Journal of Experimental Social Psychology 82 (2019) 200–207

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

The complexity of relations between dimensions of social perception: Decomposing bivariate associations with crossed random factors

Charles M. Judd[a,*], Teresa Garcia-Marques[b], Vincent Y. Yzerbyt[c]

[a] University of Colorado at Boulder, United States of America
[b] William James Center for Research, ISPA – Instituto Universitário, Portugal
[c] Catholic University of Louvain at Louvain-la-Neuve, Belgium

ARTICLE INFO

Handling editor: Nicholas Rule

ABSTRACT

Theories of social perception argue that there are two underlying dimensions of social judgment, variously labeled competence/agency and warmth/communality. How these relate to each other has been the focus of extensive empirical work with research showing both a 'halo" relation (targets rated more positively on one dimension are rated more positively on the other) and a "compensatory" relation (targets rated more positively on one dimension are rated more negatively on the other). We argue that these divergent findings result from different comparative contexts under which participants judge social stimuli and on the underlying factors that contribute to variance in the resulting judgments. In two studies, we vary the comparative context under which perceivers made judgments and we decompose the variance in such judgments (and their covariance) into components due to the random factors of participants, targets, and their interaction. Halo relations emerge for participant means, regardless of comparative context. On the other hand, the covariance between target means changes signs under different comparative contexts, as does the interaction covariance.

The literature on social perception has convincingly argued that two fundamental dimensions are used in perceiving the social world: Individuals and social groups are judged to vary in their competence/agency and in their warmth/communality (Fiske, Cuddy, Glick, & Xu, 2002; Judd, James-Hawkins, Yzerbyt, & Kashima, 2005; Rosenberg, Nelson, & Vivekananthan, 1968; for reviews, see Fiske, 2015; Yzerbyt, 2016). Each of these two fundamental dimensions has been variously defined, and related but distinct sub-dimensions have been identified (Abele, Cuddy, Judd, & Yzerbyt, 2008). Thus the warmth/communion dimension captures both a sub-dimension focusing on successful social functioning with traits such as warmth, sociability, and friendliness, and a sub-dimension focusing on moral intentions in interaction with traits such as trustworthiness, dishonesty, and benevolence (e.g. Abele et al., 2016; Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Leach, Ellemers, & Barreto, 2007). The other dimension, competence/agency, captures also two distinct sub dimensions, one focusing on ability-related traits, such as competence or intelligence, and the other focusing on assertiveness traits, such as dominance or confidence (e.g. Abele et al., 2016; Carrier, Louvet, Chauvin, & Rohmer, 2014).

These additional complexities notwithstanding, an important empirical and theoretical question has arisen in the literature about the nature of the association between these two fundamental dimensions. As we review, evidence exists for a positive association between them, for little or no association between them, and also for a negative association between them. It is our purpose in this paper to offer further evidence and empirical approaches to understand this diversity of results, even while we acknowledge that ultimately models of person perception that refine and go beyond only these two dimensions, however they are related, are necessary.

In line with theoretical ideas underlying cognitive consistency and balance theories of mid-twentieth century psychology (e.g., Asch, 1946; Heider, 1958; Rosenberg & Hovland, 1960), the predominant view has long been that evaluative inclinations dominate interpersonal judgment and that, therefore, person perception is primarily driven by motivations towards evaluative consistency. In essence, while researchers understood that warmth, competence, honesty, and other trait dimensions are semantically different dimensions of interpersonal judgment, they were thought to covary positively to reflect underlying evaluative consistencies in social perception. This gave rise to the widely documented 'halo effect' in interpersonal judgment, whereby a target seen positively on one trait dimension is also seen positively on others (Bruner & Tagiuri, 1954; Kelley, 1950; Thorndike, 1920). For instance, when observers see a given person as beautiful, they see this same person as being good (Dion, Berscheid, & Walster, 1972; Wheeler &

* Corresponding author.
E-mail address: charles.judd@colorado.edu (C.M. Judd).

https://doi.org/10.1016/j.jesp.2019.01.008
Received 13 August 2018; Received in revised form 22 January 2019; Accepted 23 January 2019

Kim, 1997; for a review, see Jackson, Hunter, & Hodge, 1995). Indeed, Osgood, Suci, & Tannenbaum (1957) argued that evaluation is *the* key dimension underlying judgment in general and social judgments in particular.

In their famous study on implicit theories of personality, Rosenberg et al. (1968; see also Rosenberg & Olshan, 1970) confirmed that a general evaluative dimension underlies social judgments but they also introduced the idea that a two-dimensional sub-structure offered a more complete description of perceivers' understanding of other people. Thus they suggested that the myriad of trait dimensions could be semantically differentiated into two clusters, even while manifesting overall evaluative consistency. To argue this, these authors handed participants a set of 64 traits, each one written on a slip of paper, and asked them to group the traits into piles according to which traits 'are likely to go together in the same individual and which traits seldom, if ever, go together in the same individual…' (1968, p. 285). Using multidimensional scaling, they found support for one underlying evaluative dimension. However, a two dimensional solution provided a better fit to the data, revealing two evaluatively loaded yet descriptively different dimensions: (1) good-intellectual versus bad-intellectual and (2) good-social versus bad-social. This advance paved the way to a fuller understanding of Asch's (1946) centrality and primacy effects (see the seminal insights here by Brown, 1986), as well as a series of important and heretofore unresolved issues in social perception (Fiske, 2015, 2018).

