"Are We Talking About the Same Person Here?": Interrater Agreement in Judgments of Personality Varies Dramatically With How Much the Perceivers Like the Targets

Daniel Leising, Olga Ostrovski and Johannes Zimmermann

Social Psychological and Personality Science 2013 4: 468 originally published online 16 October 2012
DOI: 10.1177/1948550612462414

Social Psychological and Personality Science
4(4) 468-474

DOI: 10.1177/1948550612462414
spps.sagepub.com

"Are We Talking About the Same Person Here?": Interrater Agreement in Judgments of Personality Varies Dramatically With How Much the Perceivers Like the Targets

Daniel Leising[1], Olga Ostrovski[1], and Johannes Zimmermann[2]

Abstract

We investigated how interrater agreement in personality judgment is affected by the perceivers' affection or dislike regarding the targets. A total of 209 perceivers judged the personalities of 15 targets by means of 30 adjectives. The targets were public figures (e.g., the Pope), which enabled gathering a large number of ratings by perceivers differing in liking. Shared liking was associated with strong increases, and large liking differences were associated with strong decreases, in profile correlations. Shared antipathy was also associated with lower agreement. The greater agreement between judgments of liked targets was largely due to the perceivers characterizing targets positively, whereas judgments of disliked targets were not affected by (shared) negativity to the same extent. The perceivers' attitudes toward the targets constitute an important factor in person perception and need to be taken into account more systematically in research studies.

Keywords

profile, agreement, liking, normative, distinctive

The 2012 U.S. primaries and presidential election once again provide a stunning example of how differently the same persons (i.e., the candidates) may be portrayed by people who have different attitudes toward them. Terms like "principled," "skilled," "successful," "strong," "qualified," "liberal," "socialist," or "leader" are being tossed around and sometimes applied to candidates in a seemingly random manner. Often, the best predictor of which terms are applied to which candidate seems to be whether a term entails a *positive* or *negative* overall evaluation and whether the person who provides the judgment (= perceiver) likes or dislikes the candidate (= target), not who the target person is. Accordingly, the extent to which two perceivers *agree* in their judgments of a target may depend quite strongly on how much each of them likes or dislikes the target. In the present study, we investigate this issue, using public figures as target persons.

Numerous studies have investigated the effects of "acquaintance" between target and perceiver on interrater agreement and accuracy in judgments of personality (e.g., Biesanz, West, & Millevoi, 2007; Borkenau & Liebler, 1992; Borkenau, Mauer, Riemann, Spinath & Angleitner, 2004). The term acquaintance essentially denotes the amount of information that the perceiver has about the target. Studies investigating effects of acquaintance usually find that greater acquaintance is associated with greater interrater agreement and accuracy, with the largest increases taking place at the earliest stages in the process of getting acquainted. In comparison, the number of studies addressing the effects of *liking* is yet very small. Leising, Erbs, and Fritz (2010) demonstrated that the extent to which perceivers are fond of targets is associated with higher ratings of the targets on socially desirable personality traits, and with lower ratings on undesirable traits. In other words, a perceiver's judgment of a target is partly shaped by an *interaction* between the perceiver's attitude toward the target and the positive or negative evaluation that the respective item entails. Leising et al. (2010, study 2) also found evidence that the perceiver's level of affection for the target is systematically associated with interrater agreement. Most important, informants (e.g., friends, spouses) who liked their targets described them in ways that were highly similar to how an *average* target would describe himself or herself. For informants who had a

[1] Department of Psychology, Technical University Dresden, Germany
[2] University of Kassel, Germany

Corresponding Author:
Daniel Leising, Department of Psychology, Technical University Dresden 01062 Dresden, Germany.
Email: leising@psychologie.tu-dresden.de

more critical attitude toward the same targets, this was not the case. In a similar vein, Borkenau and Zaltauskas (2009) found that participants who expressed positive regard for themselves (on the Rosenberg self-esteem scale; Rosenberg, 1965; Ferring & Filipp, 1996) rated themselves in ways that were more similar to how the average person was described by his or her peers. Both studies show that a perceiver who has a more positive attitude toward a target (e.g., himself) will provide a more normative (i.e., average) rating of that target's personality. This certainly has something to do with the fact that normative personality profiles are typically positive: Edwards (1953) was the first to demonstrate that the correlation between item means and item desirabilities (as judged by a group of raters) exceeds $r = .80$. Thus, with regard to personality profiles, what is average is considered good, and vice versa.

