

Perceiver Effects in Person Perception Reflect Acquiescence, Positivity, and Trait-Specific Content: Evidence From a Large-Scale Replication Study

Social Psychological and
Personality Science
2022, Vol. 13(4) 839–848
© The Author(s) 2021



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/19485506211039101
journals.sagepub.com/home/spp



Magdalena Heynicke¹ , Richard Rau² , Daniel Leising¹ ,
Nele Wessels³, and Anne Wiedenroth¹ 

Abstract

Person judgments reflect *perceiver effects*: differences in how perceivers judge the average person. The factorial structure of such effects is still discussed. We present a large-scale, preregistered replication study using over 1 million person judgments (different groups of 200 perceivers judged 200 targets in one of 20 situations, using 30 personality items). Results unanimously favored a model comprising three systematic components: acquiescence (endorsing all items more than other perceivers), positivity (endorsing positive over negative items), and trait specificity (endorsing items reflecting a specific trait more). The latter two factors each accounted for approximately a quarter of the variance in perceiver effects, and acquiescence accounted for less than 10%. Positivity was more influential for evaluative items and was strongly associated with how likable perceivers found their targets to be ($r = .55$). With considerable statistical power and generalizability, our findings significantly improve the knowledge base regarding the structure of perceiver effects.

Keywords

person perception, perceiver effect, impression formation, judgment

Background

What Are Perceiver Effects?

People differ systematically from one another in how they judge the average person: For instance, some perceive almost everybody as being friendly, whereas others see most people as being low on self-discipline. These patterns in how individual perceivers judge the average other person are called perceiver effects (Kenny, 1994). They constitute one of several possible reasons why perceivers tend to disagree somewhat when judging someone, even if their judgments are based on the same information.

Previous studies found that, on average, perceiver effects account for approximately 20%–30% of the variance in person judgments (Hehman et al., 2017; Kenny, 1994). Perceiver effects have been shown to exist in judgments of the Big Five personality factors (Kenny, 1994; Rau et al., 2021; Srivastava et al., 2010) as well as for other trait taxonomies (Dufner et al., 2016; Wood et al., 2010) and at different levels of acquaintance between target and perceiver (Hehman et al., 2017; Vazire, 2010). Moreover, they appear to reflect a relatively stable perceiver characteristic (Rau et al., 2020; Srivastava et al., 2010; Wetzel et al., 2016; Wood et al., 2010).

While these findings clearly emphasize the relevance of perceiver effects in person perception, the factorial structure of perceiver effects is still debated. Perceivers may differ from one another in (a) the tendency to endorse any item (i.e., acquiescence), and/or (b) the tendency to endorse positively valenced items more than negatively valenced items (i.e., positivity), and/or (c) the tendency to endorse items reflecting a specific trait (i.e., trait specificity). If one or several of these mechanisms were at work, individual items assessing perceiver effects would correlate. The present preregistered study sets out to examine which (combinations) of these tendencies represent the structure of perceiver effects best. Table 1 presents some hypothetical data to illustrate the meaning of the three potential sources of perceiver-effect (co)variance.

¹Technical University Dresden, Germany

²University of Münster, Germany

³University of Osnabrück, Germany

Corresponding Author:

Magdalena Heynicke, Technical University Dresden, Dresden 01062, Germany.
Email: magdalena.heynicke@gmail.com

Table 1. Hypothetical Data Example.

Model	Item	Trait	Valence	Perceiver	Target				PE		
					Tom	Tracy	Tina	Average	Acq	Pos	Trait Spec
A	Outgoing	E	Positive	Paul	3	4	5	4	-1	0	0
				Pam	4	5	6	5	0	0	0
				Peter	5	6	7	6	1	0	0
				<i>Average</i>	4	5	6	5			
	Loud	E	Negative	Paul	3	4	5	4	-1	0	0
				Pam	4	5	6	5	0	0	0
				Peter	5	6	7	6	1	0	0
				<i>Average</i>	4	5	6	5			
	Neat	C	Positive	Paul	3	4	5	4	-1	0	0
				Pam	4	5	6	5	0	0	0
				Peter	5	6	7	6	1	0	0
				<i>Average</i>	4	5	6	5			
Compulsive	C	Negative	Paul	3	4	5	4	-1	0	0	
			Pam	4	5	6	5	0	0	0	
			Peter	5	6	7	6	1	0	0	
			<i>Average</i>	4	5	6	5				
P	Outgoing	E	Positive	Paul	3	4	5	4	0	-1	0
				Pam	4	5	6	5	0	0	0
				Peter	5	6	7	6	0	1	0
				<i>Average</i>	4	5	6	5			
	Loud	E	Negative	Paul	5	6	7	6	0	1	0
				Pam	4	5	6	5	0	0	0
				Peter	3	4	5	4	0	-1	0
				<i>Average</i>	4	5	6	5			
	Neat	C	Positive	Paul	3	4	5	4	0	-1	0
				Pam	4	5	6	5	0	0	0
				Peter	5	6	7	6	0	1	0
				<i>Average</i>	4	5	6	5			
Compulsive	C	Negative	Paul	5	6	7	6	0	1	0	
			Pam	4	5	6	5	0	0	0	
			Peter	3	4	5	4	0	-1	0	
			<i>Average</i>	4	5	6	5				
T	Outgoing	E	Positive	Paul	4	5	6	5	0	0	0
				Pam	4	5	6	5	0	0	0
				Peter	4	5	6	5	0	0	0
				<i>Average</i>	4	5	6	5			
	Loud	E	Negative	Paul	4	5	6	5	0	0	0
				Pam	4	5	6	5	0	0	0
				Peter	4	5	6	5	0	0	0
				<i>Average</i>	4	5	6	5			
	Neat	C	Positive	Paul	3	4	5	4	0	0	-1
				Pam	4	5	6	5	0	0	0
				Peter	5	6	7	6	0	0	1
				<i>Average</i>	4	5	6	5			
Compulsive	C	Negative	Paul	3	4	5	4	0	0	-1	
			Pam	4	5	6	5	0	0	0	
			Peter	5	6	7	6	0	0	1	
			<i>Average</i>	4	5	6	5				

Note. Values may vary between 1 (very low) and 9 (very high). E = extraversion; C = conscientiousness; Acq = acquiescence; Pos = positivity; Trait Spec = trait specific. Average scores across targets and perceivers are displayed in italicized print.