Building on this work, Fiske and colleagues elaborated the Stereotype Content Model (SCM; Fiske et al., 2002; Fiske, 2018), in an attempt to account for the fact that observers often appraise groups or people in ambivalent, rather than univalent, ways. In line with earlier proposals in the gender and stereotyping literature (Bakan, 1966; Glick & Fiske, 1996; Phalet & Poppe, 1997; Reeder & Brewer, 1979; see also Abele & Wojciszke, 2014), the model posits two fundamental dimensions of warmth and competence, with groups and individuals perceived to vary on both. The dimension of warmth reflects the nature of the interdependence between the perceiver and the target, denoting whether the target comes across as having friendly or hostile intentions vis-à-vis the perceiver. The second dimension of competence refers to the social target's power, expertise, resources, motivation, and thus its ability and its willingness to act upon its friendly or hostile intentions. The SCM stresses the pragmatic concerns of perceivers as they navigate the social world and has proven very successful as a general framework for social perception (for reviews, see Fiske, 2015, 2018).

Typically, SCM work requires participants to evaluate a substantial number of social groups on a list of personality traits tapping into various facets of the two dimensions of warmth and competence (Cuddy, Fiske, & Glick, 2007; Fiske et al., 2002; Fiske, Cuddy, & Glick, 2007; Kervyn, Fiske, & Yzerbyt, 2015). Of note, unless researchers take specific steps to include groups seldom mentioned spontaneously (the participants' ingroups and a handful of derogated groups and social outcasts), SCM researchers report that a majority of groups seem to have so-called mixed or ambivalent stereotypes. In other words, a high proportion of groups seem to be perceived positively on one dimension but negatively on the other. Thus this work is suggestive of a negative association between the two fundamental dimensions.

The dimensional compensation model (DCM; for a recent review, see Yzerbyt, 2018) builds on social identity approaches (Tajfel & Turner, 1979; Turner, Hogg, Oakes, Reicher, & Wheterel, 1987) and social perception work (Fiske et al., 2002; Rosenberg et al., 1968). The DCM proposes that the two dimensions of social perception are often negatively related when perceivers face social situations that trigger comparisons across the two dimensions in judging targets. For instance in a correlational study tapping French and Belgians stereotypes, Yzerbyt, Provost, and Corneille (2005) showed that members of both of these groups rated the French as more competent than Belgians but Belgians as warmer than the French, a pattern also shown by others who were not members of these groups, i.e., Swiss participants. Judd et al. (2005) replicated this pattern in a more controlled laboratory setting using fictitious groups. In their studies, participants received behavioral information about two novel groups suggesting that one group was higher than the other group on one of the two fundamental dimensions. In addition, participants received equivalent but non-diagnostic information about the two groups on the other dimension. In spite of the absence of any informational difference on this second dimension, and in line with the compensation pattern, participants judged the group higher on the manipulated dimension to be lower on the non-manipulated one. In these studies, the judgment task invites participants to compare the relative standing of the target groups on the two dimensions simultaneously, asking for instance, whether one group is higher on one dimension but lower on the second than the other group.

Building on these initial demonstrations, an impressive number of studies have shown the robustness of compensation in a large number of situations and with a variety of targets (for reviews, see Kervyn, Yzerbyt, & Judd, 2010; Yzerbyt, 2016, 2018). For instance, Kervyn, Yzerbyt, Demoulin, and Judd (2008) showed that pairs of countries were seen in a compensatory manner although individual countries might be seen very differently depending on the comparison country. More recently, Cambon, Yzerbyt, and Yakimova (2015) confirmed the fluidity of social perception in intergroup settings. Depending on whether their ingroup enjoyed a higher or lower status than the outgroup, psychology students rated their group as more competent (and less warm) or less competent (and warmer) than the other group. Of note, compensation emerged in this comparative context only to the extent that respondents perceived the status difference as legitimate and when there was no conflict between the groups (Cambon & Yzerbyt, 2016). Elaborating an intuition put forth by Brown (1986), Kervyn, Bergsieker, and Fiske (2012) provided evidence for what they call an innuendo effect, such that people provided with information about only one of the two fundamental dimensions tend to fill in on the second dimension in a compensatory manner. Empirical work also confirms that compensation emerges in the absence of social desirability pressures. That is, compensation shows up not only on explicit ratings but also in spontaneous and implicit measures of group stereotypes, such as the linguistic category model (Kervyn, Yzerbyt, & Judd, 2011) or an implicit association task (Schmitz & Yzerbyt, 2018).

In sum, often the two fundamental dimensions of social perception are positively related, consistent with a 'halo effect', such that an overall positive appraisal of individual groups or individuals generalizes to perceptions of positivity on both fundamental dimensions. On the other hand, when comparisons across the two dimensions are made salient in judging targets, be they groups or individuals, then compensation tends to occur, with those targets that are seen as higher on one dimension judged lower on the other. The necessity of comparison for compensation to emerge suggests that different ways of framing judgment tasks may affect whether or not judgments of warmth and competence are positively or negatively related to each other.

The goal of the research that we report is to examine the covariance between judgments of targets' warmth and competence (operationalized as intelligence in our studies) under different task instructions that induce different comparative frameworks. A second goal is to explicate a more sophisticated model of the covariance between judgment dimensions than has been used in past research. This more sophisticated model takes into account for the first time in this literature that both the participants and the targets they judge are most properly treated as random factors, allowing generalization to other samples of both that might have been used (Judd, Westfall, & Kenny, 2012, 2017). Heretofore, in analytic models employed, random variances attributable to participants and targets (and their interaction) have not been separated from each other, resulting in an incomplete specification of different components of covariance between the judgment dimensions, as we will show. We will also show that the more complete specification of components of covariance between judged trait dimensions can reveal evidence of simultaneous halo and compensation associations in the same data.