In the present study, we provide another investigation of how the perceivers' liking of their targets is associated with interrater agreement in judgments of personality. In this study, we try to overcome some of the limitations of previous studies: In Leising et al.'s (2010) study 2, there were only two informants judging each target, and those informants always represented quite specific segments of the liking continuum (i.e., extreme dislike was relatively rare, and most informant pairs differed considerably in how much they liked the targets). In the present study, we investigate the effects of liking more systematically, by using a larger number of perceivers who represent the full range of the liking variable. In order to be able to obtain ratings by many perceivers differing in liking, we decided to use *public figures* as targets. The reasoning behind this is that public figures, by definition, are known to a large number of people, so the number of potential raters for each target will be very large too.

Method

Selection of Target Persons

As a first step, we compiled a list of $N = 40$ public figures (e.g., actors, politicians, singers, TV hosts) who were particularly "visible" in Germany at the time the study was conducted. This list was presented to a sample of $N = 50$ participants (32 female, age: $M = 24.6$, $SD = 1.9$) who reported how much they liked each target (1 = not at all, 5 = very much), and how well they thought they knew each target (1 = not at all, 5 = very well). From the larger pool of 40 potential targets, we selected 15 for the main study, based on three considerations: First, the targets should be relatively well known to most participants. Second, there should be sufficient variance in how much each target was liked by the participants. Third, the targets should be relatively well distinguishable from each other in terms of their "actual" personalities. The targets we selected were Gregor Gysi (a left-wing politician), Daniela Katzenberger (a Paris-Hilton-like celebrity), Oliver Kahn (the former goal keeper of the German national soccer team), Pope Benedict XVI, Helge Schneider (a comedian), Uli Hoeneß (the president of the Bayern Munich soccer team), Dieter Bohlen (a music producer), Silvio Berlusconi (the prime minister of Italy at that time), Angelina Jolie (the actress), Lena Meyer-Landrut (the German winner of the 2010 Eurovision Song Contest), Karl Lagerfeld (the clothes designer), Ursula von der Leyen (a member of the German government), Heidi Klum (a German model and TV show anchor), Nina Hagen (a singer), Michael Ballack (a soccer player), and Madonna (the singer).

Sample

A total of 209 participants (120 female, 87 male, 2 failed to report sex) took part in the main study. They were 23.1 years old on average ($SD = 4.2$). All participants were recruited by means of a snowball system via e-mail. Most were university students. Accordingly, the average level of education was high (94.3% reported having "Abitur," which is a relatively selective secondary school degree, implying permission to attend university).

Measure

The raters judged each of the 15 targets by means of 30 adjectives, using a 5-point scale ranging from 1 = Doesn't fit at all to 5 = Fits perfectly. The list of adjectives was devised by Borkenau and Ostendorf (1998) to measure the Big Five personality factors by means of six adjectives each. If a perceiver did not know a particular target at all, it was possible to skip the rating of that target. In 192 cases, a perceiver did not provide any ratings of a target. We also omitted six descriptions with more than two missing values. In addition to rating the targets' personalities, the participants also reported how much they thought they knew each target (1 = hardly, 5 = very well), and how much they liked each target (1 = not at all, 5 = very much). In 19 cases, a perceiver did not provide information regarding his or her liking of a target. Altogether, we obtained 2,918 personality profiles that were accompanied by knowing and liking ratings regarding the respective target. The average distribution of liking levels regarding the same target was $n_1 = 34.5$, $n_2 = 52.5$, $n_3 = 69.7$, $n_4 = 28.7$, $n_5 = 9.1$ (Knowing: $n_1 = 37.5$, $n_2 = 41.6$, $n_3 = 57.9$, $n_4 = 45.7$, $n_5 = 11.7$). The average correlation between liking and knowing regarding the same target was $r = .26$ (range: $-.06$ to .43).