In Table 1, three perceivers (Paul, Pam, and Peter) have each judged three targets (Tom, Tracy, and Tina) on four different items that are assumed to represent high levels on two different traits—extraversion (“outgoing,” “loud”), and conscientiousness (“neat,” “compulsive”)—yet with a different valence or

“evaluative tone” (positive vs. negative). Note that the items “outgoing” and “loud” reflect a more positive and a more negative description of the same set of (highly extraverted) target behaviors (column “Trait”), whereas the items “neat” and “compulsive” reflect a more positive and a more negative

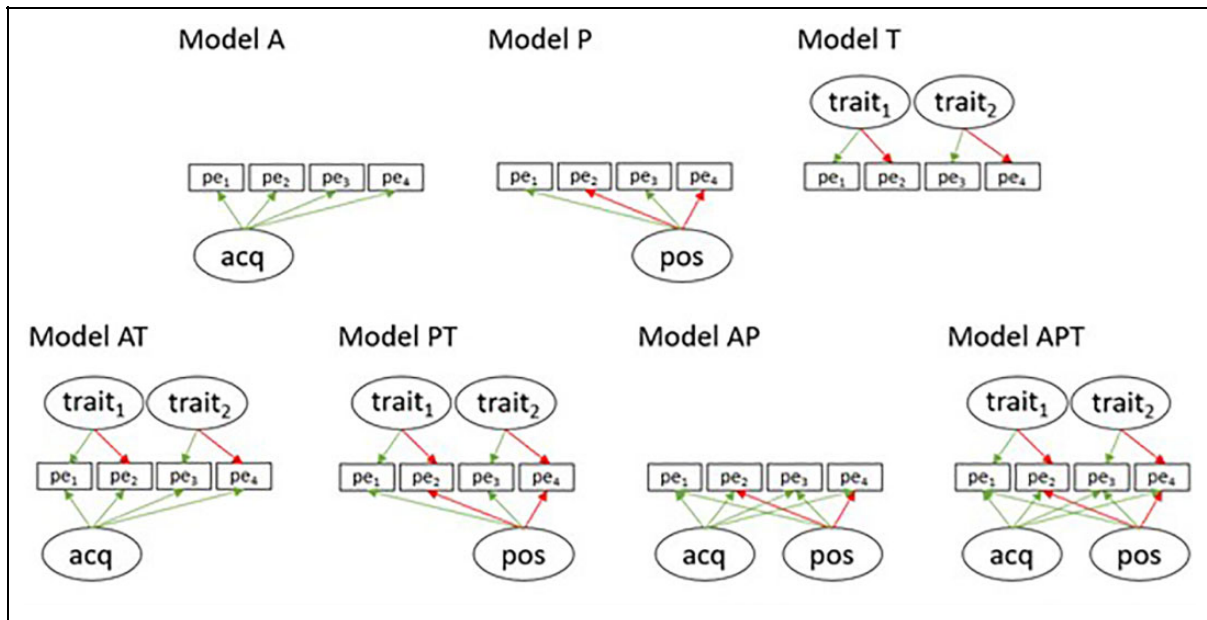


Figure 1. Factor models representing the potential structure of perceiver effects. Note. All factors are constrained to be orthogonal with items loading positively (green) or negatively (red) on the factors. For simplicity, we only display four items and two trait factors (e.g., extraversion and conscientiousness) whereas the actual data set comprises 30 items and five trait-specific scales. Note further that all variations here are between perceivers, not between targets or dyads (pe = perceiver effect, acq = acquiescence, pos = positivity).

description of another set of (highly conscientious) target behaviors.

In order to make this example as unambiguous as possible, we simplified it in several ways. First, we treat the extent to which the use of items reflects how a target is generally seen by others and the extent to which the use of the same items reflects the perceivers’ evaluative stance toward the targets as dichotomous (high/low and positive/negative, respectively). In reality, they likely vary continuously. Second, we assume that the substance and the attitude components are statistically orthogonal, which constitutes a simplification but aligns well with a vast body of literature in personality psychology (e.g., Borkenau & Ostendorf, 1989; John & Robins, 1993; McCrae & Costa, 1983; Saucier et al., 2001; see Leising et al., 2015, for an overview). Finally, Table 1 accounts for differences between the targets in terms of how they are judged by the average perceiver (target effects; Kenny, 1994): On all four items, Tina receives a higher rating from the average perceiver than Tracy, who receives a higher rating than Tom. However, target effects are not the focus of the present article and will thus be ignored in the following.

Table 1 displays three hypothetical scenarios (column “Model”) in which the same perceiver effects (column “Average”) are fueled by three different sources of variance (ignoring measurement error): In Model A, the perceivers’ average judgments of the three targets are perfectly explained in terms of acquiescence levels (column “Acq”), that is, differences between perceivers in the extent to which they endorse any item (Cronbach, 1946): Peter, Pam, and Paul respectively assign an average value of 6, 5, and 4 on all items regardless

of how positive or negative an item’s evaluative tone is (column “Valence”) or what trait it measures (column “Trait”).

In Model P, the perceiver effects are perfectly explained in terms of the perceivers’ generalized positivity toward the targets (column “Pos”), that is, differences between perceivers in the extent to which they endorse items based on how much of a positive light they will shed on targets: Peter assigns an average value of 6 to positive items (“outgoing,” “neat”) but an average value of 4 to negative items (“loud,” “compulsive”). Paul does the opposite, and Pam’s ratings are unaffected by the items’ evaluative tones. These differences between the perceivers exert their influences irrespective of acquiescence or the items’ specific trait content.

