Cognition 157 (2016) 257–267

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Integrating social and facial models of person perception: Converging and diverging dimensions

Clare A.M. Sutherland [a,b,*], Julian A. Oldmeadow [c], Andrew W. Young [a]

[a] Department of Psychology, University of York, Heslington, York YO10 5DD, UK
[b] ARC Centre of Excellence in Cognition and its Disorders, School of Psychology, University of Western Australia, Crawley 6009, WA, Australia
[c] Psychological Science and Statistics, Swinburne University of Technology, Hawthorn 3122, VIC, Australia

ARTICLE INFO

Article history:
Received 2 April 2015
Revised 8 September 2016
Accepted 11 September 2016
Available online 28 September 2016

Keywords:
Face perception
First impressions
Social cognition
Person perception

ABSTRACT

Models of first impressions from faces have consistently found two underlying dimensions of trustworthiness and dominance. These dimensions show apparent parallels to social psychological models of inter-group perception that describe dimensions of warmth (cf. trustworthiness) and competence (cf. dominance), and it has been suggested that they reflect universal dimensions of social cognition. We investigated whether the dimensions from face and inter-group social perception models are indeed equivalent by evaluating first impressions of faces. Across four studies with differing methods we consistently found that while perceptions of trustworthiness and warmth were closely related, perceptions of dominance and competence were less strongly related. Taken together, our results demonstrate strong similarity on the first dimension across facial and social models, with less similarity on the second dimension. We suggest that facial impressions of competence and dominance may represent different routes to judging a stranger's capability to help or harm.

1. Introduction

1.1. Dimensional accounts of facial first impressions

Recently, dimensional approaches have become very influential in understanding facial first impressions (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Walker & Vetter, 2009). For example, Oosterhof and Todorov (2008) used principal components analysis to reduce a variety of spontaneous trait impressions into two underlying dimensions, trustworthiness and dominance. Oosterhof and Todorov (2008) argue that these dimensions are fundamental in first impressions because together they form a judgement of threat. The first dimension approximates trustworthiness trait judgements, and concerns perceived *intention* to help or harm. This dimension is strongly (but not exclusively) influenced by cues to emotion, so that faces which appear angry are perceived as untrustworthy and to be avoided, while faces which appear happy are viewed as trustworthy and approachable (Sutherland et al., 2013; Todorov, Baron, & Oosterhof, 2008; Vernon, Sutherland, Young, & Hartley, 2014). The second dimension approximates dominance judgements and concerns perceived *capability* to carry out any helpful or harmful intentions, largely based on structural facial cues such as the masculinity of the face (Oosterhof & Todorov, 2008; Sutherland et al., 2013). More recently, others have found that three dimensions subserve first impressions made to a more varied sample of naturalistic face images (Sutherland et al., 2013; Sutherland, Liu, et al., 2015; Vernon et al., 2014). Crucially, the dimensions emerging from these naturalistic photographs replicated Oosterhof and Todorov's (2008) findings of trustworthiness and dominance dimensions, with a third youthful-attractiveness factor perhaps representing cues linked to sexual selection.

1.2. Dimensional accounts of social group and person perception

The two dimensions that have been repeatedly found in the facial first impressions literature, trustworthiness and dominance, show a strong parallel to warmth (c.f. trustworthiness) and competence (c.f. dominance) dimensions found in the perception of social groups (Cuddy, Fiske, & Glick, 2008; Fiske, Cuddy, & Glick, 2007). In these studies of inter-group perception, group membership has been defined by a wide range of characteristics, including gender, age or ethnicity, so it is impressive that the findings can be encompassed within a common overall model. Moreover, the theoretical underpinnings of dimensional models of inter-group perception

* Corresponding author at: ARC Centre of Excellence in Cognition and its Disorders, School of Psychology, University of Western Australia, Crawley 6009, WA, Australia.
E-mail address: clare.sutherland@uwa.edu.au (C.A.M. Sutherland).

http://dx.doi.org/10.1016/j.cognition.2016.09.006
0010-0277/

seem to be highly similar to models of facial first impressions. For example, in Fiske and colleagues' stereotype content model, warmth and competence together represent threat appraisal, with the warmth dimension corresponding to a judgement of the perceived intent of social groups, and the competence dimension forming a judgement of the perceived ability of that group to carry out these intentions (e.g. Fiske et al., 2007). Fiske et al. (2007) suggest that warmth and competence are universal dimensions of social cognition, which have developed over a long evolutionary history, and which represent functional adaptations that promote survival. This focus on the functional basis of social judgments as the result of evolutionary adaptations for survival closely agrees with the theoretical background of models of facial impressions (e.g. Oosterhof & Todorov, 2008; Sutherland et al., 2013).

Given that the theoretical underpinnings of the face and social group perception models are highly similar (i.e. involving approach/avoid cues and threat appraisal), there seems good reason to believe that the dimensions of impressions of real faces will closely correspond with the dimensions of abstract concepts of people and social groups. Indeed, authors are now starting to point out these potential links between facial and social models (e.g. Imhoff, Woelki, Hanke, & Dotsch, 2013; Montoya & Horton, 2014; Sutherland et al., 2013; Todorov et al., 2008; although see Dotsch & Todorov, 2011 who urge caution until this idea is tested). If the dimensions underlying facial and social judgments do correspond, this would be an elegant finding, indicating that one model can explain person perception ranging from an initial 'snap judgement' on the basis of visual information, to long-held beliefs about the nature of people and groups.

Support for the idea that the two dimensions may correspond across models comes from the observation that a substantial body of work in social psychology has implicated similar dimensions across different targets of judgement (Abele & Wojciszke, 2007). For example, communality/agency dimensions appear to underlie judgements of the self (Abele & Wojciszke, 2007; Wojciszke, 2005), whereas social/intellectual dimensions appear in judgements of abstract person perception (Rosenberg, Nelson, & Vivekananthan, 1968). The interpersonal circumplex of the perception of familiar others can also be represented with similar affiliation/dominance axes (Wiggins, 1979). This body of evidence suggests a general tendency for humans to judge others based on two broad dimensions, representing intentionality and capability, which may plausibly also encompass the dimensions of trustworthiness and dominance found in face perception.

However, there are also good reasons to question the assumption of equivalence between facial and social models. Since the two literatures use completely different stimuli (faces versus words), it is not at all clear that conceptual judgements based on words defining social groups should directly map onto the perception of non-verbal faces, or vice versa. For example, although it is clear that we can easily pick up on facial cues to certain types of social group membership (Bruce & Young, 2012), faces do not explicitly direct one's attention to social groups in the same way as group labels do. Moreover, we might pay attention to different cues in judging intentionality or capability from perceptually immediate visual stimuli (such as faces), compared to relatively abstract words describing social groups. For example, Fiske and colleagues have linked the perception of the second (competence) inter-group dimension to judgements of social structure and group power (Cuddy et al., 2008), whereas Todorov and colleagues have linked the second (dominance) facial dimension to judgements of physical strength and facial maturity (Oosterhof & Todorov, 2008). These judgements of capability do not necessarily need to be similar.

Moreover, even within social psychological studies that do not use face stimuli, there are other theoretical reasons for why dominance and competence may differ. In particular, research on the evolutionary history of human status hierarchies has shown that there are two distinct and independent ways to achieve status, with one route involving displays of dominance and the other route involving displays of competence (i.e. prestige: Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013; Henrich & Gil-White, 2001). In these evolutionary models, dominance is defined as the tendency to act aggressively to further one's interests, so that others follow out of fear (Buss & Duntley, 2013; Cheng et al., 2013). Prestige, on the other hand, is the tendency to show competent behaviour which inspires rather than compels others to follow, and is associated with admiration (Anderson & Kilduff, 2009; Cheng et al., 2013). In support of this theoretical distinction, recent studies of impression formation that have used verbally presented targets, have found that judgments of others' competence and dominance (i.e. potency or agency) also diverge (Carrier, Louvet, Chauvin, & Rohmer, 2015; Kervyn, Fiske, & Yzerbyt, 2013). This work suggests that the dimension of dominance found in face perception may not correspond with the dimension of competence found in the stereotype content model and elsewhere in social psychology.

