Social Cognition, Vol. 37, No. 5, 2019, pp. 516-545

COMBINING TRAITS INTO A FACE: A REVERSE CORRELATION APPROACH

Manuel Oliveira and Teresa Garcia-Marques
William James Center for Research, ISPA – Instituto Universitário

Ron Dotsch
Utrecht University

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; and (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.

Keywords: face perception, reverse correlation, trait integration

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).

This research had the support of Fundação para a Ciência e a Tecnologia (FCT) through 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.

Correspondence concerning this article should be addressed 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

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 & Jones, 1972; Rosenberg, Nelson, & Vivekananthan, 1968). 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). Further, 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 (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 that 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 (2IFC) task (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 (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, are publicly available online at: https://osf.io/4zncm/

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 1TCI, 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. Further, 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, $\varpi/2$) 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).

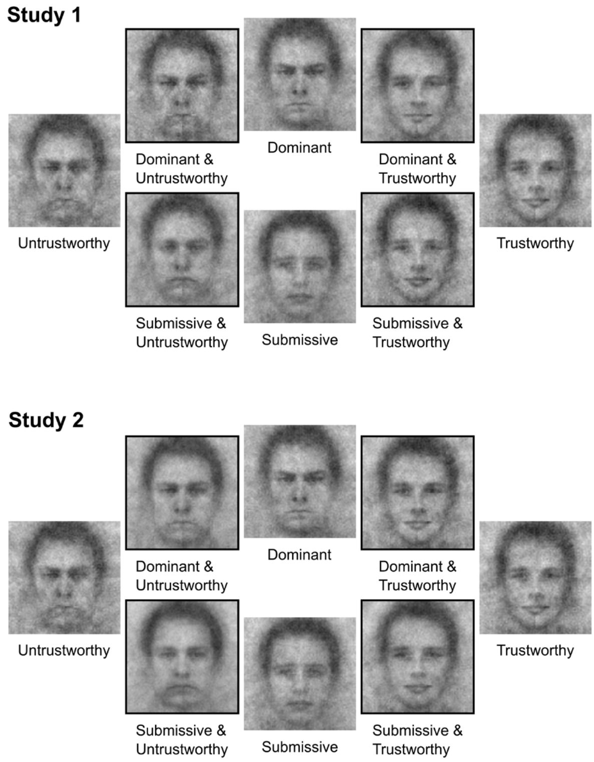

FIGURE 1. 2TCIs (framed pictures) obtained in Studies 1 and 2, for each of the four two-trait combinations of the poles of the dominance and trustworthiness dimensions, and 1TCIs (unframed pictures) extracted from Oliveira and colleagues (2019).

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 and colleagues (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] or 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 here for the sake of simplicity. However, because some of these additional ratings (i.e., threat and competence) may be of use 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' traits also inform about the extent to which one trait outweighed the other in different trait combination contexts.

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.

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).