In Model T, the perceiver effects are perfectly explained in terms of the individual perceivers’ generalized tendencies to see targets as being high or low on specific traits (column “Trait Spec”): There are no such differences between the three perceivers in responding to items capturing trait E (“outgoing,” “loud”), but there are differences in responding to items capturing trait C (“neat,” “compulsive”): Here, Peter, Pam, and Paul, respectively, assign average values of 6, 5, and 4 to the targets.

While the abovementioned Models A, P, and T consist of a single factor influencing perceiver effects, models with each combination of these factors are also conceivable (Models AT, PT, AP, and APT). All possible perceiver-effect structures in terms of acquiescence, positivity, and trait specificity are depicted in Figure 1. The focus of the present study is to analyze which of these models capture the actual structure of perceiver effects best.

Previous Empirical Studies on the Structure of Perceiver Effects

We are aware of three previous studies addressing research questions closely related to ours: Wood et al. (2010) found that the structure of perceiver effects was sufficiently captured by a model incorporating a single positivity factor. Whereas there was some indication of acquiescence, trait specificity was found to be entirely absent. In contrast, Srivastava et al. (2010) found that, besides a positivity and an acquiescence factor, trait-specific (Big Five) factors were needed to explain perceiver effects. The most comprehensive analysis to date is the one by Rau et al. (2021). These authors reanalyzed a variety of existing data sets and found that, largely in accordance with Srivastava et al. (2010), Model APT fit the data best in 14 of 17 analyses. The remaining analyses favored a model incorporating only acquiescence and positivity (i.e., Model AP). In sum, acquiescence and positivity factors were consistently found whereas the unique contribution of trait-specific factors was mostly, but not always, supported.

A relevant design feature that may explain these discrepancies is the level of acquaintance between targets and perceivers, which differed quite a bit between the various samples reanalyzed by Rau et al. (2021). Studies in which perceivers and targets were largely unacquainted seemed more likely to uncover trait-specific factors in perceiver effects. Given that the present study comprised perceiver–target dyads that were previously unacquainted, and that no interaction took place, we expected that Model APT would fit our data best, in line with the findings by Srivastava et al. (2010) and Rau et al. (2021). This hypothesis was preregistered.

In addition, we determine the relative proportions of variance in perceiver effects, and in judgments overall, that are accounted for by the three factors, respectively. Rau et al. (2021) found that the relative importance of the positivity and trait-specific factors seemed to depend on the characteristics of the scales that were used in a study. Specifically, positivity seemed to more strongly affect ratings on more evaluative scales (e.g., agreeableness) whereas trait-specific factors seemed to be more influential when items referred to highly observable target characteristics (e.g., extraversion). We look into this issue as well. Furthermore, we investigate the nature of the positivity factor in more depth by determining the correlation between its factor scores and the extent to which the perceivers said they found their targets to be “likable” on average. A nonzero correlation would be expected here if one interprets both variables as expressions of the perceivers’ evaluative attitudes toward targets (Leising et al., 2015).

The present study improves on previous research in this area in a number of respects: First, the number of items in many previous studies was relatively small (i.e., 2 items per trait factor), giving individual items a relatively strong (and probably disproportionate) influence over the content of the respective trait factors. We assess each of the Big Five traits with six items, which are also balanced in terms of valence.

Second, in most previous studies, perceivers observed the targets in just one specific social context (e.g., an icebreaker game) and based all of their ratings on this information. In the present study, we collect judgments in 20 different laboratory settings, using a different group of perceivers in each case, which should considerably improve generalizability.

Third, many previous studies involved interactions between targets and perceivers, which make it impossible to distinguish between a perceiver’s typical way of judging the average person and the average other person’s actual behavior toward the perceiver. In the present study, no interaction between perceivers and targets takes place. All perceivers judging the same set of targets receive the same information from videotapes. This enables us to study perceiver effects purely as a perception/judgment phenomenon.

Method

To address our research questions, we use data that were collected as part of a larger research project (LE2151/6-1 funded by Deutsche Forschungsgemeinschaft) on accuracy and bias in person perception (Wiedenroth & Leising, 2020). There, a large number of observers judged the behavior of the same 200 targets in 20 different situations in the lab. For the present analyses, we only used data from the “between-target” condition of the study, in which each perceiver observed and then judged 10 different targets in the same situation. As random groups of 10 targets were always to be judged (in the same situation) by random groups of 10 perceivers, this data set is well-suited for assessing perceiver effects with relatively good reliability. Sample sizes (targets, perceivers, items, situations) were determined a priori as part of the grant proposal. We report all data exclusions, all experimental manipulations, and all measures and preregistrations that are relevant to our research questions (R script, data and materials are publicly available on the Open Science Framework; <https://osf.io/58yvdl/>).

Target and Perceiver Samples

The first part of our data collection involved videotaping behavioral samples in the laboratory. Two-hundred German-speaking target persons were recruited from the general population of a German city. Their age ranged from 17 to 80 years ($M = 33.29$; $SD = 14.48$) with 51% ($n = 102$) being female.

The targets were videotaped as they engaged in 20 different situations, each lasting only a few minutes, such as telling a joke or talking about one’s personal weaknesses. The order in which the individual targets engaged with the situations was balanced using Latin squares (Williams, 1949). The situations were designed to make interindividual differences on the Big Five personality factors visible, and many of them had been successfully used for the same purpose in previous research (Back et al., 2009; Borkenau et al., 2004; Leising et al., 2014). For each situation, the total sample of 200 targets was divided into 20 subsamples of 10 targets each. Each subsample

was judged by a block of 10 perceivers, which enabled us to compute perceiver effects (i.e., how each perceiver judged the same 10 targets, on average). The order in which the 10 perceivers in a subsample watched and then judged “their” 10 targets was also systematically varied across perceivers using Latin squares (Williams, 1949). Perceivers registered for the study via email and then received a single-use link to a secure online platform where they could provide their ratings. All perceivers were compensated for their efforts with €10.