1. Components of variance and covariance in participants' judgments of targets

Consider a matrix of ratings provided by some number of participants (p) to some number of target stimuli (q). Let us assume for now that participants and targets are fully crossed, resulting in $p*q$ ratings. Henceforth we treat both participants and targets as random, rather than fixed variables, meaning that both participants and targets are considered to be sampled from larger populations to which we would like to be able to generalize results (i.e., to address whether similar results would be expected in future studies with different samples of both participants and targets; see Judd et al., 2012, 2017). From the analysis of variance, we can decompose the variability in these ratings into three independent components of variance: variance attributable to participant mean differences, variance attributable to target mean differences, and variance attributable to the participant by target interaction. This latter component of variance also includes in it error variance, unless there are multiple ratings or observations in each cell of the participant by target matrix. This decomposition, with two crossed random factors, is consistent with Kenny's exposition of actor – partner differences in his Social Relations model (Kenny, 1994; Kenny & La Voie, 1984; Snijders & Kenny, 1999).

These components of variance are readily interpretable. Participant mean variance represents variance attributable to participants in how they respond on average to all stimuli, i.e., individual participant differences in mean ratings given. Target mean variance represents stable variance attributable to the targets in how they are rated on average by all participants, i.e., target differences that are perceived on average by all participants. Finally, the interaction variance represents the unique tendency for a particular participant to respond to a particular target differently than what one might expect, given that participant's mean response across all targets and given that target's mean rating from all participants.

Now consider two matrices of such ratings, one consisting of judgments of competence (or some highly related dimension) and the other consisting of judgments of warmth (or again something highly related). If one were interested in the association between the competence and warmth judgments, one might start by examining the correlation or covariance between all the values in one of these matrices and all of those in the other, computing that correlation across all $p*q$ values in the two matrices. But just as the variance in each matrix can be decomposed into three sources, so too can one decompose the covariance between the values in the two matrices into three sorts: participant mean covariance, target mean covariance, and interaction covariance. The first of these assesses whether participants who give higher competence ratings, on average across all targets, also give higher average warmth ratings. The second assesses whether targets that are judged on average by all participants to be higher in competence are also judged on average to be higher in warmth. The interaction covariance is also meaningful but a bit more difficult to explain: If a particular target is perceived by a particular participant as surprisingly high in competence, that is, higher than what one might expect from both that participant's mean competence rating of all targets and that target's mean competence rating from all participants, is it also judged as surprisingly high in warmth?

In the right experimental context, these three sources of covariance might differ quite radically, perhaps even varying in their sign. For instance, it seems quite likely that participant means in the two matrices would manifest a positive covariance: people who tend to see all targets relatively positively on one dimension of judgment (or equivalently habitually use higher scale values in their judgments) might also do so on the second dimension of judgment. Thus, one might expect positive relations between participant means. On the other hand, in a context where targets are compared on the two dimensions, then it is possible that compensatory processes might be invoked and be manifested in the covariance of the target means: Those targets which everyone agrees are higher in competence are also, with agreement on average from everyone, judged relatively lower in warmth. And in such a comparative context, it seems plausible to us that the interaction covariance might also be negative, consistent with compensation again. As already discussed, when one simply computes one covariance or correlation among all the values in one matrix and all the values in the other, one might be mixing up very different processes underlying comparative judgments that ideally should be teased apart.

It is important to note that nowhere in this work, either theoretically or empirically, do we make claims about causal relations between the two dimensions. We are simply documenting a bivariate association in judgments and how that association may vary with both the comparative context and the component of covariance, without any additional claims about the causal mechanisms responsible for the association. It is because of this that we feel comfortable presenting both models that predict competence from warmth and warmth from competence.

We report two studies in this paper, manipulating between studies the nature of the judgment task, with the intention thereby of inducing rather different comparative processes. In both studies, male and female participants judged a set of male and female target faces on both their intelligence (our proxy for competence) and their warmth.

In the first study, each participant first judged all of the target faces on one of the two dimensions and then judged them all on the second dimension, varying between participants the order of the two judgments. In this study, with this task structure, it seems likely to us that the ratings would reflect comparative processes between faces within one and the same dimension. In other words, the participant would ask him or herself whether a particular face is higher or lower on the relevant dimension of judgment than the many other faces he or she has already seen, with little reference to the other rating task (dimension) that is already finished or yet to be done. Given that both dimensions of judgment are highly evaluative ones, the overall evaluation of individual faces should, we expect, dominate judgments on both dimensions in this first study.

In the second study, each participant judged each face simultaneously on both dimensions, first judging a face's intelligence and then its warmth, before going on to judge the next face. Here, we surmised that rather different comparative processes would be invoked. Rather than thinking about all of the faces along a single dimension, participants might well be asking themselves for each face whether their judgment should be higher on one dimension than on the other, relative to all other faces they have seen. That is, the question might be for each face, is this one that is more intelligent than warm, or more warm than intelligent. In this case, where the comparison is between dimensions for each face, we might expect compensation, at least in some components of the covariance between the two judgments, as we detail in the following paragraphs.

In both studies, it seemed likely to us that the covariance between participant means would manifest a positive association, suggesting that participants who give higher average ratings on one dimension (more positivity) give higher average ratings on the second. This would reflect individual participant differences in evaluative tendencies or simply differences in judgment scale usage.

Rather different results may emerge for target mean and interaction covariances in the two comparative situations. The comparative situation of Study 1 might induce a positive relation in all components: one is simply basing one's judgment of a particular face on whether this face seems in general to be a more positive or negative face than the many other faces one is asked to judge. On the other hand, compensation or a negative association might be found for target mean and interaction covariances in the second study, where the comparative processes demand that one differentiate the two dimensions, evaluating whether a particular face is higher in intelligence than in warmth or higher in warmth than in intelligence, relative to other faces already judged.

