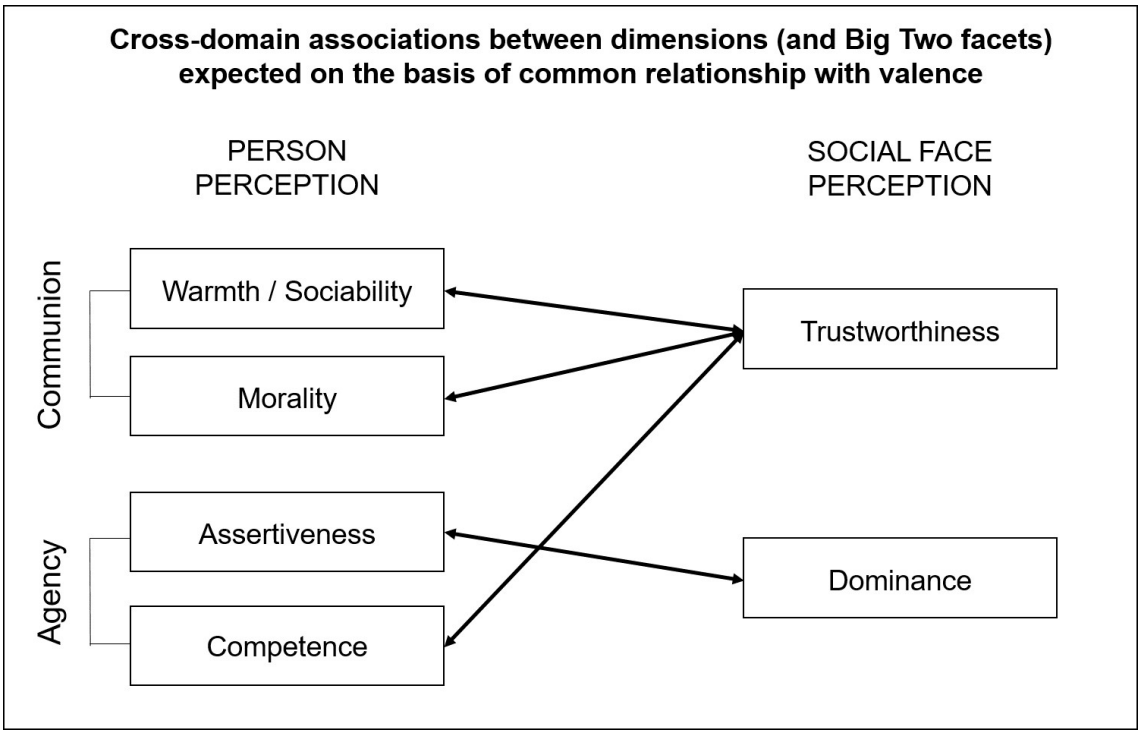

Figure 3. Hypothetical associations between the dimensions of social face perception and the facets of the Big Two dimensions of person perception, predicted on the basis of their common relationship with valence (see Chapters VI and VII).

One of the main differences we can find when comparing the interrelationships between domains (i.e., between Figures 1 and 2), is that the predicted relationship between the assertiveness facet (represented only by dominance in facial impressions; see Figure 3) with any other facet (or trait dimension in facial impressions) differs in nature across domains. Another clear pattern is how the common positive relationship that warmth, trustworthiness, and competence establish with valence predicts their positive association in both contexts. However, we can also observe that the stronger quadratic trend detected for the warmth-valence relationship (see Chapter VII), although smaller than its linear positive trend, is large enough to be non-negligible, especially in cases where warmth is understood and measured as sociability. As a result, it further complicates a prediction of how it would be related with other facets also exhibiting curvilinear relationship with valence (e.g., dominance, or assertiveness-related). Thus, whenever traits used to represent warmth emphasize sociability or closely related concepts (e.g., gregarious, talkative, or even friendly) in studies measuring correlations between the fundamental dimensions, a stronger curvilinear (less positive and

linear) trend may be potentially, and inadvertently, found between the fundamental dimensions, especially if the second dimension is positively and linearly related with valence (e.g., competence). Moreover, it is unclear whether a positive linear or a curvilinear relationship between the fundamental dimensions would emerge if they were represented by facets exhibiting both a stronger curvilinear relationship with valence, such as assertiveness and sociability. Of note, our correlation analysis in Chapter VII (Table 2) suggests that the relationship between assertiveness and sociability would be linear and positive. Sociability is the only communion-related trait with which dominance (agency assertiveness-related) establishes a high positive correlation, in terms of their perceived valence. However, we believe that further examination of the relationship between sociability and dominance is desirable for a more confident prediction about their relationship. The predicted interrelationships charted in Figures 1 and 2 are additionally helpful to predict which type of relationship pattern between the Big Two may emerge when specific pairs of communion and agency facets are used to represent them (e.g., when trait sets oversample only one facet of each of the Big Two). Lastly, the predicted cross-domain associations shown in Figure 3 provide a basis for more fine-grained predictions about how dimensions of facial impressions integrate with the dimensions of person perception and its facets. For instance, the pattern of those associations suggests that dominant-related facial features will be more strongly associated with assertiveness perceptions, and that trustworthiness-related facial features will be more relevant for both communion-related and competence perceptions due to a common relationship with valence.

Altogether, the body of findings and knowledge offered by the present work paint a clearer picture on how we should expect core dimensions of person perception and of social face perception to relate within their specific domains and across domains. Our work emphasizes that not only the direction but also the nature of the relationship that trait dimensions establish with valence is crucial for a deeper understanding of how person perception and social face perception operate.

Additional Topics and Future Directions

Throughout this thesis, we have been discussing some of the limitations of our studies and suggesting future directions to overcome them. We also point out how the hypothetical scenarios presented above can inform future research. However, several critical features of our studies that offer new topics for future research remain to be discussed. Thus, in this

section we critically discuss these features and suggest potential ways to address them in the future.

Face Materials

We worked almost exclusively with male faces. This occurred in all of our reverse correlation (RC) studies (Chapter VI and VIII). And in Chapter VII (Study 3) we found that the nature of the relationship of valence with facial trait dimensions is sensitive to gender categories. Thus, we must clarify the extent to which this gender imbalance was problematic for our research aims. We believe this was not problematic for our research aims of demonstrating that the person perception and social face perception dimensions are not (always) redundant (Chapter VI), or that trustworthiness information carries a heavier weight in facial impressions (Chapter VIII). First, because our aim is achieved when we show that the person perception and social face perception are not (always) redundant, regardless of the existence of potential circumstances under which their redundancy is observed (see also General Discussion in Chapter VI). Moreover, although our materials do not allow us to generalize the results of our RC studies across perceiver and target gender, they offer other benefits. Our materials facilitate the comparison of our results with previous studies (e.g., Dotsch & Todorov, 2012) with a similar goal to ours (i.e., visualizing mental representations of fundamental trait dimensions in faces).

But with these assertions, we do not intent to claim that an extension of the findings to female faces is not important or desirable. It is. And in fact, such an extension has recently been investigated by Oh and colleagues (Oh et al., 2019), who clarified that (a) both male and female perceivers rely on the same facial information to judge faces on trustworthiness and dominance; (b) facial impressions of female faces were less differentiated and more valence-driven than impressions of male faces (especially for female perceivers); (c) gender stereotypes moderated facial impressions (stronger negative correlation between dimensions for female faces), in agreement with our results in Chapter VII (Study 3); and that (d) similar trustworthiness-by-dominance models were obtained for male and female faces, regardless of perceiver gender (Oh et al., 2019). Moreover, the results we obtained in Study 3 of Chapter VII, also clarified that the divergence between dominance and competence in regard to their relationship with valence applies to both male and female targets, thus indirectly corroborating the findings in Chapter VI. Lastly, the finding by Oh and colleagues (Oh et al., 2019) showing that impressions of female faces are more strongly driven by valence, further

suggest that the higher weight of valence (i.e., trustworthiness) information in facial impressions found in Chapter VIII, would be intensified for female targets.

Future research investigating this topic using a reverse correlation methodology should, thus, counterbalance perceiver gender and use either a female average-face or an androgynous average-face (e.g., average of an equal amount of female and male faces) as the base face stimulus. Of note, an androgynous base face will be highly sensitive to stereotypical associations between gender and trait dimensions (Abele & Wojciszke, 2007; Sutherland et al., 2013, 2015). Given the impact of gender stereotypes in facial impressions, we advise future researchers to reflect on whether this influence is relevant for their research question or will, instead, introduce undesirable noise in their results.

Social Perspective

Judgments of the Big Two and its facets may differ when others or the self are the target of our apprehension (e.g., Abele & Wojciszke, 2014). Self-perception is likely to be more prone to desirability judgments. In some of our studies (Chapter VII) we found that the ratings of a trait's desirability for oneself were highly correlated with ratings of a trait's likeability, but also that the perceived desirability of dominance and intelligence for the self was dependent on perceiver and target gender. This may introduce a confound in studies assessing valence effects in self-perception contexts. In fact, the topic of social perspective is far from being new, and has been extensively investigated in the person perception literature (e.g., Abele & Wojciszke, 2007; Wojciszke, 2005b; Wojciszke & Abele, 2008; Wojciszke, Baryla, Parzuchowski, Szymkow, & Abele, 2011). One of the main findings in this line of research was the reversal of the primacy of morality effect into a *primacy of agency* effect in self-perception (Wojciszke, 2005b; Wojciszke & Abele, 2008). Put simply, this effect indicates a higher relevance and desirability of competence than morality in self-perception. The latest findings on this topic (Hauke & Abele, 2019) further clarified that the primacy of agency effect occurs specifically for agency assertiveness-related information, and only when the self-perspective reflects the private self as opposed to the public self (i.e., "me as I am viewed by others"), in which case communion-morality seems to be more relevant (and primacy of morality occurs). Although this line of research has been accruing a dense body of knowledge about self-perception in the context of the Big Two, its extension to the domain of social face perception remains, to the best of our knowledge, outside the radar of research in the field.

For this reason, we believe an extension of the investigation of self-other differences in social face perception would be an exciting new avenue of research. Some potential questions to be addressed could include: Do we evaluate others' faces using the impression we have of our own face as a reference? Is there a primacy of dominance over trustworthiness depending on type of self-perspective (private vs. public)? Or perhaps, what is the impact of gender identity on the desirability of specific facial attributes for oneself?

The Continuum of a Dimension

In our studies we used face continua to assess the nature of the relationship between valence and trait dimensions (Study 3, Chapter VII). The presentation of the whole continua was intentional as our aim was to activate the dimension in the mind of the perceiver so we could assess how the different levels of the (more highly salient) target dimension were evaluated in relation to each other in a dimensional context. In the case of faces, the focal dimension may even only be possible to convey through the perceived variability of faces across the continuum. If instead we presented the different faces of a face continuum one-by-one in separate trials, we would have reduced control over what dimensions are capturing the attention of the perceiver, given the vast number of dimensions that are conveyed by a single face. Nevertheless, a more severe confirmatory test of our findings could be carried out in a follow-up study to strengthen the confidence in our results.

The follow-up study could, for instance, present the faces of a continua one-by-one in separate trials of a dimension block, either sequentially (less severe test) or in random order (more severe test). In this approach, it may be crucial to include initial practice blocks with faces varying along example social dimensions (e.g., extraversion, gender, skin tone), as well as instructions emphasizing that the different faces within each block are varying along one unidentified social dimension. In this way, one could maximize the participants' focus on the set of features that vary across the trials, and minimize the focus on less relevant facial features. Moreover, the different face stimuli representing the different continuum levels would ideally convey larger variation between them (e.g., create a face continua in PsychoMorph software using larger steps between levels, or creating a larger range of faces to then select the levels at which stimuli maximally convey variation in relation to the previous level), to minimize the similarity of faces across levels promoted by the more invariant facial identity.

Flexibility in the Relationship between Dimensions across Domains

Our studies show that dominance and competence do not overlap. However, we only provide direct evidence of this divergence for the context of social face perception (Chapter VI). Our studies did not directly approach that relationship between dimensions in a person perception context. This was however done by Stolier and colleagues (Stolier et al., 2018). These authors have demonstrated that personal beliefs about a strong conceptual association between trait concepts (in conceptual trait space) predict more similar face ratings on those traits (in face space). According to these findings, someone who believes competence and dominance are highly similar and strongly related concepts (e.g., a perceiver's *lay theory of personality*), will evaluate a face similarly on competence and dominance, thus producing a positive correlation between those dimensions. However this encompasses a question regarding the stability of the dominance-competence divergence across time and/or different perceivers. Stolier and colleagues' (Stolier et al., 2018) findings seem to suggest that the relationships found between dimensions within- and between-domains (i.e., person perception and social face perception) are flexible and interact across domains.

Thus, future studies should continue addressing the stability in the relationship between trait dimensions at different levels of analysis (i.e., individual- vs. group-level), not only within each of the two domains, but also across domains. In this regard, some potential questions worth pursuing include: Are relationships between traits exhibiting a similar stability (across time and/or perceivers) across the two domains? Could such stability be more easily disturbed in any of the domains? Or what is the direction of influence between domains (conceptual knowledge to face perception, or vice-versa)?

Further studies could also examine the stability at the individual level, by addressing if the degree to which the association of traits with valence is reinterpretable or resistant to change (see also Rothbart & Park, 1986). Such studies could employ a methodology similar to that used by Garcia-Marques and colleagues (Garcia-Marques et al., 2006) in their investigations of stereotype malleability. In doing so, for instance, one could ask participants to select from a set of traits, the ones that best describe their impressions of target faces, while at the same time collecting valence ratings of those traits and/or global impressions, in two separate experimental sessions across time (e.g., several days or weeks apart). This could help to clarify the extent to which the flexibility of these associations in semantic memory is able to disrupt the relationships found between core trait dimensions of social judgment.

The Use of a Reverse Correlation Methodology

Future studies should address the implications that the use of a reverse correlation (RC) methodology had on our conclusions. One first aspect to consider is that RC methods are not capturing mental representations *per se*. At best, they capture an *approximation* of an assumed "real" mental representation of facial appearance associated with a trait concept, or the perceiver's expectation regarding the visual facial information associated with a trait. Thus, the output reflects a combination of that expectation and the participant's performance during the two-image forced-choice task (see Brinkman et al., 2017).