Because we aimed to have targets judged by independent groups of 200 perceivers (i.e., 20 subsamples of 10 perceivers) in each of the 20 situations, we attempted to recruit 4,000 perceivers altogether. After applying a priori exclusion criteria, the final perceiver sample included 3,963 perceivers, resulting in 39,630 sets of judgments of individual targets by individual perceivers altogether. Perceivers were recruited nationwide from the general German population. They had to be at least 18 years old and had to complete the entire rating session. Further reasons for exclusion were if perceivers (1) reported knowing a target, (2) reported problems with the video or audio quality, or (3) were detected for careless responding (e.g., zero rating variance). Approximately 61% ($n = 2,429$) of the perceivers were female ($n = 1,534$ male). Their age ranged from 18 to 82 years ($M = 28.56$; $SD = 10.77$). According to self-reported sociodemographics, a broad spectrum was represented regarding education (e.g. school dropouts, doctoral candidates) and occupation (e.g. student, unemployed, self-employed, clerk).

Measures

The perceivers assessed the targets’ personalities using a list of 30 adjectives that capture the Big Five personality traits (Borkenau & Ostendorf, 1998). For each trait, the measure comprises three items with a positive valence and three items with a negative valence. Items were to be answered using a 1 (*does not apply at all*) to 5 (*applies exactly*) rating scale. For the sake of consistency, ratings on neuroticism items were reversed such that the scale score represented “emotional stability.” The perceivers also reported how likable they found the targets to be, by rating an item (“how likable do you find this person?”) from 1 (*not at all*) to 5 (*very*). The latter ratings are used to further illuminate the nature of the positivity factor in perceiver effects.

Data Analytical Approach

Our data analytical approach aligns very closely with the one described in Rau et al. (2021) and was preregistered. As a descriptive statistic, we computed the proportion of perceiver variance in judgments by estimating intraclass correlations from two-way random effects models with random intercepts for perceivers and targets (Judd et al., 2017). This was first done separately for each item, situation, and target block using the lme4 package (Version 1.1-23; Bates et al., 2015) in R (Version 3.6.3; R Core Team, 2020) and then aggregated across items and situations.

We computed perceiver-effect scores for each item by averaging each perceiver’s ratings across their 10 respective targets. These scores showed no signs of nonnormality and were separately factor analyzed for each of the 20 situations, estimating all of the factor models displayed in Figure 1 using the lavaan package (Version 0.6-5; Rosseel, 2012). We specified the following orthogonal factors: (1) an acquiescence factor with equal loadings on all items, to account for variance that gets introduced when some perceivers provide higher ratings irrespective of item content; (2) a positivity factor with freely estimated loadings on all items, to account for global evaluative differences between perceivers, and note that this interpretation requires that positive loadings be found for socially desirable items and negative loadings for undesirable items; and (3) five trait factors with freely estimated loadings on the six items tapping the same personality factor and zero loadings on the other four trait factors. To examine which of these factors model the existing covariation in perceiver-effect scores between items best, we estimated and compared all seven possible combinations (Model A to APT, see Figure 1). Model selection was based on the Bayesian information criterion (BIC), which penalizes nonparsimony by controlling for the number of free parameters per model (Schwarz, 1978). To also describe the “winning” model in terms of absolute fit, we report the CFI (Comparative Fit Indices), RMSEA (Root-Mean-Square Error of Approximation), and SRMR (Standardized Root-Mean-Square Residual) and use the respective fit indices reported by Srivastava et al. (2010) as a benchmark for comparison.

A small number of models initially produced improper solutions or yielded convergence problems: There were two cases of negative residual variance (Model T for Situation 15 and Model APT for Situation 1) and one irregularly large factor loading (Model APT Situation 18). As we had anticipated such issues (based on the previous analyses by Rau et al., 2021), we were able to solve them in line with minor respecifications that we had outlined beforehand in our preregistration: We set the two negative residual variances to zero and imposed an equality constraint on the respective indicator across all models in Situation 18.

Results

Perceiver Variance

Across all situations and items, differences between perceivers accounted for approximately 14% of the overall variance in judgments, while approximately 16% were attributable to differences between targets. The remaining 70% were due to unique relationship effects between individual perceivers and targets as well as measurement error. Note that there were only small differences between the situations, as perceiver variance proportions ranged between 13% and 15% (see Table 2). However, somewhat larger differences emerged between the Big Five scales, with extraversion items having the lowest proportions of perceiver variance (4%–17%) and agreeableness items

Table 2. Proportions of Perceiver Variance.

Situation	A	C	E	ES	O	Average
1	.17	.13	.10	.16	.11	.14
2	.16	.11	.10	.15	.11	.13
3	.22	.16	.09	.14	.14	.15
4	.20	.14	.08	.15	.14	.14
5	.17	.13	.10	.14	.15	.14
6	.14	.14	.11	.14	.14	.14
7	.17	.14	.09	.13	.14	.13
8	.19	.16	.10	.14	.15	.15
9	.14	.15	.10	.15	.16	.14
10	.19	.13	.10	.14	.11	.13
11	.16	.14	.09	.13	.16	.14
12	.16	.17	.12	.14	.15	.15
13	.17	.16	.10	.16	.16	.15
14	.17	.14	.12	.15	.14	.14
15	.17	.14	.10	.10	.15	.13
16	.16	.14	.10	.14	.17	.14
17	.18	.16	.09	.13	.13	.14
18	.18	.15	.08	.12	.13	.13
19	.17	.17	.11	.14	.12	.14
20	.13	.12	.08	.12	.18	.13
Average	.17	.15	.10	.14	.14	.14

Note. Perceiver variance is the proportion of judgment variance in ratings on individual items which is attributable to systematic differences between perceivers. Scale averages represent the average values of the six items belonging to the same scale (A = agreeableness, C = conscientiousness, E = extraversion, ES = emotional Stability, and O = openness). Averages per situation and scale are displayed in bold print.

having the highest proportions (9%–31%). For most items, at least 10% of the overall variance in judgments was attributable to perceiver effects, which is generally considered a benchmark for a meaningful contribution (Kenny, 1994). We concluded that the data contained a sufficient amount of variance between perceivers to justify the subsequent analyses.