In summary, it is an open question whether the models of conceptual stereotypes of people and the facial impression models have similar dimensions (Dotsch & Todorov, 2011 also raise this question). Answering this question represents an important theoretical step forward, since it will bridge two influential research literatures that have largely remained rather distinct (Dotsch & Todorov, 2011). Establishing the correspondence between the two leading models in these fields is also especially timely given the recent call for face perception studies to investigate the perception of visual facial cues in light of insights from social psychology (Dotsch & Todorov, 2011; Quinn & Macrae, 2011). If these social and facial dimensions do correspond, these dimensions would provide a simple and elegant model for impressions of others, covering initial and rapid reactions based on the visual presentation of unfamiliar faces, through to high-level abstract concepts about people.

1.3. Overview of current studies

Our aim was to directly examine whether the two main dimensions derived from facial and social group studies are equivalent for judgements made from faces. Despite the plausibility of this assertion, no one has directly tested this. In order to test this idea, we focused on trustworthiness and dominance judgements to represent facial dimensions (Oosterhof & Todorov, 2008), and warmth and competence judgements to represent social dimensions based on the stereotype content model of social groups (Fiske et al., 2007). We chose to focus on these two models because they are highly influential in their respective fields, because they have been explicitly compared to each other (Sutherland et al., 2013; Todorov et al., 2008), and because they have the same underlying theoretical basis (threat-based appraisal with an evolutionary background).

In Study 1 we examined the relationship between trustworthiness and warmth judgements; and the relationship between dominance and competence judgements made to the same 500 male and 500 female ambient image faces. The key finding was that warmth and trustworthiness judgements were more similar than dominance and competence judgements, especially for female faces. We also entered these judgments into a factor analysis, which confirmed that dominance and competence tended to lie on separate factors. In Study 2, we replicated this finding using a different sample of faces. In Study 3, we showed that this pattern was also robust across individual participants. In Study 4, we replicated these findings in an experimental design and with a more direct measure of similarity.

2. Study 1

In order to examine the correspondence between facial models of first impressions and social models of person and group perception, we first tested the similarity between judgements of facial trustworthiness and facial warmth, and between facial dominance and competence, using correlations between ratings collected on a database of 1,000 ambient images of faces.

2.1. Methods 1

2.1.1. Stimuli 1

The stimuli used in Study 1 were a set of 1,000 highly varied "ambient image" face photographs (500 male faces, 500 female) used in previous studies (Santos & Young, 2005, 2008, 2011; Sutherland et al., 2013; Vernon et al., 2014). The concept of ambient images was introduced by Burton, Jenkins and Schweinberger (2011; see also Jenkins, White, Van Montfort, & Burton, 2011) to emphasise the potential importance of the variability between images of faces we see in everyday life. In order to represent this variability and thus allow us to examine naturalistic first impressions, faces in this ambient image database are deliberately allowed to vary on many potential cues including age, pose, expression, lighting and facial paraphernalia such as hairstyles and glasses (see Fig. 1, see Santos & Young, 2005, 2008; Sutherland et al., 2013 for further details). However, the faces only depict adults of Caucasian appearance, as cross-cultural or own-race biases were not the current focus. The photographs in this database have been cropped around the head and shoulders and are standardised to be 150 pixels in height (approx. 5 cm on screen), but vary in width to preserve aspect ratio.

2.1.2. Warmth and competence ratings

We collected ratings of warmth and competence from twenty participants (warmth ratings group: 5 female; mean age: 21.8 years, SD age = 4.8; competence ratings group: 5 female; mean age: 21.0 years, SD age = 3.7) who volunteered to take part in this study in return for course credit or a small remuneration. We based the sample size here and in all studies in the current paper on previous work showing that this sample size is enough per trait for good reliability at the level of the group (Sutherland et al., 2013). All participants were Caucasian and self-identified as culturally Western. Participants provided informed consent to procedures that were approved by the ethics committee of the University of York Psychology Department.

Participants were tested in a quiet room on a PC running E-Prime software (version 2; Psychology Software Tools, Pittsburgh, USA). Ten participants (five male) rated the 1,000 faces on their perceived warmth (1–7, with 1 being anchored as very cold, and 7 as very warm). Ten different participants (five female) rated the 1,000 faces on their perceived competence (1–7, with 1 being very incompetent, and 7 being very competent). Before rating, 10 faces were randomly pulled from the database and used as a practice. On each trial, participants saw one photograph with the Likert scale (1–7; warmth or competence) presented underneath. Participants pressed the number key that corresponded with their rating and the next face photograph, randomly selected, appeared after a blank interval of approximately 750 ms. Participants were given as much time as they wanted but were encouraged to go with their 'gut instinct' (Todorov, Mandisodza, Goren, & Hall, 2005). On average, participants took around 2 s to rate each face, corresponding with previous facial first impressions studies (Rule, Ambady, & Adams, 2009; Sutherland, Rowley, et al., 2015).

2.1.3. Trustworthiness and dominance ratings

The ambient image database has already been rated on trustworthiness and dominance and these ratings were also used in the current study (see Sutherland et al., 2013 for further details). These were each rated by at least 6 (gender-balanced) raters and inter-rater reliabilities were good, with all alphas above 0.7 (Nunnally, 1978). Scales (1–7) were anchored as (very) untrustwor thy–trustworthy or non-dominant–dominant. All other aspects of rating collection were the same as for the current study.

2.2. Results 1

The inter-rater reliabilities for warmth and competence were good (Cronbach's alphas of $\alpha = 0.93$ and $\alpha = 0.72$ respectively). In order to assess the claim that trustworthiness and warmth are highly similar trait judgements, we correlated the warmth ratings from the current study with the previously collected trustworthiness ratings (Sutherland et al., 2013; see Fig. 2). The correlation between trustworthiness and warmth was substantial and highly significant: $r = 0.78$, $p < 0.001$, $n = 1{,}000$. Similarly, in order to assess the claim that dominance and competence are highly similar trait judgements, we correlated the competence ratings from the current study with the previously collected dominance ratings (Sutherland et al., 2013). The correlation for dominance and competence was significant but only moderate in size: $r = 0.32$, $p < 0.001$, $n = 1{,}000$. The size of the correlation between trustworthiness and warmth was significantly greater than the size of the correlation for dominance and competence: $ZPF = 15.46$, $p < 0.001$, $n = 1{,}000$ (for further details on the statistical test see Raghunathan, Rosenthal, & Rubin, 1996). It therefore appears that trustworthiness and warmth facial judgments are indeed highly similar. In contrast, dominance and competence facial judgments are less strongly related (see Fig. 2).

2.2.1. Factor analyses

In order to examine the relationship between dominance and competence within a broader group of traits, we entered the ratings of dominance, trustworthiness, warmth, and competence into a factor analysis along with a wide range of fourteen other social judgments made to the same 1,000 faces, taken from previous work (see Table 1; Santos & Young, 2005, 2008, 2011; Sutherland et al., 2013; Vernon et al., 2014). These social judgments were

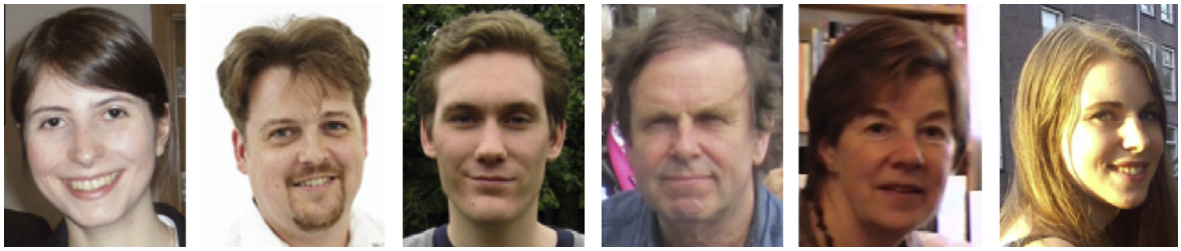

Fig. 1. Example ambient face images, reprinted with permission from Vernon et al. (2014), *Proceedings of the National Academy of Sciences*.