Data Analysis and Hypotheses

The central unit of analysis in the present study is the correlation between two personality profiles, each of which consists of 30 adjective ratings. Two profiles may refer to either the same target or two different targets, and they may be provided either by the same perceiver or by two different perceivers. We investigated how profile correlations were associated with how much the perceivers said they knew and liked the targets.

It has often been noted that profile correlations contain a "stereotype" or "normative" component, because the average person receives higher ratings on some items (e.g., "reliable")

than on other items (e.g., "hostile"; Cronbach, 1955; Furr, 2008). Thus, correlations between personality profiles will be higher than zero, on average, even if the profiles do not refer to the same target person. In the present study, we used what Furr (2008) calls the "sample level (random dyads)" strategy to deal with this issue, that is, we systematically compared profile correlations for judgments of the same targets with profile correlations for judgments of different targets (cf. Kenny, Kashy, & Cook, 2006). Altogether, we analyzed (2,918 × 2,917)/2 = 4,255,903 profile correlations.

We conducted a multiple regression analysis to predict profile correlations from (a) whether the two profiles described the same or different targets (dummy coded: 0 = different, 1 = same), (b) whether the two profiles were provided by the same or different perceivers (dummy coded: 0 = different, 1 = same), and (c) how well the two perceivers said they knew and how much they liked the targets. To test for curvilinear effects of knowing and liking, we centered the theoretical scales at zero and included (d) the respective quadratic terms in the regression model. Finally, to assess the specific effects of shared/unshared knowing and liking, we included (e) the respective interaction effects between the two perceivers. The full regression model is reported in Appendix A.[1] It should be noted that multiple regression analyses determine the *unique* contributions of each predictor, thereby unconfounding the effects of liking and knowing (cf. Leising, Erbs, & Fritz, 2010).

The intercept of this regression model represents normative agreement (i.e., profile similarity that is due to the features of the *average* target person). We expected normative agreement to be greater than zero. The regression coefficient for target sameness represents distinctive agreement (i.e., the gain in agreement that is due to two perceivers' shared views of the distinctive features of the same target). We also expected distinctive agreement to be greater than zero. The regression coefficient for perceiver sameness represents assimilation (i.e., the gain in agreement that is due to the average perceiver's viewing other people similarly; cf. Kenny, 1994). We expected assimilation to be greater than zero, as well. Moreover, we expected shared liking to have a positive effect on profile similarity, and large liking differences to have a negative effect on profile similarity. According to our knowledge, the effect of shared *antipathy* on profile similarity has never been investigated before, so we did not formulate a specific hypothesis for this particular constellation of perceiver attitudes.

Note that, as the perceivers in the present study were interchangeable, one would expect identical regression weights to emerge for the liking and knowing of both perceivers. In order to take this into account, we ran the multiple regression 100 times, each time randomly assigning the role of "perceiver 1" and "perceiver 2" to the perceivers in each perceiver pair. In Table 1, we report the *average* regression parameters resulting from this procedure.