2. Study 1: method

2.1. Participants and design

In this study and the subsequent one, we report all measures, manipulations, and exclusions. No analyses were conducted before complete data were obtained.

A total of 68 English-speaking participants (78% Females, M_{age} = 26.59; SD = 9.84) were recruited online through Mechanical Turk and received $3 for their participation. The design was a 2 (Order: Intelligence first or Warmth first) × 2 (Dimension judged: Intelligence or Warmth) mixed design with the second factor varying within participants.

2.2. Materials

We selected 50 photographs (25 young females and 25 young males) from the "Lifespan database of adult facial stimuli" (Minear & Park, 2004). We took great care to ensure that all facial expressions were emotionally neutral but that there was diversity in terms of appearance. The photographs showed faces and shoulders of all targets. They were presented in black and white and shown at the center of a computer screen within a frame of 10 cm × 7 cm.

2.3. Procedure

We implemented the study via an online survey using *Qualtrics Survey Software*. After reading and accepting an informed consent form, participants were asked to read the instructions carefully and complete the rating tasks.

We randomly assigned participants to one of the two order conditions, either completing all the intelligence ratings before doing all the warmth ratings, or the other way around. Specifically, in the first rating task, a 7-point rating scale was presented below each of the 50 consecutive photographs inviting participants to evaluate the face on the first dimension from 1 (=not at all) to 7 (=very much). Immediately after judging each face, the next one appeared on the screen and stayed there until its judgment was completed. Once participants completed the first set of ratings on all 50 pictures, the sequence of faces was repeated and participants rated the faces on the second dimension. Within each rating task, the order of presentation of the photographs was randomly determined.

2.4. Power considerations

We based our sample sizes, both in this study and Study 2, on approximate power estimates required to detect a relatively large simple correlation between the two dimensions of judgment. This expectation was based on the positive correlation between the two dimensions (0.39) found in Rosenberg et al. (1968). Sample sizes were determined based on published work that has made clear that power in the presence of multiple random factors is a complex function of the n's of both random factors (participants and targets) and the magnitude of the relevant random variances (Westfall, Kenny, & Judd, 2014). To do the power analysis, we used the power application discussed in Judd et al. (2017). This application computes power for dichotomous rather than continuous predictor variables, but we adapted it using a standardized effect size for power computations.

3. Study 1: results and discussion

We report here various univariate and bivariate models. The former models examine contributions to variance in either the warmth or intelligence ratings. The latter examine the prediction of one set of ratings from the other. In all models, both participants and targets are treated as crossed random factors (Judd et al., 2012), estimating random intercepts and slopes for each as appropriate for the specific model.[1] We also examined the effects of various categorical predictors in some of these models: order of judgment task (whether the participants first did intelligence ratings for all targets and then warmth ratings (+1) or the reverse order (−1); participant gender (+1 male; −1 female); and target gender (+1 male; −1 female). Model specifics are given as the models are presented.

3.1. Univariate models

The initial univariate models for each set of ratings (warmth or intelligence) modeled the ratings as a function of the order predictor. In this model, the intercept estimates the grand mean and the order slope estimates half the order difference in the mean ratings. We allowed random intercepts for both participants and targets and random order slopes for targets, but not participants (since targets are crossed with order but participants are not). Thus, in addition to the fixed effects, four random components of variance were estimated: intercept (mean) variance for participants and target, order slope variance for targets, and residual variance.

We estimated the mixed model using the Lmer function in R, with the model specification as follows (with the warmth judgments as the modeled variable):

$$W_{ij} = \beta_0 + \beta_1 O_{ij} + (\mu_{0i} + \mu_{1i} O_{ij}) + (\mu_{0j}) + e_{ij}$$

Here, i represents participant and j target. The beta's are the fixed effects and random effects are given in parentheses, first for participant and then for target.

Table 1 shows model results, first for the warmth ratings and then for intelligence. Given are the fixed effect estimates (grand mean and order difference) and the random variances. Additionally, we present for each set of ratings the relative magnitude (percentage) of total variance in the ratings attributable to participants, targets, and residual (including the participant by target interaction).

There was a significant order effect for the intelligence ratings (d = 0.22),[2] but not for the warmth ratings (d = 0.02). Mean intelligence ratings were slightly higher on average when it was the second dimension rated. Variances due to participants and to targets were substantial for both ratings, although in both cases they were somewhat higher for the warmth dimension than for intelligence. The majority of the variance in both ratings was associated with the residual, and this was especially true for the intelligence dimension.

We then estimated univariate models that examined gender differences in ratings on each dimension. The models included order, participant gender, target gender, and the participant gender by target gender interaction as fixed effects. Again, we treated both participants and targets as crossed random factors. We allowed random intercepts for both and random slopes for both order and participant gender for targets and for target gender for participants. These models showed no significant gender differences in either warmth or intelligence ratings (all d's < 0.12; p's > .15).

[1] The structure of the random components of all models were fully specified, including all relevant variances and covariances, following the recommendations of Barr, Levy, Scheepers, and Tilly (2013). We give the full specification of the models in the results.

[2] There is no firm consensus on appropriate effect sizes for mixed models with crossed random factors. We use here a measure analogous to Cohen's d, as developed by Judd et al. (2017). This works in the case of categorical predictors. At a later point, with continuous predictors in the model we use an effect size measure analogous to partial r-squared, similar to one outlined by Snijders and Bosker (1999).