Factor Structure of Perceiver Effects

Our series of confirmatory factor analyses yielded an unanimous result concerning the factor structure of perceiver effects on grounds of the BIC: Model APT was the favored model in all 20 situations (see Table 3), suggesting that perceiver effects were consistently fueled by acquiescence *and* global positivity *and* specific trait content.

Table 3 displays fit indices for Model APT across all 20 situations (see Supplement 4 for factor loadings and residual variances). These fit indices ($.794 \leq \text{CFIs} \leq .854$, $.090 \leq \text{RMSEAs} \leq .109$, $.096 \leq \text{SRMRs} \leq .147$) did not reach conventional criteria for “good” model fit (e.g., $\text{CFI} > .950$; $\text{RMSEA} < .080$; $\text{SRMR} < .080$) and were also somewhat lower than the indices reported by Srivastava et al. (2010), which we had preregistered as a benchmark ($\text{CFI} = .850$; $\text{RMSEA} = .094$; $\text{SRMR} = .069$).¹

There were some minor discrepancies between the expected and actual patterns of how the individual items loaded on the factors. However, the vast majority of factor loadings pointed into the expected direction, with only 0.5% of the factor

loadings on the positivity factor and 3% of the factor loadings of the trait-specific factors diverging from expectations.

Relative Contributions of Acquiescence, Positivity, and Trait Specificity

To determine the relative importance of the individual factors, we conducted separate variance decompositions for each item and each situation. Variance attributable to the individual factors was estimated by squaring the factor loadings of the standardized solution and aggregated across items and situations.

Overall, positivity and trait specificity were about equally important: The positivity factor explained 26% of the perceiver variance, while the trait-specific factors together accounted for 27%. The acquiescence factor contributed approximately another 9%, whereas 38% remained unexplained (see Figure 2A). We also related the variances accounted for by the individual factors to the overall variance in judgments (which also includes target, dyadic, and error variance). Of the 14% perceiver variance on average (see above), about 4% each was attributable to positivity and trait specificity (see Figure 2B). The few factor loadings with unexpected polarity (see above) were included in these calculations, as these deviations were too minor to have any noticeable effect on any of the conclusions.

Conceptual Meaning of the Positivity Factor

In order to explore whether the overall positivity of judgments was associated with the explicit attitude that perceivers reported having toward targets, we conducted two additional analyses, which were not preregistered: First, we correlated the items’ average factor loadings on the positivity factor with ratings of the same items’ social desirability (collected by Leising et al., 2010). The correlation was almost perfect, $r(28) = .98$, $p < .001$ with CI 95% [.97, .99], suggesting that lay raters are very well capable, when asked, to discern the extent to which the use of an item will reflect a perceiver’s more positive or negative view of a target in general.

Second, we correlated the perceivers’ average positivity factor scores with their average responses to the question “How likable do you find this person?” (across “their” 10 targets, respectively). This correlation was far from perfect but still very substantial, $r(3,961) = .55$, $p < .001$ with CI 95% [.52, .57]. Our interpretation is that both the overall positivity of judgments and the perceivers’ self-reported liking of their targets may be viewed as expressions of the perceivers’ generalized *attitudes*.

Discussion

The present study investigated the properties of perceiver effects in judgments of strangers with rigorous methodology (i.e., strict preregistration) and a data set with several important strengths: First, subsamples of 10 perceivers each judged the same 10 target persons, based on the same behavioral information, and with no interaction taking place between them. This design ensured that perceiver effects became measurable with

Table 3. Model Comparison and Fit Indices for Model APT.

Situation	BIC							$\chi^2(df)$	CFI; RMSEA; SRMR		
	Model A	Model P	Model T	Model AT	Model PT	Model AP	Model APT		Model APT		
1	7,291	5,837	5,342	4,651	4,661	5,420	4,235	1,081(375)	.832; .098; .105		
2	7,464	5,818	5,327	4,992	4,703	5,605	4,463	1,123(374)	.817; .101; .110		
3	7,398	6,435	5,867	5,201	4,943	5,819	4,756	1,193(374)	.794; .105; .129		
4	7,221	6,015	5,317	4,580	4,554	5,587	4,206	973(374)	.854; .090; .102		
5	8,070	6,331	5,683	5,220	5,124	6,087	4,740	1,164(374)	.824; .103; .108		
6	8,105	6,751	6,274	5,826	5,680	6,416	5,310	1,133(374)	.809; .101; .108		
7	7,435	6,317	5,713	5,134	5,109	5,945	4,761	1,019(374)	.830; .093; .107		
8	7,143	6,337	6,064	4,893	4,827	5,443	4,552	1,131(374)	.815; .101; .112		
9	7,948	6,044	5,772	5,261	5,071	5,719	4,714	1,109(374)	.832; .099; .097		
10	7,351	6,202	5,649	4,955	4,834	5,783	4,548	1,068(374)	.826; .097; .122		
11	8,015	6,349	5,924	5,253	5,248	5,958	4,858	1,249(374)	.803; .109; .110		
12	7,815	6,668	6,333	5,411	5,438	6,035	5,033	1,125(374)	.817; .101; .100		
13	7,666	6,133	5,802	5,167	5,028	5,697	4,700	1,177(374)	.810; .104; .108		
14	7,773	6,181	5,571	4,803	4,789	5,691	4,326	1,101(374)	.844; .099; .147		
15	7,115	6,362	6,172	4,869	4,894	5,493	4,488	1,053(374)	.839; .095; .096		
16	7,694	6,095	5,455	5,073	4,946	5,854	4,637	1,076(374)	.829; .097; .098		
17	7,478	6,249	5,756	5,044	4,906	5,857	4,734	1,137(374)	.810; .102; .128		
18	7,302	5,790	5,228	4,581	4,501	5,398	4,089	1,131(379)	.830; .100; .117		
19	7,815	6,242	5,787	5,171	5,047	5,884	4,759	1,041(374)	.840; .094; .101		
20	8,029	6,566	5,913	5,291	5,253	6,239	4,838	1,146(374)	.823; .102; .132		
Average	7,606	6,236	5,747	5,069	4,978	5,797	4,637	1,112(374)	.824; .100; .112		