A second aspect to consider relates to the stimulus set used in the task. If the features of the base face stimulus largely deviate from the features of the "true" face reflecting the target construct, the ability of the task's stimulus set to capture that construct is greatly reduced (e.g., using a smiling face as the base stimulus with the aim of capturing expectations about the facial appearance of anger). To minimize the impact of this limitation, we used an average-face derived from a previously validated face database. Average-faces are less noisy because they average out the appearance-related idiosyncrasies associated with each single stimulus of a database that, in turn, could more easily introduce bias in the base face. However, future studies can examine whether using a previously validated face expressing the target trait as the base image, would facilitate capturing the target trait in the output image, comparatively to using a base image that does not express that trait. This extension of the procedure would be informative about the difference between those two base images regarding the number of task trials each required to reproduce the original image capturing the trait (according to previous validation).

A limitation of the RC approach relates to the statistical quantification of the signal present in an output classification image. It is easy for researchers to fall prey to Type I errors while interpreting a classification image in the absence of more objective measures. Following prior RC studies similar to ours (e.g., Dotsch & Todorov, 2012; Dotsch et al., 2008; Imhoff et al., 2011, 2013), we subjectively validated classification images with trait judgments performed by raters from an independent sample of that used to generate the images. However, this approach remains suboptimal in the sense that it cannot rule out the possibility that raters are judging random noise as signal. As a result, inter-rater reliability becomes an important measure to understand, at least, whether raters agree on what they perceive as signal (a point to which we will return below). A solution to this problem was provided by Brinkman and colleagues (Brinkman et al., 2019). These authors introduced a

metric—infoVal—that quantifies the probability that a classification image was not generated by a random process. However, this measure is focused on the individual level classification images. In our approach, we were interested in classification images at the group level (i.e., average of all individual-level images in a given trait condition). Future approaches should address whether the use of the infoVal metric would, retrospectively, provide insights on the inter-rater agreement across perceivers for a given trait, with implications at the group-level. In addition, future research should invest in the further development of the infoVal metric by extending its application to group-level classification images given that, more often than not, these have been the images in which researchers place their focus.

Final Remarks

The present work sheds a new light on the interrelationships between core dimensions of social judgment, not only across the domains of person perception and social face perception, but also within each of these domains.

Our findings paint a clearer picture on the structural role played by valence in shaping the relationship between core dimensions of person perception and of facial impressions. Contrary to previous assumptions in the literature, our work in Chapters VI and VII has shown how dominance and competence are not redundant dimensions and exhibit very distinct relationships with valence. In doing so, our findings took an important step towards the integration of person perception and social face perception models of personality impressions in particular, and of these two research domains in general.

In addition, our work in Chapter VIII opens a new line of research by providing initial evidence on the visual integration of important trait dimensions in faces, while further demonstrating and extending the usefulness of reverse correlation methods in social perception research. Specifically, our results clarified how trait dimensions strongly overlapping with valence, such as trustworthiness, play a dominant role in facial impressions by determining to higher extent the expected visual facial appearance of impression formation targets, and modulating the expression of other trait dimensions, such as dominance, which establish weaker linear (sometimes more curvilinear) relationships with valence. In doing so, we additionally extended to the domain of social face perception previous findings in the person perception literature showing a higher weight of valence and communion-related information in personality impressions.

Importantly, and altogether, our empirical findings across Chapters VI to VIII are expected to have a considerable impact on future research assessing the relationships between fundamental dimensions of social judgment, both within and across the person perception and social face perception domains, by offering useful predictions about the type of relationship one should expect to find between different dimensions on the basis of the nature and direction of the relationship that they establish with valence.

We believe the present work has taken few, but nevertheless important steps, towards the integration of the social face perception and person perception domains.

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*, 768–776. doi: 10.1037/0022-3514.85.4.768

Abele, A. E., & Bruckmüller, S. (2011). The bigger one of the "Big Two"? Preferential processing of communal information. *Journal of Experimental Social Psychology*, *47*, 935–948. doi: 10.1016/j.jesp.2011.03.028

Abele, A. E., Cuddy, A. J. C., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, *38*, 1063–1065. doi: 10.1002/ejsp.574

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: 10.3389/fpsyg.2016.01810

Abele, A. E., Uchronski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *European Journal of Social Psychology*, *38*, 1202–1217. doi: 10.1002/ejsp.575

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*, 751–763. doi: 10.1037/0022-3514.93.5.751

Abele, A. E., & Wojciszke, B. (2013). The big two in social judgment and behavior. *Social Psychology*, *44*, 61–62. doi: 10.1027/1864-9335/a000137

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 50, pp. 195–255). Academic Press. doi: 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*, 1751–1756. doi: 10.1121/1.1912577

Anderson, N. H. (1962). Application of an additive model to impression formation. *Science*, *138*, 817–818. doi: 10.1126/science.138.3542.817

Anderson, N. H. (1968a). Application of a linear-serial model to a personality-impression task using serial presentation. *Journal of Personality and Social Psychology*, *10*, 354–362.

doi: 10.1037/h0026816

Anderson, N. H. (1968b). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, *9*, 272–279. doi: 10.1037/h0025907

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*, 258–290. doi: 10.1037/h0055756

Bakan, D. (1966). *The duality of human existence*. Reading, PA: Addison-Wesley.

Ballew, C. C. 2nd, & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences*, *104*, 17948–17953. doi: 10.1073/pnas.0705435104

Bijlstra, G., Holland, R. W., Dotsch, R., Hugenberg, K., & Wigboldus, D. H. J. (2014). Stereotype Associations and Emotion Recognition. *Personality and Social Psychology Bulletin*, *40*, 567–577. doi: 10.1177/0146167213520458

Bijlstra, G., Holland, R. W., & Wigboldus, D. H. J. (2010). The social face of emotion recognition: Evaluations versus stereotypes. *Journal of Experimental Social Psychology*, *46*, 657–663. doi: 10.1016/j.jesp.2010.03.006

Bodenhausen, G. V., & Macrae, C. N. (2006). Putting a face on person perception. *Social Cognition*, *24*, 511–515. doi: 10.1521/soco.2006.24.5.511

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, *41*, 135–143. doi: 10.1002/ejsp.744

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

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*, 333–361. doi: 10.1080/10463283.2017.1381469

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*. Advance online publication. doi: 10.3758/s13428-019-01232-2

Brooks, J. A., Stolier, R. M., & Freeman, J. B. (2018). Stereotypes bias visual prototypes for sex and emotion categories. *Social Cognition*, *36*, 481–496. doi: 10.1521/soco.2018.36.5.481

Bruner, J., & Tagiuri, R. (1954). The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2, pp. 634–654). Cambridge, Massachusetts: Addison-Wesley.

Carrier, A., Louvet, E., Chauvin, B., & Rohmer, O. (2014). The primacy of agency over competence in status perception. *Social Psychology*, *45*, 347–356. doi: 10.1027/1864-9335/a000176

Carter, N. T., Dalal, D. K., Boyce, A. S., O'Connell, M. S., Kung, M. C., & Delgado, K. M. (2014). Uncovering curvilinear relationships between conscientiousness and job performance: How theoretically appropriate measurement makes an empirical difference. *Journal of Applied Psychology*, *99*, 564–586. doi: 10.1037/a0034688

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. doi: 10.1016/j.jesp.2013.10.008

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, *25*, 215–224. doi: 10.1177/0146167299025002007

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*, 103–125. doi: 10.1037/a0030398

Cislak, A., & Wojciszke, B. (2008). Agency and communion are inferred from actions serving interests of self or others. *European Journal of Social Psychology*, *38*, 1103–1110. doi: 10.1002/ejsp.554

Conway, M., Pizzamiglio, M. T., & Mount, L. (1996). Status, communality, and agency: Implications for stereotypes of gender and other groups. *Journal of Personality and Social Psychology*, *71*, 25–38. doi: 10.1037/0022-3514.71.1.25

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. *Advances in Experimental Social Psychology*, *40*, 61–149. doi: 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., … Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, *48*, 1–33. doi: 10.1348/014466608X314935

Darwin, C. (1872). *The expression of the emotions in man and animals*. London: John Murray.

De Bruin, E. N. M., & Van Lange, P. A. M. (2000). What people look for in others:

Influences of the perceiver and the perceived on information selection. *Personality and Social Psychology Bulletin*, *26*, 206–219. doi: 10.1177/0146167200264007

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, *3*, 562–571. doi: 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*, 978–980. doi: 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*, 999–1014. doi: 10.1037/a0023026

Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science*, *13*, 513–519. doi: 10.1111/1467-9280.00490

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual Review of Psychology*, *30*, 527–554. doi: 10.1146/annurev.ps.30.020179.002523

Fischer, A. H., Rodriguez Mosquera, P. M., Van Vianen, A. E. M., & Manstead, A. S. R. (2004). Gender and culture differences in emotion. *Emotion*, *4*, 87–94. doi: 10.1037/1528-3542.4.1.87

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, *27*, 67–73. doi: 10.1177/0963721417738825

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*, 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*, 878–902. doi: 10.1037//0022-3514.82.6.878

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, *23*, 1–74. doi: 10.1016/S0065-2601(08)60317-2

Fiske, S. T., 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*, 473–489. doi: 10.1111/0022-4537.00128

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. New York:

Norton.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*, 247–279. doi: 10.1037/a0022327

Garcia-Marques, L., Santos, A. S. C., & Mackie, D. M. (2006). Stereotypes: Static abstractions or dynamic knowledge structures? *Journal of Personality and Social Psychology*, *91*, 814–831. doi: 10.1037/0022-3514.91.5.814

Garcia-Marques, L., Santos, A. S., Mackie, D. M., Hagá, S., & Palma, T. A. (2017). Cognitive malleability and the wisdom of independent aggregation. *Psychological Inquiry*, *28*, 262–267. doi: 10.1080/1047840X.2017.1373558

Goodwin, G. P. (2015). Moral character in person perception. *Current Directions in Psychological Science*, *24*, 38–44. doi: 10.1177/0963721414550709

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*, 148–168. doi: 10.1037/a0034726

Gosselin, F., & Schyns, P. G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychological Science*, *14*, 505–509. doi: 10.1111/1467-9280.03452

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*, 61–76. doi: 10.1177/1745691610393523

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*, 837–852. doi: 10.1037/0022-3514.78.5.837

Hauke, N., & Abele, A. E. (2019). Two faces of the self: Actor-self perspective and observer-self perspective are differentially related to agency versus communion. *Self and Identity*. Advance online publication. doi: 10.1080/15298868.2019.1584582

Hays, W. L. (1958). An approach to the study of trait implication and trait similarity. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 289–299). Stanford, California: Stanford University Press.

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. doi: 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*, 165–196. doi: 10.1016/S1090-5138(00)00071-4

Hentschel, T., Heilman, M. E., & Peus, C. V. (2019). The multiple dimensions of gender stereotypes: A current look at men's and women's characterizations of others and themselves. *Frontiers in Psychology*, *10*, 1–19. doi: 10.3389/fpsyg.2019.00011

Horstmann, G. (2003). What do facial expressions convey: Feeling states, behavioral intentions, or actions requests? *Emotion*, *3*, 150–166. doi: 10.1037/1528-3542.3.2.150

Imhoff, R., Dotsch, R., Bianchi, M., Banse, R., & Wigboldus, D. H. J. (2011). Facing Europe: Visualizing spontaneous in-group projection. *Psychological Science*, *22*. doi: 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. doi: 10.3389/fpsyg.2013.00386

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*, 122–137. doi: 10.1177/1745691616657334

Jack, R. E., & Schyns, P. G. (2017). Toward a social psychophysics of face communication. *Annual Review of Psychology*, *68*, 269–297. doi: 10.1146/annurev-psych-010416-044242

Jaeger, B., Evans, A. M., Stel, M., & van Beest, I. (2019). Explaining the persistent influence of facial cues in social decision-making. *Journal of Experimental Psychology: General*, *148*, 1008–1021. doi: 10.1037/xge0000591

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

Jones, B., DeBruine, L. M., Flake, J. K., Aczel, B., Adamkovic, A., Alaei, R., … Chartier, C. R. (2018). To which world regions does the valence-dominance model of social perception apply? Retrieved from https://psyarxiv.com/n26dy/

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. doi: 10.1016/j.jesp.2019.01.008

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*, 899–913. doi: 10.1037/0022-3514.89.6.899

Kelley, H. H. (1950). The warm-cold variable in first impressions of persons. *Journal of Personality*, *18*, 431–439. doi: 10.1111/j.1467-6494.1950.tb01260.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. doi: 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*, 673–681. doi: 10.1002/ejsp.1978

Kervyn, N., Yzerbyt, V., & 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*, 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*, 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*, 828–842. doi: 10.1037/a0013320

Klapper, A., Dotsch, R., van Rooij, I., & Wigboldus, D. H. J. (2016). Do we spontaneously form stable trustworthiness impressions from facial appearance? *Journal of Personality and Social Psychology*, *111*, 655–664. doi: 10.1037/pspa0000062

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, *20*, 165–182. doi: 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*, 675–709. doi: 10.1037/pspa0000046

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*, 284–308. doi: 10.1037/0033-295X.103.2.284

Leach, C., 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*, 234–249. doi: 10.1037/0022-3514.93.2.234

Lemann, T. B., & Solomon, R. L. (1952). Group characteristics as revealed in sociometric

patterns and personality ratings. *Sociometry*, *15*, 7–90. doi: 10.2307/2785447

Lewin, K. (1935). *A dynamic theory of personality*. New York: McGraw-Hill.

Livingston, R. W., & Pearce, N. A. (2009). The teddy-bear effect: Does having a baby face benefit Black Chief Executive Officers? *Psychological Science*, *20*, 1229–1236. doi: 10.1111/j.1467-9280.2009.02431.x

Louvet, E., Cambon, L., Milhabet, I., & Rohmer, O. (2019). The relationship between social status and the components of agency. *The Journal of Social Psychology*, *159*, 30–45. doi: 10.1080/00224545.2018.1441795

Lundqvist, D., Flykt, A., & Ohman, A. (1998). The Karolinska directed emotional faces (KDEF) [Face database]. Stockholm, Sweden: Karolinska Institutet. Retrieved from http://kdef.se/

Maner, J. K., & Case, C. R. (2016). Dominance and prestige: Dual strategies for navigating social hierarchies. *Advances in Experimental Social Psychology*, *54*, 129-180. doi: 10.1016/bs.aesp.2016.02.001

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*, *28*, 209–226. doi: 10.1016/j.cogsci.2003.11.004

Mason, M. F., Cloutier, J., & Macrae, C. N. (2006). On construing others: Category and stereotype activation from facial cues. *Social Cognition*, *24*, 540–562. doi: 10.1521/soco.2006.24.5.540

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*, 215–238. doi: 10.1037/0033-295X.90.3.215

Mollaret, P., & Miraucourt, D. (2016). Is job performance independent from career success? A conceptual distinction between competence and agency. *Scandinavian Journal of Psychology*, *57*, 607–617. doi: 10.1111/sjop.12329

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior*, *27*, 237–254. doi: 10.1023/A:1027332800296

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*, 250–256. doi: 10.1037/0022-3514.35.4.250

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2019). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General*, Advance online publication. doi: 10.1037/xge0000638

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*, 888–902. doi: 10.1002/ejsp.2569

Olivola, C. Y., Eubanks, D. L., & Lovelace, J. B. (2014). The many (distinctive) faces of leadership: Inferring leadership domain from facial appearance. *Leadership Quarterly*, *25*, 817-834. doi: 10.1016/j.leaqua.2014.06.002

Olivola, C. Y., & Todorov, A. (2010a). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, *34*, 83–110. doi: 10.1007/s10919-009-0082-1

Olivola, C. Y., & Todorov, A. (2010b). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, *46*, 315–324. doi: 10.1016/j.jesp.2009.12.002

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*, 11087–11092. doi: 10.1073/pnas.0805664105

Osgood, C. E. (1964). Semantic differential technique in the comparative study of cultures. *American Anthropologist*, *66*, 171–200. doi: 10.1515/9783110215687.109

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Oxford, England: University of Illinois Press.