Table 1
Study 1: Univariate models for warmth and intelligence ratings.

Warmth				
Fixed effects	*b*	*T*	*Df*	*p*
Intercept (mean)	3.61	30.57	93.90	< .001
Order	0.01	0.20	66.05	.845
Random variance components				
	Variance		Percent of total	
Participant	0.346		20.1%	
Target	0.419		24.3%	
Residual	0.957		55.6%	
Total	1.722			
Intelligence				
Fixed effects	*b*	*T*	*Df*	*p*
Intercept (mean)	3.89	38.85	83.25	< .001
Order	−0.14	2.41	68.39	.017
Random variance components				
	Variance		Percent of total	
Participant	0.200		12.3%	
Target	0.331		20.4%	
Residual	1.094		67.3%	
Total	1.625			

Note: Both models also included random order slope variance for targets and the covariance between the two target components. In all cases these were negligible.

3.2. Bivariate models

Our initial bivariate models did not decompose the variance in the ratings. We simply predicted one rating (either warmth or intelligence) from the other (intelligence or warmth) across all target by participant observations, treating both participants and targets as crossed random factors. Initially in these models, we included order as an additional predictor along with its interaction with the predictor. Since these interactions were not significant, we report the simpler models that do not include these additional predictors. The Lmer specification for the model where warmth was the criterion and intelligence the predictor was the following:

$$W_{ij} = \beta_0 + \beta_1 I'_{ij} + (\mu_{0i} + \mu_{1i} I'_{ij}) + (\mu_{0j} + \mu_{1j} I'_{ij}) + e_{ij}$$

with terms defined as specified earlier and the prime (′) associated with the predictor meaning that it has been mean-centered.[3]

In the model predicting warmth from intelligence, the predictor slope equaled 0.218 ($t_{(73.56)} = 8.81$; $pr^2 = .08$; $p < .001$). The random variance of the participant slopes equaled 0.020 and that of the target slopes was 0.001. Thus, the relation between the two variables was significantly positive. Additionally that relation varied more from participant to participant across targets than it did from target to target across participants. In the model predicting intelligence from warmth, the predictor slope equaled 0.253 ($t_{(88.16)} = 9.40$; $pr^2 = .08 p < .001$). Again, the variance of the random participant slopes (0.023) was considerably larger than the variance in random target slopes (0.001). In sum, when a target is given a higher rating by a participant on one dimension, it is also given a higher rating on the other dimension, consistent with a 'halo effect' across the two dimensions. Additionally, the within-participant slopes across targets were always considerably more variable than the within-target slopes across participants. This suggests that some participants manifested a stronger 'halo effect' than others whereas the positive relation between the warmth and the intelligence ratings was rather similar for all targets.

The next models decomposed the predictor variable (either intelligence when predicting warmth, or warmth when predicting intelligence) into the three components of variance defined in the introduction: participant, target, and residual (interaction). For each participant by target observation, the first was computed as that participant's mean across all targets, the second was computed as that target's mean across all participants, and the third was computed by taking the individual observation, subtracting both the participant's mean and the target's mean, and then adding back in the grand mean. These three components are necessarily independent of each other. Additionally, they can only explain variation in the dependent variable associated with the same component of that variable. For instance, the slope of the participant mean component is due solely to its covariance with the participant mean component in the dependent variable, computed across all observations. (In supplemental materials, we present the variance/covariance matrices for the two dimensions of judgment, for the participant means, the target means, and the interaction. The raw data are available from the authors.)

The model specification in Lmer was as follows (again predicting warmth from the intelligence components):

$$W_{ij} = \beta_0 + \beta_1 imean'_i + \beta_2 jmean'_j + \beta_3 ijint'_{ij} + (\mu_{0i} + \mu_{2i} jmean'_j) + (\mu_{0j} + \mu_{1j} imean'_j) + e_{ij}$$

with *i* and *j* referencing participant and target and the three predictors being the participant mean, target mean, and the residual or interaction (all centered).

Table 2 presents the models' results. In each model, the fixed effects are the grand mean (intercept) and the slopes of the three predictors (participant means, target means, and interaction or residual). Each of these has been centered to make the interpretation of the random variance components easier. The random components include participant and target intercept (mean variance), participant slope variation for the target mean predictor, and target slope variation for the participant mean predictor.

In both models, all predictors have significant positive slopes. We interpret these one at a time. The positive slopes of the participant mean predictor suggest that participants who give higher mean responses to all targets on average on one dimension give higher mean responses on the other. In other words, individual differences in average responses are positively related across the two dimensions. The positive slopes of the target mean predictor suggests that targets that are consensually judged by all participants as higher on one dimension are consensually judged as higher on the other. In other words, if everyone agrees that a target is high on one dimension, they tend to agree that it is high on the other. Finally, the positive slope of the interaction predictor means that if a given participant judges a given target as higher than one might expect on one dimension, given his or her judgment mean and that target's mean judgment, then that participant judges that same target as higher than one might expect on the other dimension. Idiosyncratic tendencies to see a target as particularly high or low are also positively associated across dimensions.

All of this suggests three different components of association that are discriminable yet consistent in the direction of their bivariate associations: higher ratings of targets on warmth are associated with higher ratings on intelligence and this is because of all three component processes that we have decomposed.

[3] A reviewer raised the issue of whether dependence is likely violated in this model since the variance decomposition that we earlier described was not implemented. But in fact, this mixed model accounts for dependencies due both the participants and targets, as laid out in the model specification for crossed random factors, as outlined by Judd et al. (2012, 2017).

Table 2
Study 1: Bivariate models with decomposition of predictor.