Note. Model comparisons, which consistently rendered Model APT (bold) superior across all twenty situations, according to the Bayesian information criterion (BIC). Absolute fit of Model G is indicated by further indices (CFI = Comparative Fit Indices; RMSEA = Root-Mean-Square Error of Approximation; SRMR = Standardized Root-Mean-Square Residual; χ^2 (df) = Chi-Square Test (degrees of freedom)).

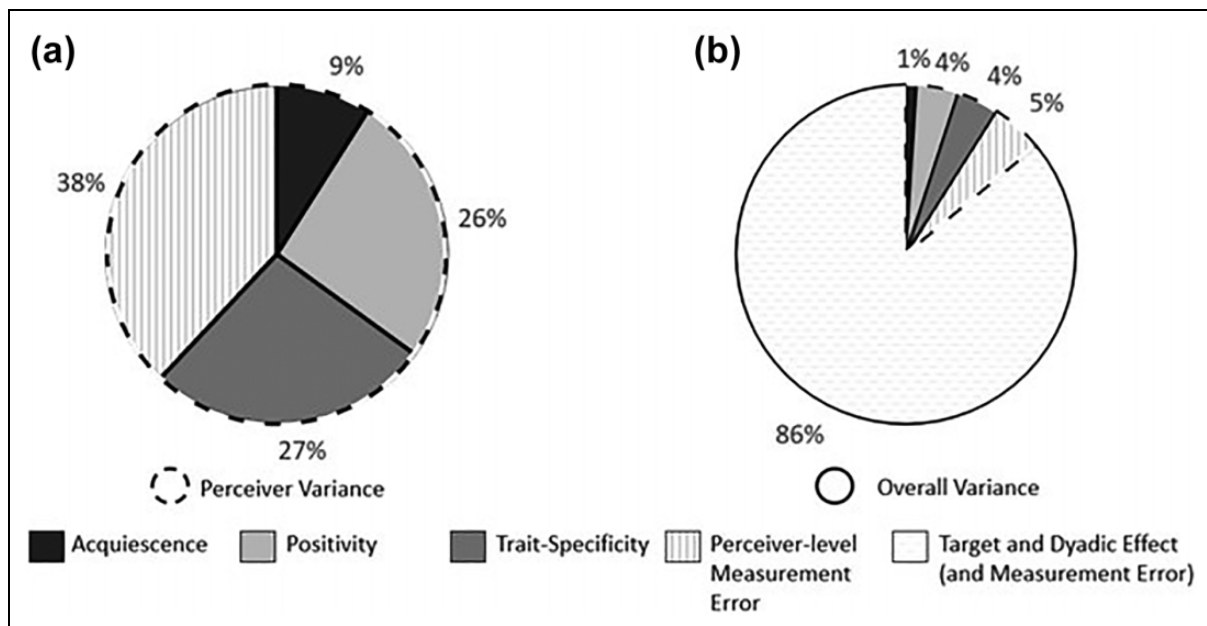


Figure 2. Relative importance of Model APT factors (averaged across all 20 situations). Note. A = Acquiescence; P = positivity; T = trait specific.

remarkable precision. Second, the overall sample comprised close to 4,000 perceivers who judged the targets in one of 20 different situations in the lab. This afforded a high degree of replicability and generalizability. In fact, results turned out to

be extremely stable across the many different perceiver samples and situations. We therefore think that the conclusions drawn from our analyses may be considered quite robust and generally accurate.

We found that perceiver effects accounted for approximately 14% of the overall variance in judgments on individual items which was somewhat smaller than the 20%–30% reported by Hehman et al. (2017) and Kenny (1994), but largely in line with other studies (e.g., Wortman & Wood, 2011), especially with other studies in which targets were judged from video (Rau et al., 2021). This is a substantial contribution and offers one explanation for the fact that consensus in person judgments is usually far from perfect (e.g., Connelly & Ones, 2010). We also found that, unequivocally, Model APT (acquiescence *plus* positivity *plus* trait-specific factors) reproduced the pattern of covariation among the perceiver effects on the 30 items best (closely replicating Rau et al., 2021). Thus, at least at zero acquaintance, all three of the mechanisms outlined in the introduction seem to be at play simultaneously and independently: Perceivers do differ from one another (a) in how much they lean toward endorsing any item (i.e., acquiescence), (b) in how much they lean toward endorsing items based on their evaluative tone (i.e., positivity), and (c) in how much they lean toward endorsing items reflecting a specific trait (i.e., trait specificity). The influence of the latter two factors was found to be considerably stronger than the first. Absolute levels of model fit were below conventional cutoffs for “good” fit, but our strictly confirmatory approach focused on model selection rather than model fit and yielded consistent support for the relevance of all three types of perceiver-effect factors that have been discussed in the literature.

Concerning the relatively large percentage of remaining unexplained variance, there may still be additional systematic influences on perceiver effects that we do not know much about yet and that have not received much attention in the literature either. For example, perceivers may differ in how much they prefer odd over even values on response scales. Or perceiver effects may just be a relatively messy phenomenon, containing many influences that do not generalize across perceivers. This may include mere random error but also more systematic sources of variation such as between-perceiver differences in generalized, learned expectations as to how various target characteristics covary with one another (Stolier et al., 2020).