Because the profile correlations in the present study were not independent of each other (i.e., they were partly provided by the same perceivers or referred to the same targets), standard inferential statistics were not applicable to the regression coefficients. To solve this problem, we used a jackknife resampling procedure (Efron & Tibshirani, 1993). Specifically, we computed the average regression parameters ($N =$) 209 times, each time omitting the data from exactly one of the 209 perceivers (the 100 random perceiver permutations described above were nested within each of these 209 analyses, requiring the computation of 20,900 regression models overall). The distribution of average regression coefficients across these 209 resamples was then used to estimate standard errors for each coefficient (for the exact formula see Efron & Tibshirani, 1993, p. 141), which enabled the application of t tests.[2] We also used this procedure to test several linear combinations of regression coefficients for statistical significance (see below). Analyses were performed with R 2.12.2 (R Development Core Team, 2011), using the Linux computer cluster at the University of Kassel.

Table 1. Multiple linear regression.

Variable	Index	b	SE_b	t	b^*
Intercept	b_0	.304	.012	26.21	***
Same Target	b_1	.140	.005	30.90	.123***
Same Perceiver	b_2	.084	.004	21.06	.020***
Knowing P1	b_3	.013	.004	3.37	.041***
Knowing P2	b_4	.013	.004	3.37	.038***
Knowing P1^2	b_5	.007	.002	3.17	.036**
Knowing P2^2	b_6	.007	.002	2.95	.036**
Knowing P1*P2	b_7	−.001	< .001	3.13	−.007**
Liking P1	b_8	.040	.004	9.79	.126***
Liking P2	b_9	.040	.005	8.79	.128***
Liking P1^2	b_{10}	−.012	.003	4.32	−.048***
Liking P2^2	b_{11}	−.011	.003	4.31	−.047***
Liking P1*P2	b_{12}	.040	.003	13.20	.166***

Note. DV = profile correlations (N = 4,255,903). b are unstandardized, and b^* are standardized regression coefficients averaged across 100 analyses, each time randomly assigning the role of Perceiver 1 (P1) and Perceiver 2 (P2) to the two perceivers in each perceiver pair. SE_b are jackknife-based standard errors for the average unstandardized regression coefficients.
*p < .05. **p < .01. ***p < .001.

Results

Effects of Liking on Interrater Agreement

Table 1 displays the results of the multiple regression analyses described above. The average overall R^2 for the complete model was .088. The average (normative) agreement between two different perceivers judging two different targets was substantial ($b_0 = .304$). When two perceivers judged the *same* target, agreement rose by $b_1 = .140$. This latter coefficient reflects distinctive agreement and is very comparable to the levels of distinctive agreement that are reported in the literature (Biesanz et al., 2007; Leising et al., 2010). Moreover, judgments of two different targets by the *same perceiver* were $b_2 = .084$ more similar than judgments of two different targets by two different perceivers. This latter coefficient reflects assimilation.