Predicting warmth from intelligence				
Fixed effects	*B*	*t*	*df*	*p*
Intercept	3.61	34.92	84.8	< .001
Participant mean	0.67	5.40	66.2	< .001
Target mean	0.43	2.87	49.1	.006
Residual (interaction)	0.20	12.47	3090	< .001
Random variances				
Participant intercept				0.235
Participant slope of target mean				0.027
Target intercept				0.360
Target slope of participant mean				0.001
Residual				0.903
Predicting intelligence from warmth				
Fixed effects	*B*	*t*	*df*	*p*
Intercept	3.89	42.77	77.3	< .001
Participant mean	0.45	5.46	66.9	< .001
Target mean	0.36	2.97	51.1	.005
Residual (interaction)	0.23	12.27	2861	< .001
Random variances				
Participant intercept				0.147
Participant slope of target mean				0.033
Target intercept				0.284
Target slope of participant mean				0.003
Residual				1.032

Note: Both models also estimated covariances between random intercepts and slopes.

4. Study 2: methods

4.1. Participants, design and procedure

A total of 33 English speaking participants (90% Females, M_{age} = 24.67; *SD* = 7.55) received $3 for evaluating online a set of 50 randomly presented consecutive photographs (the same as those used in Study 1). Below each photo two 7-point rating scales (1 = not at all; 7 = very much) asked participants to judge the face on the two dimensions of warmth and intelligence. Participants gave both ratings of each face before going on to the next face.

5. Results and discussion

We replicated in this second study all of the models that were estimated with the data of Study 1, with the exception that there was no between-participant order variable to be included in the models.

5.1. Univariate models

The initial univariate models simply estimate the mean (intercept) for each dimension, with both participants and targets having randomly varying intercepts. These models decompose the variance in each set of ratings into participant, target, and residual components. Results are reported in Table 3 first for the warmth ratings and then for intelligence.

We also estimated univariate models that examined gender of participant and gender of target effects on the two dimensions. As in Study 1, none of these differences, nor their interaction, were significant (all *d*'s < 0.12; *p*'s > .15).

Table 3
Study 2: Univariate models for warmth and intelligence ratings.

Warmth				
Fixed effects	*B*	*t*	*df*	*p*
Intercept (mean)	4.02	24.69	62.15	< .001
Random variance components				
	Variance			Percent of total
Participant	0.449			22.7%
Target	0.545			27.6%
Residual	0.980			49.6%
Total	1.974			
Intelligence				
Fixed effects	*B*	*t*	*df*	*p*
Intercept (mean)	3.96	26.12	62.22	< .001
Random variance components				
	Variance			Percent of total
Participant	0.394			20.6%
Target	0.464			24.4%
Residual	1.071			56.2%
Total	1.905			

5.2. Bivariate models

As in Study 1, the initial bivariate models regressed one set of ratings on the other, across all participant by target observations, without decomposing the predictor variable. As always, we treated both participants and targets as random, with random intercepts and slopes for the predictor variable for both. In the model predicting warmth from intelligence ratings, the estimated intelligence slope equaled −0.035 but was not significant ($t_{(33.06)} = 1.32$; $pr^2 = .01$; $p = .197$). Likewise, in the model predicting intelligence from warmth, the slope equaled −0.032 and was not significant ($t_{(40.58)} = 0.99$; $pr^2 = .01$; $p = .325$). In Study 1, both of these slopes were positive and significant. At a global level, without decomposing the variance, the bivariate data from this study where the two ratings were done simultaneously show little evidence of the positive association between the two dimensions that was observed in Study 1. At a global level it is not true that the more positively a target is rated on one dimension, the more positively it is rated on the other.

Table 4 reports the results from the bivariate models where we decomposed the predictor into participant means, target means, and the residual or interaction, using the same approach as that used for the models of Study 1 reported in Table 2. (Again raw variance/covariance matrices for the participant means, target means, and interaction for the two dimensions are reported in the supplementary materials.) The results in these models differ dramatically from what emerged in the bivariate models without the decomposition, reported in the previous paragraph. Instead of finding non-significant bivariate associations, we now have evidence of a significant positive association between the participant means on one variable and those on the second, a significant negative association between the target means on one variable and those on the second, and a marginally significant negative association between the interaction components. These results are also in stark contrast to those we obtained in the first study, where all of these slopes were positive and significant.

It appears that when participants judge targets on warmth and intelligence simultaneously, rather different processes are set in motion.

Table 4
Study 2: Bivariate models with decomposition of predictor.

Predicting warmth from intelligence				
Fixed effects	*B*	*t*	*df*	*p*
Intercept	4.02	40.70	52.8	< .001
Participant mean	1.03	18.70	31.1	< .001
Target mean	−0.38	2.57	46.0	.013
Residual (interaction)	−0.05	1.89	1503	.059
Random variances				
Participant intercept				0.027
Participant slope of target mean				0.000
Target intercept				0.040
Target slope of participant mean				0.000
Residual				0.979
Predicting intelligence from warmth				
Fixed effects	*B*	*t*	*df*	*p*
Intercept	3.96	42.84	51.0	< .001
Participant mean	0.89	17.12	33.5	< .001
Target mean	−0.37	3.01	45.8	.004
Residual (interaction)	−0.05	1.83	1334	.067
Random variances				
Participant intercept				0.019
Participant slope of target mean				0.007
Target intercept				0.350
Target slope of participant mean				0.026
Residual				1.051

Note: Both models also estimated covariances between random intercepts and slopes.