The present work should be considered largely descriptive in nature, as our study design did not permit a systematic disentangling of ways in which perceiver effects *come about*. Theoretically, two competing explanations may be considered in this regard: Either, perceiver effects are basically the varying “default values” that individual perceivers apply in their judgments of people. In this model, perceiver effects exist independently of how the targets actually behave—they are basically perceiver-specific random effects on the intercept of the regression by which the targets’ actual behaviors are translated into judgments. Or, perceiver effects reflect differences in how perceivers process the information that they receive about targets. For example, Peter may assign higher values on trait C to the average target because he more readily interprets a target’s actual behaviors in terms of evidence for high C levels. In this model, perceiver effects are rooted in perceiver-specific random effects on the slope of the regression translating target

behaviors into judgments. Although our research design does not enable us to decide which of these two mechanisms is at work (both may also operate simultaneously), we do think that the remarkable consistency of the relative proportions of perceiver-effect variance across the diverse set of 20 situations speaks in favor of the first. More research is needed to unveil how perceiver effects come about more conclusively.

We were also able to show, for the first time, that the general positivity of the perceivers’ judgments of the targets correlated substantially with the extent to which the perceivers said they found their targets likable. Furthermore, we found an almost perfect correlation between the items’ loadings on the positivity factor and ratings of the items “social desirability” (Biderman et al., 2018; Leising et al., 2021). Overall, these findings accord well with a conceptual framework positing that (a) most person judgments do reflect attitude variation between perceiver–target dyads (with only some of this variation being attributable to perceivers, however); (b) this attitude variation is reflected both in the positivity bias and perceivers’ self-reported liking of the target, though perceivers may not necessarily be aware of this bias; and (c) people in general are well aware of the ways in which the use of certain terms to describe targets reflects the respective perceiver’s attitude toward a target (see Leising et al., 2015).

The present study did have some limitations that should be acknowledged: First, the situations in which the targets were observed were specifically designed to evoke interindividual differences within the Big Five framework, and the used items were taken from a Big Five measure with well-established factorial validity (Borkenau & Ostendorf, 1998). These two design choices combined are likely to have at least somewhat increased our chances of finding trait-specific perceiver-effect components. Using more natural behavioral samples and a broader range of descriptors may certainly yield a more ecologically valid—and possibly messier—picture of the structure of perceiver effects. Also, the universal applicability and utility of the Big Five taxonomy itself (e.g., across cultures) is still, and even increasingly, being criticized (Möttus et al., 2020; Thalmeyer et al., 2020).

Second, the “one-directional” research design that we used enabled us to clearly separate perceiver variance that is rooted exclusively in perceivers’ minds from other kinds of perceiver variance. While this type of design is well-suited for analyses like the ones we intended to perform, it does systematically rule out other influences that likely play a role in everyday interpersonal perception. Specifically, perceivers likely differ in what target behaviors they typically evoke and encounter, causing perceiver effects that are not merely perceptual. The fact that such differences could not occur in the present study may explain why we observed less perceiver variance (14%) than is often observed in studies that permit interactions between perceivers and targets (20%–30%). This difference suggests that both types of perceiver variance may be substantial in size and that these two types of studies may not be treated as being interchangeable.

Furthermore, the current study investigated the influence and structure of perceiver effects at zero acquaintance, but personality judgments in real life often occur at considerably higher levels of acquaintance. Rau et al. (2021) found that the influence of trait-specific factors may diminish with increasing information about targets. In line with this reasoning, Wood et al. (2010) concluded that a single positivity factor was sufficient to model perceiver effects in judgments of previously acquainted students. From a theoretical perspective, such an effect would be plausible because the better the perceivers get to know the targets, the more their specific leanings toward rating the average target as being high (or low) on a given trait may be overridden by actual behavioral information of the average target (i.e., target effects). Whether trait-specific perceiver-effect components really become more negligible as perceivers get to know targets better will have to be addressed by future studies.


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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The data collection for this manuscript was part of the larger research project LE2151/6-1 funded by the Deutsche Forschungsgemeinschaft (DFG).

ORCID iDs

Magdalena Heynicke  <https://orcid.org/0000-0002-6328-5125>

Richard Rau  <https://orcid.org/0000-0002-1490-8584>

Daniel Leising  <https://orcid.org/0000-0001-8503-5840>

Anne Wiedenroth  <https://orcid.org/0000-0002-9282-3666>

Note

1. To learn about potential systematic sources of misfit, we explored whether there were post hoc modifications that would considerably improve model fit across situations. Indeed, we found five conceptually plausible modifications that were indexed by modification indices ≥ 15 in at least half of the situations. These modifications either captured method variance (e.g., a correlation of the items “phantasie/los” [not imaginative] and “einfalls/los” [not inventive] or accounted for unmodeled semantic overlap (e.g., a cross-loading of the emotional stability item “launisch” [moody] on the agreeableness factor). Allowing for these modifications yielded levels of absolute fit comparable to preregistered benchmarks, $.827 \leq CFI_s \leq .894$, $.077 \leq RMSEAs \leq .103$, and $.082 \leq SRMRs \leq .123$.