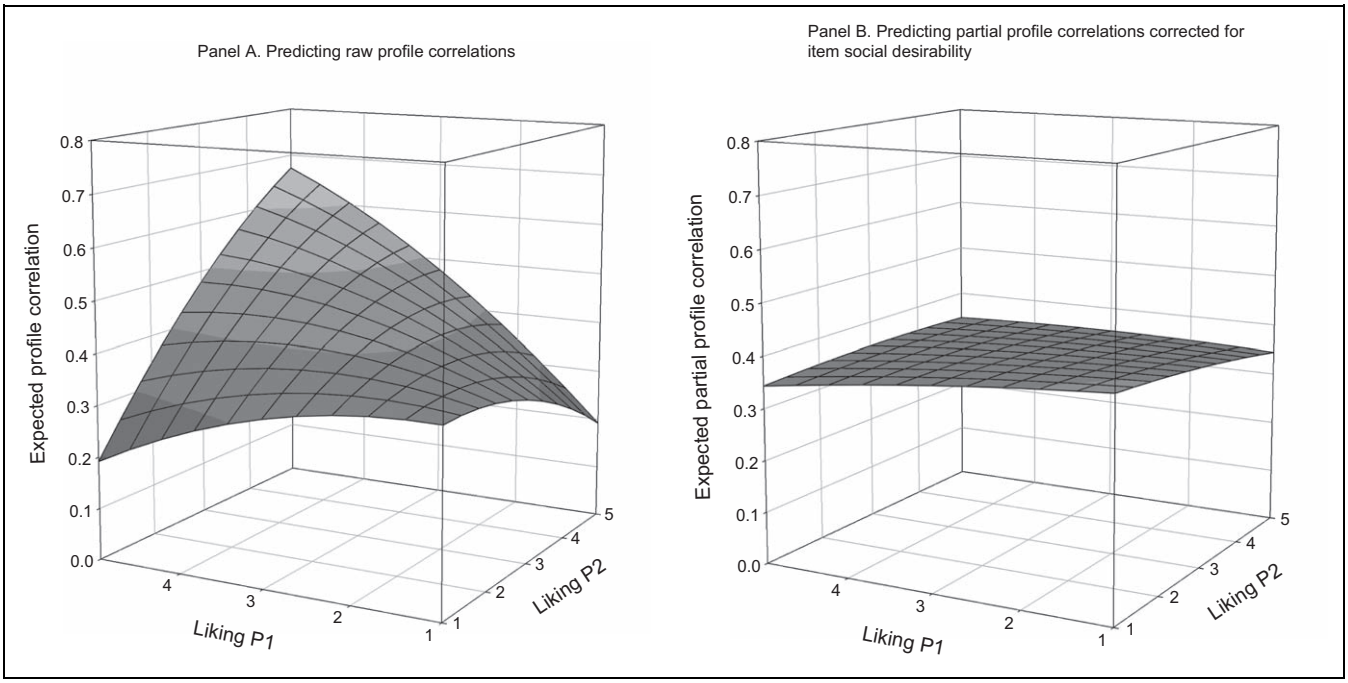

Figure 1. Effects of the perceivers' liking on profile similarity when two perceivers describe the same target.

Although the perceivers' subjective knowing of the targets contributed significantly to profile similarity, the respective regression coefficients were generally small (all *b*s < .015). In contrast, the perceivers' liking had quite dramatic effects. Panel A in Figure 1 displays the regression plane for the unique influence of the liking levels of two different perceivers (i.e., controlling for knowing) on profile correlations between judgments of the *same* target: If both perceivers liked the same target very much, the average profile correlation was substantially higher (expected $r = .67$; back corner of the regression plane in Panel A), as compared to judgments of the same target by two neutral perceivers, $p < .001$.

If both perceivers were neutral (center of the plane) or disliked the target very much (front corner), the expected profile correlations were .44 and .35, respectively. Thus, similar liking levels of the perceivers were not associated with interrater agreement in a symmetric fashion along the liking continuum: Shared liking was associated with *higher* interrater agreement, but shared antipathy was associated with *lower* interrater agreement (the difference between .44 and .35 was statistically significant, $p < .001$). However, the latter profile correlation ($r = .35$) was significantly higher than the overall intercept of the model, $p = .013$, suggesting that judgments by two perceivers who disliked and judged the same target person were still recognizable as referring to the *same* target person.

Dissimilar liking levels of the perceivers were associated with another marked drop in the expected profile correlation ($r = .19$; left and right corners of the regression plane in Panel A). Notably, this latter correlation was significantly lower than the overall intercept of the model, $p < .001$, suggesting that two perceivers who greatly differed in their attitudes toward the same target described that target in more dissimilar ways than did two neutral perceivers who described two different targets. For example, if one perceiver loved Silvio Berlusconi and another perceiver hated Silvio Berlusconi, their portrayals of Berlusconi's personality resembled each other *less* than did a portrayal of Silvio Berlusconi by one neutral perceiver and a portrayal of Angelina Jolie by another neutral perceiver. In other words, the effect of highly different attitudes toward the same target did override the effect of target "sameness."