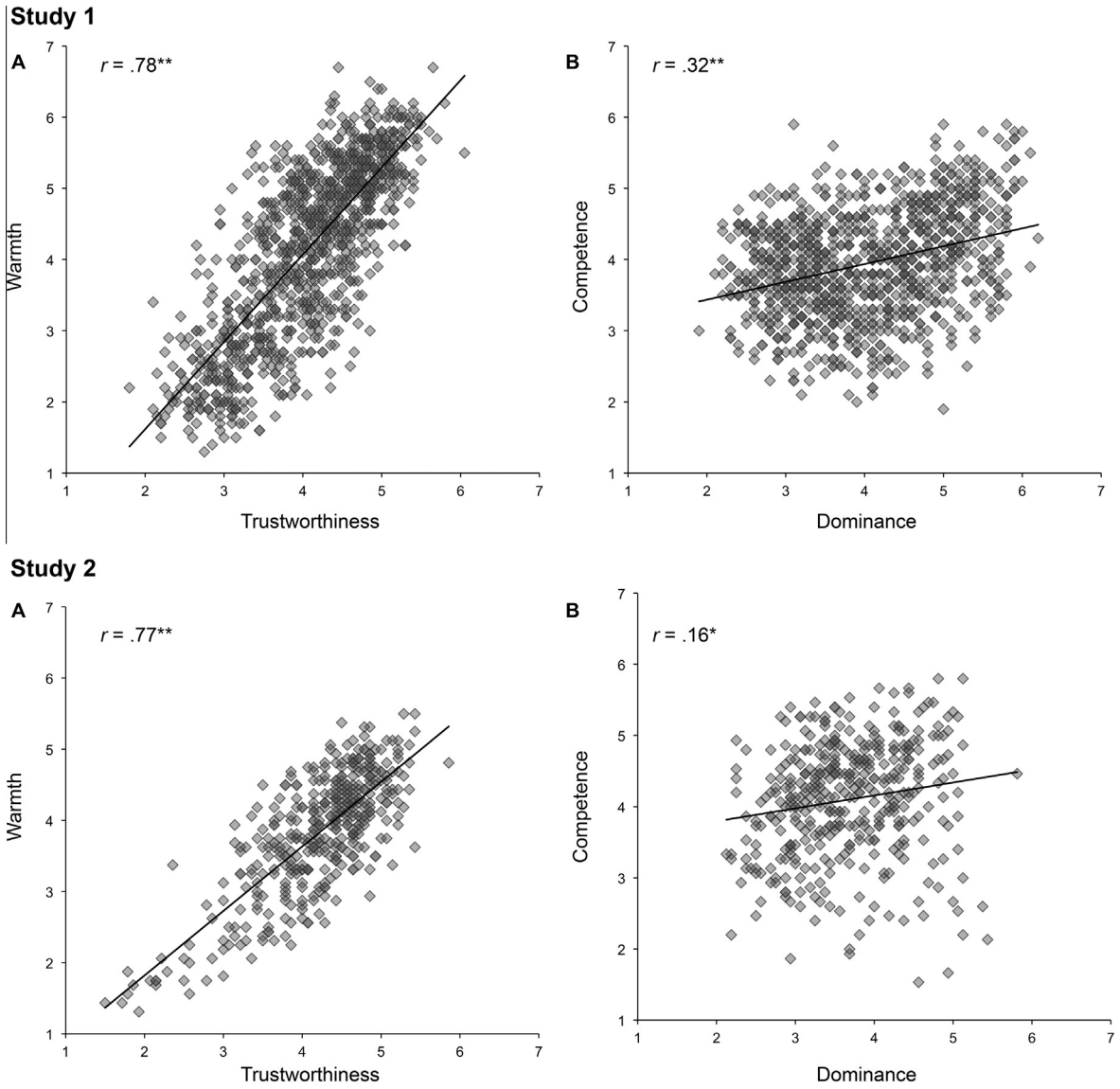

Fig. 2. The correlation between (A) average trustworthiness and warmth judgements; and (B) average competence and dominance judgements, separately in Study 1 and Study 2. $^{**}p < 0.001$, $^{*}p < 0.05$. Each point represents a single face image.

Table 1
Principal axis factor analysis: structure matrices (direct oblimin rotation). These can be interpreted as akin to correlations between the factors and variables. Trait loadings above 0.3 are highlighted in bold.

	Four factor solution				Three factor solution		
	F1	F2	F3	F4	F1	F2	F3
Smiling	**0.96**	−0.15	−0.10	−0.15	**0.95**	0.13	0.10
Warmth	**0.96**	−0.05	−0.05	−0.21	**0.95**	0.05	0.17
Pleasantness of expression	**0.96**	−0.22	−0.08	−0.21	**0.95**	0.19	0.16
Approachability	**0.93**	−0.12	−0.08	**−0.32**	**0.93**	0.15	0.25
Aggressiveness	**−0.89**	0.07	**0.35**	0.21	**−0.92**	−0.20	−0.01
Arousal of expression	**0.88**	−0.16	0.04	−0.18	**0.85**	0.10	0.20
Trustworthiness	**0.80**	−0.13	**−0.37**	**−0.48**	**0.84**	**0.33**	0.22
Attractiveness	0.26	**−0.75**	−0.23	**−0.68**	**0.33**	**0.85**	**0.47**
Health	0.24	**−0.77**	−0.06	**−0.66**	0.27	**0.77**	**0.53**
Age	0.07	**0.89**	**0.36**	−0.08	0.01	**−0.73**	0.22
Masculinity	−0.15	**0.31**	**0.86**	0.01	−0.26	**−0.55**	**0.40**
Babyfacedness	0.19	**−0.50**	−0.19	−0.01	0.21	**0.43**	−0.08
Facial adiposity	0.10	0.23	0.18	0.23	0.05	**−0.32**	−0.09
Competence	0.19	−0.05	0.21	**−0.85**	0.20	0.17	**0.79**
Dominance	−0.24	**0.33**	**0.93**	−0.22	**−0.35**	**−0.54**	**0.62**
Intelligence	0.27	0.04	0.09	**−0.72**	0.29	0.11	**0.61**
Confidence	**0.55**	**−0.30**	0.10	**−0.60**	**0.55**	**0.32**	**0.57**
Skintone (tanned)	0.14	−0.05	0.22	**−0.31**	0.12	0.02	**0.38**

included for their theoretical importance across the field of facial impressions (Oosterhof & Todorov, 2008; Santos & Young, 2005; Sutherland et al., 2013; Vernon et al., 2014; Zebrowitz, 2005). Bartlett's test of sphericity indicated that the correlations were large enough that a factor analysis was appropriate: X^2 (153) = 19,424, $p < 0.001$; and Kaiser's criterion and scree test, a parallel analysis and minimum average partial analysis were carried out in order to determine the number of factors (as in Sutherland et al., 2013). Kaiser's criterion and the MAP test returned five factors (the fifth factor was not stable), the parallel analysis returned four factors and the scree test indicated that two to four factors were present.

In both four and three factor models, the first factor seemed to index approachability (see Table 1). Both trustworthiness and warmth loaded strongly on this approachability factor, which clearly approximated the first dimension from both facial (trustworthiness) and social group (warmth) models. However, in the four-factor model, (in)competence and dominance formed separate factors (see Table 1). In the three-factor model, the third factor appeared to be competence. Although dominance also loaded on competence in the three-factor model, it also loaded highly on the third (youthful-attractiveness) factor in this solution (see Table 1). Thus, our finding of a greater separation between dominance and competence than warmth and trustworthiness remained when other theoretically important traits were also examined.