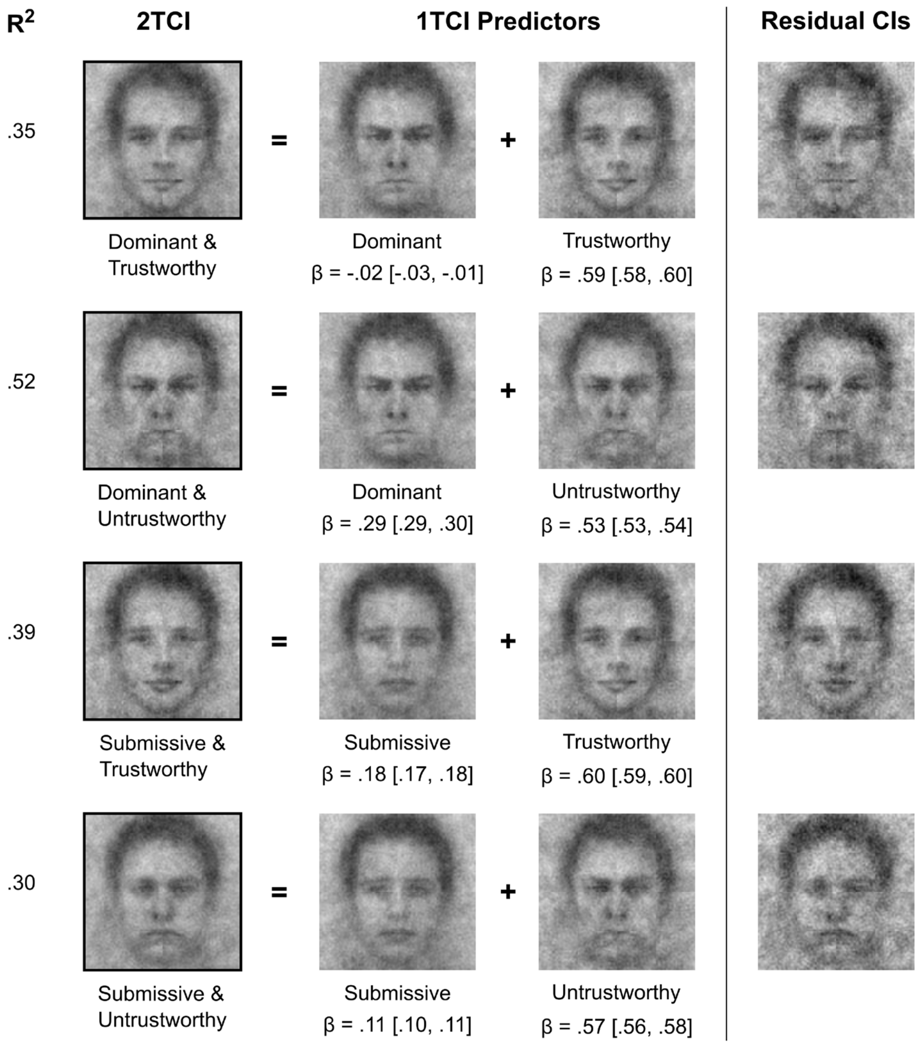

FIGURE 2. 2TCIs (framed pictures) obtained in Study 1 for each of the four two-trait combinations of the poles of the dominance and trustworthiness dimensions, and 1TCIs (unframed pictures) from Oliveira and colleagues (2019). All results presented here are for the unmasked CIs. Each row illustrates 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 *p*s < .001.

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.

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

			R^2 [95% CI]	
Study	**CI masking**	**Trait combination**	**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.

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 *R2*, 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 *R2* 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 *R2* 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 assessing 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

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. & Trust.	0.35	0.011	0.010	0.008	0.030	0.033
Dom. & Untrust.	0.52	0.013	0.013	0.010	0.041	0.050
Sub. & Trust.	0.39	0.015	0.013	0.015	0.036	0.043
Sub. & Untrust.	0.30	0.014	0.014	0.011	0.042	0.054

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.

R-Squared Simulations. To ascertain whether the observed *R2* 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 number 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 *R2* ranges obtained for each predicted 2TCI simulation are described in Table 2. The resulting ranges represent the interval of *R2* values that could be expected to be obtained by CIs without trait signal (i.e., random noise). Since none of the *R2* values of the four 2TCI regression models falls within the simulations' *R2* 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.

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

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

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

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).

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).

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 in 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).