Over, H., & Cook, R. (2018). Where do spontaneous first impressions of faces come from? *Cognition*, *170*, 190–200. doi: 10.1016/j.cognition.2017.10.002

Peabody, D. (1967). Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology*, *7*, 1–18. doi: 10.1037/h0025230

Peabody, D. (1970). Evaluative and descriptive aspects in personality perception: A reappraisal. *Journal of Personality & Social Psychology*, *16*, 639–646. doi: 10.1037/h0030259

Peabody, D. (1984). Personality dimensions through trait inferences. *Journal of Personality and Social Psychology*, *46*, 384–403. doi: 10.1037/0022-3514.46.2.384

Peeters, G. (2002). From good and bad to can and must: Subjective necessity of acts associated with positively and negatively valued stimuli. *European Journal of Social Psychology*, *32*, 125–136. doi: 10.1002/ejsp.70

Plant, E. A., Hyde, J. S., Keltner, D., & Devine, P. G. (2000). The gender stereotyping of emotions. *Psychology of Women Quarterly*, *24*, 81–92. doi: 10.1111/j.1471-6402.2000.tb01024.x

Quinn, K. A., & Macrae, C. N. (2011). The face and person perception: Insights from social

cognition. *British Journal of Psychology*, *102*, 849–867. doi: 10.1111/j.2044-8295.2011.02030.x

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*, 897–911. doi: 10.1037/a0036498

Rhodes, N., & Wood, W. (1992). Self-esteem and intelligence affect influenceability: The mediating role of message reception. *Psychological Bulletin*, *111*, 156–171. doi: 10.1037/0033-2909.111.1.156

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*, 283–294. doi: 10.1037/h0026086

Rosenberg, S., & Sedlak, A. (1972). Structural representations of implicit personality theory. *Advances in Experimental Social Psychology*, *6*, 235–297. doi: 10.1016/S0065-2601(08)60029-5

Rosenberg, S., & Olshan, K. (1970). Evaluative and descriptive aspects in personality perception. *Journal of Personality and Social Psychology*, *16*, 619–626. doi: 10.1037/h0030081

Rothbart, M., & Park, B. (1986). On the confirmability and disconfirmability of trait concepts. *Journal of Personality and Social Psychology*, *50*, 131–142. doi: 10.1037/0022-3514.50.1.131

Rotteveel, M., Gierholz, A., Koch, G., van Aalst, C., Pinto, Y., Matzke, D., … Wagenmakers, E. J. (2015). On the automatic link between affect and tendencies to approach and avoid: Chen and Bargh (1999) revisited. *Frontiers in Psychology*, *6*, 1–12. doi: 10.3389/fpsyg.2015.00335

Rule, N. O., Ambady, N., Adams, R. B., Ozono, H., Nakashima, S., Yoshikawa, S., & Watabe, M. (2010). Polling the face: Prediction and consensus across cultures. *Journal of Personality and Social Psychology*, *98*, 1–15. doi: 10.1037/a0017673

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, *9*, 260–264. doi: 10.1037/a0014681

Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin*, *79*, 294-309. doi: 10.1037/h0034496

Secord, P. F. (1958). Facial features and inference processes in interpersonal perception. In R.

Tagiuri & L. Petrullo (Eds.), *Person Perception and Interpersonal Behaviour* (pp. 300–315). Stanford, California: Stanford University Press.

Secord, P. F., & Bevan, W. (1956). Personalities in faces III: A cross-cultural comparison of impressions of physiognomy and personality in faces. *Journal of Social Psychology*, *43*, 283–288. doi: 10.1080/00224545.1956.9919224

Singular Inversions. (2005). Facegen Main Software Development Kit [Software]. Vancouver, BC, Canada: Singular Inversions.

Skowronski, J. J., McCarthy, R. J., & Wells, B. M. (2013). Person memory: Past, perspectives, and prospects. In D. Carlston (Ed.), *The Oxford Handbook of Social Cognition* (pp. 352–374). New York: Oxford University Press. doi: 10.1093/oxfordhb/9780199730018.013.0017

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*, 9210–9215. doi: 10.1073/pnas.1807222115

Suitner, C., & Maass, A. (2008). The role of valence in the perception of agency and communion. *European Journal of Social Psychology*, *38*, 1073–1082. doi: 10.1002/ejsp.525

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*, 105–118. doi: 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. doi: 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*, 186–208. doi: 10.1111/bjop.12085

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*, 25–29. doi: 10.1037/h0071663

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*, 775–791. doi: 10.1111/j.1751-9004.2011.00389.x

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*, 519-545. doi: 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*, 455–460. doi: 10.1016/j.tics.2008.10.001

Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. New Jersey: Princeton University Press.

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. doi: 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*, 813–833. doi: 10.1521/soco.2009.27.6.813

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*, 1603–1616. doi: 10.1177/0146167216666266

Toscano, H., Schubert, T. W., & Sell, A. N. (2014). Judgments of dominance from the face track physical strength. *Evolutionary Psychology*, *12*, 1–18. doi: 10.1177/147470491401200101

Uleman, J., & Kressel, L. M. (2013). A brief history of theory and research on impression formation. In D. Carlston (Ed.), *The Oxford Handbook of Social Cognition* (pp. 53–73). New York: Oxford University Press. doi: 10.1093/oxfordhb/9780199730018.013.0004

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, *43*, 161–204. doi: 10.1080/14640749108400966

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

Walker, M., & Vetter, T. (2009). Portraits made to measure : Manipulating social judgments about individuals with a statistical face model. *Journal of Vision*, *9*, 1–13. doi: 10.1167/9.11.12

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, *37*, 395–412. doi: 10.1037/0022-3514.37.3.395

Wiggins, J. S., Trapnell, P., & Phillips, N. (1988). Psychometric and geometric characteristics

of the Revised Interpersonal Adjective Scales (IAS-R). *Multivariate Behavioral Research*, *23*, 517–530. doi: 10.1207/s15327906mbr2304_8

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

Wojciszke, B. (1994). Multiple meanings of behavior: Construing actions in terms of competence or morality. *Journal of Personality and Social Psychology*, *67*, 222–232. doi: 10.1037/0022-3514.67.2.222

Wojciszke, B. (2005a). Affective concomitants of information on morality and competence. *European Psychologist*, *10*, 60–70. doi: 10.1027/1016-9040.10.1.60

Wojciszke, B. (2005b). Morality and competence in person- and self-perception. *European Review of Social Psychology*, *16*, 155–188. doi: 10.1080/10463280500229619

Wojciszke, B., & Abele, A. E. (2008). The primacy of communion over agency and its reversals in evaluations. *European Journal of Social Psychology*, *38*, 1139–1147. doi: 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*, 617–627. doi: 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*, 1251–1263. doi: 10.1177/01461672982412001

Wojciszke, B., & Szymkow, A. (2003). Emotions related to others' competence and morality. *Polish Psychological Bulletin*, *34*(3), 135–142.

Ybarra, O., Chan, E., & Park, D. (2001). Young and old adults' concerns about morality. *Motivation and Emotion*, *25*, 85–100. doi: 10.1023/A:1010633908298

Yzerbyt, V., Provost, V., & Corneille, O. (2005). Not competent but warm . . . really? Compensatory stereotypes in the french-speaking world. *Group Processes & Intergroup Relations*, *8*, 291–308. doi: 10.1177/1368430205053944

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*, 1110–1123. doi: 10.1177/0146167208318602

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*, 151–175. doi: 10.1037/0003-066X.35.2.151

Zanna, M. P., & Hamilton, D. L. (1972). Attribute dimensions and patterns of trait inferences.

Psychonomic Science, *27*, 353–354. doi: 10.3758/BF03328989

Zebrowitz, L. A. (1997). *Reading faces: A window to the soul?* Boulder, Colorado: Westview Press.

Zebrowitz, L. A. (2006). Finally, faces find favor. *Social Cognition*, *24*, 657–701. doi: 10.1521/soco.2006.24.5.657

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science*, *26*, 237–242. doi: 10.1177/0963721416683996

Zebrowitz, L. A., Andreoletti, C., Collins, M. A., Lee, S. Y., & Blumenthal, J. (1998). Bright, bad, babyfaced boys: Appearance stereotypes do not always yield self-fulfilling prophecy effects. *Journal of Personality and Social Psychology*, *75*, 1300–1320. doi: 10.1037/0022-3514.75.5.1300

Zebrowitz, L. A., Kikuchi, M., & Fellous, J. M. (2010). Facial resemblance to emotions: Group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology*, *98*, 175–189. doi: 10.1037/a0017990

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*, 1497–1517. doi: 10.1111/j.1751-9004.2008.00109.x

Zebrowitz, L. A., Wang, R., Bronstad, P. M., Eisenberg, D., Undurraga, E., Reyes-García, V., & Godoy, R. (2012). First impressions from faces among U.S. and culturally isolated tsimane' people in the Bolivian rainforest. *Journal of Cross-Cultural Psychology*, *43*, 119–134. doi: 10.1177/0022022111411386

Section IV

Appendices

Appendix A: Supporting Information for Chapter VI

Statistical analyses related:

Outputs of statistical analyses can be obtained by running the R scripts publicly available online at: https://osf.io/hr5pd/

Interactive version of the heatmap (Chapter VI, Figure 2) that allows to inspect the effect size values in each cell: https://osf.io/rs65b/

(Supporting Information initiates in the next page.)

Supporting Information

Also available online at https://osf.io/kpmjb/

Study 1

Noise generation procedure for the 2IFC task

The random noise patterns superimposed on the base-image consisted of truncated sinusoid patches in six orientations (0º, 30º, 60º, 90º, 120º, and 150º), five spatial frequency scales (2, 4, 8, 16, and 32 cycles per image), and two phases (0, π/2) with random amplitudes. As such, the random noise was a function of 4092 parameters where each defines the amplitude of one truncated sinusoid across two cycles (for a visual explanation see Dotsch & Todorov, 2012, Figures 1 and 2). For each of the 300 pair of images used in our task, the noise superimposed in one of the images had the exact opposite values of pixel luminance (i.e., negative image or inverted noise) of its counterpart. This maximized the differences between the paired stimuli, while minimizing the number of trials required for the task.

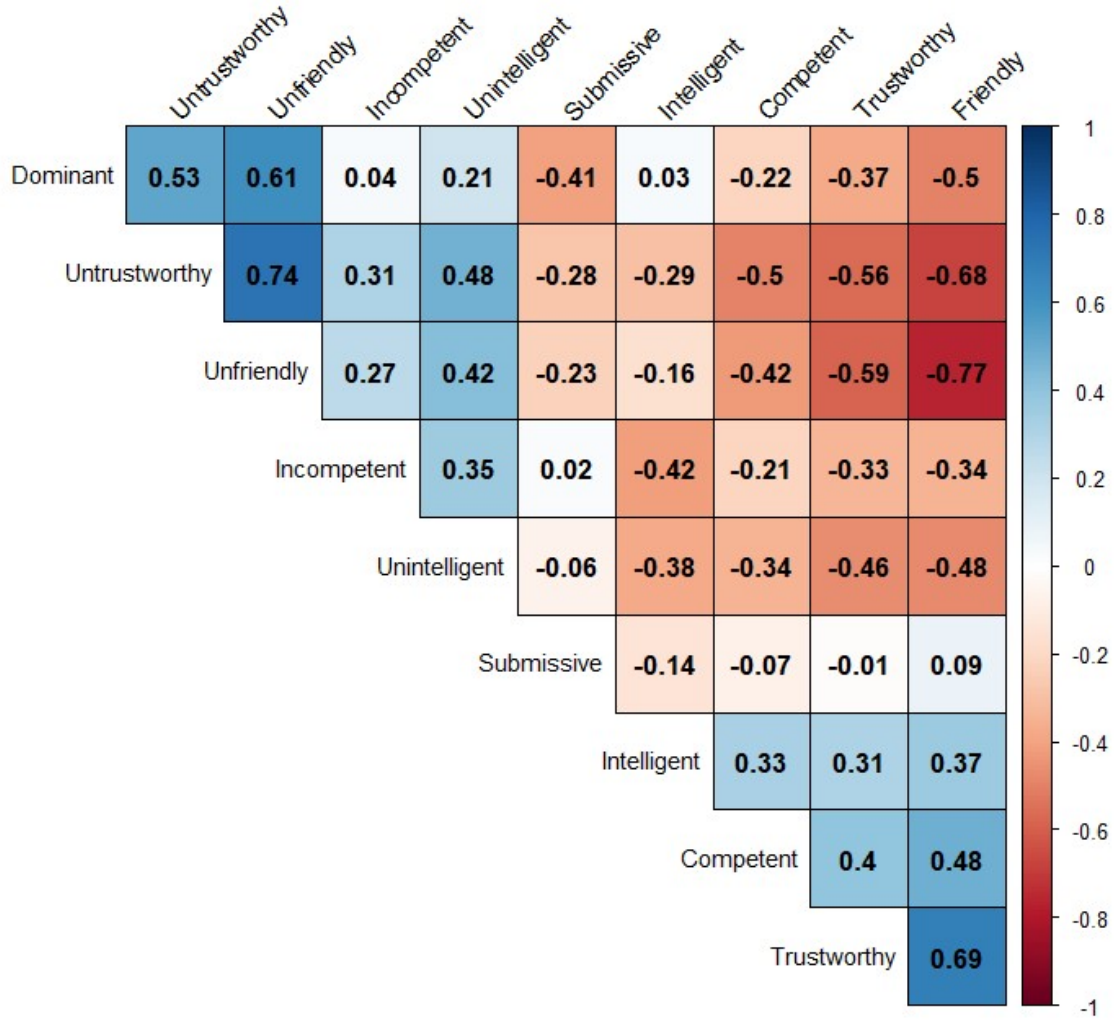

Figure S1. Color version of the matrix of pixel-wise correlations between all group-level CIs obtained in Study 1.