At the participant level, it is still the case that those participants with higher mean ratings on one dimension also have higher mean ratings on the second. In other words, those participants who judge all of the targets more positively on one dimension judge them all more positively on the second. But at the level of the target means and the interaction, things are quite different. Targets that are consensually rated more positively on one dimension are consensually rated significantly more negatively on the second. This is clear evidence of compensation at the target level. Likewise, although not quite significant, the interaction has a negative slope, meaning that if a particular participant sees a particular target as surprisingly positive on one dimension, more than one might expect given that participant's mean rating and that target's mean rating, then the judgment on the other dimension is surprisingly negative. Again, albeit not significant, compensation tends to characterize the relation between the two judgments at the level of the participant by target interaction.

6. General discussion

The message of these two studies, and the results we have presented, is that the general question of how judgments of warmth and competence covary requires a complex answer that depends both on the nature of the comparative context in which those judgments are rendered and on the components of covariance that are examined. Recognizing that variation in participants' judgments of targets may be due to at least three sources (participant mean variance, target mean variance, and interaction or residual variance) allows the possibility that two sets of judgments may manifest different relations depending on the source of the variance (and covariance) one is examining.

Across both of our studies, we have found that participant means show a positive covariance across the two judgment dimensions. Thus, regardless of our manipulation of comparative context, participants who, on average, give more positive ratings than other participants on one dimension also give more positive ratings on the other dimension. This participant difference may reflect genuine differences in the positivity of outlook and it may also arise simply from individual differences in scale usage. Regardless, ratings of targets by participants show evidence of a halo effect when one only examines this component of variance: those judges who generally see more warmth in faces generally see more intelligence in faces.

On the other hand, the association between target mean differences on one dimension and those on the other seems to depend dramatically on the comparative context under which judgments are rendered. When, as in Study 1, the task is to judge all the targets on a given dimension before going on to the other dimension, then the task becomes one of determining which targets deserve more positive regard than other targets on a given dimension. And to the extent that this task definition holds for the second rated dimension as well as the first and that there is some agreement across participants in which targets deserve more positive regard, then target means also manifest a halo effect: targets rated by all participants more positively on one dimension are rated more positively on the other. When, however, the task changes, as in Study 2, and becomes one where participants are asking themselves whether a particular target should be rated higher on warmth than intelligence, or the other way around, compared to other targets, then compensation emerges. A target judged relatively high on one dimension on average tends to be judged lower on the second on average. Thus, when considering covariation due to target mean ratings, a 'halo effect' is found in Study 1, where the goal is the judgment of all targets on one dimension and then of all targets on the second, but compensation is found in Study 2, where the goal is the judgment of each target on the two dimensions simultaneously, asking whether a target is relatively higher on one of the two dimensions than the other, relative to other targets one has seen.

Like the variance attributable to target means, the variance in the participant by target interaction positively covaries across dimensions in Study 1, whereas its sign becomes negative in the comparative context of Study 2 (albeit not significantly). This residual interaction variance represents idiosyncratic ways in which an individual target is judged by an individual participant, either surprisingly high or surprisingly low, given that target and that participant. These surprising departures in ratings manifest a positive or halo association in the comparative context of Study 1 but a negative or compensatory association in the comparative context of Study 2. A more complete look at this participant by target interaction variance would require separating it out from residual error variance. Future research should ask participants for multiple ratings of each dimension in order to accomplish this.

The value of teasing apart the sources of variance (and hence covariance) in these ratings becomes clear if one compares these component results to what emerges from the analyses that simply regressed one set of ratings, on one dimension, on the other set, simultaneously across all participants and all targets (the initial bivariate analyses we reported in both studies). Because all three components manifest positive covariance in the comparative context of Study 1, the model that collapses across all components finds a large and significant positive association between the two dimensions. On the other hand, in Study 2, where the component covariances go in opposite directions, the net association if one ignores the variance decomposition is no apparent relation at all. Clearly, this latter result is misleading as it masks a significant 'halo effect' at the level of the participant means and significant compensation at the level of the target means (and marginally significant at the interaction level).

So the lesson from this work is that the simple question of how the two fundamental dimensions relate to each other demands a complex answer. The association depends both on the comparative context under which participants rate the targets and on the source of the variation in those ratings. Positive associations emerge regardless of comparative context when examining participant mean covariance. In contrast, under the right comparative context, compensation emerges for target mean and interaction covariances.

Another lesson stemming from this work, from a methodological perspective, is that one must think about sources of variance in data due to underlying random factors, beyond our usual focus on variance arising from participants (which is the typical error term in most analyses). In general, variance in dependent variables can accrue from multiple sources, both random ones and fixed ones, and our analytic models need to account for these. The question of whether the fundamental dimensions of judgment manifest a positive or negative association demands an answer that demonstrates the importance of this perspective.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2019.01.008.

References

Abele, A. E., Cuddy, A. J., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology, 38*(7), 1063–1065. https://doi.org/10.1002/ejsp.574.

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness – Communion with warmth and morality. *Frontiers in Psychology, 7*, 1810. https://doi.org/10.3389/fpsyg.2016.01810.

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content. A dual perspective model. *Advances in Experimental Social Psychology, 50*, 195–255. https://doi.org/10.1016/B978-0-12-800,284-1.00004-7.

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

Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion.* Chicago, IL: Rand McNally.

Barr, D. J., Levy, R., Scheepers, C., & Tilly, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

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.

Brown, R. (1986). *Social psychology: The second edition.* New York: Free Press.

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

Cambon, L., & Yzerbyt, V. Y. (2016). Compensation is for real: Evidence from existing groups in the context of actual relations. *Group Processes & Intergroup Relations, 20*, 745–756.