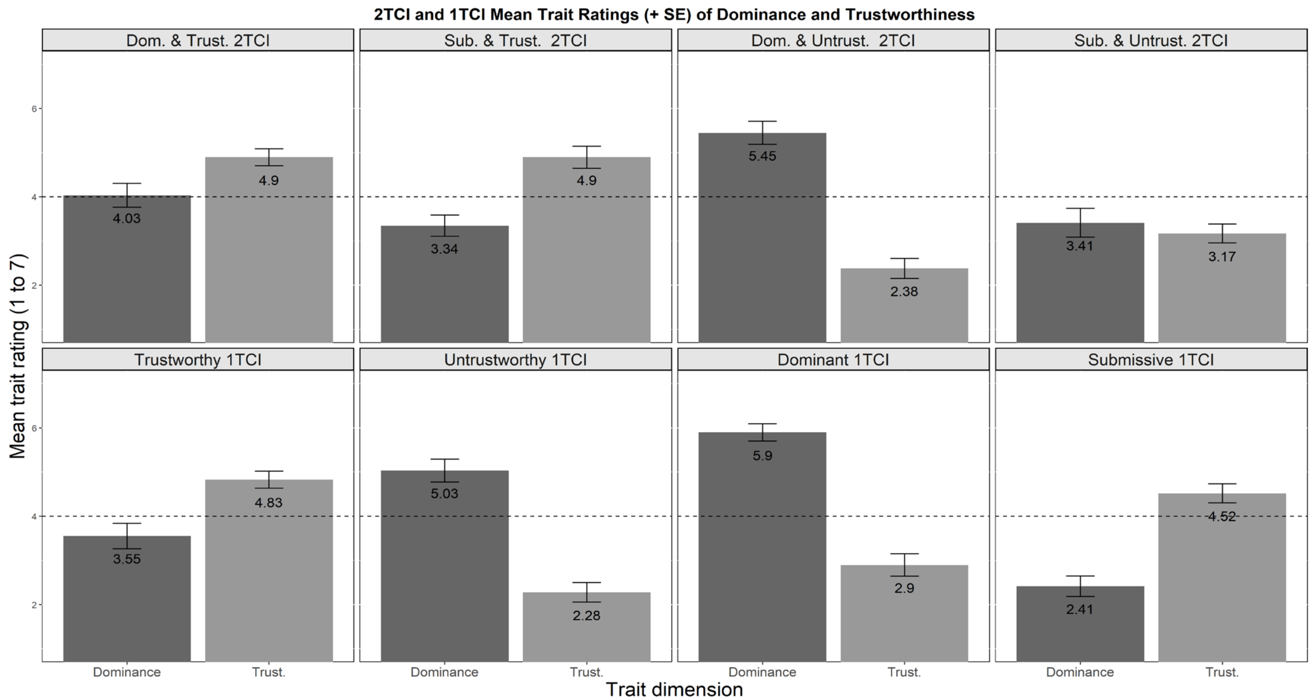

FIGURE 3. Mean trait ratings of dominance and trustworthiness obtained for the 2TCIs generated in Study 1, and for the 1TCIs obtained from Oliveira and colleagues (2019). The horizontal line intercepting the y-axis at value 4 represents the midpoint of the rating scale which was labeled in the task as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

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-β = 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 in 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-β = 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

Altogether, Study 1's results seem to support that trustworthiness overweighs dominance in trait integration, and that this integration does not result from a simple additive linear model. In Study 2, we aimed to obtain stronger support for this conclusion by replicating the main results of Study 1 with more stable CI estimates. With this goal in mind, we increased the number of judges per 2TCI, and additionally controlled for other potential sources of noise, such as word order in the target trait combinations.

In addition, Study 2 adds two tasks to the 2IFC task used in Study 1. These tasks aimed to assess some properties of the type of trait integration participants were making during the 2IFC task (from which 2TCIs are derived). The first task used as targets both the 2TCIs generated in Study 1 and the 1TCIs. In this task, we assessed participants' performance in detecting the signal of the two-trait combination (i.e., 2TCI generated in Study 1) when it is presented in a trial of the 2IFC task, and how the correct detection of that signal suffers interference from the signal of each of the four possible 1TCIs (paired with the 2TCI). An analysis of how participants reacted to these stimulus pairings informs us about the extent to which they were relying on the trait combination, as opposed to each trait in isolation, to perform the task. In the second task, only 1TCIs were used as targets, each corresponding to one of the traits in the target combination (e.g., Dominant 1TCI vs. Trustworthy 1TCI with Dominant & Trustworthy as the target instruction). This task addressed

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.

the type of bias that participants were prone to when no proper signal was presented for the detection of the simultaneous presence of the two traits.