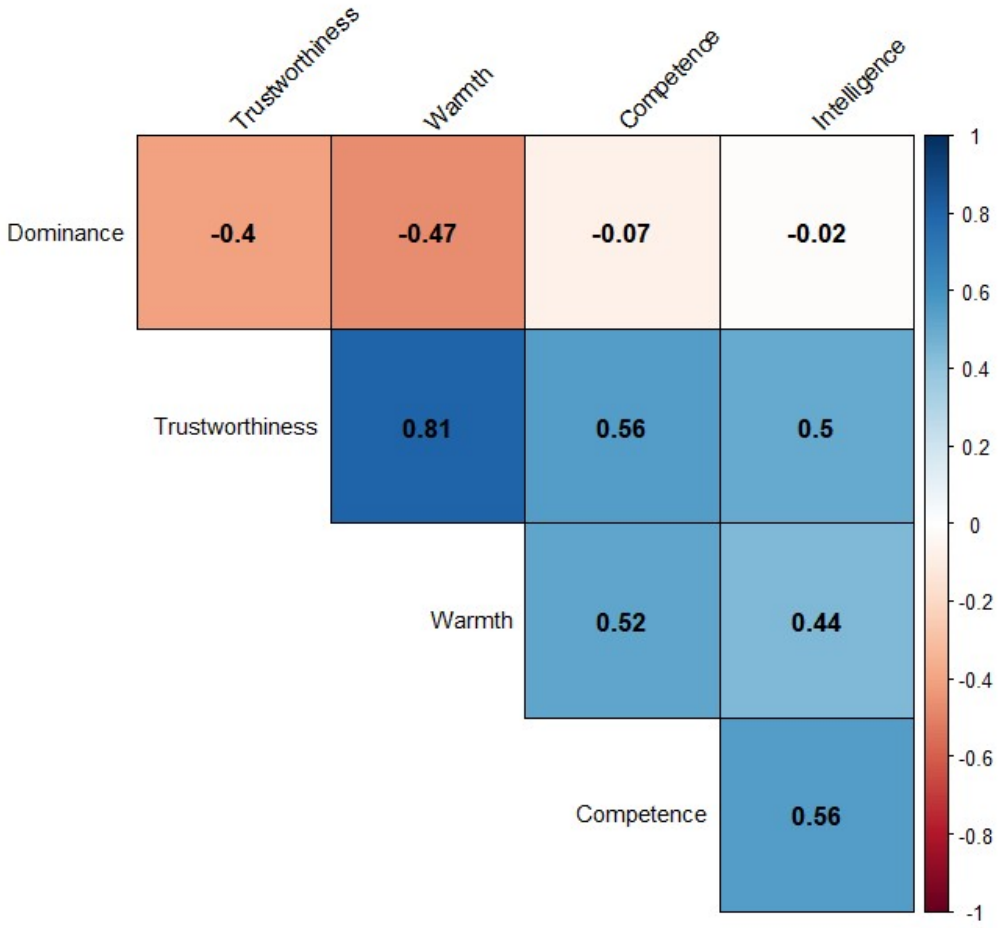

Figure S2. Pixel-wise correlations between all dimension classification images (CIs) computed for the analysis reported in the results section of Study 1. Each dimension-CI was obtained by subtracting the CI of the low pole of a dimension (e.g., Untrustworthy CI) from the CI of the high pole of a dimension (e.g., Trustworthy CI).

Study 2

Table S1

Participants' Agreement Measures for Judgments of all Group-level CIs on Eighteen Trait Dimensions Quantified as Intra-class Correlation (in Descending Order).

Trait Judgment	ICC (2, *k*)[a]	95% C.I.
Warm	.99 ***	[.97, 1.00]
Unsociable	.99 ***	[.97, 1.00]
Likeable	.99 ***	[.97, 1.00]
Friendly	.98 ***	[.96, .99]
Unfriendly	.99 ***	[.97, 1.00]
Cold	.98 ***	[.97, 1.00]
Sociable	.99 ***	[.97, 1.00]
Trustworthy	.97 ***	[.94, .99]
Untrustworthy	.97 ***	[.93, .99]
Honest	.97 ***	[.93, .99]
Dominant	.97 ***	[.94, .99]
Submissive	.97 ***	[.94, .99]
Unlikeable	.97 ***	[.93, .99]
Intelligent	.96 ***	[.92, .99]
Dishonest	.94 ***	[.88, .98]
Competent	.94 ***	[.87, .98]
Unintelligent	.94 ***	[.87, .98]
Incompetent	.93 ***	[.84, .98]

Note: [a] *k* represents the number of raters; $k = 31$ for all trait judgments; C.I. = confidence interval. Agreement is for each trait judged across all 10 group-level CIs. Degrees of freedom for the *F*-statistic were (2, 270). *** $p < .001$

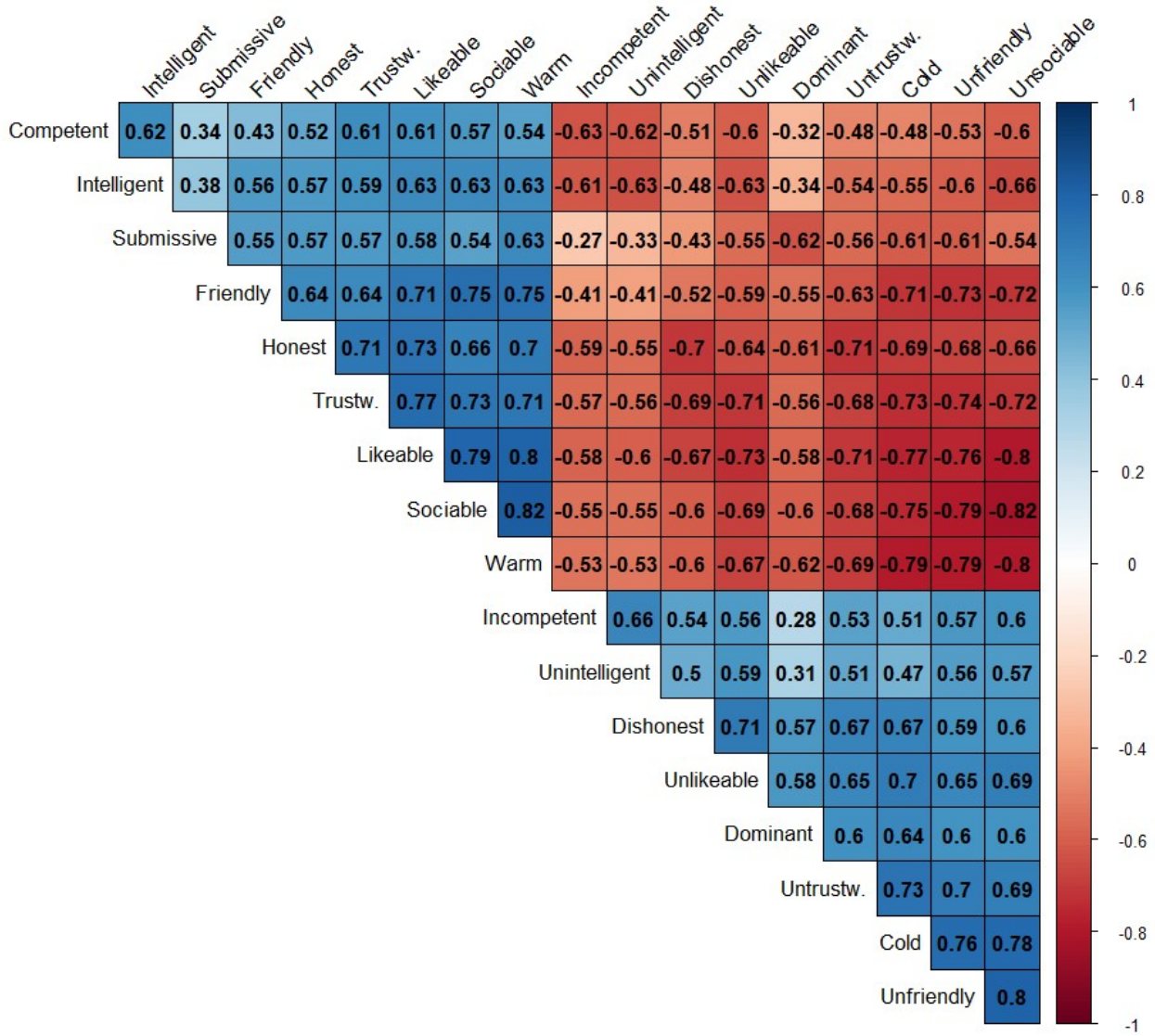

Figure S3. Correlations between trait ratings across all group-level CIs.

Studies 1 and 2: Multidimensional Scaling and Property Fitting Analysis

Table S2

Multidimensional Scaling Dimensionality Results.

Dimensionality	*Stress-I*	Multiple *R*	R^2
2	.082	.975	.952
3	.043	.979	.959
4	.025	.980	.961
5	.015	.980	.961

Note: *Stress-I* is a measure of the badness-of-fit between the input proximity data and the computed distances for the data in a given *n*-dimensional MDS solution. Rules of thumb for *Stress-I* values are: Bad [> .20]; Acceptable [0.10, 0.20]; Good [.05, .10]; Excellent [.025, .05]; Zero indicates perfect fit (Kruskal & Wish, 1978). Multiple R indicates the correlation between the input proximity data and the yielded *n*-dimensional distance coordinates. R^2 indicates the proportion of explained variance of the output distances.

Table S3

Goodness-Of-Fit Values for the Multiple Regression of the 2D Space Coordinates on Each Trait Judgment, Listed in Descending Order.

Trait Judgment	R^2	$F(2, 307)$
Unsociable	.67	317.04***
Warm	.67	310.67***
Unfriendly	.65	284.10***
Likeable	.65	281.53***
Cold	.63	263.90***
Sociable	.63	260.33***
Friendly	.58	216.31***
Dominant	.50	155.56***
Trustworthy	.50	150.92***
Submissive	.49	149.38***
Honest	.47	138.34***
Untrustworthy	.47	133.62***
Unlikeable	.45	125.01***
Intelligent	.38	93.92***
Dishonest	.33	75.39***
Competent	.31	67.52***
Unintelligent	.28	60.42***
Incompetent	.25	51.83***

Note: *** $p < .001$

Additional property fitting analyses

In Figure 3 of the manuscript, the angles between warmth and friendliness (3º for high-pole, 6º for low-pole) ascertained a high overlap between these judgments. The same occurred for intelligence and competence judgments (0º at high-pole, 8º at low-pole). The angles between judgments of warmth and competence suggest that these are not completely orthogonal (in face judgments), but slightly positively correlated (24º at high-pole, 32º at low-pole). We also did not find orthogonality between dominance and trustworthiness, but a high negative correlation instead (145º at high-pole, 157º at low-pole).

Appendix B: Supporting Information for Chapter VII

Statistical analyses related:

Outputs of statistical analyses can be obtained by running the R scripts publicly available online at: https://osf.io/gc4d8/

Materials

Task instructions

Study 1: https://osf.io/wqh6f/

Study 2: https://osf.io/k86wu/

Study 3: https://osf.io/ksqvu/

Face stimuli (path: Study 3 / Face stimuli materials)**:** https://osf.io/gc4d8/

Supporting Information

Includes additional Tables, Figures, and additional exploratory analyses.

Also available online at https://osf.io/dh74n/

Study 1

Table S1

Study 1 Linear Mixed-Effect Models, Mean Valence Peak, and Inter-Rater Agreement Detailed Results for the Valence-Score.

Big Two	Trait	Mean Peak (*SD*)	Intercept (*SD*)	Linear *b* [95% CI]	Quadratic *b* [95% CI]	*ICC* (2, *k*) [95% CI]	*AIC*
Communion	Benevolence	2.30 (1.18)	3.77 (0.37)	**0.85** *** [0.79, 0.90]	**-0.04** ** [-0.07, -0.01]	.994*** [.986, .999]	.948
	Honesty	2.65 (0.58)	3.55 (0.31)	**0.96** *** [0.92, 1.00]	0.01 [-0.02, 0.03]	.997*** [.993, .999]	.960
	Sincerity	2.71 (0.59)	3.69 (0.29)	**0.91** *** [0.87, 0.95]	0.00 [-0.03, 0.02]	.997*** [.993, .999]	.961
	Trustw.	2.95 (0.22)	3.37 (0.38)	**1.01** *** [0.97, 1.05]	**0.05** *** [0.03, 0.07]	.997*** [.993, .999]	.974
	Sociability	1.52 (1.16)	4.30 (0.40)	**0.61** *** [0.55, 0.67]	**-0.14** *** [-0.17, -0.10]	.986*** [.966, .997]	.827
	Warmth	2.35 (0.77)	3.93 (0.24)	**0.83** *** [0.79, 0.88]	**-0.05** *** [-0.08, -0.02]	.996*** [.990, .999]	.937
Agency	Confidence	1.80 (0.97)	4.36 (0.24)	**0.54** *** [0.48, 0.60]	**-0.13** *** [-0.17, -0.10]	.986*** [.965, .997]	.809
	Dominance	0.32 (1.25)	4.25 (0.46)	**0.14** *** [0.07, 0.20]	**-0.22** *** [-0.26, -0.18]	.966*** [.916, .993]	.627
	Power †	1.40 (1.17)	4.16 (0.33)	**0.35** *** [0.28, 0.41]	**-0.13** *** [-0.17, -0.09]	.967*** [.918, .993]	.651
	Status †	1.45 (1.22)	4.22 (0.32)	**0.35** *** [0.30, 0.41]	**-0.11** *** [-0.14, -0.07]	.971*** [.929, .994]	.658
	Competence	2.85 (0.43)	3.74 (0.35)	**0.90** *** [0.86, 0.94]	0.02 [-0.01, 0.04]	.996*** [.991, .999]	.961
	Intelligence	2.50 (0.64)	3.99 (0.21)	**0.78** *** [0.73, 0.82]	-0.02 [-0.05, 0.01]	.995*** [.987, .999]	.929

Note: Valence-score values range from 1 to 7. Mean peak locations for the valence-score range from -3 to 3 (i.e., continuum point values).Significant unstandardized regression coefficients are in bold, and correspond to the fixed effect predictors (Linear *b* = continuum points; Quadratic *b* = squared continuum points). Intercept and its standard deviation refer to the random-intercept by participant effect and represent the between-participant variability of the mean valence-score per trait. *ICC*s indicate inter-rater agreement for the target judgments

(k = 40, i.e. number of raters). *AIC* = Average inter-rater correlation (i.e., zero-order correlation of all possible raters within trait). † Power and Status are not traits per se, but have been used to measure agency in previous research (see Study 1 Method section). A Benjamini-Hochberg (*FDR*) correction was applied to all p-values of linear and quadratic estimates. α = .05; **p < .01; ***p < .001

Study 2

Table S2

Study 2 Linear Mixed-Effect Models, Mean Valence Peak, and Inter-Rater Agreement Detailed Results for the Valence-Score in the Bipolar Continuum Type.