Cambon, L., Yzerbyt, V. Y., & Yakimova, S. (2015). Compensation in intergroup relations: An investigation of its structural and strategic foundations. *British Journal of Social Psychology, 54*(1), 140–158. https://doi.org/10.1111/bjso.12067.

Carrier, A., Louvet, E., Chauvin, B., & Rohmer, O. (2014). The primacy of agency over competence in status perception. *Social Psychology, 45*(5), 347–356. https://doi.org/10.1027/1864-9335/a000176.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS Map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology, 92*, 631–648.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Experimental Social Psychology, 24*, 285–290.

Fiske, S. T. (2015). Intergroup biases: A focus on stereotype content. *Current Opinion in Behavioral Sciences, 3*, 45–50. https://doi.org/10.1016/j.cobeha.2015.01.010.

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science, 27*(2), 67–73. https://doi.org/10.1177/0963721417738825.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*, 77–83.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*(3), 491. https://doi.org/10.1037/0022-3514.70.3.491.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Jackson, L. A., Hunter, J. E., & Hodge, C. N. (1995). Physical attractiveness and intellectual competence: A meta-analytic review. *Social Psychology Quarterly,* 108–122.

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology, 89*, 899–913. https://doi.org/10.1037/0022-3514.89.6.899.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*, 54–69.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68*, 601–625.

Kelley, H. H. (1950). The warm-cold variable in first impressions of persons. *Journal of Personality, 18*, 431–439.

Kenny, D. (1994). *Interpersonal perception: A social relations analysis.* New York: Guilford Press.

Kenny, D. A., & La Voie, L. (1984). The social relations model. In D. L. Berkowitz (Vol. Ed.), *Advances in experimental social psychology. vol. 18. Advances in experimental social psychology* (pp. 141–182). New York: Academic Press.

Kervyn, N., Bergsieker, H. B., & Fiske, S. T. (2012). The innuendo effect: Hearing the positive but inferring the negative. *Journal of Experimental Social Psychology, 48*(1), 77–85. https://doi.org/10.1016/j.jesp.2011.08.001.

Kervyn, N., Fiske, S. T., & Yzerbyt, Y. (2015). Forecasting the primary dimension of social cognition. *Social Psychology, 46*(1), 36–45. https://doi.org/10.1027/1864-9335/a000219.

Kervyn, N., Yzerbyt, V., & Judd, C. (2010). Compensation between warmth and competence: Antecedents and consequences of a negative relation between the two fundamental dimensions of social perception. *European Review of Social Psychology, 21*(1), 155–187. https://doi.org/10.1080/13546805.2010.517997.

Kervyn, N., Yzerbyt, V. Y., Demoulin, S., & Judd, C. M. (2008). Competence and warmth in context: The compensatory nature of stereotypic views of national groups. *European Journal of Social Psychology, 38*(7), 1175–1183. https://doi.org/10.1002/ejsp.526.

Kervyn, N., Yzerbyt, V. Y., & Judd, C. M. (2011). When compensation guides inferences: Indirect and implicit measures of the compensation effect. *European Journal of Social Psychology, 41*(2), 144–150. https://doi.org/10.1002/ejsp.748.

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

Minear, M., & Park, D. C. (2004). A life-space database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers, 36*, 630–633.

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

Phalet, K., & Poppe, E. (1997). Competence and morality dimensions of national and ethnic stereotypes: A study in six eastern-European countries. *European Journal of Social Psychology, 27*(6), 703–723 (doi:10.1002/(SICI)1099-0992(199711/12)27:6 < 703::AID-EJSP841 > 3.0.CO;2-K).

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review, 86*, 61–79.

Rosenberg, M. J., & Hovland, M. J. (1960). Cognitive, affective, and behavioral components of attitudes. In C. I. Hovland, & M. J. Rosenberg (Eds.). *Attitude organization and change* (pp. 1–14). New Haven, CT: Yale University Press.

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

Rosenberg, S., & Olshan, K. (1970). Evaluative and descriptive aspects in personality perception. *Journal of Personality and Social Psychology, 16*, 619–626.

Schmitz, M., & Yzerbyt, V. (2018). *Implicit and explicit compensation from the perspective of insider and outsiders* (Manuscript submitted for publication)Louvain-la-Neuve: Catholic University of Louvain.

Snijders, T. A., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: Sage.

Snijders, T. A., & Kenny, D. A. (1999). The social relations model for family data: A multilevel approach. *Personal Relationships, 6*(4), 471–486.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (Eds.). *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks/Cole.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology, 4*, 25–29.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wheterel, M. S. (1987). *Rediscovering the social group. A self-categorization theory.* Oxford, UK: Blackwell.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General, 143*, 2020–2045.

Wheeler, L., & Kim, Y. (1997). What is beautiful is culturally good: The physical attractiveness stereotype has different content in collectivistic cultures. *Personality and Social Psychology Bulletin, 23*(8), 795–800.

Yzerbyt, V., Provost, V., & Corneille, O. (2005). Not so competent but warm… Really? Compensatory stereotypes in the French-speaking world. *Group Processes & Intergroup Relations, 8*, 291–308. https://doi.org/10.1177/1368430205053944.

Yzerbyt, V. Y. (2016). Intergroup stereotyping. *Current Opinion in Psychology, 11*, 90–95. https://doi.org/10.1016/j.copsyc.2016.06.009c.

Yzerbyt, V. Y. (2018). The dimensional compensation model: Reality and strategic constraints on warmth and competence in intergroup perceptions. In A. Abele, & B. Wojciszke (Eds.). *Agency and communion in social psychology* (pp. 126–141). London, UK: Routledge.