PARTICIPANTS AND DESIGN

A total of 160 undergraduate students from ISPA – University Institute (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 were identical to those in Study 1 until the 300$^{\text{th}}$ trial. After the 300$^{\text{th}}$ 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 the 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 a 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.

RESULTS

Our analytical approach for Study 2 was similar to the one in Study 1. We expected to find: (a) that 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 the CI masking procedure on our results, and we 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 again indicated 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 results similar to those in Study 1, and suggests that the assumption of a context-independent linear integration of individual traits is unlikely to have occurred.

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 adds 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

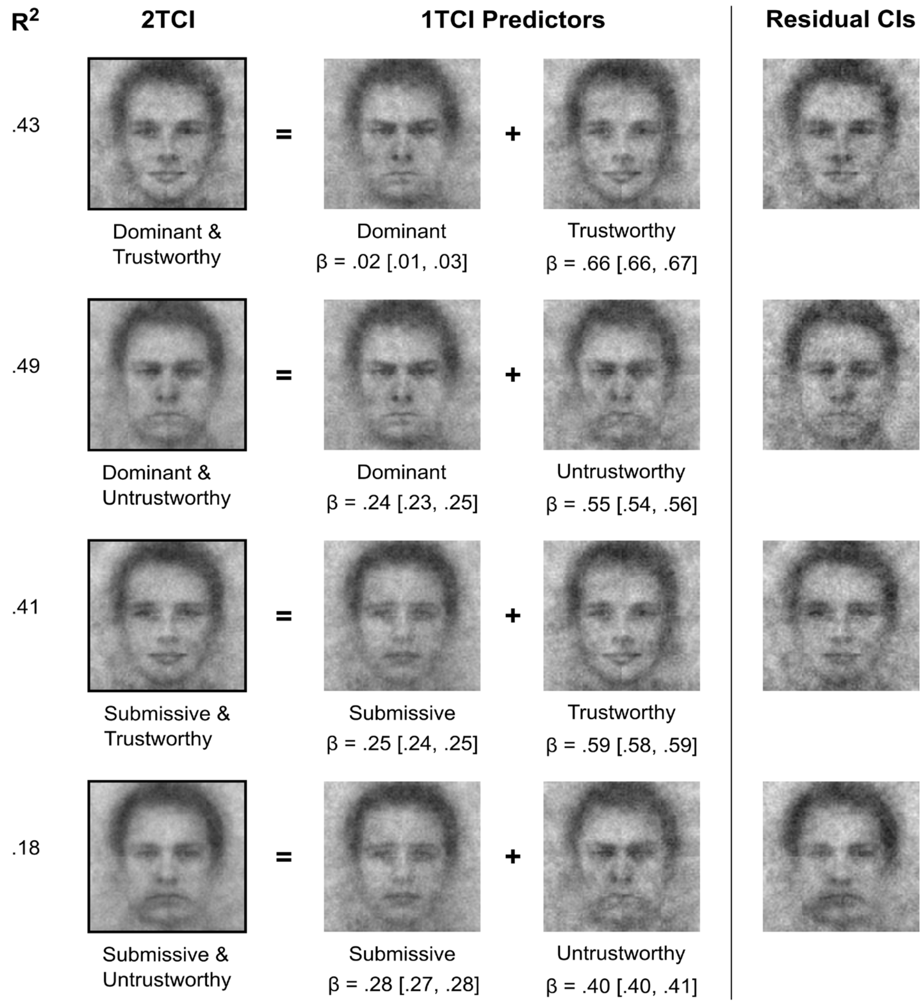

FIGURE 4. 2TCIs (framed pictures) obtained in Study 2 for each of the four two-trait combinations of the poles of the dominance and trustworthiness dimensions, collapsed across order of traits in the task instruction, and 1TCIs (from Oliveira et al., 2019). All results presented here are for the unmasked 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 *p*s < .001.