Big Two Facet	Trait	Mean Peak (*SD*)	Intercept (*SD*)	Linear *b* [95% CI]	Quadratic *b* [95% CI]	*ICC* (2, *k*) [95% CI]	*AIC*
CM	Honesty	2.58 (0.70)	3.71 (0.32)	**0.96** *** [0.92, 1.00]	0.00 [-0.03, 0.02]	.997*** [.993, .999]	.978
	Trustw.	2.88 (0.32)	3.53 (0.42)	**1.00** *** [0.96, 1.03]	**0.04** *** [0.02, 0.06]	.998*** [.995, 1.000]	.994*
CW	Friendliness	2.46 (0.77)	4.19 (0.34)	**0.92** *** [0.89, 0.96]	**-0.05** *** [-0.07, -0.03]	.998*** [.994, 1.000]	.979
	Sociability	1.90 (1.16)	4.45 (0.39)	**0.69** *** [0.65, 0.73]	**-0.09** *** [-0.11, -0.07]	.994*** [.985, .999]	.925
AA	Confidence	2.00 (0.97)	4.47 (0.32)	**0.64** *** [0.60, 0.69]	**-0.10** *** [-0.13, -0.08]	.993*** [.983, .999]	.863
	Dominance	0.74 (1.47)	4.60 (0.45)	**0.22** *** [0.16, 0.28]	**-0.22** *** [-0.25, -0.18]	.972*** [.932, .994]	.593
AC	Competence	2.87 (0.34)	3.86 (0.11)	**0.91** *** [0.88, 0.95]	0.02 [0.00, 0.04]	.998*** [.996, 1.000]	.984*
	Intelligence	2.62 (0.61)	4.14 (0.23)	**0.80** *** [0.77, 0.84]	**-0.02** * [-0.04, 0.00]	.997*** [.993, .999]	.961
CM	-	2.65 (0.66)	3.62 (0.40)	**0.98** *** [0.95, 1.00]	**0.02** ** [0.01, 0.04]	.997*** [.995, .999]	.933**
CW	-	2.30 (0.82)	4.32 (0.37)	**0.81** *** [0.78, 0.83]	**-0.07** *** [-0.09, -0.05]	.996*** [.992, .998]	.888**
AA	-	1.29 (1.13)	4.53 (0.38)	**0.43** *** [0.39, 0.47]	**-0.16** *** [-0.18, -0.14]	.986*** [.973, .994]	.666*
AC	-	2.73 (0.52)	4.00 (0.20)	**0.86** *** [0.83, 0.88]	0.00 [-0.02, 0.01]	.998*** [.995, .999]	.936**

Note: CM = Communion Morality; CW = Communion Warmth; AA = Agency Assertiveness; AC = Agency Competence. Valence-score values range from 1 to 7. Mean peak locations for the valence-score range from -3 to 3 (i.e., continuum point values). Significant unstandardized regression coefficients are in bold, and correspond to the fixed effect predictors (Linear *b* = continuum points; Quadratic *b* = squared continuum points). Intercept and its standard deviation refer to the random-intercept by participant effect and represent the between-participant variability of the mean valence-score per trait. *ICC*s indicate inter-rater agreement for the target judgments (*k* = 60, i.e. number of raters). *AIC* = Average inter-rater correlation

(i.e., zero-order correlation of all possible raters within trait/facet). A Benjamini-Hochberg (*FDR*) correction was applied to all *p*-values of linear and quadratic estimates. $\alpha = .05$; $*p < .05$, $**p < .01$; $***p < .001$

Table S3

Study 2 Linear Mixed-Effect Models, Mean Valence Peak, and Inter-Rater Agreement Detailed Results for the Valence-Score in the Composite Bipolar Continuum Type.

Big Two Facet	Trait	Mean Peak (*SD*)	Intercept (*SD*)	Linear *b* [95% CI]	Quadratic *b* [95% CI]	*ICC* (2, *k*) [95% CI]	*AIC*
CM	Honesty	2.55 (0.77)	3.75 (0.28)	**0.89** *** [0.85, 0.92]	0.00 [-0.02, 0.02]	.997*** [.993, .999]	.963
	Trustw.	2.83 (0.56)	3.70 (0.30)	**0.93** *** [0.90, 0.96]	**0.02** * [0.00, 0.04]	.998*** [.995, 1.000]	.983*
CW	Friendliness	2.46 (0.81)	4.14 (0.15)	**0.84** *** [0.80, 0.87]	**-0.04** *** [-0.06, -0.02]	.997*** [.993, .999]	.968
	Sociability	1.93 (1.04)	4.43 (0.13)	**0.63** *** [0.59, 0.67]	**-0.08** *** [-0.10, -0.06]	.994*** [.986, .999]	.912
AA	Confidence	2.00 (1.05)	4.47 (0.19)	**0.57** *** [0.53, 0.61]	**-0.09** *** [-0.11, -0.07]	.993*** [.984, .999]	.884
	Dominance	0.70 (1.31)	4.60 (0.26)	**0.17** *** [0.12, 0.21]	**-0.17** *** [-0.20, -0.14]	.973*** [.935, .995]	.581
AC	Competence	2.65 (0.78)	4.02 (0.15)	**0.79** *** [0.75, 0.82]	0.00 [-0.02, 0.02]	.997*** [.992, .999]	.969
	Intelligence	2.61 (0.70)	4.22 (0.29)	**0.73** *** [0.70, 0.77]	**-0.03** * [-0.04, -0.01]	.996*** [.991, .999]	.964
CM	-	2.73 (0.60)	3.73 (0.35)	**0.91** *** [0.89, 0.93]	0.01 [0.00, 0.02]	.997*** [.995, .999]	.950**
CW	-	2.37 (0.76)	4.28 (0.21)	**0.73** *** [0.71, 0.76]	**-0.06** *** [-0.08, -0.05]	.996*** [.992, .998]	.903**
AA	-	1.41 (1.03)	4.53 (0.24)	**0.37** *** [0.34, 0.40]	**-0.13** *** [-0.15, -0.11]	.987*** [.976, .995]	.699*
AC	-	2.75 (0.60)	4.12 (0.27)	**0.76** *** [0.74, 0.78]	-0.01 [-0.03, 0.00]	.996*** [.993, .999]	.933**

Note: CM = Communion Morality; CW = Communion Warmth; AA = Agency Assertiveness; AC = Agency Competence. Valence-score values range from 1 to 7. Mean peak locations for the valence-score range from -3 to 3 (i.e., continuum point values). Significant unstandardized regression coefficients are in bold, and correspond to the fixed effect predictors (Linear *b* = continuum points; Quadratic *b* = squared continuum points). Intercept and its standard deviation refer to the random-intercept by participant effect and represent the between-participant variability of the mean valence-score per trait. *ICC*s indicate inter-rater agreement for the target judgments ($k = 60$, i.e. number of raters). *AIC* = Average inter-rater correlation (i.e., zero-order correlation of all possible raters within trait/facet). A Benjamini-Hochberg (*FDR*) correction was applied to all *p*-values of linear and quadratic estimates. $\alpha = .05$; *$p < .05$, ***$p < .001$

Plots for Bipolar, High Trait, and Low Trait Continuum Types

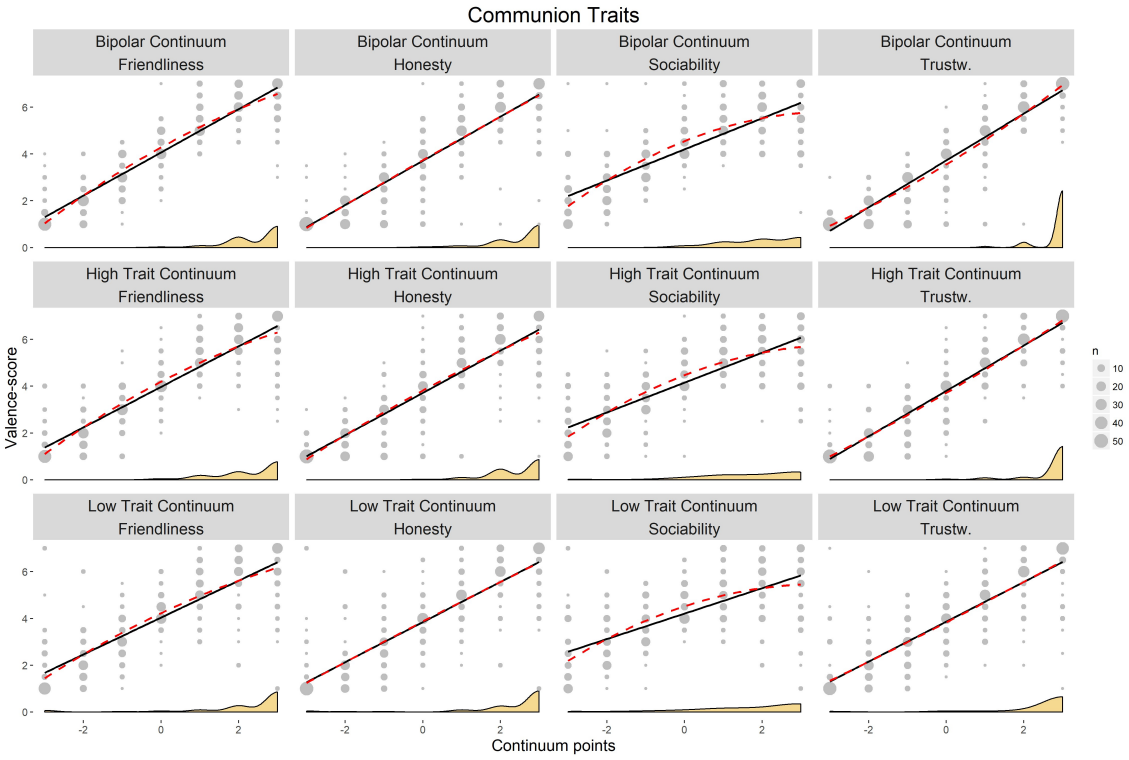

Figure S1. Linear and quadratic regression lines for each communion-related trait dimension by continuum type. Black lines represent linear fit. Red dashed lines represent quadratic fit. Each gray dot represents the number of observed valence ratings (n) and their density along the valence-score scale, across the continuum points of the trait dimensions. The density distributions of valence-score peaks across the seven continuum points of a trait dimension are shown at the bottom of each plot (in yellow).

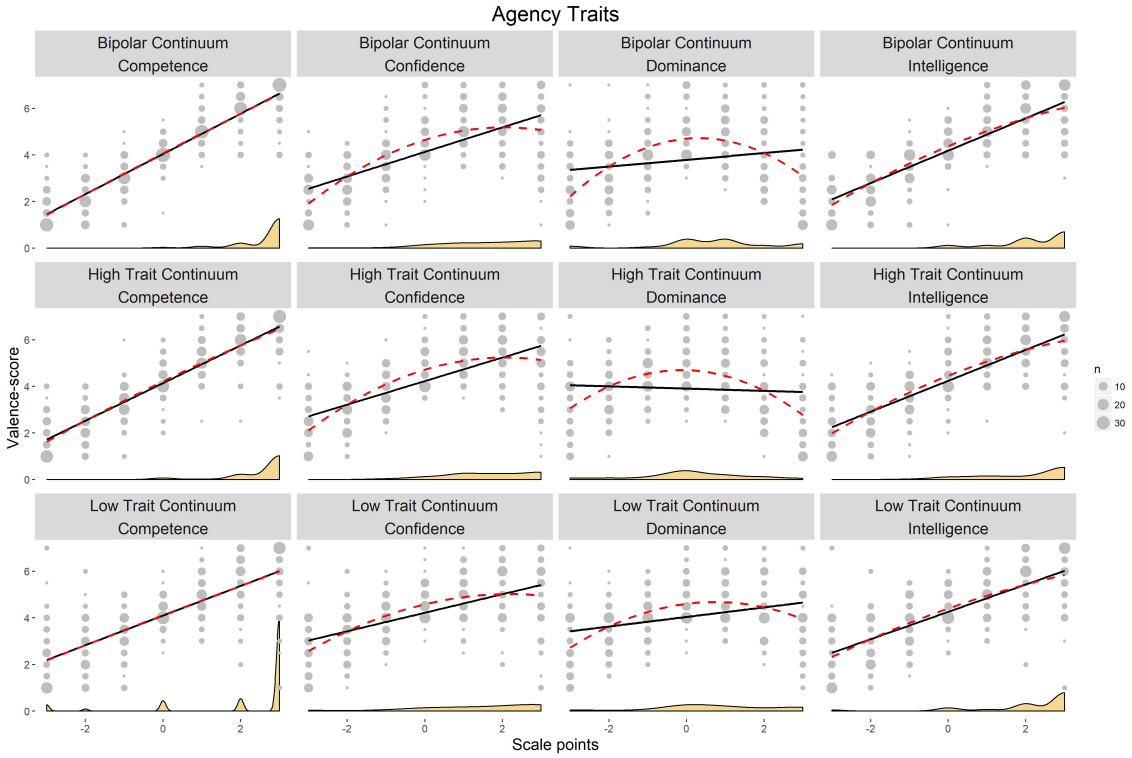

Figure S2. Linear and quadratic regression lines for each agency-related trait dimension by continuum type. Black lines represent linear fit. Red dashed lines represent quadratic fit. Each gray dot represents the number of observed valence ratings (n) and their density along the valence-score scale, across the continuum points of the trait dimensions. The density distributions of valence-score peaks across the seven continuum points of a trait dimension are shown at the bottom of each plot (in yellow).

Table S4

Study 2 Linear Mixed-Effect Models and Mean Valence Peak Results For High Trait Continuum.