We will now report a few findings for judgments of *different* targets. For such judgments, a parallel plane lying ($b_1 =$) .140 below the one shown in Panel A could be displayed. Notably, descriptions of two different targets by two perceivers who liked those targets were significantly more similar to each other ($r = .53$) than descriptions of the same target by two neutral perceivers ($r = .44$), $p = .001$. For example, if one perceiver loved Silvio Berlusconi and another perceiver loved Angelina Jolie, their portrayals of these two targets resembled each other *more* than a portrayal of Angelina Jolie by one neutral perceiver and a portrayal of Angelina Jolie by another neutral perceiver. In other words, the effect of highly similar attitudes toward different targets did override the effect of target "differentness."

If two perceivers shared an *antipathy* regarding their respective targets, the expected profile correlation was $r = .21$, which is significantly lower than the overall intercept of the model, $p < .001$. This means that judgments of two different targets by two perceivers who disliked those targets were less similar than judgments of two different targets by two neutral perceivers. Finally, if one perceiver liked his or her target and the other perceiver disliked his or her target, the expected profile

correlation dropped even more to $r = .05$. However, this latter correlation was still significantly higher than zero, $p = .029$. Thus, if one perceiver hated Silvio Berlusconi and judged him, and another perceiver loved Angelina Jolie and judged her, the correlation between the respective personality profiles was still positive.

Effects of Item Desirability

In Leising et al.'s (2010) study, the extent to which a perceiver would assign a personality attribute to a target was strongly predicted by the interaction between that perceiver's liking of the target and the desirability/undesirability of the respective personality attribute. We attempted to replicate this finding, by correlating Leising et al.'s (2010) desirability ratings for the 30 items with the average itemwise βs resulting from simultaneously regressing the perceivers' ratings of each target on the respective perceivers' knowing and liking levels regarding that target.[3] We did not find a significant association between the average knowing βs and item desirability, $r(28) = .01$, $p = .983$, but we did find an extremely strong association between the average liking βs and item desirability, $r(28) = .95$, $p < .001$. This clearly confirms that person descriptions using items with more evaluative (i.e., positive or negative) content are more prone to being affected by the respective perceiver's attitude toward the target (cf. John & Robins, 1993; Vazire, 2010).

Across targets and knowing levels, the average profile correlations between target characterizations and item desirabilities were $r = -.10$ for liking level 1, $r = .15$ for liking level 2, $r = .38$ for liking level 3, $r = .53$ for liking level 4, and $r = .46$ for liking level 5. That is, perceivers who disliked their targets characterized them in a more or less *neutral* fashion, whereas perceivers who neither liked nor disliked their targets described them in a more positive fashion, and perceivers who liked their targets a lot described them in an even more positive fashion. This linear increase probably explains some of the asymmetric association between shared levels of liking and interrater agreement (cf. Figure 1, Panel A): The more two perceivers like their targets, the more they will hold shared positive views of those targets, which will contribute to a higher profile correlation. Of note, perceivers who *dislike* targets do not seem to share strong *negative* views to the same extent.

As a final step, we directly investigated the extent to which the dramatic effects of Liking on interrater agreement could be explained by the perceivers' tendency to respond in a socially desirable manner when judging liked targets. For this purpose, we reran the regression analyses, using *partial* profile correlations as the dependent variable (i.e., correlations between profiles from which the item desirabilities had been statistically removed; cf. Leising et al., 2010). To the extent that the effect of Liking on interrater agreement is driven by an interaction with item desirability, the effect should be reduced when predicting these partial correlations. Panel B in Figure 1 shows that, when controlling for item desirability, the effect of Liking on profile correlations vanished almost completely (all bs < |.01|).

Discussion

The present study demonstrates how dramatically interrater agreement in judgments of personality varies with how much the perceivers like the targets: Shared liking is associated with substantially higher profile correlations, whereas large discrepancies in liking are associated with lower profile correlations. If two different targets are described by perceivers who like them very much, the resulting personality profiles will be even more similar to each other than if the *same* target is judged by two neutral perceivers. When two perceivers differ greatly in how much they like the same target, their characterizations of that target's personality will be even less similar than characterizations of two different targets by neutral perceivers. Thus, the present study suggests that judgments of personality may sometimes depend more strongly on how much the perceivers like the targets than on who the targets are.