2.2.2. Face gender

Finally, we examined these relationships for male and female faces separately, since dominance is highly sexually dimorphic (Oosterhof & Todorov, 2008) and since previous studies have found that dominance is evaluated differently for men and women (e.g. Rudman & Glick, 2001; Sutherland, Young, Mootz, & Oldmeadow, 2015). We wanted to ascertain that our current finding of a low correlation between competence and dominance was not an artefact introduced by correlating across distinct populations (i.e. male and female faces). Interestingly, we found that dominance judgments were more related to competence judgments for male faces ($r_{male} = 0.50$, $p < 0.001$, $n = 500$) compared to female faces ($r_{female} = 0.12$, $p < 0.01$, $n = 500$; significant difference: $Z = 6.63$, $p < 0.001$). Trustworthiness and warmth were equally highly related for male and female faces ($r_{male} = 0.82$, $p < 0.001$; $r_{female} = 0.81$, $p < 0.001$; difference $Z = 0.23$, $p = 0.82$). Crucially, for both male and female faces separately, trustworthiness and warmth were clearly more related than dominance and competence (both ZPF > 8.87, $p < 0.001$).

2.3. Discussion 1

In Study 1 we found that while trustworthiness and warmth judgements were highly related, dominance and competence judgements were only moderately related. This pattern was affected by face sex, so that the relationship between dominance and competence was especially weak for female faces. Moreover, this separation between dominance and competence remained when other traits were included in a factor analytic approach. This pattern is consistent with the idea that the first dimension described in facial first impressions models (e.g. trustworthiness: Oosterhof & Todorov, 2008) is essentially the same as the first dimension found in social psychology models of inter-group perception (e.g. warmth: Fiske et al., 2007). However, our findings show that the second, dominance dimension found in face perception models (Oosterhof & Todorov, 2008) is not identical to the competence dimension found in social psychology (e.g. Fiske et al., 2007; see the General discussion).

In Study 2, we decided to replicate the findings using a different independently collected sample of 400 ambient images and a different group of participants. As for Study 1, we collected ratings of warmth, trustworthiness, dominance and competence from separate groups of participants.

3. Study 2

3.1. Stimuli 2

The stimuli used in Study 2 were a set of 400 highly varied "ambient image" face photographs (200 male faces, 200 female) taken from a public database (Bainbridge, Isola, Blank, & Oliva, 2012; Bainbridge, Isola, & Oliva, 2013; available at http://wilmabainbridge.com/facememorability2.html). We chose to use another ambient image database to give us an independent face sample that was as large as possible, so that we had the best possibility of finding relationships between the traits from different models. Four hundred faces were chosen as the largest feasible face sample, given that the study was run online. We randomly sampled faces from the overall database of 2,000 face images which had been previously categorised on face sex, age, ethnicity and expression (see Bainbridge et al., 2012 for details) with the constraints that half of the sample faces were male, none were celebrities, and all were Caucasian, since race biases were not the focus of this study. We also ensured that we had a similar age range and range of expressions for male and female faces (ratings taken from Bainbridge et al., 2012), since the female faces in this database appeared less varied than the male faces.

3.2. Participants and procedure 2

Sixty-one participants (35 female; mean age = 23.7 years, SD age = 3.0) were recruited through an academic crowdsourcing website (Prolific Academic: http://www.prolific.ac/) and tested online using Qualtrics (2015, www.qualtrics.com). Thirteen additional participants initially started the experiment but either dropped out (six) or were otherwise excluded (two were not native British, four could not complete the task since we blocked mobile responding, and one pressed the same button to all faces). Participants were all young adults (18–30), native British, Caucasian and located within the UK. Participants took approximately 40 min to complete this task and received a small remuneration.

Participants either rated the faces on their trustworthiness ($n = 14$), warmth ($n = 16$), dominance ($n = 16$) or competence ($n = 15$) on a 1–7 scale (not at all to very trustworthy, warm, dominant or competent). Before the actual experiment, participants rated 10 other faces that were randomly pulled from the database as a practice. Faces were viewed in a random order, and all other aspects were as Study 1.

3.3. Results & discussion 2

The inter-rater reliabilities were good (trustworthiness: $\alpha = 0.83$, warmth: $\alpha = 0.90$, dominance: $\alpha = 0.81$ and competence: $\alpha = 0.87$) so we averaged across individual participant responses. As in Study 1, we correlated the warmth and trustworthiness ratings and found that this correlation was substantial and highly significant: $r = 0.77$, $p < 0.001$, $n = 400$. We then correlated the dominance and competence ratings and again this correlation was substantially lower, although still significant: $r = 0.16$, $p < 0.005$, $n = 400$. Again, the size of the correlation between trustworthiness and warmth was significantly greater than the size of the correlation for dominance and competence: $ZPF = 10.62$, $p < 0.001$, $n = 400$. We thus replicated the finding that

trustworthiness and warmth facial judgments are indeed highly similar while dominance and competence facial judgments are less strongly related (see Fig. 2).

As in Study 1, we also examined the male and female faces separately, to ensure that our results were not due to face gender. As for Study 1, the correlation between dominance and competence ($r_{male} = 0.23$, $r_{female} = 0.21$, both $p < 0.005$) was significantly lower than the correlation between trustworthiness and warmth ($r_{male} = 0.80$, $r_{female} = 0.75$, both $p < 0.001$) for both male and female faces (both $ZPF > 6.88$, both $p < 0.001$, both $n = 200$). Unlike in Study 1, however, the correlation for male faces for dominance and competence was now as low as for female faces: $Z = 0.21$, $p = 0.834$.

Study 2 results again suggest that the second dimension of facial and social models (Cuddy et al., 2008; Oosterhof & Todorov, 2008) is not identical, since dominance and competence seem to be judged differently when impressions of these traits are made from faces. In Study 2 this effect was even more marked and extended to both male and female faces; this is perhaps due to sampling differences, since in the second stimuli set we carefully controlled the facial expression and age of the faces across face sex.

4. Study 3

Since Study 1 and 2 were focused on relationships between ratings at the level of the faces themselves, different participants rated the different social traits. This was done to avoid carryover effects (i.e. where correlations between social judgments are inflated due to the participants being influenced by their previous judgments: Rhodes, 2006). However, this leaves open the question of whether this finding would also remain at the individual level, when the same participants made these judgments. In Study 3 we thus attempted to ascertain whether our previous findings were robust at the individual participant level. Based on Studies 1 and 2, we expected to find higher agreement (correlations) between warmth and trustworthiness judgements than competence and dominance judgements, now when the same participants judged all traits.

4.1. Methods 3

Twenty-four participants (12 female; mean age: 20.6 years, SD age = 2.5) volunteered to take part in the second study in return for course credit or a small remuneration. Two additional participants only took part in the first session: their data were excluded. Participants provided informed consent to procedures that were approved by the ethics committee of the University of York Psychology Department. All participants were Caucasian and self-identified as culturally Western. Participants did not take part in the other currently reported experiments.

4.1.1. Stimuli & procedure 3

Two hundred face stimuli (100 female) were chosen at random from the original 1,000-strong ambient image dataset (Sutherland et al., 2013). Participants were tested in a quiet room on a PC running E-Prime software (version 2; Psychology Software Tools, Pittsburgh, USA). Participants rated all 200 faces on 12 different traits, each in a separate block with 24 practice faces (not analysed) presented before each new block (order of blocks counterbalanced across participants). We included trustworthiness, warmth, dominance and competence, as key traits relevant to the current paper. Participants also rated the faces on a number of other traits, as part of a different study, which was not analysed here.