Big Two Facet	Trait	Mean Peak (*SD*)	Intercept (*SD*)	Linear *b*	Quadratic *b*
CM	Honesty	2.52 (0.72)	3.72 (0.32)	**0.91** ***	-0.01
	Trustw.	2.82 (050)	3.67 (0.38)	**0.98** ***	**0.03** **
CW	Friendliness	2.38 (0.74)	4.13 (0.26)	**0.87** ***	**-0.05** ***
	Sociability	1.86 (1.19)	4.42 (0.17)	**0.66** ***	**-0.09** ***
AA	Confidence	1.97 (0.92)	4.54 (0.25)	**0.62** ***	**-0.11** ***
	Dominance	0.25 (1.42)	4.66 (0.18)	0.03	**-0.20** ***
AC	Competence	2.75 (0.60)	4.07 (0.34)	**0.86** ***	0.00
	Intelligence	2.47 (0.83)	4.25 (0.33)	**0.78** ***	**-0.03** **

Note: CM = Communion Morality; CW = Communion Warmth; AA = Agency Assertiveness; AC = Agency Competence. Valence-score values range from 1 to 7. Mean peak locations for the valence-score range from -3 to 3 (i.e., continuum point values). Significant unstandardized regression coefficients are in bold, and correspond to the fixed effect predictors (Linear *b* = continuum points; Quadratic *b* = squared continuum points). Intercept and its standard deviation refer to the random-intercept by participant effect and represent the between-participant variability of the mean valence-score per trait. A Benjamini-Hochberg (*FDR*) correction was applied to all *p*-values of linear and quadratic estimates. $\alpha = .05$; $^{**}p < .01$; $^{***}p < .001$

Table S5

Study 2 Linear Mixed-Effect Models and Mean Valence Peak Results For Low Trait Continuum.

Big Two Facet	Trait	Mean Peak (*SD*)	Intercept (*SD*)	Linear *b*	Quadratic *b*
CM	Honesty	2.38 (1.29)	3.78 (0.29)	**0.87** ***	0.00
	Trustw.	2.75 (0.91)	3.73 (0.38)	**0.88** ***	0.02
CW	Friendliness	2.36 (1.23)	4.14 (0.21)	**0.80** ***	**-0.04** **
	Sociability	2.01 (1.45)	4.44 (0.00)	**0.61** ***	**-0.07** ***
AA	Confidence	1.85 (1.43)	4.40 (0.25)	**0.52** ***	**-0.07** ***
	Dominance	0.99 (1.59)	4.54 (0.29)	**0.30** ***	**-0.14** ***
AC	Competence	2.45 (1.31)	3.97 (0.12)	**0.72** ***	0.00
	Intelligence	2.50 (1.16)	4.20 (0.34)	**0.69** ***	-0.02

Note: CM = Communion Morality; CW = Communion Warmth; AA = Agency Assertiveness; AC = Agency Competence. Valence-score values range from 1 to 7. Mean peak locations for the valence-score range from -3 to 3 (i.e., continuum point values). Significant unstandardized regression coefficients are in bold, and correspond to the fixed effect predictors (Linear *b* = continuum points; Quadratic *b* = squared continuum points). Intercept and its standard deviation refer to the random-intercept by participant effect and represent the between-participant variability of the mean valence-score per trait. A Benjamini-Hochberg (*FDR*) correction was applied to all *p*-values of linear and quadratic estimates. $\alpha = .05$; $^{**}p < .01$; $^{***}p < .001$

Participant Gender by Trait Plots

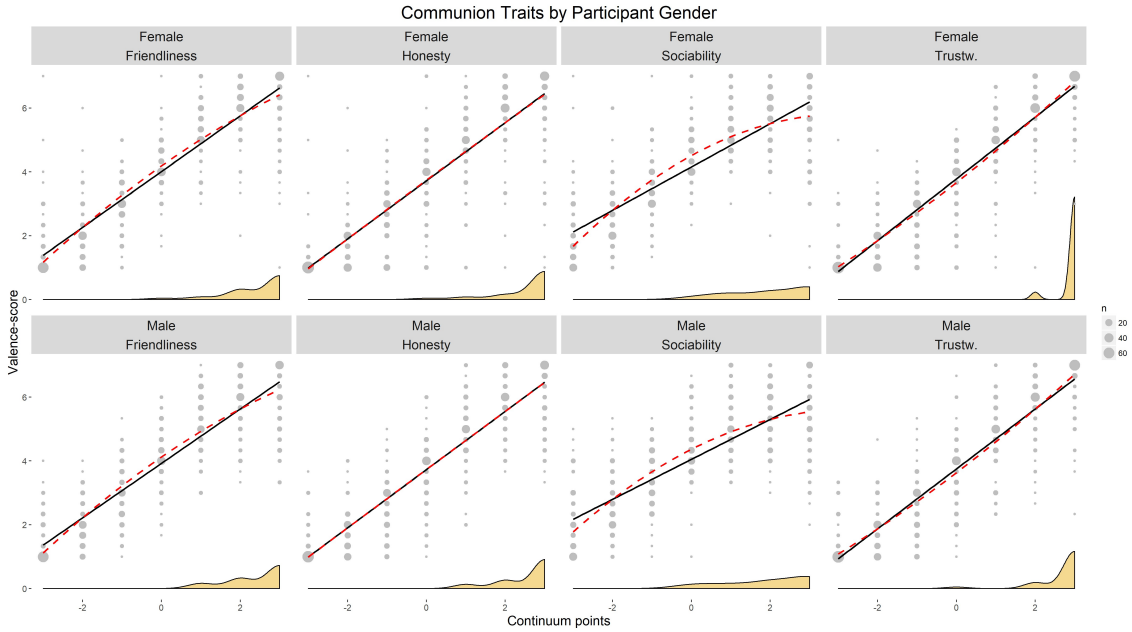

Figure S3. Linear and quadratic regression lines for each communion-related trait dimension by participant gender (collapsed across continuum type). Black lines represent linear fit. Red dashed lines represent quadratic fit. Each gray dot represents the number of observed valence ratings (n) and their density along the valence-score scale, across the continuum points of the trait dimensions. The density distributions of valence-score peaks across the seven continuum points of a trait dimension are shown at the bottom of each plot (in yellow).

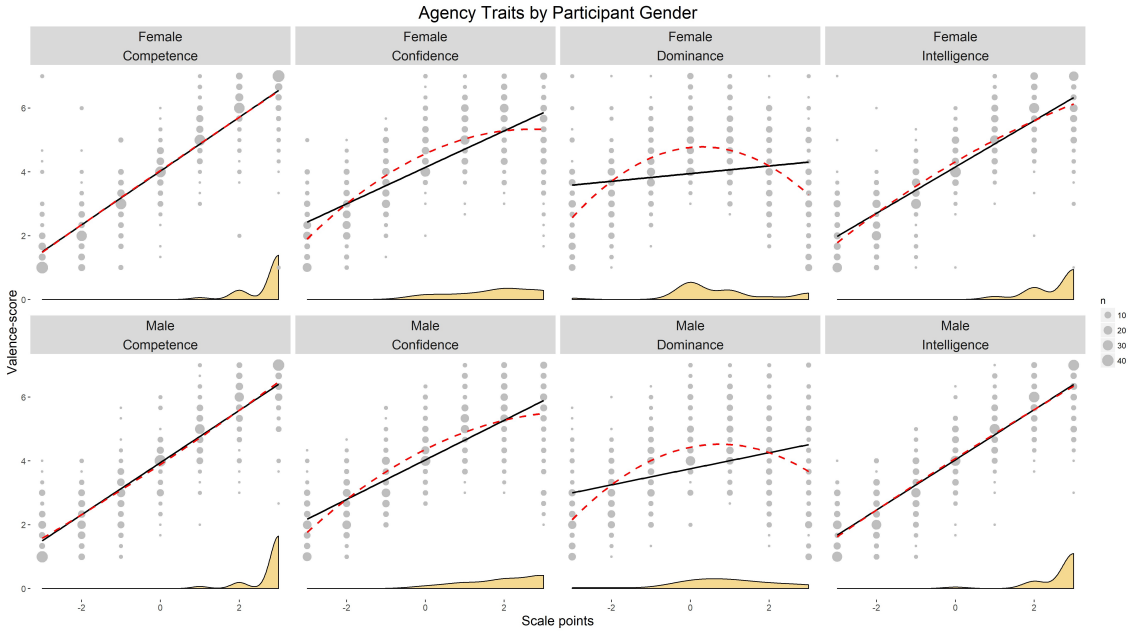

Figure S4. Linear and quadratic regression lines for each agency-related trait dimension by participant gender (collapsed across continuum type). Black lines represent linear fit. Red dashed lines represent quadratic fit. Each gray dot represents the number of observed valence ratings (n) and their density along the valence-score scale, across the continuum points of the trait dimensions. The density distributions of valence-score peaks across the seven continuum points of a trait dimension are shown at the bottom of each plot (in yellow).

Big Two facets plots

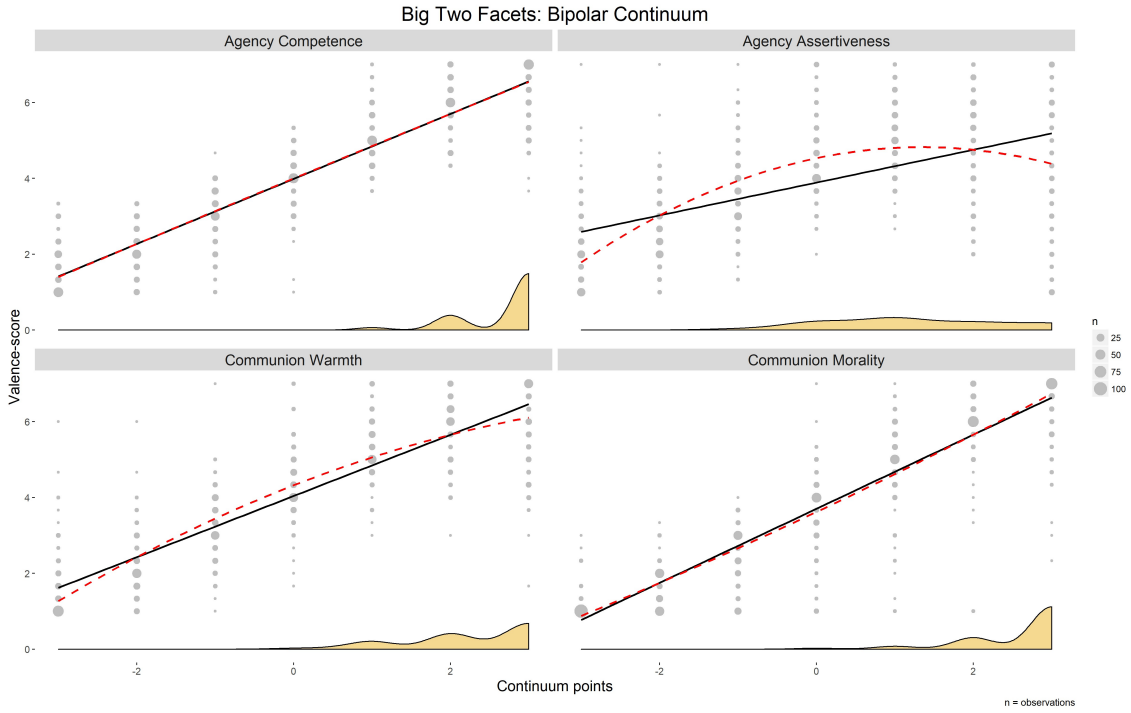

Figure S5. Linear and quadratic regression lines fitted to the valence-scores of each Big Two facet (two traits per facet), for the Bipolar Continuum used in Study 2. Black lines represent linear fit. Red dashed lines represent quadratic fit. The gray dots represent the number of observed valence-scores (*n*) and their density along the valence-score scale, for every continuum point (ranging from -3 to +3) of the facet dimension. The density distributions of valence-score peaks across the seven continuum points of a facet dimension are shown at the bottom of each plot (in yellow).

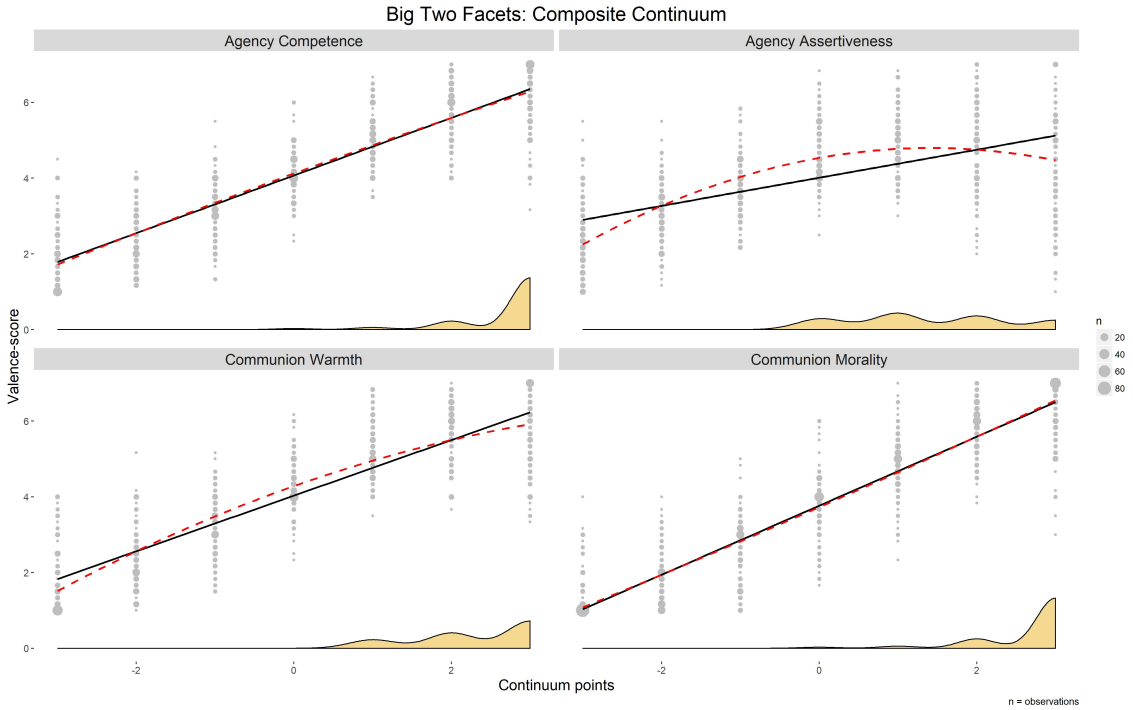

Figure S6. Linear and quadratic regression lines fitted to the valence-scores of each Big Two facet (two traits per facet), for the Composite Bipolar Continuum used in Study 2. Black lines represent linear fit. Red dashed lines represent quadratic fit. The gray dots represent the number of observed valence-scores (*n*) and their density along the valence-score scale, for every continuum point (ranging from -3 to +3) of the facet dimension. The density distributions of valence-score peaks across the seven continuum points of a facet dimension are shown at the bottom of each plot (in yellow).