The association between shared liking and higher profile agreement was largely driven by the perceivers' describing the targets in a more socially desirable manner. Of note, this effect was strong even when two perceivers judged the *same* target person, so it may not be accounted for by actual personality differences between (e.g., more or less likeable) targets. Given the strength of these associations, future studies of personality judgment should take the perceivers' attitudes toward the targets, and the interactions between those attitudes and the evaluative content of the items, into account more systematically.

Shared antipathy was generally associated with lower profile correlations. Thus, the effect of similar liking levels on interrater agreement was *asymmetric* along the liking continuum. This finding is quite remarkable because it would have been just as conceivable that shared antipathy is also associated with *higher* interrater agreement (e.g., if perceivers portrayed persons they dislike in stereotypically negative ways). However, this seems not to be the case. As indicated by our analysis using partial profile correlations, perceivers judging liked targets seem to share positive views of the targets, which contributes to better interrater agreement, but agreement between perceivers who judge disliked targets is not augmented to the same extent by shared negative views. A recent study by Leising, Ostrovski, and Borkenau (2012) suggests that negative person descriptions tend to be more idiosyncratic than positive person descriptions. The findings of the present study point into the same direction and also align well with findings from studies in experimental psychology which demonstrate that positively evaluated stimuli tend to be perceived as being more similar to each other than negatively evaluated stimuli (cf. Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008).

Knowing did not have strong effects in the present study. We suspect that this may have to do with the nature of our target sample, which comprised public figures only. The perceivers in our study did not have any personal contact or personal relationships with the targets. Thus, greater subjective

knowing did not necessarily imply greater access to private information, which it does when ordinary people get better acquainted with each other. In other words, relatively higher levels of knowing in the present study may still have been relatively low absolute levels of knowing, as compared to how well people may actually get to know each other. Likewise, high or low levels of liking may affect judgments of public figures somewhat differently, as compared to judgments of people one personally knows. Therefore, future research will have to demonstrate that the conclusions of the present study do apply to personality judgments that are derived under more typical circumstances (i.e., situations in which the perceivers and the targets do have personal contact with each other).

Appendix A

Regression model

Profile correlation = $\beta_0 + \beta_1$ * Same Target + β_2 * Same Perceiver + β_3 * Knowing_P1 + β_4 * Knowing_P2 + β_5 * Knowing_P1^2 + β_6 * Knowing_P2^2 + β_7 * Knowing_P1_P2 + β_8 * Liking_P1 + β_9 * Liking_P2 + β_{10} * Liking_P1^2 + β_{11} * Liking_P2^2 + β_{12} * Liking_P1_P2.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. We also tested models including several additional predictors: We included 14 code variables representing the specific targets that were involved in a profile correlation. Doing so enabled us to determine the effects of the other predictors while controlling for differences in the ''normativeness'' of individual targets. We also included interactions between target sameness or perceiver sameness and the perceivers' liking and knowing, because it was conceivable that the impact of liking and knowing might partly depend on whether the perceiver/target is the same person. The interaction effects were very small in size ($b^* < .01$) and did not substantially improve the overall model ($\Delta R^2 < .001$). The individual targets (expectably) differed in normativeness, but the effects of knowing and liking remained virtually the same in both alternative models. Thus, we present the simpler model here.
2. An anonymous reviewer noted that the jackknife approach has not yet been validated for data with more than one random factor and thus might not be an appropriate way to handle the complex nonindependence in our data. In order to rule out this concern, we ran a Monte Carlo simulation assessing the empirical Type-I error rates for jackknife based *t* tests in data with a similar structure: We generated a population of 10,000 perceivers assessing four targets on 30 items and providing ratings on liking and knowing. To make the population data as realistic as possible, we used the item means, standard deviations, and intercorrelations from our sample. The effects of target and perceiver sameness, knowing and liking were constrained to zero. To assess empirical Type-I error rates for each regression coefficient, we drew 1,000 random samples of $n = 50$ perceivers from this population, each time applying the jackknife procedure outlined above, computing jackknife-based *t* tests with a nominal Type-I error rate of .05, and counting the number of samples in which regression coefficients became (erroneously) significant. Ideally, the percentage of significant coefficients (i.e., the ''empirical'' Type-I error rate) should be close to .05. The Type-I error rates that we actually found were slightly lower for the knowing and liking coefficients (ranging from .020 to .043), quite accurate for perceiver sameness (.057), and considerably lower for target sameness (.009). Thus, our simulation study confirms that the jackknife approach yields quite realistic *p* values for data structures such as ours, although tests for the effect of target sameness might be overly conservative.
3. Specifically, we split the data set by target and then separately regressed the perceivers' ratings of the targets on each of the 30 items on the perceivers' liking and knowing levels. That way, we determined the extent to which ratings of a given target on a given item could be predicted from the perceivers' liking and knowing. Then we averaged the resulting β coefficients across targets, to determine how much ratings of an *average* target on a given item could be predicted from liking and knowing. Then we correlated these averaged βs with the ratings of item desirability.