4.2. Results 3

4.2.1. Reliabilities

Reliability coefficients (Cronbach's alpha) across all participants were high (trustworthiness: 0.92, warmth: 0.96, dominance: 0.91, and competence: 0.89). We then sought to determine whether the data would replicate the finding in Study 1 and 2 that warmth/trustworthiness judgements were more similar than dominance/competence judgments. Since participants rated the same faces several times, we could now correlate participants' ratings on one trait with their rating on another and then compare the strength of these correlations at the individual level. Correlations were transformed using Fisher's r-to-z transform before analysis.

4.2.2. Correlations between dimensions

Averaged individual participants' correlations are shown in Fig. 3. There was a significant main effect of dimension, so that traits on the proposed first dimension (trustworthiness with warmth) were more correlated (average $r = 0.53$) than traits on the proposed second dimension (dominance with competence: average $r = 0.19$) on average across participants: $F(1,22) = 34.43$, $p < 0.001$, $\eta_p^2 = 0.61$. There was no main effect of participant gender, face gender or any interactions between these factors: all $F(1,22) \leqslant 2.93$, all $p \geqslant 0.10$, all $\eta_p^2 \leqslant 0.12$. The effect of dimension also held at the aggregate level, as in the previous studies (trustworthiness and warmth: $r = 0.88$, dominance and competence: $r = 0.33$, $ZPF = 8.69$, $p < 0.001$). Thus, as in Study 1 and 2, dominance is less highly correlated with competence than warmth is with trustworthiness, and this pattern is reliable across individual participants as well as at the aggregate level (see Fig. 3).

4.3. Discussion 3

Study 3 again found a greater separation between dominance and competence than trustworthiness and warmth, now at the individual perceiver level. The results of Study 3 indicate that our finding of a separation between dominance and competence was not an artefact of averaging judgments at the group level, but held when the same individuals made these impressions on the same faces. In Study 4, we tested the similarity of the dimensions using an experimental manipulation of these traits and a direct measure of similarity.

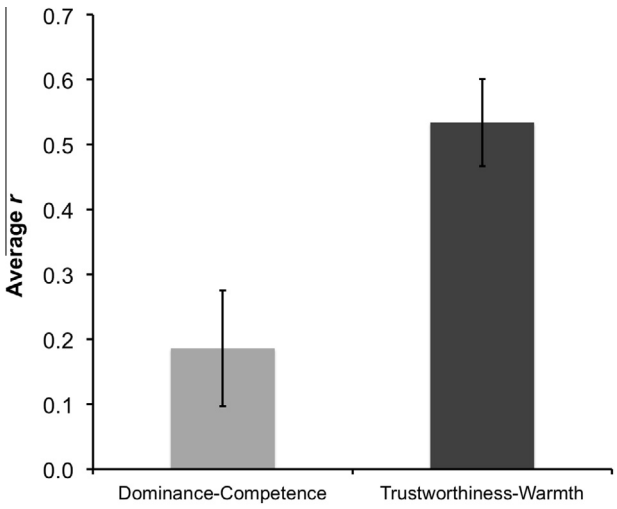

Fig. 3. Average correlations between individual participants' trustworthiness and warmth ratings, and between their dominance and competence ratings. Error bars are 95% CIs.

5. Study 4

In Study 4, we asked participants to evaluate the similarity between pairs of average images of faces that were either high or low in warmth and trustworthiness, or high or low in dominance and competence. In this way, we could directly examine how similar warmth and trustworthiness were perceived to be, and compare this similarity to dominance and competence. Using average face images also offered a well-controlled method of examining these differences, since only cues to trait judgements which are consistent across the individual faces remain in the face averages.

We sought to use this novel experimental method to extend the main findings of the previous three studies. We predicted that the dominance and competence average images would be perceived as less similar to each other than the warmth and trustworthiness average images. We also predicted that the high and low warmth faces would also be perceived as high and low on trustworthiness (and vice versa), while the high and low dominance and competence faces would not change as much on the equivalent trait.

5.1. Methods 4

Twenty-four participants (12 male; mean age: 23.2 years, SD age = 10.9) volunteered to take part in the third study in return for course credit or a small remuneration. Participants provided electronic informed consent to procedures that were approved by the ethics committee of the University of York Psychology Department. The experiment was hosted online by Qualtrics (www.qualtrics.com) although participants were recruited through the University. All participants were Caucasian and self-identified as culturally Western. Participants did not take part in the other currently reported experiments.

5.1.1. Stimuli & procedure 4

Stimuli were created using Psychomorph (version 5, Tiddeman, Burt, & Perrett, 2001) by averaging the 20 male and the 20 female faces rated highest and lowest in Study 1 on warmth, trustworthiness, competence and dominance (see Sutherland et al., 2013 for a full description of averaging procedures). The stimuli are depicted in Fig. 4. Stimuli were cropped around the face and aligned so that their eyes were horizontal in order to stop head tilt affecting the similarity judgements, since the pairs were to be presented side by side. We kept the gender of the face consistent within a pair to avoid people using this as a cue. As in Sutherland et al. (2013), nine faces were removed either because closer inspection they seemed to be of another race, a celebrity, or they were very difficult to delineate due to head pose. By averaging across individual exemplars, the resulting face-like average images (referred to as 'faces' for simplicity) should represent the consistent cues underlying the perception of these traits. We allowed the original face images to covary across the averaged faces, to pick up on the natural overlap between facial cues without introducing researcher bias. Overall, this method allowed us to examine the four judgements directly and with different stimuli than those used in the previous studies.

In a first block of trials, pairs of averaged faces were simultaneously presented and participants rated how similar the faces in each pair were (1 not very similar - 7 very similar). Eight face pairs were used, which contrasted either dominance and competence average images or warmth and trustworthiness average images, for a given gender and for either a high or low level of trait. For example, the low warmth male average image was compared to the low trustworthiness male average image, and so forth. Pairs were rated for their similarity three times, with the first set of ratings as practice trials (not analysed). Within the experiment, the order of the side of presentation of each of the average images in a pair was counterbalanced across trials.

In second and third blocks, each average image face was rated along the two model dimensions, with half of the participants rating the faces on the face perception model dimensions (trustworthiness and dominance, order counterbalanced) and the other half on the social group model dimensions (warmth and competence, order counterbalanced). Again, participants rated the faces three times, with the first time as a practice (not analysed). These additional ratings were collected to confirm that the average faces were indeed perceived as high or low on the manipulated trait, so that any lack of similarity between pairs of average faces could not be explained by a failure of the stimuli to represent these characteristics adequately. The ratings were also used to establish to what extent the faces were also rated as high or low on the parallel (but not manipulated) trait, as a second independent test of our hypothesis.

5.2. Results 4

5.2.1. Similarity ratings

The second and third similarity ratings were averaged together. A three-way ANOVA was run ($n = 24$) on the similarity ratings (at the level of the participants) with the dimension of the face (2 levels: trustworthiness/warmth, or dominance/competence), the dimensional position of the face (2 levels: high or low) and the gender of the face pair (female or male) as within-subjects factors. The three-way interaction was significant: $F(1,23) = 7.91$, $p = 0.010$, $\eta_p^2 = 0.26$; and there were significant two-way interactions between the gender of the face and the dimension: $F(1,23) = 13.33$, $p < 0.001$, $\eta_p^2 = 0.37$; as well as between the dimension and the position of the face: $F(1,23) = 76.75$, $p < 0.001$, $\eta_p^2 = 0.77$. Importantly, as predicted, there was a main effect of the dimension of the face, so that the trustworthiness/warmth pairs were rated as significantly more similar than the dominance/competence pairs: $F(1,23) = 217.13$, $p < 0.001$, $\eta_p^2 = 0.90$ (see Fig. 4). There was also a main effect of the position of the face pairs, so that face pairs high on a trait were rated as more similar than low face pairs: $F(1,23) = 8.68$, $p < 0.01$, $\eta_p^2 = 0.27$; and a main effect of face gender, so that male average face pairs were rated as more similar than female pairs: $F(1,23) = 19.44$, $p < 0.001$, $\eta_p^2 = 0.46$ (see Fig. 4).