Alternative analysis suggested by reviewer:

Separate analyses for Desirability for self ratings and Valence-score (Aggregating Likeability and Valence ratings)

Question: Would the results with trait words in Study 1 show a similar pattern to the ones in Study 3 where only faces were used, if the desirability for self ratings were dropped from the Valence-score, and analyzed separately?

Procedure: We dropped the desirability for self ratings from the valence-score and analyzed them separately (with the valence-score now aggregating only the likeability and valence ratings). We conducted this analysis using only the data from the bipolar continuum in Study 2, for the sake of consistency and comparability across studies 1 and 2. Results for the desirability measure and for the alternative valence-score are listed in Table S6.

Table S6

Results of the Alternative Analysis for the Desirability for Self Ratings and Alternative Valence-Score (Aggregating Likeability and Valence Ratings).

Facet	Trait	Desirability for self			Valence-score (Likeability & Valence)		
		Intercept (*SD*)	Linear *b*	Quadratic *b*	Intercept (*SD*)	Linear *b*	Quadratic *b*
CM	Honesty	3.65 (0.41)	**0.98** ***	0.01	3.74 (0.31)	**0.94** ***	-0.01
	Trustw.	3.50 (0.46)	**0.98** ***	**0.04** **	3.55 (0.43)	**1.00** ***	**0.04** ***
CW	Friendliness	4.00 (0.52)	**0.92** ***	**-0.04** **	4.28 (0.31)	**0.92** ***	**-0.05** ***
	Sociability	4.27 (0.48)	**0.75** ***	**-0.10** ***	4.55 (0.44)	**0.66** ***	**-0.09** ***
AA	Confidence	4.11 (0.46)	**0.88** ***	**-0.06** ***	4.64 (0.33)	**0.53** ***	**-0.13** ***
	Dominance	4.40 (0.45)	**0.37** ***	**-0.20** ***	4.70 (0.50)	**0.15** ***	**-0.23** ***
AC	Competence	3.43 (0.35)	**1.01** ***	**0.06** ***	4.07 (0.06)	**0.86** ***	-0.01
	Intelligence	3.67 (0.42)	**1.01** ***	**0.03** *	4.38 (0.31)	**0.70** ***	**-0.05** ***

Note: CM = Communion Morality; CW = Communion Warmth; AA = Agency Assertiveness; AC = Agency Competence. Desirability for self and Valence-score values range from 1 to 7. Significant unstandardized regression coefficients are in bold, and correspond to the fixed effect predictors (Linear *b* = continuum points; Quadratic *b* = squared continuum points). Intercept and its standard deviation refer to the random-intercept by participant effect and represent the between-participant variability of the mean valence-score per trait. $\alpha = .05$; *$p < .05$; **$p < .01$; ***$p < .001$

Study 3

Table S7

Study 3 Detailed Results With 95% Confidence Intervals For the Unstandardized Regression Coefficients and ICCs Obtained for the Valence-Score.

Stimuli	Face Continuum	Mean Peak (*SD*)	Intercept (*SD*)	Linear *b* [95% C.I.]	Quadratic *b* [95% C.I.]	*ICC* (2, *k*) [95% CI]	*AIC*
Photograph-based	Female Caring	2.31 (1.02)	4.19 (0.62)	**0.52** *** [.47, .58]	-0.03 [-0.06, 0.00]	.977 *** [.943, .995]	.815
	Male Caring	2.42 (1.17)	3.94 (0.56)	**0.46** *** [.40, .51]	0.02 [0.00, 0.05]	.972 *** [.932, .994]	.754
	Female Trustw.	1.91 (1.59)	4.27 (0.50)	**0.40** *** [.34, .46]	**-0.04** * [-0.07, -0.01]	.961 *** [.905, .992]	.656
	Male Trustw.	2.15 (1.42)	4.18 (0.52)	**0.41** *** [.36, .47]	**0.00** * [-0.04, 0.03]	.962 *** [.907, .992]	.679
	Female Dominance	-1.68 (1.61)	4.22 (0.65)	**-0.46** *** [-.52, -.40]	-0.03 [-0.06, 0.00]	.966 *** [.916, .993]	.700
	Male Dominance	-0.26 (2.13)	4.29 (0.44)	**-0.16** ** [-.23, -.10]	**-0.06** ** [-0.10, -0.03]	.811 *** [.551, .961]	.163
	Female Intelligence	1.90 (1.19)	4.34 (0.64)	**0.47** *** [0.42, 0.52]	**-0.07** *** [-0.10, -0.04]	.972 *** [.931, .994]	.763
	Male Intelligence	1.10 (1.63)	4.37 (0.52)	**0.21** ** [0.16, 0.26]	**-0.05** ** [-0.08, -0.02]	.892 *** [.743, .977]	.396
FaceGen	Likeability	1.90 (1.65)	3.91 (0.80)	**0.33** *** [0.27, 0.39]	-0.01 [-0.04, 0.03]	.925 *** [.821, .984]	.532
	Trustw.	1.85 (1.67)	3.72 (0.94)	**0.34** *** [0.28, 0.39]	-0.01 [-0.04, 0.02]	.916 *** [.801, .982]	.586
	Competence	1.10 (1.93)	3.94 (0.83)	**0.17** *** [0.11, 0.24]	-0.03 [-0.07, 0.01]	.740 *** [.431, .943]	.247
	Dominance	-0.47 (2.11)	4.00 (0.85)	**-0.17** *** [-0.23, -0.11]	**-0.07** *** [-0.11, -0.04]	.809 *** [.571, .958]	.272

Note: Valence-score values range from 1 to 7. Mean peak locations for the valence-score range from -3 to 3 (i.e., continuum point values). Significant unstandardized regression coefficients are in bold, and correspond to the fixed effect predictors (Linear b = continuum points; Quadratic b = squared continuum points). Intercept and its standard deviation refer to the random-intercept by participant effect and represent the between-participant variability of the mean valence-score per trait. *ICC*s indicate inter-rater agreement in the target trait judgments (k = 40, i.e. number of raters). *AIC* = Average inter-rater correlation (i.e., zero-order correlation of all possible raters within face continuum). A Benjamini-Hochberg (*FDR*) correction was applied to all p-values of linear and quadratic estimates. α = .05; *p < .05; **p < .01; ***p < .001

Table S8

Study 3 Detailed Results With 95% Confidence Intervals For the Unstandardized Regression Coefficients and ICCs Obtained for the Desirability For Self Ratings.

Face Continuum	Participant Gender	Mean Peak (*SD*)	Intercept (*SD*)	Linear *b* [95% CI]	Quadratic *b* [95% CI]	*ICC* (2, *k*) [95% CI]	*AIC*
Caring	Female	1.35 (1.76)	4.51 (0.56)	**0.43** *** [0.31, 0.54]	-0.07 [-0.13, 0.00]	.881 *** [.707, .975]	.421
	Male	1.82 (1.94)	3.79 (0.61)	**0.48** *** [0.38, 0.57]	0.01 [-0.05, 0.06]	.923 *** [.809, .984]	.670
Trustw.	Female	1.60 (1.24)	4.22 (0.91)	**0.41** *** [0.30, 0.52]	**-0.07** * [-0.13, -0.01]	.875 *** [.698, .974]	.604
	Male	2.02 (1.44)	4.15 (0.36)	**0.59** *** [0.49, 0.69]	-0.05 [-0.11, 0.00]	.953 *** [.882, .990]	.651
Dominance	Female	-1.20 (1.91)	4.22 (0.71)	**-0.35** *** [-0.47, -0.23]	-0.05 [-0.12, 0.02]	.788 *** [.495, .955]	.212
	Male	0.37 (1.91)	4.46 (0.53)	0.06 [-0.06, 0.17]	**-0.11** ** [-0.17, -0.04]	.491 [-.170, .891]	.129
Intelligence	Female	1.42 (1.16)	4.57 (0.63)	**0.45** *** [0.34, 0.56]	**-0.11** ** [-0.17, -0.05]	.924 *** [.812, .984]	.534
	Male	1.12 (2.18)	4.10 (0.51)	**0.24** *** [0.13, 0.35]	-0.03 [-0.09, 0.04]	.652 ** [.175, .927]	.201

Note: Desirability ratings range from 1 to 7. Mean peak locations for the desirability ratings range from -3 to 3 (i.e., continuum point values). Significant unstandardized regression coefficients are in bold, and correspond to the fixed effect predictors (Linear b = continuum points; Quadratic b = squared continuum points). Intercept and its standard deviation refer to the random-intercept by participant effect and represent the between-participant variability of the mean valence-score per trait. *ICC*s indicate inter-rater agreement in the target trait judgments (k = 20, i.e. number of raters). *AIC* = Average inter-rater correlation (i.e., zero-order correlation of all possible raters within face continuum). A Benjamini-Hochberg (*FDR*) correction was applied to all p-values of linear and quadratic estimates. α = .05; *p < .05; **p < .01; ***p < .001

Appendix C: Supporting Information for Chapter VIII

Statistical analyses related:

Outputs of statistical analyses can be obtained by running the R scripts publicly available online at: https://osf.io/4zncm/

(Supporting Information initiates in the next page.)

Supporting information

Also available online at https://osf.io/p5wmn/

Study 1

Additional results for the Pure-Artificial 2TCI relationship analysis

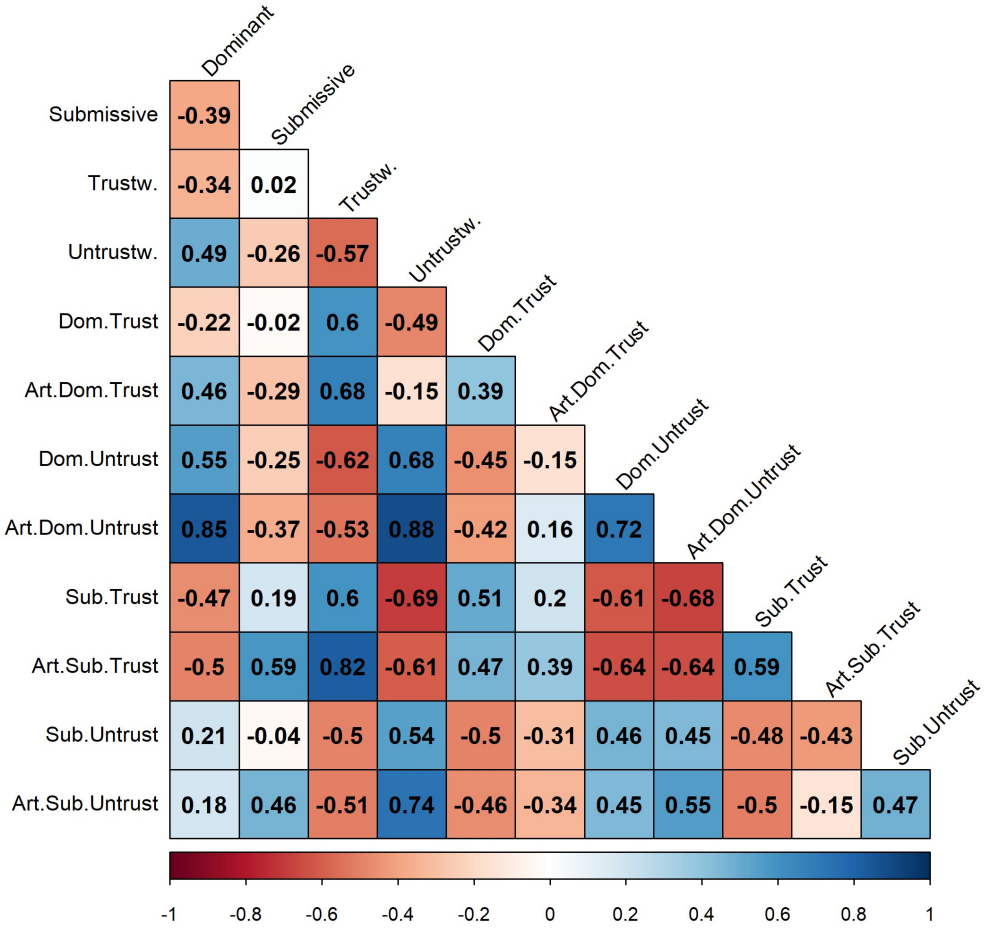

Figure S1. Pearson correlations between pixel luminance values of pure 2TCIs, artificial 2TCIs (averages of two 1TCIs), and 1TCIs (obtained from Oliveira et al., 2019), in Study 1.

Additional exploratory CI ratings

The descriptive results of the additional ratings collected in the CI rating task (see Procedure section of Study 1) are presented in Figure S2. We collected these ratings to explore how the 2TCIs were perceived in traits that have been found to be associated with face content resulting from linear combinations between features of trustworthiness and features of dominance, such as threat and competence (see Todorov, Said, Engell, & Oosterhof, 2008). Perceived threat has been previously shown to correspond to a linear combination between high dominance and untrustworthiness facial features. Faces with these features are located in the "dominant and untrustworthy" and "submissive and trustworthy" quadrants of the 2D dominance-by-trustworthiness face evaluation space (Oosterhof & Todorov, 2008). On the other hand, the face content located in the "dominant and trustworthy" and "submissive and untrustworthy" quadrants were found to be correlated with judgments of competence (as well as attractiveness and extraversion). Oosterhof and Todorov's (2008) results would be supported by a pattern of ratings showing higher (lower) threat ratings for 2TCIs generated with high (low) dominance and low (high) trustworthiness traits. And by higher (lower) ratings of competence for 2TCIs generated with high (low) dominance and high (low) trustworthiness traits. The pattern of descriptive results for the 2TCIs regarding the threat and the competence ratings seems to be consistent with the findings reported by Oosterhof and Todorov (2008). The additional 1TCIs ratings are useful in that they can serve as a reference for the 2TCI ratings.

We did not include these ratings in our main analyses as they were deemed secondary to our main focus. Nevertheless, we provide these ratings in our online data repository, recognizing their value to researchers in the field.