References

- Back, M., Schmukle, S., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology, 97*, 533–548. <https://doi.org/10.1037/a0016229>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Biderman, M. D., McAbee, S. T., Job Chen, Z., & Hendy, N. T. (2018). Assessing the evaluative content of personality questionnaires using bifactor models. *Journal of Personality Assessment, 100*(4), 375–388. <https://doi.org/10.1080/00223891.2017.1406362>
- 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*(4), 599–614. <https://doi.org/10.1037/0022-3514.86.4.599>
- Borkenau, P., & Ostendorf, F. (1989). Descriptive consistency and social desirability in self- and peer reports. *European Journal of Personality, 3*(1), 31–45. <https://doi.org/10.1002/per.2410030105>
- Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor model to describe intra-individual variations over time? *Journal of Research in Personality, 32*(2), 202–221. <https://doi.org/10.1006/jrpe.1997.2206>
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers’ accuracy and predictive validity. *Psychological Bulletin, 136*(6), 1092–1122. <https://doi.org/10.1037/a0021212>
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement, 6*(4), 475–494. <https://doi.org/10.1177/001316444600600405>
- Dufner, M., Leising, D., & Gebauer, J. E. (2016). Which basic rules underlie social judgments? Agency follows a zero-sum principle and communion follows a non-zero-sum principle. *Personality and Social Psychology Bulletin, 42*(5), 677–687. <https://doi.org/10.1177/0146167216640902>
- Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology, 113*(4), 513–529. <https://doi.org/10.1037/pspa0000090>
- 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*(4), 521–551. <https://doi.org/10.1111/j.1467-6494.1993.tb00781.x>
- 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. <https://doi.org/10.1146/annurev-psych-122414-033702>
- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. 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*(4), 668–682. <https://doi.org/10.1037/a0018771>
- Leising, D., Gallrein, A. M., & Dufner, M. (2014). Judging the behavior of people we know: Objective assessment, confirmation of pre-existing views, or both? *Personality and Social Psychology Bulletin, 40*(2), 153–163. <https://doi.org/10.1177/0146167213507287>
- Leising, D., Scherbaum, S., Locke, K., & Zimmermann, J. (2015). A model of “substance” and “evaluation” in person judgments.

- Journal of Research in Personality*, 57, 61–71. <https://doi.org/10.1016/j.jrp.2015.04.002>
- Leising, D., Vogel, D., Waller, V., & Zimmermann, J. (2021). Correlations between person-descriptive items are predictable from the product of their mid-point-centered social desirability values. *European Journal of Personality*, 35(5), 667–689. <https://doi.org/10.1177/0890207020962331>
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51, 882–888. <https://doi.org/10.1037/0022-006X.51.6.882>
- Möttus, R., Wood, D., Condon, D. M., Back, M., Baumert, A., Costantini, G., Epskamp, S., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A. G. C., Yarkoni, T., Ziegler, M., & Zimmermann, J. (2020). Descriptive, predictive and explanatory personality research: Different goals, different approaches, but a shared need to move beyond the big few traits. *European Journal of Personality*, 34(6), 1175–1201. <https://doi.org/10.1002/per.2311>
- Rau, R., Carlson, E. N., Back, M. D., Barranti, M., Gebauer, J. E., Human, L. J., Leising, D., & Nestler, S. (2021). What is the structure of perceiver effects? On the importance of global positivity and trait-specificity across personality domains and judgment contexts. *Journal of Personality and Social Psychology*, 120(3), 745–764. <https://doi.org/10.1037/pspp0000278>
- Rau, R., Nestler, W., Dufner, M., & Nestler, S. (2020). Seeing the best or worst in others: A measure of generalized other-perceptions. *Assessment*. <https://doi.org/10.1177/1073191120905015>
- R Core Team. (2020). *R: A language and environment for statistical computing*. <https://www.R-project.org>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <http://www.jstatsoft.org/v48/i02/>
- Saucier, G., Ostendorf, F., & Peabody, D. (2001). The non-evaluative circumplex of personality adjectives. *Journal of Personality*, 69, 537–582. <https://doi.org/10.1111/1467-6494.694155>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464. <https://doi.org/10.1214/aos/1176344136>
- Srivastava, S., Guglielmo, S., & Beer, J. S. (2010). Perceiving others' personalities: Examining the dimensionality, assumed similarity to the self, and stability of perceiver effects. *Journal of Personality and Social Psychology*, 98(3), 520–534. <https://doi.org/10.1037/a0017057>
- Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour*, 4(4), 361–371. <https://doi.org/10.1038/s41562-019-0800-6>
- Thalmeyer, A. G., Saucier, G., Ole-Kotikash, L., & Payne, D. (2020). Personality structure in East and West Africa: Lexical studies of personality in Maa and Supyire-Senufo. *Journal of Personality and Social Psychology*, 119(5), 1132–1152. <https://doi.org/10.1037/pspp0000264>
- Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98(2), 281–300. <https://doi.org/10.1037/a0017908>
- Wetzel, E., Lüdtke, O., Zettler, I., & Böhnke, J. R. (2016). The stability of extreme response style and acquiescence over 8 years. *Assessment*, 23(3), 279–291. <https://doi.org/10.1177/1073191115583714>
- Wiedenroth, A., & Leising, D. (2020). The more the better—But more of which? Information quantity and shared meaning as predictors of consistency and accuracy in person judgment. *Journal of Research in Personality*, 87, Article 103968. <https://doi.org/10.1016/j.jrp.2020.103968>
- Williams, E. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Chemistry*, 2(2), 149–168. <https://doi.org/10.1071/ch9490149>
- Wood, D., Harms, P., & Vazire, S. (2010). Perceiver effects as projective tests: What your perceptions of others say about you. *Journal of Personality and Social Psychology*, 99(1), 174–190. <https://doi.org/10.1037/a0019390>
- Wortman, J., & Wood, D. (2011). The personality traits of liked people. *Journal of Research in Personality*, 45(6), 519–528. <https://doi.org/10.1016/j.jrp.2011.06.006>

Author Biographies

Magdalena Heynicke pursued her bachelor's degree in Psychology at the Technische Universität Dresden. In her thesis she focused on the structure of perceiver effects.

Richard Rau works as a post-doctoral researcher in the Personality and Assessment Lab and the Statistics and Psychological Methods Lab at University of Münster. He studies the psychometric properties and social consequences of interpersonal perception.

Daniel Leising is a professor of diagnostics and intervention research at Technische Universität Dresden in Germany. His research focuses on person judgement, interpersonal behavior and personality pathology, and the abstract conceptual modelling of these phenomena in particular.

Nele Wessels is postdoctoral researcher at University of Osnabrueck. She is a social and personality psychologist who studies interpersonal judgment and accuracy, as well as their consequences.

Anne Wiedenroth is a Ph.D. student at Technische Universität Dresden in Germany. In her research she focuses on accuracy, consistency, and consensus in interpersonal judgments and behavior observation.

Handling Editor: Lisa Libby