Planned comparisons examined the similarity of average faces on each proposed dimension separately. For trustworthiness/warmth average face pairs, there was no main effect of face gender: $F(1,23) = 0.82$, $p = 0.373$, $\eta_p^2 = 0.04$, and face gender did not interact with the position of the face: $F(1,23) = 2.28$, $p = 0.144$, $\eta_p^2 = 0.09$. There was only a main effect of the position of the face, so that low average warmth or trustworthiness faces were seen as more similar than high average warmth or trustworthiness faces: $F(1,23) = 6.87$, $p = 0.015$, $\eta_p^2 = 0.23$. However, for the dominance/competence average face pairs there was a main effect of face gender; so that female competence and dominance face average pairs were seen as significantly less similar than male competence and dominance pairs: $F(1,23) = 31.49$, $p < 0.001$, $\eta_p^2 = 0.58$. This was mediated by a significant interaction between face gender and the position of the average image: $F(1,23) = 7.36$, $p = 0.012$, $\eta_p^2 = 0.24$. The low competence/dominance female average image pairs were rated as significantly less similar than the male average pairs: $t(23) = 7.52$, $p < 0.001$, $d = 1.54$ (see Fig. 4); but the high competence/dominance male and female pairs did not significantly differ: $t(23) = 1.48$, $p = 0.152$, $d = 0.30$ (see Fig. 4). There was also a main effect of the dimensional position of the face, so that high

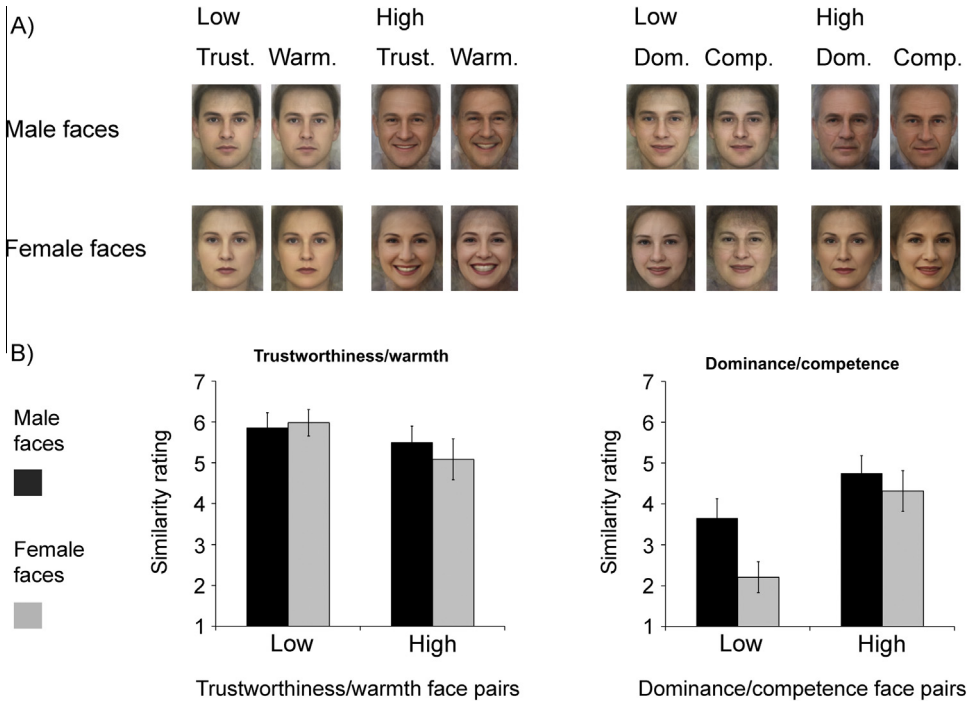

Fig. 4. (A) Pairs of male and female face averages made from the 20 faces rated highest and lowest for trustworthiness and warmth; or dominance and competence. (B) Similarity ratings (1–7) for trustworthiness and warmth pairs; and dominance and competence pairs of male and female face averages. Error bars are 95% CIs.

face dominance and competence averages were seen as more similar than low pairs: $F(1,23) = 83.61$, $p < 0.001$, $\eta_p^2 = 0.78$.

Finally, as expected, participant sex did not significantly interact or show main effects when entered as between-participant factor: all $p > 0.10$, all $\eta_p^2 < 0.12$. The theoretically interesting three-way interaction between dimension, face gender and position, the two-way interaction between dimension and face gender, and the main effect of dimension were all still significant (all $p < 0.01$, all $\eta_p^2 \geq 0.27$).

5.2.2. Trait ratings

After discarding the first ratings as practice trials, the second and third trait ratings were averaged together for each participant, rated trait and stimulus face. The trait ratings showed that the pairs of averaged images (high and low, within gender and trait) differed on the manipulated trait, as expected ($n = 12$; all $p \leqslant 0.011$; see Table 2). All were still significant after Bonferroni correction except the comparison between the trustworthiness high and low male faces, although the means were in the expected direction. Importantly, the male and female dominance and competence trait face averages were rated as expected on these traits; thus our results on the similarity measure were not merely due to a failure of the average images to represent the manipulated trait dimension.

Finally, we also used the trait ratings as an additional method to test our overall question about the overlap in the dimensions. Specifically, we examined how strongly the high and low warmth face averages (within face gender) also differed on perceptions of trustworthiness, and vice versa; and compared this to the equivalent measure for dominance and competence.

This analysis revealed an interaction between the dimension and the trait level as predicted. The faces manipulated on their warmth and trustworthiness were also perceived as differing on their trustworthiness and warmth respectively (mean high and low difference: 2.63), and this correspondence was greater than for dominance with competence (mean high and low difference: 1.68; two-way interaction: $F(1,23) = 5.37$, $p = 0.03$, $\eta_p^2 = 0.19$). The three-way interaction with face gender was not significant: $F(1,23) = 1.65$, $p = 0.21$, $\eta_p^2 = 0.07$; however, again the effect of the impression dimension on the difference between high and low faces was only significant for female faces: $t(23) = 2.48$, $p = 0.021$, $d = 0.51$, mean dimension difference = 1.25, but not for male faces: $t(23) = 1.48$, $p = 0.15$, $d = 0.30$, mean dimension difference = 0.65.

5.3. Discussion 4

Study 4 found that the dominance and competence face averages were perceived as less similar to each other than the warmth

Table 2
Mean trait ratings (and standard deviations) for high and low male or female average faces, for each manipulated trait.

	Face gender	High average face	Low average face	High vs. low comparison	Effect size
Trustworthiness faces/rating	Female	5.92 (0.87)	3.67 (1.48)	$p < 0.001$	$d = 1.28$
	Male	5.29 (1.18)	3.38 (1.65)	$p = 0.011$	$d = 0.88$
Warmth faces/rating	Female	6.04 (1.08)	2.42 (1.51)	$p < 0.001$	$d = 1.68$
	Male	6.08 (1.08)	1.96 (0.96)	$p < 0.001$	$d = 2.34$
Dominance faces/rating	Female	5.50 (0.83)	2.50 (1.02)	$p < 0.001$	$d = 2.45$
	Male	5.33 (0.54)	3.17 (0.91)	$p < 0.001$	$d = 1.69$
Competence faces/rating	Female	5.42 (0.73)	3.46 (1.10)	$p < 0.001$	$d = 1.59$
	Male	5.71 (1.01)	3.67 (0.98)	$p < 0.001$	$d = 2.29$

and trustworthiness face averages. This replicates the previous studies, using a more direct measure of similarity. We also found that this difference between the dimensions was greater for female faces than male faces, as in Study 1. This pattern was qualified by a three-way interaction with dimensional position so that the low dominance and competence female faces were seen as significantly less similar to each other than the low dominance and competence male faces. Finally, manipulating the faces to be high or low on warmth or trustworthiness was also more successful at changing perceptions of the proposed equivalent trait (i.e. trustworthiness or warmth respectively), than was the case for dominance and competence.