cannot be accounted for by any of the 1TCIs in the model. We discuss these results further 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. 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 regression[1] 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 (ORs) 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). ORs 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 the 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, *pFDR* = 1.00, 95% CI: 0.53, 1.85; in both 2TCI conditions). Moreover, compared 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 (ORs were, respectively, 9.00, 12.33, and 19.00, all *psFDR* < .001, in the DomTrust condition; and 4.71, 4.71, and 9.00, all *psFDR* < .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).

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 because it not only provided the same information as a signal detection analysis, but it also facilitated a clearer and more 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."

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, *pFDR* = .489, 95% CI: 0.72, 2.53; SubTrust condition: OR = 1.00, *pFDR* = 1.00, 95% CI: 0.54, 1.85) did not differ across the 2TCI conditions (OR = 0.74, *pFDR* = .656, 95% CI: 0.31, 1.78). In the DomUntrust 2TCI condition, compared to the Dominance 1TCI, 2TCI hits were more likely for any other reference 1TCIs (ORs were, respectively, 9.12, *pFDR* =.002, and 5.17, *pFDR* =.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, *pFDR* =.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, *pFDR* =.002) compared 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, *pFDR* =.194, 95% CI: 0.84, 5.14, and OR = 0.82, *pFDR* =.794, 95% CI: 0.34, 1.97).

From those results, we can infer that the generally high probabilities for 2TCI hits (ORs > 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 on 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 the methodological relevance of this evidence, these results also suggest that the weight 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.

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 either 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 log-linear regression analysis. Specifically, we entered 2TCI Condition and 1TCI Dimension as variables in the regression. Results are shown in Figure 5. There

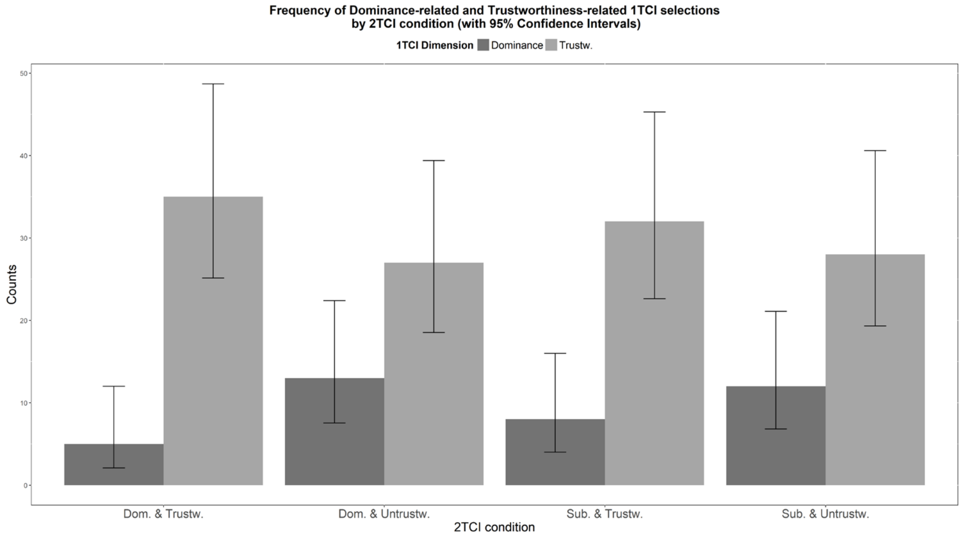

FIGURE 5. Plot exhibits absolute frequencies (counts) of dominance-related and trustworthiness-related 1TCI selections in the trial of the 2IFC extended reverse correlation task in which only 1TCIs were presented (task II), by 2TCI condition. Error bars denote 95% confidence intervals.