References

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology*, *92*, 119–135. Doi:10.1037/0022-3514.92.1.119

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, *62*, 645–657. Doi:10.1037//0022-3514.62.4.645

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, *86*, 599–614. Doi: 10.1037/0022-3514.86.4.599

Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor model to describe intraindividual variations over time? *Journal of Research in Personality*, *32*, 202–221. Doi:10.1006/jrpe.1997.2206

Borkenau, P., & Zaltauskas, K. (2009). Effects of self-enhancement on agreement on personality profiles. *European Journal of Personality*, *23*, 107–123. Doi:10.1002/per.707

Cronbach, L. J. (1955). Processes affecting scores on ''understanding of others'' and ''assumed similarity''. *Psychological Bulletin*, *52*, 177–193. Doi:10.1037/h0044919

Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, *37*, 90–93. Doi: 10.1037/h0058073

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. London, England: Chapman & Hall.

Ferring, D., & Filipp, S.-H. (1996). Messung des Selbstwertgefühls: Befunde zu Reliabilität, Validität und Stabilität der Rosenberg-Skala [Measurement of self-esteem: Reliability, validity, and stability of the Rosenberg scale]. *Diagnostica*, *42*, 284–292.

Furr, R. M. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality*, *76*, 1267–1316. Doi:10.1111/j.1467-6494.2008.00521.x

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, *61*, 521–551. Doi:10.1111/j.1467-6494.1993.tb00781.x

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

Kenny, D.A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford Press.

Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology*, *98*, 668–682. Doi:10.1037/a0018771

Leising, D., Ostrovski, O., & Borkenau, P. (2012). Vocabulary for describing disliked others is more differentiated than vocabulary for describing liked others. *Journal of Research in Personality*, *46*, 393–396. Doi: 10.1016/j.jrp.2012.03.006

R Development Core Team (2011). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

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

Vazire, S. (2010). Who knows what about a person? The Self-Other Knowledge Asymmetry Model (SOKA). *Journal of Personality and Social Psychology*, *98*, 281–300. Doi:10.1037/a0017908

Author Biographies

Daniel Leising is a professor at the Psychology Department of Technical University Dresden (TUD). His research focuses on person perception, interpersonal behavior, and personality pathology.

Olga Ostrovski is a PhD student at the Psychology Department of Technical University Dresden (TUD). Her research focuses on comparisons between judgments of positively and negatively evaluated target persons, in terms of validity.

Johannes Zimmermann is a research associate at the Psychology Department of the University of Kassel. His research focuses on personality assessment, interpersonal processes in depression, and psychotherapy.