6. General discussion

Our principal aim was to explore the parallel between trustworthiness and dominance dimensions identified in facial first impressions research (Oosterhof & Todorov, 2008) and warmth and competence dimensions found in studies of social group perception (Fiske et al., 2007). In Study 1, using trait ratings and face photographs, we found that while trustworthiness and warmth ratings of 1,000 ambient face images are highly related, dominance and competence ratings of these face images were less strongly related, especially for female faces. Moreover, a factor analysis demonstrated that dominance and competence tended to lie on different factors. In Study 2, we replicated the overall difference in the dimensions for both male and female faces, using a different sample of 400 ambient image faces and a larger sample of participants, while in Study 3, we extended this finding to show that it held at the individual participant level. Finally, in Study 4, we used carefully controlled face averages and found that dominance and competence pairs of averaged face images were also rated as less similar than warmth and trustworthiness pairs (especially for female faces). We found the same pattern when the participants also rated the face averages individually on the social traits.

Together, these findings support the claim that the first dimension found in facial first impressions research (trustworthiness: Oosterhof & Todorov, 2008) and the first dimension found in social psychology studies (warmth: Fiske et al., 2007) are very similar for judgements made from faces. However, our current findings indicate that the second dimension found in studies of face perception (dominance: Oosterhof & Todorov, 2008) is not identical to the second dimension found in research on social groups (competence: Fiske et al., 2007).

6.1. Dominance and competence dimensions of person perception

There are two related ways to resolve the current findings with the suggestion that the dimensions underlying social group and facial first impression models are similar (Montoya & Horton, 2014; Sutherland et al., 2013; Todorov et al., 2008). First, it is worth stressing that the similarity between the face perception and social psychological models also rests on their theoretical underpinnings. Specifically, both Oosterhof and Todorov (2008) and Fiske and colleagues (Fiske et al., 2007) relate the first dimension to a person or a group's intentions, with the second dimension being their capability to carry out their intentions. Therefore, it is possible that the cues underlying capability could differ between abstract (e.g. verbal) or facial stimuli. Potentially, faces offer more cues to physical capability (i.e. dominance) rather than social competence or status, as would be the case for group labels (Fiske, Cuddy, Glick, & Xu, 2002). Contextual differences could also explain the finding in Studies 1 and 4 that competence and dominance were even less related for female faces (although we note that this face gender difference was slight, and was not found in Studies 2 or 3). Capability appraisal based on physical dominance may simply be less useful in judging female targets, given that cues to dominance include masculinity and physical strength (Oosterhof & Todorov, 2008) and that perceivers are inaccurate at identifying physical strength from female faces (Sell et al., 2009). This first account therefore resolves the differences we noted between face and social models by highlighting the models' similar theoretical focus on intentions and capability, and then emphasising that the specific cues to capability might vary depending on the stimuli (e.g. verbal or visual), or target group (e.g. male or female).

A second (related) explanation, which we at present prefer, is that dominance and competence may reflect different routes to achieving the capability to be able to help or harm the observer. This suggestion is based on a theoretical literature on the evolutionary development of power and status, which has demonstrated that high status in humans can be derived through prestige (due to competence) or dominance, depending on which type of behaviour is visible and rewarded in a given context (Cheng et al., 2013). This divergence between competence and dominance as different routes to status is not only found in Western, industrialised nations, but also in the Tsimane people who live in isolation in the Bolivian jungle (Reyes-Garcia et al., 2008). Both aspects of status are hypothesised to have a long evolutionary history (see Cheng et al., 2013; Henrich & Gil-White, 2001 for reviews). Crucially, other social psychological studies that examine impression formation, using verbally described targets, have found that judgments of the dominance (i.e. potency, power) and competence of these abstract targets also diverge (Carrier et al., 2015; Kervyn et al., 2013); more so than traits on the warmth dimension (Kervyn et al., 2013). This line of work on verbally presented stimuli agrees with our current findings with impressions of real faces.

A theoretical analysis of the underpinning of these impression formation dimensions would also predict the current separation between dominance and competence. Clearly, what makes someone competent in a given situation depends on the task at hand. This suggestion is supported by other studies of facial judgments, so that perceived facial competence but not dominance predicts the success of a target in political contexts (Chiao, Bowman, & Gill, 2008), whereas perceived facial dominance predicts success in the military (Mueller & Mazur, 1996). In some situations, competence may even rely on being sociable or trustworthy. We think that this distinction can be resolved with the dimensional approach by suggesting that while a broad dimension of capability clearly underlies facial impressions, the context will determine the specific form this dimension takes (i.e. as competence or dominance), as well as how differentiated these traits are from each other, and even from traits on the warmth dimension. In the current studies, which like much of the research on first impressions had no specified context, competence seemed to best reflect the capability dimension, since competence formed the largest and most distinct contribution to the third factor. We also note that participants do not spontaneously mention dominance when freely describing their first impressions of faces without a context, but they do mention capability or intelligence (Oosterhof & Todorov, 2008; Sutherland, Liu, et al., 2015).

6.2. Trustworthiness and warmth dimensions

In the current study we found that trustworthiness and warmth judgements made from faces were highly related (Studies 1–4) and that they clearly loaded on the same factor (Study 1). This pattern agrees with a number of other similar findings for abstract judgements of these traits (Abele & Wojciszke, 2007; Kervyn et al., 2013). However, the current results do not fit with other studies that find that trustworthiness and warmth judgements can be dissociated in extra-facial person perception (Brambilla, Rusconi,

Sacchi, & Cherubini, 2011; Goodwin, Piazza, & Rozin, 2014; Leach, Ellemers, & Barreto, 2007). The discrepancy might occur because the current study used facial stimuli. Potentially, people might rely on a more rudimentary approach/avoid judgement for faces, based on salient cues, since they lack information needed to disambiguate subtle differences. Alternatively, specific contextual manipulations might find situations where facial trustworthiness and warmth judgements diverge, since they can be conceptually dissociated (c.f. Goodwin et al., 2014).

Nevertheless, we still suggest that the second dimension (competence or dominance) will be more variable than the first dimension (warmth or trustworthiness), across a range of contexts. This hypothesis is based on our current results and from our theoretical analysis of the functional basis of these dimensions. In particular, we predict that cues underlying competence judgements would differ in a top-down way depending on competencies perceived as necessary in a given context, based on studies showing that perceivers use facial cues that they expect to be most relevant (Oldmeadow, Sutherland, & Young, 2013; Olivola & Todorov, 2010). We are currently testing this hypothesis by examining judgments across cultural contexts (Sutherland, Liu, et al., 2015). While dimensional approaches in facial impressions have been a fruitful way to understand a broad range of social judgments, we hope that the current results and theorising encourages future work to further examine the relationships between these broad factors with judgments made in a specific context. In particular, we suggest that the next stage of theoretical development of models of facial impressions needs to explicitly address the context in which the face is perceived, by developing a program of research that quantifies everyday contexts in terms of their implications for facial impression formation.

7. Conclusions

To the best of our knowledge, this is the first set of studies to directly test the equivalency of facial (trustworthiness and dominance) and social group (warmth and competence) dimensions in terms of first impressions of faces. We found that while trustworthiness and warmth dimensions were highly similar, dominance and competence dimensions were not so closely related. The current results therefore highlight interesting differences between social psychological models of extra-facial person perception, and facial models of first impressions. Our suggestion is that these differences can be resolved by examining the functional basis of dimensions of facial impression and social stereotype models, while acknowledging that the specific traits and cues used to form these dimensions will vary depending on their utility for the context at hand. More broadly, the current results highlight the benefit of integrating models of face perception based on visual stimuli, with social psychological theories that attempt to understand our conceptual knowledge of people and groups.