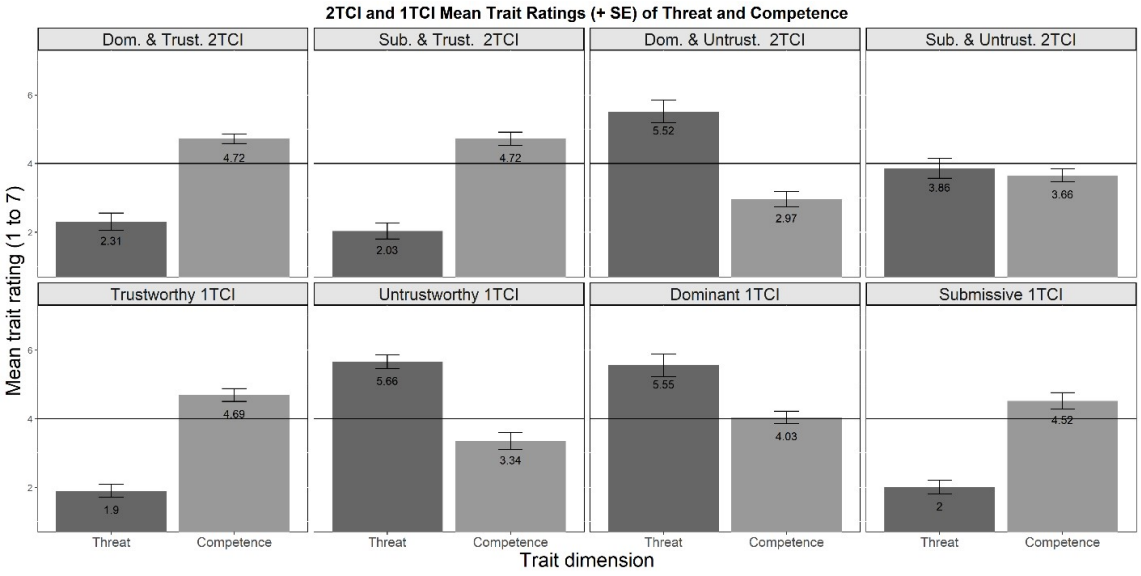

Figure S2. Mean trait ratings of threat and competence obtained for the 2TCIs generated in Study 1, and for the 1TCIs (Oliveira, Garcia-Marques, Dotsch, & Garcia-Marques, 2019). The horizontal line intercepting the y-axis at value 4 represents the midpoint of the rating scale which was labeled in the task "not very [trait] nor not at all [trait]". Error bars denote within-subjects standard errors.

Study 2

Additional results for the Pure-Artificial 2TCI relationship analysis

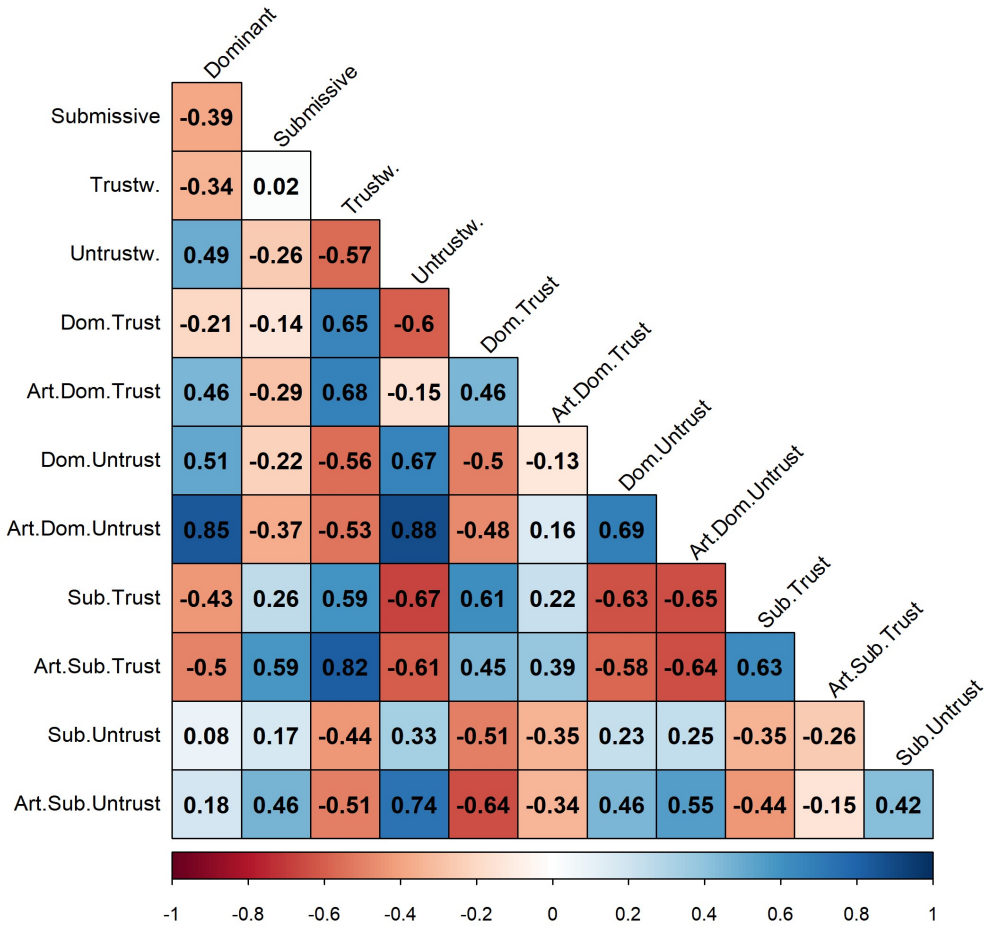

Figure S3. Pearson correlations between pixel luminance values of pure 2TCIs, artificial 2TCIs (averages of two 1TCIs), and 1TCIs (obtained from Oliveira et al., 2019), in Study 2.

Study 2: Task I

Binomial logistic regression analysis

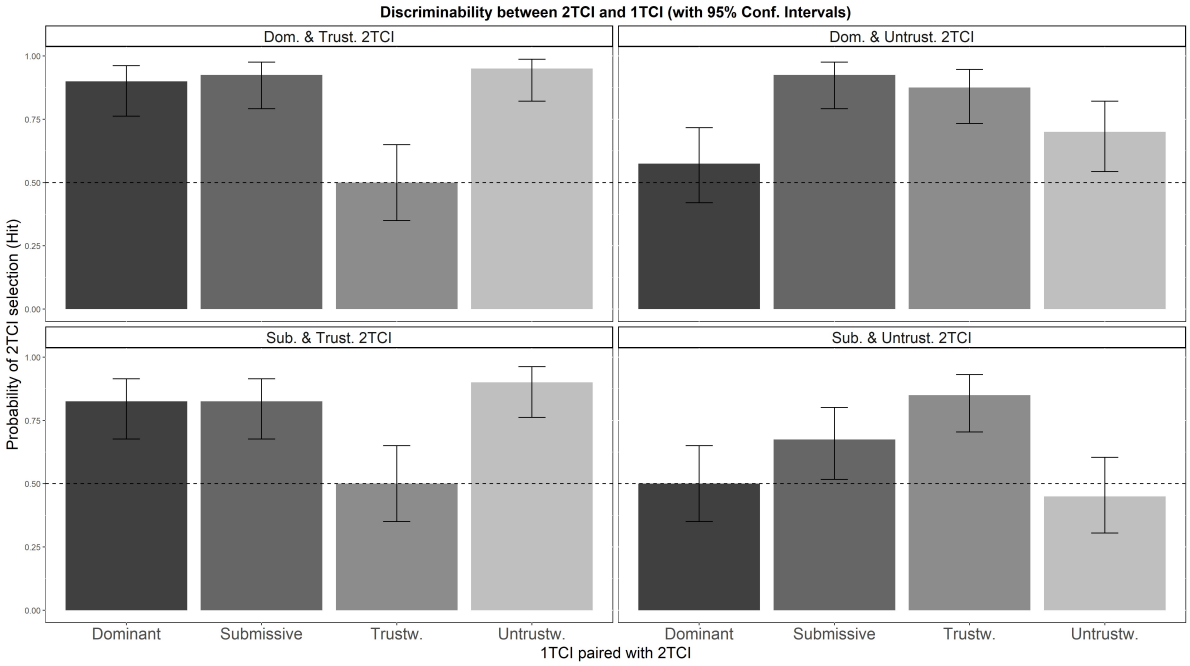

Figure S4. Results for Task I in Study 2. The plot shows the probability of a 2TCI hit in trials where a (reference) 1TCI was paired with a 2TCI, in all four 2TCI conditions. The dashed line represents chance level (i.e., probability of hit equal to probability of false alarm). Lower probability values indicate lower discriminability between the 2TCI and the 1TCI. Higher probability values indicate higher discriminability between 2TCI and 1TCI, as well as superior performance in correctly detecting the signal of both traits in the 2TCI (as opposed to the signal of just one of the traits).

Alternative signal detection analysis

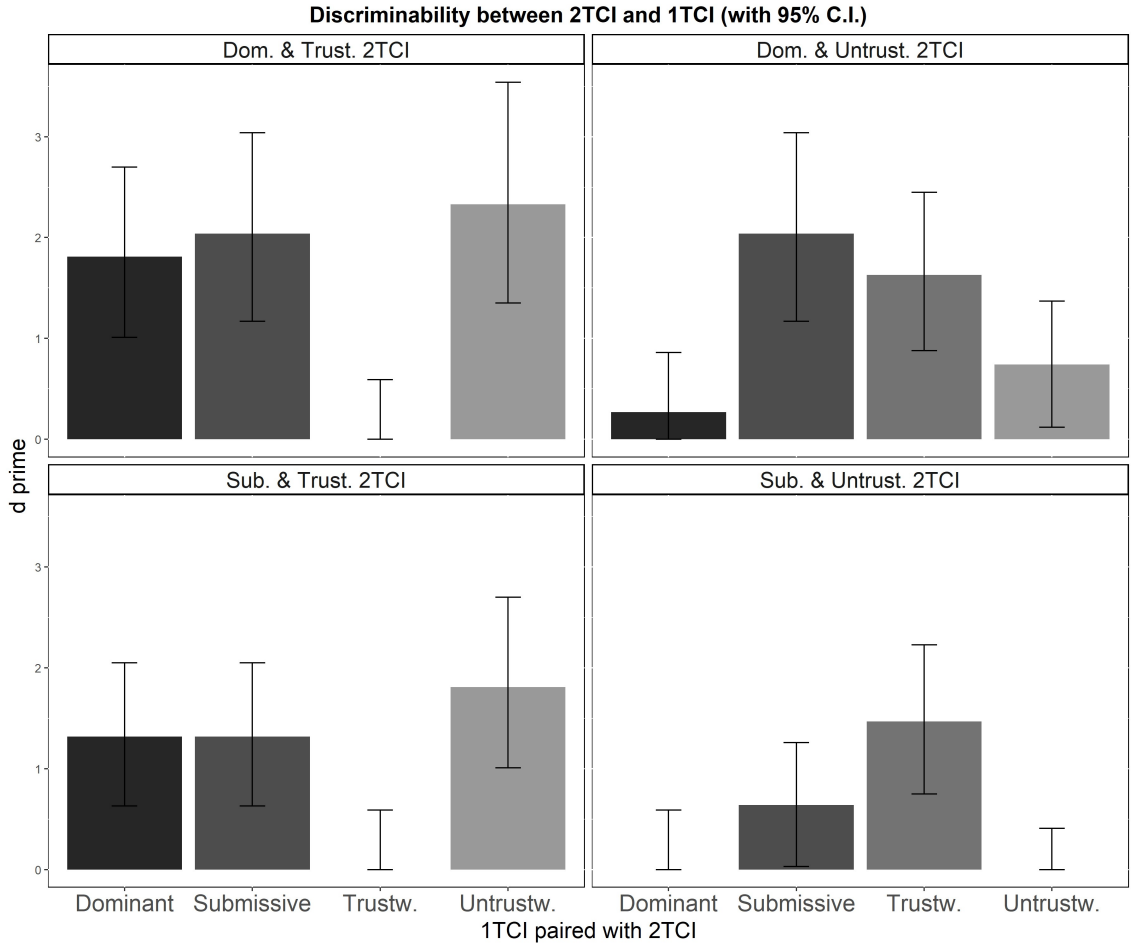

Figure S5. D-prime scores for each CI pairing (2TCI vs. 1TCI) in the trials of the Task I in Study 2, by 2TCI condition. Error bars denote the 95% confidence intervals for the d-prime score, and have a minimum possible value of zero (no discrimination between paired CIs) in this analysis.

Table S1

Study 2 Task I: D-prime Scores, Hits, and False Alarms by CI Pairing.

CI Pairing						*d′* 95% CI	
2TCI	1TCI	Hits	FA	*d′*	*p*	Lower	Upper
DomTrust	Dominant	36	4	1.81	.000	1.01	2.7
	Submissive	37	3	2.04	.000	1.17	3.04
	Trustw.	20	20	0	.563	0	0.59
	Untrustw.	38	2	2.33	.000	1.35	3.54
DomUntrust	Dominant	23	17	0.27	.215	0	0.86
	Submissive	37	3	2.04	.000	1.17	3.04
	Trustw.	35	5	1.63	.000	0.88	2.45
	Untrustw.	28	12	0.74	.008	0.12	1.37
SubTrust	Dominant	33	7	1.32	.000	0.63	2.05
	Submissive	33	7	1.32	.000	0.63	2.05
	Trustw.	20	20	0	.563	0	0.59
	Untrustw.	36	4	1.81	.000	1.01	2.7
SubUntrust	Dominant	20	20	0	.563	0	0.59
	Submissive	27	13	0.64	.019	0.03	1.26
	Trustw.	34	6	1.47	.000	0.75	2.23
	Untrustw.	18	22	0	.785	0	0.41

Study 2: Task II

Table S2

Absolute Frequencies of Dominance-Related and Trustworthiness-Related 1TCI Selections in the Trial of the 2IFC Extended Reverse Correlation Task in Which Only 1TCIs Were Presented (Task II), by 2TCI Condition.

			95% CI		
2TCI Condition	1TCI Dimension	Counts	Lower	Upper	*n*
Dom. & Trust.	Dominance-related	5	2.08	12.0	40
	Trustw.-related	35	25.13	48.7	
Dom. & Untrust.	Dominance-related	13	7.55	22.4	40
	Trustw.-related	27	18.52	39.4	
Sub. & Trust.	Dominance-related	8	4.00	16.0	40
	Trustw.-related	32	22.63	45.3	
Sub. & Untrust.	Dominance-related	12	6.81	21.1	40
	Trustw.-related	28	19.33	40.6	

Note: Dominance-related 1TCIs include the Dominant 1TCI and the Submissive 1TCI. Trustworthiness-related 1TCIs include the Trustworthy 1TCI and the Untrustworthy 1TCI. *n* indicates the number of participants in each 2TCI condition, which is identical to the total number of trials in the condition (one trial per participant).