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 the 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 the 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 conditions. 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 onto 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, suggesting 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 impression formation that shows 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, 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 relative 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 that relies on 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 suggest that dominance is more heavily mapped in a face when coupled with untrustworthiness than when it is coupled with trustworthiness. Hence, our data fit well with Asch's 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" (1946, p. 268).

Besides offering new evidence supporting a gestaltic integration of trait information in face perception, we believe that this article 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 methodologic approach opens a new perspective to the study of how information is integrated in our minds, which can go beyond the focus of the discussion in this article. 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.

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*(5), 935–948. https://doi.org/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). Cambridge, MA: Academic Press. https://doi.org/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*(4), 394–400. https://doi.org/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*(4), 354–362. https://doi.org/10.1037/h0026816

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, *41*(3), 258–290. https://doi.org/http://dx.doi.org/10.1037/h0055756

Asch, S. E., & Zukier, H. (1984). Thinking about persons. *Journal of Personality and Social Psychology*, *46*(6), 1230–1240. https://doi.org/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*(1), 289–300. https://doi.org/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*(4), 397–408. https://doi.org/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*(1), 333–361. https://doi.org/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, CA: 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. https://doi.org/10.1016/S0065-2601(07)00002-0

Dotsch, R. (2015). *rcicr: Reverse correlation image classification toolbox. R package version 0.3.0*. Retrieved March 17, 2015 from https://cran.r-project.org/src/contrib/Archive/rcicr/

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, *3*(5), 562–571. https://doi.org/10.1177/1948550611430272

Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of Personality and Social Psychology*, *100*(6), 999–1014. https://doi.org/10.1037/a0023026

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/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: 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*(1), 148–168. https://doi.org/10.1037/a0034726

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, *103*(2), 336–355. https://doi.org/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*(2), 157–172. https://doi.org/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*(5), 837–852. https://doi.org/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*(3), 309–313. https://doi.org/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*. https://doi.org/10.3389/fpsyg.2013.00386

Jack, R. E., & Schyns, P. G. (2017). Toward a social psychophysics of face communication. *Annual Review of Psychology*, *68*(1), 269–297. https://doi.org/10.1146/annurev-psych-010416-044242

Johnson, D. H. (1999). The insignificance of statistical significance testing. *Journal of Wildlife Management*, *63*(3), 763–772. https://doi.org/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. https://doi.org/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*(3). https://doi.org/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*. https://doi.org/10.3389/fpsyg.2013.00863

Lundqvist, D., & Litton, J. E. (1998). *The Averaged Karolinska Directed Emotional Faces: AKDEF* [CD-ROM]. Stockholm, Sweden: Karolinska Institutet.

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

Murayama, K., Pekrun, R., & Fiedler, K. (2014). Research practices that can prevent an inflation of false-positive rates. *Personality and Social Psychology Review*, *18*(2), 107–118. https://doi.org/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*(5), 888-902. https://doi.org/10.1002/ejsp.2569

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

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

Rosenberg, S., & Jones, R. (1972). A method for investigating and representing a person's implicit theory of personality: Theodore Dreiser's view of people. *Journal of Personality and Social Psychology*, *22*(3), 372–386. https://doi.org/10.1037/h0032891

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*(4), 283–294. https://doi.org/10.1037/h0026086

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*(1), 105–118. https://doi.org/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*(1), 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001

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. https://doi.org/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*. https://doi.org/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*. https://doi.org/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*(7), 592–598. https://doi.org/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*(12), 1251–1263. https://doi.org/10.1177/01461672982412001

Ybarra, O., Chan, E., & Park, D. (2001). Young and old adults' concerns about morality. *Motivation and Emotion*, *25*(2), 85–100. https://doi.org/10.1023/A:1010633908298

Zebrowitz, L. A. (2006). Finally, faces find favor. *Social Cognition*, *24*(5), 657–701. https://doi.org/10.1521/soco.2006.24.5.657