Acknowledgements

The author(s) declared no potential conflicts of interests. The first author received the following financial support: an ESRC studentship [ES/I900748/1] and postdoctoral research support from the Australian Research Council Centre of Excellence in Cognition and its Disorders, University of Western Australia [CE110001021]. The funding sources had no influence on the research. Parts of this paper are based on portions of the first author's doctoral thesis at the Department of Psychology, York, UK. We thank Alex Koch as well as three anonymous reviewers for their insightful comments.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.09.006.

References

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93*(5), 751–763. http://dx.doi.org/10.1037/0022-3514.93.5.751.

Anderson, C., & Kilduff, G. J. (2009). The pursuit of status in social groups. *Current Directions in Psychological Science, 18*(5), 295–298. http://dx.doi.org/10.1111/j.1467-8721.2009.01655.x.

Bainbridge, W. A., Isola, P., Blank, I., & Oliva, A. (2012). Establishing a database for studying human face photograph memory. In *Proceedings of the 34th annual meeting cognitibe science society* (pp. 1302–1307).

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General, 142*(4), 1323. http://dx.doi.org/10.1037/a0033872.

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*(2), 135–143. http://dx.doi.org/10.1002/ejsp.744.

Bruce, V., & Young, A. W. (2012). *Face perception*. London; New York: Psychology Press.

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology, 102*, 943–958. http://dx.doi.org/10.1111/j.2044-8295.2011.02039.

Buss, D. M., & Duntley, J. D. (2013). The evolution of aggression. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology*. New York: Psychology Press.

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

Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology, 104*(1), 103. http://dx.doi.org/10.1037/a0030398.

Chiao, J. Y., Bowman, N. E., & Gill, H. (2008). The political gender gap: Gender bias in facial inferences that predict voting behavior. *PLoS ONE, 3*(10), e3666. http://dx.doi.org/10.1371/journal.pone.0003666.

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. http://dx.doi.org/10.1016/S0065-2601(07)00002-0.

Dotsch, R., & Todorov, A. (2011). Reverse correlating social face perception. *Social Psychological and Personality Science, 3*(5), 562–571. http://dx.doi.org/10.1177/1948550611430272.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77–83. http://dx.doi.org/10.1016/j.tics.2006.11.005.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902. http://dx.doi.org/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*(1), 148–168. http://dx.doi.org/10.1037/a0034726.

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior, 22*(3), 165–196. http://dx.doi.org/10.1016/S1090-5138(00)00071-4.

Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in Psychology, 4*(386), 1–8. http://dx.doi.org/10.3389/fpsyg.2013.00386.

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

Kervyn, N., Fiske, S. T., & Yzerbyt, V. Y. (2013). Integrating the stereotype content model (warmth and competence) and the Osgood semantic differential (evaluation, potency, and activity). *European Journal of Social Psychology, 43*(7), 673–681. http://dx.doi.org/10.1002/ejsp.1978.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology, 93*(2), 234–249. http://dx.doi.org/10.1037/0022-3514.93.2.234.

Montoya, R. M., & Horton, R. S. (2014). A two-dimensional model for the study of interpersonal attraction. *Personality and Social Psychology Review, 18*(1), 59–86. http://dx.doi.org/10.1177/1088868313501887.

Mueller, U., & Mazur, A. (1996). Facial dominance of West Point cadets as a predictor of later military rank. *Social Forces, 74*(3), 823–850. http://dx.doi.org/10.1093/sf/74.3.823.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
Oldmeadow, J. A., Sutherland, C. A. M., & Young, A. W. (2013). Facial stereotype visualization through image averaging. *Social Psychological and Personality Science, 4*(5), 615–623. http://dx.doi.org/10.1177/1948550612469820.
Olivola, C. Y., & Todorov, A. (2010). Elected in 100 ms: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior, 34*(2), 83–110. http://dx.doi.org/10.1007/s10919-009-0082-1.
Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *PNAS, 105*(32), 11087–11092. http://dx.doi.org/10.1073/pnas.0805664105.
Quinn, K. A., & Macrae, C. N. (2011). The face and person perception: Insights from social cognition. *British Journal of Psychology, 102*(4), 849–867. http://dx.doi.org/10.1111/j.2044-8295.2011.02030.x.
Raghunathan, T. E., Rosenthal, R., & Rubin, D. B. (1996). Comparing correlated but nonoverlapping correlations. *Psychological Methods, 1*(2), 178. http://dx.doi.org/10.1037/1082-989X.1.2.178.
Reyes-Garcia, V., Molina, J. L., Broesch, J., Calvet, L., Huanca, T., Saus, J., et al. (2008). Do the aged and knowledgeable men enjoy more prestige? A test of predictions from the prestige-bias model of cultural transmission. *Evolution and Human Behavior, 29*(4), 275–281. http://dx.doi.org/10.1016/j.evolhumbehav.2008.02.002.
Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57*(1), 199–226. http://dx.doi.org/10.1146/annurev.psych.57.102904.190208.
Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9*(4), 283–294. http://dx.doi.org/10.1037/h0026086.
Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57*(4), 743–762. http://dx.doi.org/10.1111/0022-4537.00239.
Rule, N. O., Ambady, N., & Adams, R. B. Jr., (2009). Personality in perspective: Judgmental consistency across orientations of the face. *Perception, 38*(11), 1688–1699. http://dx.doi.org/10.1068/p6384.
Santos, I. M., & Young, A. W. (2005). Exploring the perception of social characteristics in faces using the isolation effect. *Visual Cognition, 12*(1), 213–247. http://dx.doi.org/10.1080/13506280444000102.
Santos, I. M., & Young, A. W. (2008). Effects of inversion and negation on social inferences from faces. *Perception, 37*(7), 1061–1078. http://dx.doi.org/10.1068/p5278.
Santos, I. M., & Young, A. W. (2011). Inferring social attributes from different face regions: Evidence for holistic processing. *The Quarterly Journal of Experimental Psychology, 64*(4), 751–766. http://dx.doi.org/10.1080/17470218.2010.519779.
Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., & Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings of the Royal Society B: Biological Sciences, 276*(1656), 575–584. http://dx.doi.org/10.1098/rspb.2008.1177.
Sutherland, C. A. M., Liu, X., Chu, Y., Zhang, L., Oldmeadow, J., & Young, A. (2015). Chinese perceivers' facial first impressions. *Journal of Vision, 15*(12), 1218.
Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition, 127*(1), 105–118. http://dx.doi.org/10.1016/j.cognition.2012.12.001.
Sutherland, C. A. M., Rowley, L., Amoaku, U., Daguzan, E., Kidd-Rossiter, K., Maceviciute, U., et al. (2015). Personality judgements from everyday images of faces. *Frontiers in Psychology, 6*, 1–11. http://dx.doi.org/10.3389/fpsyg.2015.01616.
Sutherland, C. A. M., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology, 106*(2), 186–208. http://dx.doi.org/10.1111/bjop.12085.
Tiddeman, B., Burt, M., & Perrett, D. I. (2001). Prototyping and transforming facial textures for perception research. *Computer Graphics and Applications, IEEE, 21*(5), 42–50. http://dx.doi.org/10.1109/38.946630.
Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social Cognitive and Affective Neuroscience, 3*(2), 119–127. http://dx.doi.org/10.1093/scan/nsn009.
Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623–1626. http://dx.doi.org/10.1126/science.1110589.
Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *PNAS, 111*(32), E3353–E3361. http://dx.doi.org/10.1073/pnas.1409860111.
Walker, M., & Vetter, T. (2009). Portraits made to measure: Manipulating social judgments about individuals with a statistical face model. *Journal of Vision, 9*(11), 1–13. http://dx.doi.org/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*(3), 395–412. http://dx.doi.org/10.1037/0022-3514.37.3.395.
Wojciszke, B. (2005). Morality and competence in person-and self-perception. *European Review of Social Psychology, 16*(1), 155–188. http://dx.doi.org/10.1080/10463280500229619.
Zebrowitz, L. A. (2005). Appearance DOES Matter. *Science, 308*, 1565–1566. http://dx.doi.org/10.1126/science.1114170.